Steve Pepper

The typology and semantics of binominal lexemes

Noun-noun compounds and their functional equivalents

Faculty of Humanities
Submitted 2020-05-26
Colophon
“Some people seem to think that there is one correct set of optimal comparative concepts, and that comparative concepts should not be based on intuition or chosen “arbitrarily” (in Gilbert Lazard’s words). But this is wrong: There are myriad ways of comparing languages, and thus myriad possible comparative concepts. Which kinds of concepts are the most productive concepts, most likely to yield deeper insights, is a matter for research. In fact, finding good comparative concepts is one of the most important ingredients of the creative process for successful comparative research.”

(Martin Haspelmath, email to the LingTyp mailing list, 2016-01-20)
Contents

Figures xi
Tables xv
Abstract xix
Acknowledgements xxi
Abbreviations xxiii
Typographical and naming conventions xxv

1 Towards a comparative concept 1
1.1 Background 2
1.1.1 Topic Maps and associative thought 2
1.1.2 Nominal compounding in Nizaa 5
1.2 Binominal lexemes as a comparative concept 7
1.2.1 The limitations of compounding 7
1.2.2 The potential of binominal lexemes 8
1.2.3 An onomasiological perspective 10
1.2.4 Defining the object of study 12
1.3 Theoretical framework 14
1.4 Design of the study 16
1.4.1 An empirical, data-driven approach 16
1.4.2 Research questions 17
1.5 Structure of this work 17

2 Earlier work 21
2.1 Compounding 22
2.1.1 Bauer (2001) 22
2.1.2 Arnaud (2004) 29
2.1.3 Morbo/Comp (2004-2006) 32
2.1.4 Scalise & Bisetto (2009) 34
2.1.5 Guevara & Scalise (2009) 37
2.2 Word-formation 41
   2.2.1 Aikhenvald (2007) 41
   2.2.2 Štekauer, Valera & Körtvélyessy (2012) 44
2.3 Prefiguring binominals 48
   2.3.1 Levi (1978) – Complex nominals 49
   2.3.2 Rainer (2013) – Relational adjectives, etc. 50
   2.3.3 Bauer & Tarasova (2013) – Adnominal modification 51
2.4 Morphological complexity 53
   2.4.1 Haspelmath & Tadmor (2009) – WOLD 53
   2.4.2 Urban (2012) 56
2.5 Chapter summary 61

3 Meanings, languages and data 62
3.1 The meaning list 62
   3.1.1 Initial extraction of 159 meanings 63
   3.1.2 Evaluation against five languages 66
   3.1.3 Expansion to 201 meanings 70
   3.1.4 Reduction to 100 meanings 70
   3.1.5 Overall evaluation 81
3.2 The language sample 85
   3.2.1 Types of sample 85
   3.2.2 The present sample 88
   3.2.3 Overall evaluation 91
3.3 Data collection 92
   3.3.1 Online database 93
   3.3.2 Questionnaires 95
   3.3.3 Dictionaries 96
   3.3.4 Data cleansing 98
3.4 Chapter summary 99

4 Data annotation 101
4.1 Identifying binominals 101
   4.1.1 Identifying thing-roots 104
   4.1.2 Identifying thing-affixes 108
4.2 Determining the head 109
  4.2.1 Binominals consisting of two thing-roots 110
  4.2.2 Binominals consisting of a thing-root and a thing-affix 113
4.3 Defining the construction 116
4.4 Data analytics 119
  4.4.1 Data analytics by language 120
  4.4.2 Data analytics by meaning 125
4.5 Chapter summary 130

5 Typological classification 131
  5.1 Theoretical prerequisites 132
    5.1.1 Koptjevskaja-Tamm (2002; 2003) 132
    5.1.2 Croft (2003) 134
  5.2 A non-hierarchical alternative 141
  5.3 Binominal types 145
    5.3.1 No additional marker: jxt, cmp, der and cls 145
    5.3.2 One additional marker: prp, gen, adj and con 154
    5.3.3 Two additional markers: dbl 159
  5.4 Unattested strategies 160
    5.4.1 Head-marking correlate of prp (prn) 161
    5.4.2 Head-marking correlate of adj (nml) 162
  5.5 Gradience 163
  5.6 Data analytics 169
    5.6.1 Distribution of binominals by type 169
    5.6.2 Intralingual competition 173
    5.6.3 Compounding as a universal 175
    5.6.4 Genealogical and areal patterning 177
  5.7 Chapter summary 180

6 Semantic relations 183
  6.1 Theoretical prerequisites 184
    6.1.1 Background 184
    6.1.2 Nominal modification (Bauer & Tarasova 2013) 188
    6.1.3 Metonymy and derivation (Janda 2011) 193
<table>
<thead>
<tr>
<th>6.2</th>
<th>An integrated approach to classification</th>
<th>195</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.1</td>
<td>A low-level classification</td>
<td>195</td>
</tr>
<tr>
<td>6.2.2</td>
<td>A high-level classification</td>
<td>198</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Extending Hatcher’s classification</td>
<td>204</td>
</tr>
<tr>
<td>6.3</td>
<td>Classifying binominals</td>
<td>216</td>
</tr>
<tr>
<td>6.3.1</td>
<td>The Bourquifier</td>
<td>218</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Bourque2</td>
<td>224</td>
</tr>
<tr>
<td>6.3.3</td>
<td>The Hatcher-Bourque classification</td>
<td>236</td>
</tr>
<tr>
<td>6.4</td>
<td>Data analytics</td>
<td>239</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Low-level semantic relations</td>
<td>240</td>
</tr>
<tr>
<td>6.4.2</td>
<td>High-level semantic relations</td>
<td>247</td>
</tr>
<tr>
<td>6.5</td>
<td>Chapter summary</td>
<td>252</td>
</tr>
<tr>
<td>7</td>
<td>Typological generalizations</td>
<td>255</td>
</tr>
<tr>
<td>7.1</td>
<td>Word order typology</td>
<td>255</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Constituent order</td>
<td>255</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Consistency</td>
<td>261</td>
</tr>
<tr>
<td>7.1.3</td>
<td>Mixed languages</td>
<td>264</td>
</tr>
<tr>
<td>7.2</td>
<td>Binominals and possessives</td>
<td>268</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Anchoring and typifying possessive constructions</td>
<td>270</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Comparing non-binary typologies</td>
<td>273</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Data analytics</td>
<td>276</td>
</tr>
<tr>
<td>7.3</td>
<td>Strategies and semantics</td>
<td>279</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Intralinguistic patterns</td>
<td>280</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Cross-linguistic patterns</td>
<td>285</td>
</tr>
<tr>
<td>7.4</td>
<td>Chapter summary</td>
<td>291</td>
</tr>
<tr>
<td>8</td>
<td>Conceptual generalizations</td>
<td>295</td>
</tr>
<tr>
<td>8.1</td>
<td>The two-paths hypothesis</td>
<td>295</td>
</tr>
<tr>
<td>8.2</td>
<td>Head-framing vs. modifier-framing</td>
<td>302</td>
</tr>
<tr>
<td>8.3</td>
<td>E-sites and the relational/sortal cline</td>
<td>308</td>
</tr>
<tr>
<td>8.4</td>
<td>Species-framing vs. attribute-framing</td>
<td>312</td>
</tr>
<tr>
<td>8.5</td>
<td>Chapter summary</td>
<td>315</td>
</tr>
</tbody>
</table>
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 A model of associative relations</td>
<td>317</td>
</tr>
<tr>
<td>9.1 Metonymic relations (Peirsman &amp; Geeraerts 2006)</td>
<td>319</td>
</tr>
<tr>
<td>9.2 Cognitive relations (Koch and Blank)</td>
<td>324</td>
</tr>
<tr>
<td>9.3 Topic Maps: roles and granularity</td>
<td>326</td>
</tr>
<tr>
<td>9.4 Putting it all together: the PHAB model</td>
<td>331</td>
</tr>
<tr>
<td>9.5 Chapter summary</td>
<td>333</td>
</tr>
<tr>
<td>10 Conclusion</td>
<td>335</td>
</tr>
<tr>
<td>10.1 Summary</td>
<td>335</td>
</tr>
<tr>
<td>10.2 Contribution to science</td>
<td>336</td>
</tr>
<tr>
<td>10.2.1 Typology</td>
<td>337</td>
</tr>
<tr>
<td>10.2.2 Semantics</td>
<td>339</td>
</tr>
<tr>
<td>10.2.3 Methods</td>
<td>343</td>
</tr>
<tr>
<td>10.3 Further research</td>
<td>346</td>
</tr>
<tr>
<td>10.4 Envoie</td>
<td>348</td>
</tr>
<tr>
<td>References</td>
<td>349</td>
</tr>
<tr>
<td>Appendices</td>
<td>381</td>
</tr>
<tr>
<td>A. Languages</td>
<td>383</td>
</tr>
<tr>
<td>B. Sources</td>
<td>387</td>
</tr>
<tr>
<td>C. Meanings</td>
<td>393</td>
</tr>
<tr>
<td>D. Strategies and constructions</td>
<td>395</td>
</tr>
<tr>
<td>E. Binominal data set</td>
<td>448</td>
</tr>
<tr>
<td>F. Database structures</td>
<td>485</td>
</tr>
<tr>
<td>G. Questionnaire</td>
<td>487</td>
</tr>
<tr>
<td>H. Additional tables</td>
<td>489</td>
</tr>
<tr>
<td>Name index</td>
<td>499</td>
</tr>
<tr>
<td>Language index</td>
<td>505</td>
</tr>
<tr>
<td>Subject index</td>
<td>511</td>
</tr>
</tbody>
</table>
Figures

Figure 1: The TAO model ................................................................. 3
Figure 2: The anatomy of an association ............................................. 3
Figure 3: The five basic onomasiological types .................................... 11
Figure 4: Extent of compounding in 13 languages .................................. 32
Figure 5: Classifying compounds (Bisetto & Scalise 2005) ....................... 35
Figure 6. Classifying compounds (Scalise & Bisetto 2009) ....................... 36
Figure 7: Scales of preference in compounding (Guevara & Scalise 2009) .... 40
Figure 8: Extract from questionnaire (Štekauer et al 2012) .................... 45
Figure 9: MCS analysis of compounding (Štekauer & al. 2012) ................. 47
Figure 10: Pepper as a loanword in WOLD ....................................... 53
Figure 11: View of the table 'meaning' in its original form ...................... 65
Figure 12: View of the table 'meaning' showing frequencies ..................... 71
Figure 13: Extract of the table 'word' showing constructions .................... 72
Figure 14: Number of constructions as a function of number of meanings ...... 73
Figure 15: Cross-linguistic representation of meanings .......................... 83
Figure 16: Distribution of meanings across semantic fields ..................... 84
Figure 17: Distribution of meanings across semantic types ...................... 85
Figure 18: Areal distribution of language sample .................................. 90
Figure 19: Extract of the data after import ....................................... 94
Figure 20: Subset of the Polish data .................................................. 103
Figure 21: Givón’s scale of temporal stability ...................................... 106
Figure 22: Subset of the Polish binominal data showing constructions ........ 118
Figure 23: Distribution of data by language and meaning ...................... 120
Figure 24: Basic data summary (by language) .................................... 121
Figure 25: Morphological complexity (A/S) by language ....................... 123
Figure 26: Binominal frequency (NN/S) by language ............................ 124
Figure 27: Basic data summary (by meaning) ..................................... 126
Figure 28: Vocabulary sizes across semantic fields .............................. 127
Figure 29: Morphological complexity (A/S) by meaning ........................................ 128
Figure 30: Binominal frequency (NN/S) by meaning .............................................. 129
Figure 31: Binominal frequency (NN/A) by meaning .............................................. 130
Figure 32: Major structural types of PNPs in Europe ............................................. 132
Figure 33: Strategies for relating concepts and their diachronic relationships ........ 139
Figure 34: Typological classification of binominal lexemes .................................. 143
Figure 35: Degree of fusion (Bybee 1985) ............................................................ 144
Figure 36: Formal classification showing gradient phenomena ................................ 163
Figure 37: Number of binominals by type ............................................................. 170
Figure 38: Areal distribution of binominal types ..................................................... 170
Figure 39: Preferred binominal types by number of languages ............................... 172
Figure 40: Geographical distribution of binominal type preferences ..................... 173
Figure 41: Intralingual competition ........................................................................ 173
Figure 42: Correlations between binominal types ................................................... 174
Figure 43: Attested, common and dominant binominal types ................................. 176
Figure 44: The Indo-European spectrogram ......................................................... 178
Figure 45: The Indo-European scatter plot ............................................................ 179
Figure 46: Hatcher’s reworking of Jespersen’s classification .................................. 200
Figure 47: Bourque’s template for POSSESSION .................................................. 218
Figure 48: The Bourquifier – FAMILY ESTATE ..................................................... 219
Figure 49: The Bourquifier – BEESWAX .............................................................. 220
Figure 50: The Bourquifier – OAK TREE ............................................................. 231
Figure 51: The Bourquifier – BEAR CUB ............................................................. 232
Figure 52: The Bourquifier – SUNBURN ............................................................... 233
Figure 53: Boathouse words .................................................................................. 234
Figure 54: The Bourquifier v 3.0 ......................................................................... 237
Figure 55: Overall frequency of low-level semantic relations ................................. 240
Figure 56: Number of languages that exhibit a particular relation ......................... 242
Figure 57: Number of meanings that exhibit a particular relation ........................... 243
Figure 58: Number of binominal types that exhibit each relation ........................... 244
Figure 59: Low-level semantic relations across binominal types ............................ 245
Figure 60: Semantic relations and semantic types ................................................. 245
Figure 61: How low-level relations vary across semantic fields ............................. 246
Figure 62: Binominals, languages, meanings and strategies by htype .......... 248
Figure 63: High-level semantic relations across binominal types .............. 249
Figure 64: How high-level relations vary across semantic types .............. 251
Figure 65: How high-level relations vary across semantic fields ............. 251
Figure 66: Constituent order by binominal type .................................. 257
Figure 67: Constituent order in compounds by area .............................. 258
Figure 69: Distribution of right-headed and left-headed binominals ........ 260
Figure 69: Grades of similarity across anchoring and binominal strategies ..... 278
Figure 70: Residuals for cells in the Polish contingency table .................. 283
Figure 71: Non-normal distribution of marking across languages .......... 287
Figure 72: Normal distribution of marking across meanings ................... 289
Figure 73: The Bourquifier – WINDMILL ........................................ 297
Figure 74: Conceptual variability of head and modifier .......................... 304
Figure 75: Scatterplot of heads and modifiers ..................................... 306
Figure 76: Heatmap of heads and modifiers by semantic type ................ 307
Figure 77: Species- vs. attribute-framing .......................................... 314
Figure 78: Semantic relations as a hierarchy ...................................... 317
Figure 79: The anatomy of an association ........................................... 327
Figure 80: The PHAB model of associative relations ............................ 331
## Tables

Table 1: Croft’s grid of basic cross-linguistic constructions 15
Table 2: Language sample (Bauer 2001) 22
Table 3: Order of noun-adjective and noun-modifier (Bauer 2001) 25
Table 4: Noun + Adjective tetrachoric tables 26
Table 5: Order of noun possessor and noun-modifier (Bauer 2001) 26
Table 6: Noun + Possessor tetrachoric tables 27
Table 7: Consistency of head-modifier ordering (Bauer 2001) 27
Table 8: Consistency of head-modifier ordering (Pepper 2015) 28
Table 9: Languages covered in Arnaud (2004) 30
Table 10: Onomasiological cross-linguistic comparison (Arnaud 2004) 31
Table 11: Languages represented in Morbo/Comp 33
Table 12: Morphological classification (Štekauer et al. 2012) 44
Table 13: Examples of WOLD meanings 55
Table 14: Binominal constructions yielded by NOSTRIL 73
Table 15: Additional constructions yielded by EARLOBE 74
Table 16: Constructions lost with a sample of 84 meanings (extract) 75
Table 17: Meanings represented by lost constructions 76
Table 18: Final set of 100 meanings 78
Table 19: Complete list of languages in the sample 90
Table 20: Genetic distribution of languages 91
Table 21: Areal distribution of languages 92
Table 22: Preliminary classification of structural types 102
Table 23: Overall database statistics 119
Table 24: Types of PNP and binominal 134
Table 25: Reordered binominal types 142
Table 26: Preferred binominal types by language 172
Table 27: Levi’s (1978) classification applied to six binominal types 191
Table 28: Bourque’s original system of semantic relations 197
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Variety of word-formation types – professions</td>
<td>222</td>
</tr>
<tr>
<td>30</td>
<td>Summary of changes to Bourque’s system</td>
<td>236</td>
</tr>
<tr>
<td>31</td>
<td>Hatcher-Bourque classification</td>
<td>238</td>
</tr>
<tr>
<td>32</td>
<td>Most frequent low-level semantic relations</td>
<td>243</td>
</tr>
<tr>
<td>33</td>
<td>Summary of high- and low-level semantic relations</td>
<td>247</td>
</tr>
<tr>
<td>34</td>
<td>Constituent order pace Morbo/Comp</td>
<td>256</td>
</tr>
<tr>
<td>35</td>
<td>Languages with significant degree of mixed order</td>
<td>263</td>
</tr>
<tr>
<td>36</td>
<td>Strategies and high-level relations in Mapudungun</td>
<td>265</td>
</tr>
<tr>
<td>37</td>
<td>Strategies and high-level relations in Greek</td>
<td>265</td>
</tr>
<tr>
<td>38</td>
<td>Strategies and high-level relations in Barain</td>
<td>265</td>
</tr>
<tr>
<td>39</td>
<td>Strategies and low-level relations in Barain</td>
<td>266</td>
</tr>
<tr>
<td>40</td>
<td>Strategies and low-level relations in Barain (rev.)</td>
<td>267</td>
</tr>
<tr>
<td>41</td>
<td>Heine’s set of eight event schemes</td>
<td>269</td>
</tr>
<tr>
<td>42</td>
<td>Anchoring/non-anchoring comparison table</td>
<td>272</td>
</tr>
<tr>
<td>43</td>
<td>Noun-possessor tetrachoric tables (Pepper 2015)</td>
<td>274</td>
</tr>
<tr>
<td>44</td>
<td>Pepper scale for non-binary typological comparison</td>
<td>275</td>
</tr>
<tr>
<td>45</td>
<td>Comparing anchoring and binominal strategies</td>
<td>277</td>
</tr>
<tr>
<td>46</td>
<td>Koch’s motivational grid</td>
<td>281</td>
</tr>
<tr>
<td>47</td>
<td>Strategies and high-level relations in Polish</td>
<td>282</td>
</tr>
<tr>
<td>48</td>
<td>$\chi^2$-squared and Fisher tests for 22 languages</td>
<td>285</td>
</tr>
<tr>
<td>49</td>
<td>Number of markers per binominal type</td>
<td>285</td>
</tr>
<tr>
<td>50</td>
<td>Coding of binominals for number of markers</td>
<td>286</td>
</tr>
<tr>
<td>51</td>
<td>Coding of languages for weighted average of markers (excerpt)</td>
<td>286</td>
</tr>
<tr>
<td>52</td>
<td>Coding of meanings for weighted average of markers (excerpt)</td>
<td>286</td>
</tr>
<tr>
<td>53</td>
<td>Semantic relations in Nizaa</td>
<td>296</td>
</tr>
<tr>
<td>54</td>
<td>Motivational grid for Kalamang with row sums</td>
<td>299</td>
</tr>
<tr>
<td>55</td>
<td>Simplified motivational grid for Kalamang</td>
<td>300</td>
</tr>
<tr>
<td>56</td>
<td>Competition between cmbR and conR in Kalamang</td>
<td>301</td>
</tr>
<tr>
<td>57</td>
<td>Conceptual variation in heads and modifiers</td>
<td>303</td>
</tr>
<tr>
<td>58</td>
<td>Salience of elaboration sites (by relation)</td>
<td>309</td>
</tr>
<tr>
<td>59</td>
<td>Salience of elaboration sites (by role)</td>
<td>310</td>
</tr>
<tr>
<td>60</td>
<td>Components that elaborate the concept HOUSE</td>
<td>310</td>
</tr>
<tr>
<td>61</td>
<td>Peirsman &amp; Geeraerts’ metonymical patterns</td>
<td>320</td>
</tr>
</tbody>
</table>
Table 62: Bourquifier template for PRODUCTION 321
Table 63: Metonymic patterns and semantic relations 322
Table 64: Blank’s ten cognitive relations 325
Table 65: Association types in the Italian Opera Topic Map 329
Table 66: Morbo/Comp database structure 489
Table 67: Language sample (Štekauer, Valera & Körtvélyessy 2012) 490
Table 68: The original 201 meanings (alphabetical order) 491
Table 69: The original 201 meanings (“binominality” order) 493
Table 70: Constructions lost with a sample of 84 meanings 495
Abstract

This dissertation establishes ‘binominal lexeme’ as a comparative concept and discusses its cross-linguistic typology and semantics. Informally, a binominal lexeme is a noun-noun compound or functional equivalent; more precisely, it is a lexical item that consists primarily of two thing-morphs between which there exists an unstated semantic relation.

Examples of binominals include Mandarin Chinese tie3 lu4 [iron road], French chemin de fer [way of iron] and Russian želez.naja doroga [iron.adjz road]. All of these combine a word denoting ‘iron’ and a word denoting ‘road’ or ‘way’ to denote the meaning ‘railway’. In each case, the unstated semantic relation is one of composition: a railway is conceptualized as a road that is composed, or made, of iron. However, three different morphosyntactic strategies are employed: compounding, prepositional phrase and relational adjective. In this study, I explore the range of such strategies used by a worldwide sample of languages to express a set of 100 meanings from various semantic domains, resulting in a classification consisting of nine different morphosyntactic types.

I also investigate the semantic relations found in the data and develop a classification called the Hatcher-Bourque system that operates at two levels of granularity, together with a tool for classifying binominals, the Bourquifier. The classification is extended to other subfields of language, including metonymy and lexical semantics, and beyond language to the domain of Topic Maps and knowledge representation, resulting in a proposal for a general model of associative relations called the PHAB model.

Among the other findings of the research are: universals concerning the recruitment of anchoring nominal modification strategies; a method for comparing non-binary typologies; the non-universality (despite its predominance) of compounding; and a scale of frequencies for semantic relations which may provide insights into the associative nature of human thought.
Samandrag

Denne avhandlingen definerer eit komparativt omgrep ‘binominalt leksem’ og handsamar omgrepet ut frå tverrspråkleg typologi og semantikk. Uformelt er eit binominalt leksem ei substantiv-substantiv-samansetning eller funksjonelt likeverdig med ei slik; meir presist er det ei leksikalsk eining som i fyrste rekkje består av to ting-morfar med ein implisitt semantisk relasjon mellom dei.


Eg undersøkjer òg dei semantiske relasjonane eg finn i dataa mine og utviklar ei klassifisering eg kallar Hatcher-Bourque-systemet, som opererer på to nivå, saman med ein reiskap til å klassifisere binominale leksem. Denne klassifiseringa blir så utvida til å gjelde andre område innanfor språk, som metonymi og leksikalsk semantikk, og utover språk til domene som emnekart og kunnskapsrepresentasjon. Resultatet er eit framlegg til ei allmenn modell av assosiative relasjonar som eg kallar PHAB-modellen.

Dette høyrer òg med til resultata av denne forsking: ein metode for samanlikning av ikkje-binære typologiar; universale som gjeld rekruttering av possessive strategiar; at samansetning ikkje er universell, sjølv om ho dominerer i høve til andre strategiar; og ein frekvensskala for semantiske relasjonar som kan gje innsikt i den assosiative naturen til mennesketanken.
I owe a major debt of thanks to my supervisors, Professor Emeritus Rolf Theil and Professor Åshild Næss, both of the University of Oslo: Rolf, for believing in my somewhat unorthodox ideas, and for his polyglot knowledge of a phenomenal number of languages; Åshild for her attention to terminological detail and for trying, usually in vain, to make me stick to my schedule. I also thank the Department of Linguistics at the University of Oslo, in particular Helge Lødrup and Andreas Sveen, for betting on such an ageing horse.

The initial inspiration for this research stems from my work with Topic Maps, and I would like to take this opportunity to acknowledge all of my co-conspirators in that venture, in particular Steve Newcomb. The world may not yet have appreciated the value of what we were doing, but history will be our judge.

In the field of linguistics, three eminent scholars, Bill Croft, Martin Haspelmath and Ron Langacker, have provided much of my inspiration. I take my hat off to them, for their integrity and for the quality of their respective life works.

In addition to the above, I have drawn particular inspiration from the work of Pierre Arnaud, Laurie Bauer, Yves Bourque, Anna Granville Hatcher†, Laura Janda, Masja Koptjevskaja-Tamm, Pavol Štekauer and Mathias Urban. Many thanks to all of you. I hope any criticisms I have voiced are taken in the spirit in which they are intended: as a reflection of my admiration, and a sincere desire to advance our scientific understanding of language.

Many colleagues provided the encouragement I needed from time to time to keep going through the ups and downs of writing a dissertation; they include Kjell-Magne Yri, Sonia Cristofaro, Martin Haspelmath (again), Pavol Štekauer (again) and Pierre Arnaud (again). I owe Pierre a special debt of gratitude for being my sparring partner as I developed the Hatcher-Bourque classification and the PHAB model.

Thanks also to Bård Uri Jensen and Natalia Levshina for checking my statistics, to Robert Forkel for helping me decipher the WOLD data, and to Lynn Rosentrater for generating the maps.
Many people have contributed data to my database, and I thank them all profusely. Without you guys, I would not have had the empirical basis on which to develop my ideas. I acknowledge you in Appendix B, Sources, along with the language(s) you assisted me with. You are, in alphabetical order:


My biggest debt of all is to my life partner, Sylvia Schwab, without whose love, understanding and encouragement this work would never have seen the light of day in any form whatsoever.

---

1 Especial thanks to the contributor who wrote “I’ve learnt much from your project and collecting the data for you”. As typologists we rely immensely on language documenters. It’s nice to know that the debt is not always one-sided.
### Abbreviations

Note: Abbreviations whose description contains a page reference (e.g. 740) denote postbases in Central Yupik (ESU); references are then to Jacobson (2013).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st person</td>
</tr>
<tr>
<td>3</td>
<td>3rd person</td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ABS</td>
<td>absolute</td>
</tr>
<tr>
<td>ABST</td>
<td>abstract</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ACT</td>
<td>action</td>
</tr>
<tr>
<td>ADIZ</td>
<td>adjectivizer</td>
</tr>
<tr>
<td>ADLT</td>
<td>adult</td>
</tr>
<tr>
<td>AG</td>
<td>agreement marker</td>
</tr>
<tr>
<td>AGT</td>
<td>agent</td>
</tr>
<tr>
<td>AL</td>
<td>alienable possession</td>
</tr>
<tr>
<td>ANAPH</td>
<td>anaphoric</td>
</tr>
<tr>
<td>ANTIP</td>
<td>antipassive</td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
</tr>
<tr>
<td>AQ3</td>
<td>thing that resembles N in some respect (740)</td>
</tr>
<tr>
<td>AR(AQ)</td>
<td>little piece of N (741)</td>
</tr>
<tr>
<td>ASS</td>
<td>associative</td>
</tr>
<tr>
<td>ATTR</td>
<td>attributive</td>
</tr>
<tr>
<td>AUG</td>
<td>augmentative</td>
</tr>
<tr>
<td>BN</td>
<td>bound noun</td>
</tr>
<tr>
<td>CENGAQ</td>
<td>one with a small N (748)</td>
</tr>
<tr>
<td>CIRC</td>
<td>circumfix</td>
</tr>
<tr>
<td>CL</td>
<td>class marker</td>
</tr>
<tr>
<td>CLF</td>
<td>classifier</td>
</tr>
<tr>
<td>COLL</td>
<td>collective</td>
</tr>
<tr>
<td>CON</td>
<td>connective</td>
</tr>
<tr>
<td>CUUN</td>
<td>device for V-ing; device associated with N (758)</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>DEP</td>
<td>dependency marker</td>
</tr>
<tr>
<td>DER</td>
<td>derivational affix</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>DEV</td>
<td>devalued</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>DUAL</td>
<td>dual</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
</tr>
<tr>
<td>ESS</td>
<td>essive</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>FAM</td>
<td>familiar</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>GNL</td>
<td>general</td>
</tr>
<tr>
<td>ILITAQ</td>
<td>device for protecting N (764)</td>
</tr>
<tr>
<td>INAL</td>
<td>inalienable possession</td>
</tr>
<tr>
<td>INDF</td>
<td>indefinite</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INS</td>
<td>instrumental</td>
</tr>
<tr>
<td>IRIN</td>
<td>forms names of weekdays (767)</td>
</tr>
<tr>
<td>LAT</td>
<td>lative</td>
</tr>
<tr>
<td>LE</td>
<td>linking element</td>
</tr>
<tr>
<td>LEK</td>
<td>one with N or Ns, one having N (786)</td>
</tr>
<tr>
<td>LIG</td>
<td>ligature</td>
</tr>
<tr>
<td>LK</td>
<td>linker</td>
</tr>
<tr>
<td>LLEQ1</td>
<td>former N (796)</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>NFE</td>
<td>noun-forming enclitic</td>
</tr>
<tr>
<td>NH</td>
<td>nonhuman</td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NONF</td>
<td>nonfinite</td>
</tr>
<tr>
<td>NONS</td>
<td>nonsubject</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>OWN</td>
<td>owner</td>
</tr>
<tr>
<td>PAUC</td>
<td>paucal</td>
</tr>
<tr>
<td>PERF</td>
<td>perfective</td>
</tr>
<tr>
<td>PER</td>
<td>pertensive</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PREF</td>
<td>prefix</td>
</tr>
<tr>
<td>PREP</td>
<td>preposition</td>
</tr>
<tr>
<td>PRON</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROP</td>
<td>proprietive</td>
</tr>
<tr>
<td>PROX</td>
<td>proximate</td>
</tr>
<tr>
<td>PSR</td>
<td>possessor</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>PURP</td>
<td>purposive</td>
</tr>
<tr>
<td>QLIQ</td>
<td>the one located far in the area of space denoted by N (848)</td>
</tr>
<tr>
<td>QUQ</td>
<td>one that is V, one that is like N (851)</td>
</tr>
<tr>
<td>REC</td>
<td>receptacle</td>
</tr>
<tr>
<td>RED</td>
<td>reduplication</td>
</tr>
<tr>
<td>REL</td>
<td>relative</td>
</tr>
<tr>
<td>RELN</td>
<td>relational noun</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SGLT</td>
<td>singulative</td>
</tr>
<tr>
<td>SPEC</td>
<td>specific</td>
</tr>
</tbody>
</table>
Typographical and naming conventions

The following typographical conventions are used in this work:

- **N PREP N, Mod. ADJZ Head** – a (binominal) construction.
- **[Gloss]** – a morpheme gloss.
  
  **Note**: Leipzig Glossing Rules are followed, EXCEPT THAT, for improved readability, I use periods instead of hyphens for morpheme breaks (Rule 2), and colons instead of periods for one-to-many correspondences (Rule 4). Where hyphens appear in glosses, they reflect the presence of a hyphen in the original orthography.

- **MEANING** – a concept or language-independent meaning, in particular one that belongs to the set of 100 meanings used as a basis for the data collection.

- **vernacular** – a linguistic item in an object language; in lower case throughout, irrespective of language-specific casing conventions.

- **DEU** – an ISO 639-3 language code.
  
  **Note**: In my database languages are identified by glottocode, as defined in Glottolog 2.7 (Hammarström et al. 2016). However, the corresponding ISO 639-3 code is used in the text and in tables since it is both shorter and more transparent. Languages can be looked up by ISO code in Appendix A (page 383) and by name in the Index of Languages (page 505). To find the ISO code for a language in the database, use either the index or Appendix B (page 387). One language, Caijia does not have an ISO code; to save space in tables I have taken the liberty of assigning it the unused code CAI, but database applications should use the glottocode caij1234.

- Page numbers given in the form page ## (as above) refer to the present work; those given as p. ## refer to a page in another, recently referenced work.

- **code** – file names, variable names and computer code, including SQL queries and R scripts (R Core Team 2018).

- ‘She’ is used throughout for the gender neutral pronoun in preference to s/he.

Glottolog 2.7 is taken as the authority for language names and genetic affiliations (except for Rif Tarifit, which has been assigned the name Tarifiyt-Beni-Iznasen-Eastern Middle Atlas). Where I know of a pending update in Glottolog, as is the case with Äiwoo (formerly Ayiwo), I use the updated form. For languages mentioned in the text I use the full name as given in the appendices and index, except in the case of familiar languages such as Eng., Ger., Rus., Jap. etc. Some principles for language names are proposed in Haspelmath (2017a).
Chapter 1. Towards a comparative concept

1 Towards a comparative concept

The primary goal of this work is to present a functional-typological, empirically-based, cross-linguistic account of binominal lexemes. These are lexical items that consist primarily of two nominal constituents and whose function is to name a (complex) concept that involves an unstated (or underspecified) relation between two entities. The most familiar strategy that serves this purpose is the Germanic Noun-Noun Compound, e.g. German eisen.bahn [iron.way], but there are many others, including French Prepositional constructions – chemin de fer [road PREP iron]; Turkish Izafet constructions – demir yol.u [iron road.3SG]; and Russian Relational Adjective constructions – želez.naja doroga [iron.ADIZ road]. All of the above combine the meanings IRON and ROAD/WAY to denote the meaning RAILWAY, but do so using quite different morphosyntactic strategies.

Starting from a set of 100 meanings, I repurpose data from the World Loanword Database (WOLD) – supplemented by data collected specifically for this project – to develop a classification of morphosyntactic strategies (four of them exemplified above), and a two-tiered taxonomy of semantic relations. Both of these are applied to a set of nearly 4,000 binominals (as I call them for short) from 106 languages, in order to reveal linguistic universals. My framework is that of traditional Greenbergian typology, as elaborated by Bill Croft in his theory of Radical Construction Grammar (2001). The annotation of the data, however, follows the principle of framework-free grammar (Haspelmath 2015) – that languages should be described in their own terms and not in terms of aprioristic assumptions – and should therefore be amenable to linguists of all theoretical persuasions.

While the main purpose of the study is to chart the morphosyntactic and semantic diversity of binominal lexemes, a secondary goal is to develop a cross-linguistically valid classification of associative relations that has applicability beyond the immediate scope of binominals, to metonymy, lexical semantics and beyond. As such, I make a contribution to cognitive linguistics as well as linguistic typology.

---

1 A more precise definition of binominal lexeme is developed in §1.2.4.
2 The term ‘meaning’ as used here reflects the usage in WOLD. See §3.1.
1.1 Background

In this introductory chapter I describe the genesis of the present study in personal terms and, at the risk of taking my linguist readers out of their comfort zone, I start in the context of my earlier work in the field of information technology. Hopefully the relevance will soon become apparent; if not, it will definitely do so later.

1.1.1 Topic Maps and associative thought

Before becoming a linguist I devoted ten years of my life to developing, promoting and implementing a radically new approach to information management called Topic Maps (Pepper 2002; 2010a). Topic Maps\(^1\) is based on a simple model that emerged from an attempt to formalize the structure implicit in finding aids such as back-of-book indexes, glossaries and thesauri, all of which involve some form of knowledge representation. (I return to this topic in §8.3.) The core of the Topic Maps model consists of topics, associations and occurrences (hence the title of my 2002 paper, *The TAO of Topic Maps*). TOPICS represent the subjects of interest in the domain covered by the topic map; ASSOCIATIONS represent relationships between those subjects; and OCCURRENCES are a special kind of association that links information about the subject to the topic that represents it.

For example, in the domain of Italian opera, some key subjects are the composer Puccini, his operas *Tosca* and *Madame Butterfly*, and the city Lucca, where he was born, all of which can be represented by topics. Various relationships between these subjects, such as the fact that *Tosca* and *Madame Butterfly* were composed by Puccini, or that the composer was born in Lucca, can be expressed using associations; and information that pertains to these subjects, such as a biography of Puccini, a map of Lucca, or the libretto of *Madame Butterfly*, can be linked to the relevant topics as occurrences (Figure 1).

Topics, associations and occurrences can all be classified by type: Puccini can be assigned to the type ‘person’ or ‘composer’; the nature of his relationship with *Madame Butterfly* specified as ‘composed by’; information resources characterized as ‘biography’, ‘libretto’, ‘map’, and so on. The concepts TOPIC TYPE, ASSOCIATION TYPE and OCCURRENCE TYPE are all part of the core Topic Maps model (and incidentally, they are also topics).

\(^1\) It is the convention to use initial capitals to refer to the technology itself or the ISO specification (in the singular; hence, “Topic Maps is”), and all lower case when referring to the document-like artefacts, a kind of semantic map, that the standard describes (hence, “topic maps are”).
The relationship between a topic and its type is actually a built-in association type (‘instance of’), which is privileged in the model because of its ubiquity and importance in knowledge modelling. Another predefined association type, ‘subtype of’, represents the relationship between types at different levels of schematicity, such as those between the topic types ‘opera’→‘work of art’→‘product’, or between the association types ‘composed by’→‘created by’→‘produced by’ (Figure 2). Notice that the kind (or type) of role played by a topic in an association (here, ‘work’ and ‘composer’) can also be specified explicitly. There is more to the model, including facilities for handling context, naming and identification, but these need not concern us here.
In my work with Topic Maps I was continually struck by parallels with natural language. On reflection, this should not come as any surprise. After all, Topic Maps is a way of representing human knowledge, and natural language – in addition to its others functions – is also a form of knowledge representation. I often wondered how the one might inform our understanding of the other, and in particular, how an understanding of language might inform the ways in which we use Topic Maps, and the further development of the standard. Viewing computer-oriented models such as this from the perspective of language seemed to me a much more exciting and worthwhile endeavour than the mainstream approach of viewing language from a computational perspective.

Some of the parallels are obvious. Topics are like nouns in that they prototypically denote objects or ‘things’, while associations are like verbs in that they represent various kinds of relationship; associations of different arities (unary, binary, ternary) resemble clauses of different valencies (intransitive, transitive, ditransitive); role types correspond (albeit at a finer level of granularity) to semantic roles (agent, patient, etc.); the ability to view and traverse an association from different directions is reminiscent of profiling in active and passive constructions; the ability to reify associations (and treat them as topics) is analogous to nominalization; and so it goes on.

The Topic Maps model turned out to be extremely intuitive and very easy for users to understand. I believe the reason for this is because it reflects the way people think. This was eloquently expressed by Vannevar Bush in 1945 in his seminal paper, *As we may think:*

> The human mind…operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature (Bush 1945).

Bush’s paper drew attention to the importance of associative relations in the field of information management, where hierarchical classification had hitherto ruled the roost. It inspired much subsequent work on hypertext, including the ideas of Ted Nelson (who coined the term hypertext), Doug Engelbart (who implemented it in Augment) and Bill Atkinson (developer of Apple’s Hypercard application), and eventually it played a central role in Tim Berners Lee’s invention of the World Wide Web (Pepper 2007). The models that underlie today’s cutting-edge semantic
technologies – such as RDF (Shadbolt & Gibbins 2010) and Topic Maps (Pepper 2010a) – are based on associative relations between entities and directly or indirectly inspired by Bush and his reflections on how the mind works.\(^1\)

So what does all of this have to do with binominals? Well, it explains my interest in compounding, in particular noun-noun compounding, as I explain below, and it plays an important role in the discussion of associative relations in Chapter 8.

### 1.1.2 Nominal compounding in Nizaa

Fast forward to 2010. At the ripe old age of 57, Steve has finally figured out what to be when he grows up: he wants to be a linguist. Inspired by his encounter with Rolf Theil, a professor of linguistics at the University of Oslo with an extraordinary ability to infect his students with the passion he has for his subject, Steve has finished a BA and is now casting around for a suitable thesis topic for his MA in Language Documentation and Description at the School of Oriental and African Studies in London. Given the focus of the course, he decides to write about a “lesser-studied” language and approaches Rolf with the idea of using the latter’s unpublished field notes on the Cameroonian language Nizaa, collected during the 1980s. Rolf agrees and suggests a list of possible topics, including one which immediately resonates: nominal compounds in Nizaa. “There are quite a few,” writes Rolf, “and the weird thing is that there are both head first and head last compounds” (p.c. 2010-03-25, my translation). The presence of both left- and right-headed compounds is very unusual cross-linguistically and thus deserving of study.\(^2\)

It struck me that nominal compounds have something in common with Topic Maps. A noun denotes a thing, which would be represented in a topic map by a topic (like the topics ‘person’ and ‘opera’ in Figure 2). Furthermore, there is a relationship between the two constituents of a compound that resembles an association (such

---

\(^{1}\) This despite my claim (Pepper 2008) that some of Bush’s ideas led people up the garden path.

\(^{2}\) The other topics suggested by Rolf, for the benefit of anyone on the look-out for an MA topic, were:

1. **Verbal inflection in Nizaa** The language has primarily aspectual categories, plus something that resembles free/conjunct in Fulfulde.
2. **Verbal derivation in Nizaa** Many of the same categories as in Fulfulde, and then some.
3. **Noun inflection in Nizaa** The language has primarily basic form, locative, plural and something that is either “definite” or “specific”. The latter will likely be the most challenging; it is expressed through a low tone on the end of the word.
4. **Word order in Nizaa** Lots of fun to be had here. The basic structure is SVO, but with SOV when the verb is negative, and moreover with SxOV (x = aux).
5. **Adjectives in Nizaa** Like many other languages it doesn’t have that many, but it is interesting to study the kind of semantic domains they cover.
as ‘composed by’) between two topics. But that relationship is unstated and therefore implicit; it is therefore more like a generic, untyped association: the “see also” relation in a back-of-book index, or the “related term” (RT) relation in a thesaurus. The research question I posed was: Could an understanding of the nature of the relationships inherent in Nizaa compounds help explain the presence of two different compounding strategies in Nizaa?

It turned out that it could. Rolf’s handwritten word lists yielded over 500 likely compounds, 200 or so of them noun-noun compounds, with a 7:5 split between head-initial and head-final, thus confirming the two original claims. An analysis of the semantic relations then led to the striking discovery that while left-headed compounds exhibit one set of relations, right-headed compounds exhibit another and completely orthogonal set of relations (see Pepper 2010b: 41; 2016: 300). In all, 15 different kinds of semantic relation were found among left-headed compounds and seven among right-headed compounds, but none of these relations occurred across both types of compound. In other words, left-headed compounds are built from a completely different set of semantic relations than right-headed compounds. The findings can be summarized as follows:

- In right-headed compounds, relations labelled PART, KIN and POSSESSION predominate, e.g. cam buu [finger head] ‘fingertip’ (part-whole) and dàáŋ njew [horse iron] ‘bit’ (possession).
- In left-headed compounds, there is a greater range of relations, many of which are more-or-less attributive, including LOCATION, RESEMBLANCE, PURPOSE, OCCUPATION and MATERIAL, e.g. nii cún [person tree] ‘carpenter’ (purpose or occupation) and cam ȳunnam [finger child: DIM] ‘little finger’ (resemblance).

An analysis of these results within the framework of Cognitive Grammar led to the postulation of my “two-paths hypothesis”: namely that the two types of compound in Nizaa reflect two quite different “paths of mental access” to the target concept: the one, in right-headed compounds, via a related concept, exploits what Langacker (1993) calls our “reference point ability”; the other, in left-headed compounds, via a superordinate concept, employs our general cognitive ability to categorize and sub-categorize. Thus FINGERTIP is conceptualized via the more salient concept of FINGER using a relation of contiguity, while CARPENTER is conceptualized as a sub-type of the more salient concept of PERSON, suitably restricted by reference to the material used to carry out the profession.

In Pepper (2010b:51) I also hypothesized that “both compounding strategies are in fact employed by most – if not all – languages; they just do not usually surface
quite so clearly [in the grammar] as in Nizaa”, and furthermore that “the same
duality probably … underlies the widely accepted distinction between subordinate
and attributive compounds” proposed by Bisetto & Scalise (2005) (see Pepper
2016 for further development of the latter idea).

For a naïve MA student, this seemed like a major discovery; something that was
worth following up in a doctoral project. For, if the study of compounding in a
single, little known African language could reveal such an insight, what might not a
large-scale cross-linguistic study of compounding bring to light?

1.2 Binominal lexemes as a comparative concept

1.2.1 The limitations of compounding

Such was the genesis of the present project, whose initial goals were “to document
the cross-linguistic diversity of phenomena in nominal compounding, to test exist-
ing hypotheses regarding universals of compounding, and where possible to pro-
pose new generalizations” (from the original project description). Phenomena to
be studied included “formal marking, head position (and its correlation with con-
stituent order in the clause and noun phrase) and semantic relations.”

Despite the enormous interest in compounding over recent decades, culminating
in the publication of the Oxford Handbook of Compounding (Lieber & Štekauer
2009a), very little typological work had been done. The one notable exception is
Bauer’s (2001) study of a genealogically and areally well-balanced sample of 36
languages (§2.1.1). Therefore, as a journeyman piece, and to test the project idea,
my project plan called for a pilot study to replicate Bauer’s work. This was in the
spirit of the “re-doing typology” debate in Linguistic Typology 10(1), except that I
chose to focus on nominal compounding rather than compounding in general. The
reason for this departure from Bauer’s design was a feeling that neither his paper
nor other cross-linguistic studies of compounding, specifically Guevara & Scalise
(2009) and Štekauer, Valera & Körtvélyessy (2012), had lived up to their potential
in terms of producing new and compelling insights. My preliminary diagnosis for
this was that the attempt to cover the whole gamut of compounding had obscured
some really interesting cross-linguistic patterns (see §2.1). Restricting the object
of study to noun-noun compounds (or more precisely, determinative noun-noun
compounds, see below) might be more fruitful, and would also fit better with the
Topic Maps-inspired approach I was thinking of taking.
The pilot study provided interesting insights into the issues involved in replicating a typological survey, but it also revealed the need to define the object of study in terms of a semantically or functionally defined ‘comparative concept’. It was apparent that the consequence of not doing so would be to restrict the scope of the investigation to something formally precise and uncontroversial but rather innocuous and uninteresting (e.g. “root compounds”), and/or risk having to continually defend a very broad notion of compounding against those who will argue that most of my examples are not compounds at all but something else. In other words, a new approach was required.

1.2.2 The potential of binominal lexemes

My intention had all along been to conduct a typological study in line with the kind of best practices advocated by functionalists like Matthew Dryer and Martin Haspelmath. One such best practice is to start out from a comparative concept that is not based on formal, language-specific descriptive categories (Haspelmath 2010). My starting point was noun-noun compounds, but it is well-known that compounding is notoriously hard to define, especially cross-linguistically (see inter alia Bauer 2001; Lieber & Štekauer 2009b; Bauer 2017). Linguists disagree profoundly on what is and what is not a compound. An extreme example of this was Paolo Ramat’s statement in his opening keynote at the Word-Formation Theories II conference in Košice, Slovakia in June 2015 that German Regierungschef ‘head of government’ is not a “true” compound because it contains a linking element, -s-. 1 Other linguists dispute whether a construction involving prepositions, such as French chemin de fer, is a compound (some might admit them as compound phrases or prepositional compounds, cf. Bauer 2001: 705), and no-one – to my knowledge – has ever entertained the idea that Russian želez. naja doroga might be a compound.

Now, what interests me, as is surely apparent by now, is the way in which speakers bring together two nominal concepts in order to name a new concept. That being the case, a lexical unit like chemin de fer is just as interesting as a “true” compound like English railway, German Eisenbahn and Norwegian jernbane. Furthermore, English solar energy and Czech sluneční energie [sun:ADJZ energy] are just as worthy of investigation as Norwegian solenergi [sun:energy] and German Sonnenenergie [sun:LE:energy], since they involve the same relation (usually denoted ɍ) – in this case “from” or “produced by” – between the same two concepts – SUN and

---

1 Paolo has since informed me that he was being deliberately provocative, but the point stands.
Chapter 1. Towards a comparative concept

ENERGY – to denote the same target concept – SOLAR ENERGY, and thus presumably involve the same underlying associative processes. All of these examples can be reduced to three basic constructions: **NN, N PREP N** and **N ADJZ N**. What they have in common (over and above their function as naming units) is that their major constituents represent two nominal concepts (RAIL, IRON, WAY; SUN, ENERGY; GOVERNMENT, CHIEF) and that 9 is unstated (or underspecified). This solves the above-mentioned definitional impasse by pointing towards a cognitive-functional comparative concept. If we ask ourselves, what is the primary function of noun-noun compounds, the answer seems to be: to provide generic names for complex concepts, utilizing the names of two existing concepts, between which there is an implicit, but unstated, relation. Noun-noun compounding can thus be characterized as a *binominal naming strategy* and my project becomes, in informal terms, a cross-linguistic study of noun-noun compounds and their functional equivalents, with my comparative concept ‘binominal lexeme’ (or just ‘binominal’), provisionally defined as:

(1) **binominal lexeme** (provisional)

*a lexical item that consists primarily of two nominal constituents and whose function is to name a complex concept that involves an unstated (or underspecified) relation between two entities*

The term ‘lexeme’ is used in the sense of a lexical item that has a naming, rather than a descriptive function, cf. Booij (2009). Štekauer (1998: 165, fn.2) prefers the term ‘naming unit’ for what is essentially the same thing:

I consistently use the term *naming unit* when referring to units generated within my approach to word-formation. This term was first suggested by V. Mathesius (1975). In my approach, it substitutes for terms like *word, lexeme, lexical unit*, etc. because of their inconsistent use and varying connotations in linguistic literature.

However, the term ‘binominal naming unit’ is rather unwieldy, and so I have opted for binominal lexeme instead.¹

The definition above is a good first approximation, but several issues remain. It had all along been my intention to exclude coordinative compounds, such as Hmong Daw *zaub-mov* [vegetable-rice] FOOD or Vietnamese *bô mẹ* [father mother] PARENTS from my study, since their typology has been described by Wälchli (2005). This is accomplished by clarifying that the relation should not be what Koch (2001: 1144) describes as a relation of “co-taxonomic similarity between subordinate concepts of

¹ Unlike Aronoff (1976:xi), I had no personal reason to avoid the term ‘lexeme’.
the same superordinate concept”; that is, by specifying that my research topic is restricted to *determinative* noun-noun compounds and their functional equivalents. Another, more serious, problem is that I also want to exclude synthetic compounds like *truck driver*, in which the head is a deverbal noun. There are two reasons for this. First of all, such constructions are inherently less interesting in terms of their semantic relations, because the relation between the two nominal constituents is stated explicitly: a truck driver is an agent who drives trucks; they correspond to the typed association ‘composed by’ in Figure 2. Secondly, there is evidence from previous work that the presence of a verbal element may involve a different set of properties, related to argument structure; this, again, would complicate the typology unnecessarily.¹

1.2.3 An onomasiological perspective

An answer to the dilemma concerning synthetic compounds is to be found via Pavol Štekauer’s (1998) classification of “onomasiological types”. The onomasiological approach to linguistics in the field of word-formation was pioneered within the Prague school of linguistics by Miloš Dokulil (1962; 1994).² Its purpose, according to Štekauer, is to reveal “how cognitively grounded categories are linguistically represented through the word-formation processes” (Štekauer, Valera & Körtvélyessy 2012: 237). In considering the product of word-formation, Štekauer (1998: 10) discerns five Onomasiological Types of naming unit, based on the presence or absence of the ‘onomasiological mark’, and the status of the latter. The ‘onomasiological base’ (B) is essentially equivalent to the semantic head and is assumed always to be present. The mark is a conceptual modifier that can be either simple (M) or complex; the latter consists of a determined (or actional) constituent (MA) and a determining (qualifying) constituent (MQ).

The five types can be briefly characterized as follows (see also Figure 3):³

---

¹ The evidence includes Levi’s (1978) recourse to a separate syntactic process in order to account for compounds whose head is a ‘nominalization’; Jackendoff’s (2009) need for two different compound schemata for noun-noun compounds (the argument schema and the modifier schema); and Tratz & Hovy’s (2010) experience with “significant overlap” between their PURPOSE and OBJECT categories (and their consequent decision to remove the latter from their taxonomy).

² Unfortunately, much of the literature is in Czech or Slovak and inaccessible to many linguists.

³ Štekauer has since extended this model, first with a sixth type (Körtvélyessy, Štekauer & Zimmermann 2015) and then to an eight-type model (Štekauer 2016). I have argued (Pepper 2018) that these changes are inconsistent and that they destroy backwards compatibility unnecessarily. However, the first three types, which are most relevant to the present discussion, are the same in all three versions of the model, so I refrain from discussing the matter further here.
OT1 all three constituents are present in the naming unit → M_Q+M_A+B  
  e.g. truck driver < TRUCK_Q+DRIVE_A+AGENT_B;  

OT2 the determining element of the mark M_Q is omitted → M_A+B  
  e.g. driver < Ø_Q+DRIVE_A+AGENT_B;  

OT3 the determined element M_A is omitted → M_Q+B  
  e.g. trucker < TRUCK_Q+Ø_A+AGENT_B;  

OT4 the mark is simple → M+B  
  e.g. blackbird < BLACK_M+BIRD_B;  

OT5 no mark; the absence of onomasiological structure  
  e.g. time < TIME_NOUN

Figure 3: The five basic onomasiological types  
(after Štekauer 1998)

My comparative concept of binominal lexeme is thus identical to Onomasiological Type 3: binominals are complex naming units consisting of an onomasiological base and the determining element of the onomasiological mark, but without the determined, i.e. actional, element (hence the unspecified nature of the semantic relation). Adopting this onomasiological perspective has a number of important consequences. Firstly, synthetic compounds are ruled out of scope. Because of the presence of the actional element (DRIVE), they are Onomasiological Type 1, not Type 3. The onomasiological perspective thus provides a theoretical underpinning and further justification for the decision to exclude such compounds from the study. Secondly, as derivational affixes and lexical roots are accorded the same status in the onomasiological model, nouns derived from other nouns, such as Slovak železnica [iron.ADJZ.NMLZ] RAILWAY, must be included. So, too, must noun classifier constructions, such as Bora tíu.řeju [nose.CM(hole)] NOSTRIL. This fits nicely with the constructionist view underlying my choice of research topic, in that it opens up the
possibility of investigating some aspects of the syntax-morphology-lexicon continuum. And thirdly, colour terms and other words representing qualities are out of scope when they function as the onomasiological mark (as in blackbird), since such constructions are of Type 4. However, when they represent the base, as in Takia patun kdabogan [egg:3SG yellow:3SG] YOLK, lit. “yellow of egg”, they are in scope (see further page 105 ff).

1.2.4 Defining the object of study

Štekauer’s model of onomasiological types provides a satisfying motivation for regarding binominal lexemes as a cross-linguistic category, and for justifying the exclusion of synthetic compounds and the inclusion of both denominal derivations and classifier constructions. Unfortunately, though, this model is not widely known or generally subscribed to, so there is a need to define my object of study in more theoretically neutral terms. This can be accomplished by refining the provisional definition given in (1) in such a way as to render synthetic compounds out of scope and denominal derivations and classifier constructions in scope.

The first objective could be achieved by adopting Haspelmath’s (2012) term ‘thing-root’, defined as “a root that denotes a physical object (animate or inanimate)”, in place of ‘nominal constituent’. This obviously excludes driver, since DRIVE is an ‘action-root’, “a root that denotes a volitional action” in Haspelmath’s terms. However, the term ‘thing-root’ does not include nominalizing affixes like the -ica in železnica. In order to include these, I require a subdivision of affixes parallel to Haspelmath’s subdivision of roots into thing-root, action-root and property-root. I therefore propose the terms ‘thing-affix’, ‘action-affix’ and ‘property-affix’, and define the first of these as “an affix that denotes a physical object (animate or inanimate)”. Then, since roots and affixes are both morphs (i.e. minimal linguistic forms, Haspelmath 2020), I propose the superordinate concept of ‘thing-morph’ to cover both (2).

(2) **thing-morph**

*a morph that denotes a thing (prototypically a physical object, animate or inanimate)*

The definition in (2) differs significantly from Haspelmath’s definition of thing-root (over and above replacing root with morph) in that it allows for the inclusion of non-prototypical thing-morphs that profile more abstract entities than physical objects.
Having thus introduced the notion of thing-morph, I can now amend the provisional definition in (1) in such a way that nouns derived from other nouns (denominal derivation) and noun classifier constructions come within its scope, and synthetic compounds (and other forms involving an actional element) are excluded (3).

(3) **binominal lexeme** (final)

*a lexical item that consists primarily of two thing-morphs and whose function is to name a complex concept that involves an unstated (or under-specified) relation between two entities*

The word ‘primarily’ makes it clear that additional morphological material may be present, provided that its function is grammatical. The functional part of the definition is actually redundant, since there will always be some kind of relation between the entities profiled by the thing-morphs in such a lexical item. However, the additional clarification does no harm, and serves to make the underlying concept clearer, so I choose to leave it in. It also serves to direct attention to the semantic relation, which will become a major concern from Chapter 6 onwards.

Henceforth the term ‘binominal’ will be used as a shorthand for binominal lexeme, and the term ‘binominal construction’ will refer to schemas that are instantiated by individual binominals, such as **Mod Head** for typical Germanic “root” compounds and **Head PREP Mod** for Romance “prepositional compounds”. Constructions that do not name a generic concept are not covered, even if the term used by other linguists includes the word ‘binominal’ (4); see also **Masini (2016)**.

(4) a. **binominal quantifier constructions**, such as Sp. *un montón de amigas* ‘a heap of friends’ (Verveckken 2015)

b. **expressive binominal NPs**, like *an angel of a child* (Foolen 2004)

c. **type binominals**, such as Fr. *une espèce de baleine* ‘a kind of whale’ (Mihatsch 2016)

Informally, I describe binominals as **noun-noun compounds and their functional equivalents**. The examples in (5), all of which mean RAILWAY unless otherwise stated, illustrate some of the variety and offers a taste of things to come. As noted above, synthetic compounds are out of scope; so too are NVN constructions such as Vietnamese *bữa ăn sáng* [meal eat morning] BREAKFAST, in which the determined element of the onomasiological mark is also present; contrast this with Kildin Sami *inc.es’pierrk* [morning.ATTR.meal] which has the same nominal constituents but lacks the actional constituent. Also out of scope are compounds composed of a noun plus an adjective (unless the adjective is denominal, as in the case of *železnaja doroga*).
The typology and semantics of binominal lexemes

(5) a. root compounds: German eisen.bahn [iron.track]
   b. compounds with linking elements: Plains Cree piwāpisk.o.mēskanaw
      [iron.CON.road]
   c. prepositional compounds: French chemin de fer [way of iron]
   d. relational compounds: Russian želez.naja doroga [iron.ROAD]
   e. genitival lexemes: Bezhta kil.o.s hino [iron.OBL.ROAD]
   f. construct case lexemes: Hebrew mesila.t barzel [track.STC.iron]
   g. izafet constructions: Turkish demir.yol.u [iron.ROAD.POSS:3SG]
   h. denominal nominalizations: Slovak želez.n.ica [iron.ADJZ.NMLZ]
   i. double-marking: Western Farsi surāx.e bin.i [hole.EZ.nose.ADJZ.NOSTRIL]
   j. classifier constructions: Bora tiū.ēheju [nose.CM(hole)] NOSTRIL

The comparative concept of binominals as used in the present work is novel, but it is not entirely without precedent. It is in some sense present, lurking (so to speak) in the background and waiting to be discovered, in three studies discussed in the next chapter, viz. Levi (1978) on ‘complex nominals’ (§2.3.1), Rainer (2013) on ‘relational adjectives and their competitors’ (§2.3.2), and Bauer & Tarasova (2013) on ‘adnominal nominal modification’ (§2.3.3).

1.3 Theoretical framework

The present study is conducted within the framework of Bill Croft’s elaboration of traditional Greenbergian typology, known as Radical Construction Grammar. This framework proceeds from three basic assumptions regarding morphosyntax:

The first of these is that the proper unit for grammatical analysis is a (morphosyntactic) construction… The second assumption is that one must always investigate a construction with respect to how its morphosyntactic form expresses its function, which in our analysis includes both meaning and information packaging. These first two assumptions are shared by construction grammar… and the second assumption is characteristic of functionalist theories of grammatical structure… The third assumption is that one must always examine how the morphosyntactic expression of a function varies across languages. The third assumption, combined with the first two, is the hallmark of linguistic typology (Croft forthc.).

In describing the function of constructions, Croft advocates separating semantic content from what he calls ‘information packaging’. Semantic content is described in terms of three basic semantic classes: objects, properties and actions; information packaging is organized around the following skeletal structure:
• **reference** – what the speaker is talking about

• **predication** – what the speaker is asserting about the referents in a particular utterance

• **modification** – additional information provided about the referent.

<table>
<thead>
<tr>
<th>Semantic class</th>
<th>Propositional act</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td>modification</td>
</tr>
<tr>
<td>UNMARKED NOUNS</td>
<td>genitive, adjectivizations, PP’s on nouns</td>
</tr>
<tr>
<td>property</td>
<td>deadjectival nouns</td>
</tr>
<tr>
<td>action</td>
<td>action nominals, complements, infinitives, gerunds</td>
</tr>
</tbody>
</table>

*Table 1: Croft’s grid of basic cross-linguistic constructions*

Since all three semantic classes can refer, modify or predicate, a 3×3 grid of basic cross-linguistic constructions is obtained (Table 1); cf. Croft (1991: 67; 2001: 88; 2003: 185; forthc.: 13, 29). In this model,

The nominal modifier construction (cxn) expresses modification with an object concept. The most common type of nominal attributive phrase is the possessive or genitive phrase as in *The boy’s bicycle*. English uses a distinct construction with the clitic -’s (p. 40).

Binominals are a special kind of nominal modifier construction in which a process of lexicalization is underway (and which may proceed as far as univerbation). In terms of Koptjevskaja-Tamm’s distinction between anchoring and nonanchoring relations in adnominal possession, discussed in §7.2.1, binominals correspond to the nonanchoring type. Again, though, they are situated towards the lexical end of the syntax-morphology-lexicon continuum. They are therefore ideal for exploring this continuum: as naming units they are all part of the lexicon, but while some (such as *chemin de fer*) are ‘syntactic’ or ‘phrasal’ in nature, some (like *železnica*) are ‘morphological’, while others (*eisenbahn*) are “problematic” (Jackendoff 2009) in a theory with a strict division between lexicon and grammar.

The present study is firmly situated within traditional Greenbergian typology, but it does not belong to either morphological or syntactic typology as traditionally understood. It belongs more properly to lexical typology, despite the fact that it is
broader in many respects than most studies within that subfield. As Kibrik (2012) points out, the latter “usually focuses on rather restricted and specialized domains such as color terms, kinship terms, body part terms, or motion-in-water verbs”. In his own work Kibrik aims to pose more general questions, and the paper cited here proposes “an approach to profiling the verbal lexical system of a language in its entirety” (p. 496). I do not presume to suggest that my study does the same with respect to the nominal lexical system, but perhaps it is a contribution to such a goal.

1.4 Design of the study

1.4.1 An empirical, data-driven approach

The experience of replicating Bauer (2001) through the pilot study mentioned in §1.2.1 made it clear that descriptive grammars would not be the best source for the kind of data needed for a broad cross-linguistic study of binominals. Most grammars cover compounding in one way or another, albeit often briefly and with few examples, as witness the experiences reported by Bauer (page 24) and Guevara et al (page 34). But very few grammars make specific reference to other kinds of binominal word-formation, let alone describe them in any detail or discuss how they compete with one another within the language in question. It is as if the functional equivalents of compounding fall between two stools: they belong neither to ‘morphology’ (and its subdomain, word-formation), because they have a phrasal aspect, nor to ‘syntax’, because they are lexical. Grammars would do well to start including a separate chapter on the lexicon and the strategies by which it is enriched.

The seeds of an alternative approach to the use of grammars were sown for me by Pierre Arnaud’s (2004a) study, which compares compounding in 13 languages by first establishing a list of 29 concepts (or meanings) and then investigating how these are named. That such an onomasiological approach could work on a larger scale was confirmed by Matthias Urban’s (2012) dissertation, and when I then came across the World Loanword Database (Haspelmath & Tadmor 2009a), I was presented with both a principled method of constructing a list of meanings (§3.1) and a way to kick-start my data collection (§3.3). All of the above-mentioned studies are described more fully in Chapter 2, along with seminal studies of compounding (§2.1) and word-formation (§2.2) in cross-linguistic perspective. Taken together, these works inspired the design of the present study, which is a detailed analysis of binominal lexemes representing 100 meanings across 106 languages, based on 10,754 data points. Details of how the meanings and languages were selected, how the data was gathered, and how it was analysed are given in Chapter 3.
1.4.2 Research questions

Since binominals as such have not previously been identified as an object of study, it was only possible at the outset to formulate very general research questions, which reflect the exploratory, data-driven nature of the present study:

- What is the extent and diversity of binominal word-formation in the world’s languages? In other words, what are the functional equivalents of noun-noun compounds in the world’s languages?
- How can binominals be classified typologically, in terms of morphosyntactic structure and semantic relations?
- What generalizations can be made and how can these be explained?
- How do the preference patterns exhibited by individual languages correlate with areal, genetic and typological features?
- How do binominal strategies relate to strategies for expressing attributive possession?

More precise research questions emerged from the data as the study progressed and are elaborated in Chapters 7 and 8.

1.5 Structure of this work

Croft (2003: 2) offers three linguistic definitions of typology that correspond to “the three stages of any empirical scientific analysis”, viz. classification (based on observations of empirical phenomena), generalization (in this case, the formulation of language universals) and functional-typological explanation. These are expanded by Song (2007: 9) into five stages of “doing typology” as follows:

(a) identification of a phenomenon to be investigated
(b) generation of a language sample
(c) creation of a typological classification
(d) formulation of a typological generalization
(e) explanation of the typological generalization

While this scheme is slightly artificial, in the sense that scientific research does not proceed in such discrete steps, I largely follow this scheme in the structure of the present work. The first stage has been covered in this chapter and will be further elucidated in the literature review that follows (Chapter 2, Earlier work). There I discuss work that influenced my choice of comparative concept and research design. This includes cross-linguistic studies of word-formation (especially compounding), studies that prefigure the concept of binominals, and large-scale, typological studies.
that take a broadly onomasiological approach. (Other work that pertains to more specific topics is discussed in the appropriate context).

Chapter 3, *Meanings, languages and data*, covers the second of Song’s five stages. I start out by treating in some detail the important methodological issue of how I selected the meanings and to what extent they can be considered representative (§3.1). I then describe and evaluate the language sample (§3.2), my sources of data (open database, questionnaires, dictionaries and grammars) and the challenges posed by each of them (§3.3). The final section covers the preliminary data annotation (“coding”) that preceded the development of the typological classification (§4).

In Chapter 5, *Typological classification*, I present a classification of morphosyntactic strategies (Song’s third stage), following a discussion of a number of theoretical prerequisites in which I have recourse to the work of Koptjevskaja-Tamm and Croft. Among other things, I raise issues associated with the use of hierarchical classifications and put forward an alternative approach, using a two-dimensional grid. I also consider the issue of gradience, including how it can be captured in a visual representation. Finally I present statistics regarding the distribution of the nine basic strategies that I have identified in the data.

In Chapter 6, *Semantic relations*, I develop a second classification, based on the unstated (or underspecified) relation between the two nominal constituents of a binominal. Again I start out with theoretical prerequisites, this time invoking Bauer & Tarasova and Janda. After reviewing the literature and having a rant, I make a point of not reinventing the wheel: instead I reuse two pre-existing systems: Hatcher’s high-level (schematic) system of four relations, and Bourque’s low-level system of 25 relations. I suggest minor amendments to Bourque and a significant extension to Hatcher, and then follow Arnaud’s example by integrating the two into a single whole: the Hatcher-Bourque classification. The chapter concludes with statistics regarding the distribution of semantic relations in the data.

In Chapter 7, *Typological generalizations*, my goal is to formulate generalizations of the kind appropriate to Song’s fourth stage. Three topics are considered: word order (§7.1), the relationship between possessives and binominals (§7.2), and the hypothesis of a correlation between binominal strategies and semantic relations (§7.3).

In Chapter 8, *Conceptual generalizations*, I depart temporarily from traditional Greenbergian typology in order to investigate whether the data can provide insights into matters more related to conceptual organisation. First I look for evidence to
support my two-paths hypothesis (§8.1) and then present two dichotomies and a cline that are suggested by the data: head-framing versus modifier-framing (§8.2), relational and sortal nouns (§8.3) and species-framing versus attribute-framing (§8.4).

In Chapter 9, *A model of associative relations*, I probe deeper into the realm of Cognitive Linguistics in an attempt to develop an overarching model of associative relations that can encompass not just the semantic relations discussed in Chapter 6, but also metonymic relations (§9.1) and cognitive relations as understood in lexical typology (§9.2). I then once again force the linguist reader out of her comfort zone in order to show how Topic Maps can enrich the discussion (§9.3), before proposing an overarching model of associative relations called the PHAB model (§9.4).

Chapter 10, *Conclusion*, offers a brief summary, discusses the contribution to science of the present work, and indicates areas for further research.

The appendices are as follows:

A A list of languages cited (ordered by ISO code for ease of reference), together with the family and genus to which they belong and the area in which they are spoken (page 383).

B A list of every source of data and grammatical information for each of the languages in the sample, ordered by language name (page 387).

C The list of meanings, how they are categorised, and various statistics associated with them (page 393).

D An inventory of possessive and binominal constructions and the strategies they embody, ordered by area, genus and language, with an example of each (page 395).

E The complete binominal data set (page 448).

F A summary of the database structure (page 485).

G The questionnaire sent to contributors (page 487).

H Various tables that were too large for the main body of the work (page 489).

In the spirit of the “re-doing typology” debate in Linguistic Typology 10(1), and in order to encourage reuse and replication, the data and scripts used in this project are all available for free download from the Tromsø Repository of Language and Linguistics, https://dataverse.no/dataverse/trolling.
2 Earlier work

The topic of binominal lexemes as conceived in this study has not previously been investigated as such, and there are no cross-linguistic studies of binominals from either an onomasiological nor any other perspective. The most relevant work, in terms of helping me arrive at my comparative concept and the onomasiological methodology, falls into four categories:

- Cross-linguistic studies of specific types of binominal, in particular noun-noun compounds
- More general cross-linguistic studies of word-formation
- Studies that prefigure the concept of binominals
- Large-scale typological studies that employ an onomasiological approach

In this chapter I discuss each of these in turn, in particular those aspects that helped shape the present work. In §2.1 Compounding I consider Bauer (2001), Arnaud (2004b), Scalise & Bisetto (2009), the Morbo/Comp project directed by Sergio Scalise at the University of Bologna, and Guevara & Scalise (2009). In §2.2 Word-formation I discuss Aikhenvald (2007) and Štekauer, Valera & Körtvélyessy (2012). In §2.3 Prefiguring binominals I present three studies that in one way or another anticipate the concept of binominals without actually recognizing it as a category: Levi (1978), Rainer (2013), and Bauer & Tarasova (2013). Finally, in §2.4 Morphological complexity I describe two studies – Haspelmath & Tadmor (2009) and Urban (2012) – that helped me fine-tune the onomasiological methodology that I wanted to adopt.

Other literature, some of it of crucial importance to my work, will be presented and discussed in later chapters where it is most relevant: in Chapter 5, Typological classification, Koptjevskaja-Tamm’s (2002; 2003) and Croft’s (2003) typologies of possessive constructions; in Chapter 6, Semantic relations, the work of Hatcher (1960), Bauer & Tarasova (2013), Bourque (2014) and Arnaud (2016) on the semantics of compounding, and of Janda (2011) on metonymy in word-formation; in Chapter 7, Typological generalizations, Koptjevskaja-Tamm (2004) on anchoring and non-anchoring relations and Koch’s (2001) idea of motivational grids; and in Chapter 8, Conceptual generalizations, Peirsman & Geeraert’s (2006) inventory of metonymic relations, and Blank’s (2003) work on conceptual associations.
2.1 Compounding

2.1.1 Bauer (2001)

Bauer (2001) is a cross-linguistic survey of compounding based on a genetically and areally diverse sample of 36 languages (see Table 2). The sample comprises six languages from each of Dryer’s (1992) large linguistic areas (A, E, O, G, N, S), with each language belonging to a different genus.\(^1\)

<table>
<thead>
<tr>
<th>Africa (A)</th>
<th>Australia / New Guinea (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hebrew <strong>HEB</strong> (Semitic)(^2)</td>
<td>Yimas <strong>YEE</strong> (Lower Sepik-Ramu)</td>
</tr>
<tr>
<td>Tswana <strong>TSN</strong> (Bantoid)</td>
<td>Kobon <strong>KPW</strong> (Madang)</td>
</tr>
<tr>
<td>Yoruba <strong>YOR</strong> (Defoid)</td>
<td>Siroi <strong>SSD</strong> (Madang)</td>
</tr>
<tr>
<td>Ewe <strong>EWE</strong> (Kwa)</td>
<td>Waskia <strong>WSK</strong> (Madang)</td>
</tr>
<tr>
<td>Turkana <strong>TUV</strong> (Niloctic)</td>
<td>Mara <strong>MEC</strong> (Mangarrayi-Maran)</td>
</tr>
<tr>
<td>Kanuri <strong>KNC</strong> (Saharan)</td>
<td>Arabana <strong>ARD</strong> (Karnic)</td>
</tr>
<tr>
<td></td>
<td><strong>Table 2: Language sample (Bauer 2001)</strong></td>
</tr>
</tbody>
</table>

| Eurasia (E)                         |                                              |
|-------------------------------------|                                              |
| Abkhaz-Adyge **ADY** (NW Caucasian) | Kalaallisut **KAL** (Eskimo-Aleut)          |
| Chukchi **CKT** (Chukotko-Kamchatkan)| Kiowa **KIO** (Kiowa-Tanoan)                |
| Tamil **TAM** (Dravidian)           | Tz’utujil **TZI** (Mayan)                    |
| Danish **DAN** (Germanic)           | Dakota **DAK** (Siouan)                     |
| Basque **EUS** (Basque)             | Takelma **TKM** (Takelma)                   |
| Finnish **FIN** (Finnic)            | Shoshone **SHH** (Northern Uto-Aztecan)      |
| **South America & Oceania (O)**     |                                              |
| Khmer **KHM** (Khmeric)             | Paumarí **PAD** (Arawan)                    |
| Vietnamese **VIE** (Vietic)          | Hixkaryána **HIX** (Parukotoan)             |
| Maori **MRI** (Eastern Malayo-Polynesian) | Cayubaba **CVB** (Cayubaba)              |
| Batak Toba **BBC** (NW Sumatra-Barrier Isl) | Pirahã **MYP** (Pirahã)           |
| Yue Chinese **YUE** (Sinitic)       | Imbabura Highl. Quichua **QVI** (Quechua II) |
| Thai **THA** (Kam-Tai)              | Paraguayan Guaraní **GUG** (Tupi-Guarani)    |

1 According to Glottolog 2.7, three of the languages chosen to represent Australia/New Guinea (Kobon, Siroi and Waskia) are now considered to belong to the same genus (Madang).

2 In Dryer’s classification, Africa includes Semitic languages of southwest Asia (e.g. Hebrew).
Bauer starts out by developing a definition of compound (6) which is only intended to present a “focal notion" of the way in which the term compound is used in the paper, since the author acknowledges that neither phonological, grammatical nor semantic isolation are necessary or sufficient criteria for compoundhood.

(6) **compound**  
*a lexical unit made up of two or more elements, each of which can function as a lexeme independent of the other(s) in other contexts, and which shows some phonological and/or grammatical isolation from normal syntactic usage* (p. 695).

Topics covered by Bauer include compound types, the order of elements, semantic relations, morphological and phonological effects, and how to delimit compounds from other multi-word lexical items, such as lexicalised phrases (Fr. *comme il faut* [as it is_necessary] ‘proper’; Fr. *pomme de terre* [apple of earth] ‘potato’; Eng. *women’s liberation* and *cat’s paw*; and Ger. *Vergiß.mein.nicht* [forget.me.not] ‘forget-me-not’). The discussion of compound types is based on Pāṇini’s classification of Sanskrit compounds:

- **tatpuruṣa** (determinative) compounds in which one element modifies the other;
- **karmadhāraya** compounds are either adjective-noun (e.g. Eng. *black.bird*) or two nouns in apposition (e.g. Eng. *fighter-bomber*);
- **dvandva** (copulative, aggregative, coordinative) compounds “have two or more words in a coordinate relation, such that the entity denoted is the totality of the entities denoted by each of the elements”;
- **bahuvrīhi** (possessive, exocentric) compounds are exemplified by the name of the type, *bahu.vrīhi* [much.rice] ‘a rich person (i.e. someone who owns a lot of rice)’, cf. Eng. *red.head*;
- **avyayībhāva** compounds\(^1\) are mentioned by Bauer for the sake of completeness, but not discussed since the term “is not used by recent scholars”;
- **upapada-samāsa**, (synthetic, verbal, verbal-nexus) compounds.\(^2\)

According to Bauer, the synthetic (or verbal) compound type is “not particularly well-defined”. While it has mostly been discussed with reference to the Germanic languages, such compounds are “much more widespread”. Bauer cites Lieber’s (1994) definition (“compounds whose head elements are derived from verbs”), but

---

\(^1\) Uninflected adverbial compounds (http://learnsanskrit.org/nouns/compounds/avyayibhava)

\(^2\) Bauer does not use the Sanskrit term for this type of compound.
The typology and semantics of binominal lexemes

points to a lack of agreement concerning the kinds of derivation to be included. According to Bauer, “much of the discussion of these compounds in the literature has centred on the fact that the modifying element in the compound is (usually) interpreted as an argument of the verb from which the head element is derived.” This observation alerted me to the fact such compounds may involve a different set of properties than root compounds and prompted me to exclude them from this study (cf. §1.2.2 and §1.2.3).

Under the rubric “morphological effects”, Bauer declares stem juxtaposition to be “the norm” in compounding. Items linked by prepositions (e.g. Fr. chemin de fer [road of iron] ‘railway’) are not considered here. Otherwise the constituents may be linked by “some kind of linking element” (e.g. Khmer yian.ə.thaan [vehicle.LK.place] ‘garage’), or through “some inflectional form of one of the elements” (e.g. Yimas num.n numpran [village.OBL pig] ‘domesticated pig’), and sometimes it may be unclear which of these is involved. Inflectional forms are usually case-markers and the most common are those used for possession, whether by marking the possessor (e.g. Fin. auto.n.ikkuna [car.GEN.window] ‘car window’) or the possessum (e.g. Takelma pliyin sgeh’xabā: [deer its:hat] ‘deerskin hat’). However, other case markers are also found, including nominative, accusative, dative, ablative, instrumental, oblique, adessive and more.

As for phonological effects, in addition to morphophonemic and morphotonemic changes that are “concomitants of the compounding process in languages such as Japanese and Nama”, Bauer provides examples of a number of processes in which phonological material is elided, ranging from the merger of two vowels, to the shortening of the first or even both elements, as in the Hebrew blend rakevel < rakevet + kevel [train cable] ‘cable car’.

In his brief discussion of “meanings in tatpuruṣa compounds”, Bauer laments the lack of information in most descriptions regarding the kinds of semantic relations exhibited by compounds, but says that the available evidence suggests – for some languages at least – that there may not be any finite list of relationships. In his own sample, “underlying semantic relationships of location” appear to be most common (e.g. Eng. furniture store, bone cancer), and the next most frequent type is where the head is made from the material in the modifier (e.g. Eng. sandcastle).

1 Lieber classifies speech-synthesizer as a synthetic compound but not speech synthesis, on the grounds that synthesis is not (overtly) derived from synthesize. In Štekauer’s scheme (§1.2.3), these would be classified as Onomasiological Types 1 and 3, respectively.

2 I return to this issue in Chapter 6, Semantic relations.
The evidence, he concludes, suggests that “compounds may be used prototypically to indicate location or source (especially if ‘made from’, ‘made by’, ‘belonging to’ and ‘coming from’ are all interpreted as sources).”¹

Also of relevance to the present work is Bauer’s discussion of the correlations between the order of head noun and modifier in compounds with the order of (i) noun and adjective, and (ii) noun and possessor. Table 3 shows the results obtained for the noun-adjective comparison.² Bauer comments that it is “not necessarily the case” that the order of head and modifier nouns in a nominal compound reflect the order of noun and adjective, and he observes “a slight preference” for modifier noun + head noun structures (right-headed compounds), independent of the syntactic order of adjective and noun.

<table>
<thead>
<tr>
<th>Word Order</th>
<th>A</th>
<th>E</th>
<th>O</th>
<th>G</th>
<th>N</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Adj &amp; N-Mod</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>N-Adj &amp; Mod-N</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Adj-N &amp; N-Mod</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Adj-N &amp; Mod-N</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>insufficient data</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3: Order of noun-adjective and noun-modifier (Bauer 2001)

What Bauer fails to observe is that his data actually reveal a very common kind of distribution in which “three types are attested and one type is not (or is extremely rare)” (Croft 1990: 56). This becomes very clear if the data are represented in the form of a tetrachoric table, as in Table 4a. From this we can derive the implicational universal \( \text{Adj-N} \supset \text{Mod-N} \) (that is, adjective-noun order implies modifier-head noun order in compounds). Furthermore, it can be concluded that \( \text{Mod-N} \) (i.e. right-headedness) is the dominant order cross-linguistically, and that \( \text{N-Mod} \) (i.e. left-headedness) is the recessive order. The numbers come out slightly differently in my replication of Bauer’s study (mentioned earlier on page 7), which was based on the same sources but restricted to nominal compounds (Pepper 2015), but they still support the same implicational universal (see Table 4b). It is no longer possible to ascertain the reason for the discrepancy in the numbers, since the data points from which Bauer derived his tables are no longer extant (Bauer, p.c.). As for the

¹ My own results, presented in §6.4, suggest a different scale of frequency.
replication study, no unambiguous order of head and modifier could be determined for six languages; three of these (Kanuri, Yue Chinese and Tz'utujil) have both left- and right-headed compounds, and three (Mara, Kalaallisut and Hixkaryána) appear not to have compounds at all. Moreover, in Tz'utujil adjectives may appear either before or after the noun. (The language that exhibits the “extremely rare” combination of adjective-noun and head-modifier orders is Cayubaba.)

<table>
<thead>
<tr>
<th></th>
<th>N-Mod</th>
<th>Mod-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Adj</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Adj-N</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

(a) Bauer 2001

<table>
<thead>
<tr>
<th></th>
<th>N-Mod</th>
<th>Mod-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Adj</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Adj-N</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

(b) Pepper 2015

Table 4: Noun + Adjective tetrachoric tables
(Bauer 2001; Pepper 2015)

Turning to the correlation between the order of head and modifier and the order of possessor and possessum, Bauer observes a “slightly better match” (shown in Table 5). Again, representing the data as a tetrachoric table (Table 6a) reveals a distribution from which it is possible to derive the implicational universal Poss-N \(\Rightarrow\) Mod-N, with Mod-N (right-headedness) again emerging as dominant. In this case, however, only the two harmonic correlations (N-Poss & N-Mod and Poss-N & Mod-N) can really be said to be frequent. And while the Poss-N & N-Mod can be characterized as “extremely rare”, the other disharmonic pattern, N-Poss & Mod-N, is also rather infrequent. The data thus tend toward a biconditional universal of the type Poss-N \(\equiv\) Mod-N.

<table>
<thead>
<tr>
<th>Word Order</th>
<th>A</th>
<th>E</th>
<th>O</th>
<th>G</th>
<th>N</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Poss &amp; N-Mod</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>N-Poss &amp; Mod-N</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Poss-N &amp; N-Mod</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Poss-N &amp; Mod-N</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>insufficient data</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5: Order of noun-possessor and noun-modifier (Bauer 2001)

In my replication study, this tendency turned into an exceptionless pattern (Table 6b). No disharmonic patterns were found at all: either the head is on the left in both compounds and possessive constructions, or it is on the right in both.
Once again, the reasons for this discrepancy between the two studies cannot be ascertained for certain because Bauer’s data are no longer extant. However, a clue can be found in a comment made by Bauer concerning the numbers he arrived at:

> It is not entirely clear how much weight can be attributed to such figures, given the lack of consistency across languages in the ordering of modifier and head in compounds [i.e. that many languages have both head-initial and head-final compounds]. Although it might be expected that this would be fixed in any individual language, that is the case only in about half of my sample from any of the areas used. The figures are given below in [Table 7]. The figures given in this table show inconsistencies across compounds of all word-classes, but even if only noun compounds are considered, there is considerable inconsistency. The figures for nouns alone are parenthesised in [Table 7]. It must be recalled that many languages are consistent because only one pattern of compound is reported (p. 697).

This carries the very strong implication that disharmonic (i.e. mixed) order of head and modifier is the norm and that harmonic ordering is the exception.

<table>
<thead>
<tr>
<th>Word Order</th>
<th>A</th>
<th>E</th>
<th>O</th>
<th>G</th>
<th>N</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consistent ordering</strong></td>
<td>3 (3)</td>
<td>3 (5)</td>
<td>4 (4)</td>
<td>4 (6)</td>
<td>2 (3)</td>
<td>2 (3)</td>
<td>18 (24)</td>
</tr>
<tr>
<td><strong>Inconsistent ordering</strong></td>
<td>3 (3)</td>
<td>3 (1)</td>
<td>2 (2)</td>
<td>2 (0)</td>
<td>3 (3)</td>
<td>3 (2)</td>
<td>16 (11)</td>
</tr>
<tr>
<td><strong>Unclear or missing</strong></td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0)</td>
<td>1 (1)</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

*Table 7: Consistency of head-modifier ordering (Bauer 2001)*

Bauer’s figures can be compared with those obtained in my replication study. In Table 7, the numbers for what Bauer terms ‘consistent ordering’ are split across the rows labelled *ModN* and *NMod* (representing right-headed and left-headed compounds respectively); these are collated in row 3. Since the replication study only investigated nominal compounds, the numbers in Table 8 correspond to those in parentheses in Table 7. The relevant comparison is thus between the numbers in parentheses in both tables (shown in boldface).
The typology and semantics of binominal lexemes

Observe that 30 languages were considered to have consistent ordering in the replication study, as against just 24 in the original; the corresponding numbers for inconsistent ordering are three and eleven. Again, it is no longer possible to trace the reasons for the discrepancy between the results obtained by Bauer and myself from investigating the same languages using the same sources, but certain hints can be obtained from a detailed examination of the two sets of data. Looking at the bottom row in each table, we observe first of all that while I conclude that three languages (Mara, Hixkaryana and Kalaallisut) do not have compounds, Bauer, based on the same sources, concludes that only one of them does not. Then, comparing the numbers for each linguistic area, we can observe that

- only one of the African languages (Kanuri) was found by me to have both right-headed and left-headed compounds, as against three according to Bauer;
- the Eurasian language that Bauer considered to have inconsistent ordering was found to be head-final (ModN) by me;
- only one language from Southeast Asia and Oceania (Yue Chinese) was found by me to have both orderings, as against Bauer’s two;¹
- one of the Australia/New Guinea languages (Mara) was found by me not to have noun compounds, whereas Bauer considers it to have consistent ordering;²

---

¹ It is not unlikely that the second language considered by Bauer to be mixed was Vietnamese, which is sometimes reported to have both left-headed and right-headed compounds. However, every native Vietnamese compound is left-headed; only compounds loaned from Chinese are right-headed. The former predominate and the latter are less transparent for native speakers: “A native speaker may not be aware of the etymology of each element within the [Sino-Vietnamese compound] construction” (Nguyên 1997: 72, 77). Dinh (2002: 150) does not mention right-headed compounds and states that the head noun “is always in the first position.” On this basis I assigned the code NMod to Vietnamese; Bauer’s criteria may have been different.

² Heath (1981) does not explicitly state that noun-noun compounds do not occur in the language, but this is strongly implied by its omission from his discussion of compounds.
• in North America, I found one language (Kalaallisut) did not have compounds, while Bauer is of a different opinion;¹
• only for South America do the two analyses coincide.

The most important conclusion to be drawn from this (apart from the need to make one’s data available for future researchers) is that different types of compound should be treated separately in order for patterns to emerge clearly. In summary, Bauer (2001) provides a good overview of the many issues involved in the study of compounding, but the study does not lead to any new insights. This is probably because compounding as a whole is too heterogeneous, in which case a study that focuses on nominal compounds only (or perhaps just noun-noun compounds, or even determinative noun-noun compounds) might bear more interesting typological fruit. It could also be because Bauer’s study was merely a limited pilot that did not offer the scope for investigating specific issues (such as semantic relations and word order correlations) in more depth.

2.1.2 Arnaud (2004)

Arnaud (2004a) is an edited collection of studies of compounding in the sixteen languages listed in Table 9. In a short concluding chapter, entitled Problématique du nom composé, Arnaud discusses a range of general issues, including those of definition, ambiguity, headedness, demarcation, semantic relations, prosody and borrowing, many of them barely touched on by Bauer, before finally presenting a short onomasiological study in which 29 meanings are examined across 13 languages in order to assess the extent of compounding in each language. It was this study that first gave me the idea of applying the onomasiological method in my own research. Arnaud describes the method as follows:

Pour comparer les langues, on peut, dans le sens onomasiologique, établir une liste de concepts et voir comment ceux-ci sont dénommés. Il s’agit évidemment d’établir une liste qui réduise les différences culturelles au maximum, c’est-à-dire comportant des concepts de parties du corps, espèces naturelles, phénomènes météorologiques, artefacts répandus … (Arnaud 2004a:347).²

¹ The possibilities for compounding more than one independent lexical stem are “extremely limited” (Fortescue 1984). Sadock (2003) mentions “a few sporadic forms that can be considered compounds”, but both his examples appear to be calques.
² To compare languages, one can, in the onomasiological sense, establish a list of concepts and look at how these are named. It is obviously a question of establishing a list that reduces cultural differences to a minimum, that is to say, comprising concepts for body parts, natural species, meteorological phenomena, widely used artefacts …/
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Africa (A)</th>
<th>Southeast Asia &amp; Oceania (O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Atlas Tamazight TZM (Berber)</td>
<td>Khmer KHM (Khmeric)</td>
</tr>
<tr>
<td>Bambara BAM (Mande)</td>
<td>Kumak NEE (Eastern Malayo-Polynesian)</td>
</tr>
<tr>
<td><strong>Eurasia (E)</strong></td>
<td>Angami Naga NJM (Kuki-Chin-Naga)</td>
</tr>
<tr>
<td>Turkish TUR (Turkic)</td>
<td>Galo ADL (Macro-Tani)</td>
</tr>
<tr>
<td>Basque EUS (Basque)</td>
<td><strong>Australia / New Guinea (G)</strong></td>
</tr>
<tr>
<td>Modern Armenian HYE (Armenian)</td>
<td>Gunwinggu GUP (Gunwinyguan)</td>
</tr>
<tr>
<td>Welsh CYM (Celtic)</td>
<td><strong>North America (N)</strong></td>
</tr>
<tr>
<td>Udi UDI (Lezgic)</td>
<td>Southern East Cree CRJ (Algonquian)</td>
</tr>
<tr>
<td>Hungarian HUN (Hungarian)</td>
<td><strong>South America (S)</strong></td>
</tr>
<tr>
<td><strong>Pidgins &amp; Creoles (P)</strong></td>
<td>Santiago del Estero Quichua QUS (Quechua II)</td>
</tr>
<tr>
<td>Tok Pisin TPI (English-based Creoles)</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Languages covered in Arnaud (2004)

Arnaud’s results are reproduced below as Table 10 and summarized in Figure 4, in which the vertical axis displays the number of compounds per language out of a possible total of 29. Clearly, the extent of compounding varies greatly from one language to another, and this raises the question of what languages that disfavour compounding do instead. In the case of French, the answer is well-known: complex concepts, such as RAILWAY, that are typically expressed through compounding in, say, English and German (and, to judge by Figure 4, Basque, Cambodian and Welsh as well) are often expressed using a prepositional construction, as in chemin de fer [road of iron]. This prompted the central research question of the present work: What are the functional equivalents of noun-noun compounds in the world’s languages?

| +   | composé     | compound      |
| (+) | trés probablement un composé | very probably a compound |
| +’  | exocentrique secondaire | secondary exocentric |
| +,  | il existe une autre dénomination non composée | non-compositional alternative exists |
| −   | donnée non disponible | data not available |
| /   | ne s’applique pas à la culture | not applicable in the culture |
| x   | cranberry compound | cranberry compound |

Legend for Table 10
<table>
<thead>
<tr>
<th>meaning (English)</th>
<th>meaning (French)</th>
<th>ENG</th>
<th>DEU</th>
<th>HYE</th>
<th>EUS</th>
<th>GUP</th>
<th>GRU</th>
<th>FRA</th>
<th>CYM</th>
<th>HUN</th>
<th>KHM</th>
<th>NEE</th>
<th>TZM</th>
<th>TUR</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>skull</td>
<td>CRÂNE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>×</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forehead</td>
<td>FRONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eyebrow</td>
<td>SOURCIL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>×</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eyelash</td>
<td>CIL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(+)</td>
<td>+</td>
<td>(+)</td>
<td>+</td>
<td>+</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pupil</td>
<td>PUPILLE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nostril</td>
<td>NARINE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>canine</td>
<td>CANINE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elbow</td>
<td>COUDE</td>
<td>×</td>
<td>+</td>
<td>(+)</td>
<td></td>
<td>×</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wrist</td>
<td>POIGNET</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thumb</td>
<td>POUCÉ</td>
<td>(+)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knee</td>
<td>GENOU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>calf</td>
<td>MOLLET</td>
<td>,+</td>
<td>+</td>
<td>×</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ankle</td>
<td>CHEVILLE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>toe</td>
<td>ORTEIL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>navel</td>
<td>NOMBRIL</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anus</td>
<td>ANUS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>butterfly</td>
<td>PAPILLON</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dragonfly</td>
<td>LIBELLULE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grasshopper</td>
<td>SAUTERELLE</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>turtle</td>
<td>TORTUE</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bat</td>
<td>CHAUVE-SOURIS</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rainbow</td>
<td>ARC-EN-CIEL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(+)</td>
<td>+</td>
<td>(+)</td>
<td>+</td>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hail</td>
<td>GRÊLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>necklace</td>
<td>COLLIER</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bracelet</td>
<td>BRACELET</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cradle</td>
<td>BERCEAU</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>home</td>
<td>FOYER</td>
<td>,+</td>
<td>,+</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hook</td>
<td>HAMEÇON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quiver</td>
<td>CARQUOIS</td>
<td>+</td>
<td>/</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Onomasiological cross-linguistic comparison (Arnaud 2004)
2.1.3 Morbo/Comp (2004-2006)

Morbo/Comp was an international research project on compounding, devised and directed by Sergio Scalise and based at the University of Bologna from 2004 to 2006. Its aim was to collect compounding data in a standardized manner in order to facilitate cross-linguistic comparison:

A systematic compilation of compounding data allowing interlinguistic comparison does not exist. As a result, every hypothesis proposed so far is descriptively inadequate and language-specific. For instance, data on the degree of endocentricity/exocentricity in the world’s languages is not available yet. There is no reliable source of data describing the different attested types of compounds, the structural complexity of possible compound words, the presence and typology of linking elements, plural formation, distribution of different structures in the world’s languages, whether categorial and semantic head coincide, etc. (Guevara et al. 2006).

The project website at http://morbocomp.sslmit.unibo.it/ (accessed 2017-12-07) lists, among other things, 16 papers dating from 2004 to 2006 and another three in preparazione, but the site appears not to have been updated since late 2006. Papers continued to appear for another four years or so, however, including the two discussed below, along with a special issue of the journal Lingue e linguaggio.

---

1 No publications or presentations are listed in the activities section after that date, and an event scheduled for January 2007 is shown as “forthcoming” (accessed 2018-06-10).
(2/2009) containing papers on compounding in Russian, Chinese, Turkish, Finnish, Swedish, Czech and Portuguese. To judge by the authorship of these papers, the principal participants in the project, besides Scalise himself, were Antonietta Bisetto, Antonella Ceccagno, Antonio Fábregas, Emiliano Guevara and Chiara Melloni. The database was reported by Guevara et al. (2006) to include data from the 25 languages shown in Table 11, and by Guevara & Scalise (2009) to contain around 80,000 compounds from 21 languages.1 The plan was for the data to be published online “soon” (Guevara et al. 2006) but unfortunately, as is so often the case, this never came to pass.

According to the website, the data were taken from the following sources:

- specific studies
- existing corpora (such as e.g. CELEX for Dutch)
- grammars and dictionaries
- competence of native speakers
- Internet
- semi-automated extraction from textual corpora

<table>
<thead>
<tr>
<th>Dutch NLD (Germanic)</th>
<th>Catalan CAT (Romance)</th>
<th>Belarusan BEL (Slavic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English ENG (Germanic)</td>
<td>French FRA (Romance)</td>
<td>Bulgarian BUL (Slavic)</td>
</tr>
<tr>
<td>German DEU (Germanic)</td>
<td>Italian ITA (Romance)</td>
<td>Polish POL (Slavic)</td>
</tr>
<tr>
<td>Norwegian NOR (Germanic)</td>
<td>Latin LAT (Romance)</td>
<td>Russian RUS (Slavic)</td>
</tr>
<tr>
<td>Swedish SWE (Germanic)</td>
<td>Spanish SPA (Romance)</td>
<td>Serbo-Croatian HBS² (Slavic)</td>
</tr>
<tr>
<td>Finnish FIN (Finnic)</td>
<td>Portuguese POR (Romance)</td>
<td>Japanese JPN (Japanese)</td>
</tr>
<tr>
<td>Hungarian HUN (Hungarian)</td>
<td>Basque EUS (Basque)</td>
<td>Korean KOR (Koreanic)</td>
</tr>
<tr>
<td>Turkish TUR (Turkic)</td>
<td>Greek ELL (Greek)</td>
<td>Mandarin CMN (Sinitic)</td>
</tr>
</tbody>
</table>

| Hebrew HEB (Semitic) |

Table 11: Languages represented in Morbo/Comp

The structure of the database is shown in Table 66 (see page 489). The project design called for each compound to be annotated with various properties, including its word class (or “output category”); internal structure (based on the word classes or “input categories” of the constituents); head position; linking element(s); locus of morphological marking; gender of constituents and compound; and English gloss.

---

1 Languages shown in italics in Table 11 are not mentioned in Guevara & Scalise (2009).

2 The ISO 639-3 code point HBS now has the name South-Western Slavic.
In fact only 4,000 or so of the 80,000 compounds were ever analysed in such detail (Sergio Scalise, p.c.). The effort required was simply too much for the resources available and the group “experienced great difficulties in obtaining enough data to achieve an adequate description of compounding phenomena”:

Traditionally, typological surveys are based on written sources: dictionaries and grammars. In this way, a high number of languages, well-balanced from the typological and areal point of view, is relatively easy to achieve. [However, this methodology] proved to be useless to collect compounding data: traditional written sources usually do not include enough examples of the various structural patterns and/or classes. The Morbo/Comp database has relied heavily on native speakers’ work to collect, classify and analyze manually all the represented examples. Unfortunately, this approach turns out to be quite slow and costly (Guevara et al. 2006).

As a result, the database is very far from being genetically and areally balanced, as the authors readily admit: 21 of the 25 languages are spoken in Europe, 17 are Indo-European and all but one of them belong to the same linguistic area, Dryer’s (1992) Eurasia. (The exception, Mandarin Chinese, is also spoken on that continent, but is assigned by Dryer, along with the rest of Sino-Tibetan, to Southeast Asia & Oceania.) This suggests that another approach is required in order to build a database suitable for use in cross-linguistic comparison and provides support for the decision taken in the present study to adopt Arnaud’s onomasiological method.

Despite these limitations, the Morbo/Comp project produced a number of useful results. In the present context the two most important are Scalise and Bisetto’s classification of compounds (§2.1.4) and the investigation into the “universals of compounding” by Guevara and Scalise (§2.1.5), which are discussed in the next two sections.

2.1.4 Scalise & Bisetto (2009)

Scalise and Bisetto’s (2009) paper is a reworking and further refinement of an earlier paper (Bisetto & Scalise 2005) in which the authors address the problem of how to classify compounds. They start by reviewing nine classification schemes, ranging from Bloomfield (1933) and Marchand (1969) to Bauer (2001) and Booij (2005). The authors point out a lack of “interlinguistic homogeneity” in most of these schemes, which they trace to three causes: (a) language-specific terminology, (b) neglect of certain categories (e.g. adjectival compounds), and (c) inconsistent definitional criteria. The latter concerns the unsystematic combination of the criteria “presence of head” and “semantic relation”.
As an alternative, they propose a novel typological classification based on the “very simple” assumption that what is special about compounds is the fact that the two constituents are linked by a grammatical relation which is not overtly expressed (cf. *apron string* vs. *string of the apron*). Therefore, we would like to suggest that the classification of compounds be uniquely and consistently based on this criterion. The possible grammatical relations holding between the two constituents of a compound are basically the relations that hold in syntactic constructions: subordination, coordination and attribution (Bisetto & Scalise 2005: 326).

Accordingly, they propose a three-way top level distinction between subordinate, attributive and coordinate compounds (none of which are defined), with a further subdivision of each into endocentric and exocentric, dependent on “the presence or absence of a head constituent” (Figure 5).

![Figure 5: Classifying compounds (Bisetto & Scalise 2005)](image)

Four years later this model was extended through the introduction of an intermediate level, as shown in Figure 6 (Scalise & Bisetto 2009: 49ff). In the new proposal, subordinate compounds are subdivided into verbal-nexus and ground, ostensibly on the basis of whether or not the head is “deverbal or non-deverbal”.¹ In addition, attributive compounds are split into attributive and appositive, on the basis of whether the non-head is an adjective (or verb) expressing a property of the head directly, or a noun specifying such a property indirectly (“in apposition”) through one of its own properties. As in the 2005 proposal, each of the resulting classes is subdivided into endocentric and exocentric.

¹ The exocentric *lavapiatti* ‘dishwasher’ < ‘wash-v’ + ‘dishes’ should, of course, be classified with *pickpocket* under verbal-nexus, not under ground. I assume this is a printing error.
Two things may be observed about the resulting classifications: Firstly, there is a logical inconsistency in the use of criteria for the second-level divisions that involve the status of a head (deverbal or not) or non-head (adjectival or nominal) when the further subdivision into endocentric and exocentric suggests that there may not be a head. Secondly, the distinction between endocentric and exocentric is really orthogonal to the rest of the classification, in the same way as, say, the output category. Indeed, “exocentricity” is in many cases better explained as a function of metonymy (and/or metaphor), rather than as a property peculiar to compounds (Bauer 2008).

![Figure 6. Classifying compounds (Scalise & Bisetto 2009)](image)

Whatever the merits or demerits of the later classification, the basic tripartite distinction between subordinate, attributive and coordinate compounds, common to both proposals, has been widely adopted, for example by Lieber in her introductory textbook on morphology (Lieber 2010). Lieber characterizes the three compound types as follows:

“In an attributive compound the non-head acts as a modifier of the head. So snail mail is (metaphorically) a kind of mail that moves like a snail, and a windmill is a kind of mill that is activated by wind […] In coordinative compounds, the first element of the compound does not modify the second; instead, the two have equal weight […] In subordinative (sic) compounds one element is interpreted as the argument of the other, usually as its object” (op.cit.)

Thus for Lieber, windmill is a prototypical attributive compound, while for Bisetto and Scalise it is a prototypical subordinate compound, as shown in Figure 6. This discrepancy is unlikely to be due to an unintentional error, as was probably the case
with lavapiatti (see footnote 1 above). In both of their papers, Bisetto and Scalise classify steamboat – another compound with the semantic structure X powered by Y – as a subordinate compound, so we can assume that their classification of windmill was intentional. And since Lieber uses windmill as a prototypical example throughout her chapter on compounding in all three editions of her book, we can safely assume that her characterization of it as an attributive compound is equally deliberate. In short, Lieber considers the powered by relation to be one of attribution, whereas Bisetto and Scalise consider it to be one of subordination. This raises the question: what does “subordination” actually mean in the context of the relation between two nominals? As noted above, Scalise and Bisetto do not provide any definition. In the case of a verb and a nominal its meaning is clear: if the nominal can be regarded as an argument of the verb, then it is subordinate to it. But in what sense is mill subordinate (or not, as the case may be) to wind in windmill? I address this issue in Pepper (2016) in the light of data from Nizaa (see §1.1.2) and suggest that subordination is best understood in terms of the reference point ability (Langacker 1993) and involves one of two different paths of mental access to the target concept (the other being access via a superordinate concept). These issues will come to the fore later in the investigation of semantic relations and conceptual universals (Chapters 6 and 8).

2.1.5 Guevara & Scalise (2009)

The most important attempt from a typological perspective to synthesize the results of the Morbo/Comp project is Guevara and Scalise’s (2009) paper, Searching for universals of compounding. The paper starts by introducing the project and justifying the search for universal properties in compounding. It then discusses the four “important issues for the typological study of compounding”: the definition of compound, the classification of compounds, the position of the head constituent, and the definition of compound type. Existing definitions, such as those of Bauer (quoted above) and Olsen (2000) (“the combining of two free forms or stems to form a new complex word”) are found to be “neither totally satisfactory nor sufficiently comprehensive”. The many definitions of compounding that one finds in the literature are “tightly predetermined by the theoretical choices made by the author(s)

Consequently, one’s views and beliefs regarding the fundamental notions of morphology – and of linguistics in general – are critical in shaping a working definition for compounding. In other terms, one’s conception of hotly debated (and never agreed
The typology and semantics of binominal lexemes

upon) issues such as word, morpheme, stem, root, lexicon, concatenation, etc., will contribute in shaping one’s definition of compound.

It was in order to solve this dilemma that I developed the comparative concept of binominal lexeme. Guevara and Scalise, for their part, simply set the definitional issue to one side and look for general tendencies in the world’s languages. This leads them to identify the presence of a relation $\mathcal{R}$ between the constituents of a compound, which is not explicitly (phonetically) realized, as a key defining feature. The authors then invoke canonical typology (Brown, Chumakina & Corbett 2013) and offer a definition of canonical compound in terms of the following four criteria:

a) syntactic atomicity (no anaphoric relations between an internal constituent of a compound and an external element);

b) lexical integrity;

c) lexical nature of constituents (lexemes, i.e. words, stems or roots), members of one of the major lexical categories;

d) the whole is a member of one of the major lexical categories.

Defined in this way, the canon matches “the most productive compound-types” of well-studied languages (i.e. Germanic, Romance and Chinese). So-called ‘phrasal compounds’,¹ “which diverge from the canonical” in that one of the constituents is syntactic not lexical in nature, are not excluded from the domain of compounding but regarded as a less-canonical subtype.²

Regarding the classification of compounds, Guevara and Scalise reproduce and adopt the 2005 version of the scheme proposed by Bisetto and Scalise (§2.1.4), with its tripartite top-level division into subordinate (SUB), attributive (ATT) and coordinate (CRD), which they contend fits phrasal and neoclassical compounds just as well as ‘normal’ compounds, but again, no definitions are provided.

A number of issues are addressed in the section devoted to the notion of the “head” and the position of the head constituent, including the distinction between what they call the formal head and the semantic head, and the difficulty of determining the head in coordinate compounds like It. studente-lavoratore ‘student-worker’, and verbal nexus compounds with the structure $[V+N]_N$, such as It. rompi.ghiaccio

¹ Exemplified on p. 111 by Eng. floor-of-a-birdcage taste and Nld. lach-of-ik-schiet humor [laugh-or-I-shoot humour].

² Unfortunately, Guevara and Scalise do not go on to develop their four criteria in such a way that a theoretical space of possibilities emerges clearly, as one would in “mainstream” canonical typology. There is therefore no way of determining how closely any specific compound conforms to the canon.
[break-ice] ‘icebreaker’. They conclude (contra Haspelmath 2002) that the former should be regarded as having two heads rather than none. As for the latter, the upshot appears to be that the word has a formal head (since the features of ghiaccio “percolate” to the compound as a whole), but no semantic head (since an icebreaker is neither a kind of break nor a kind of ice). This leads to the following definitions:

An endocentric compound has at least one formal head and at least one semantic head. If an endocentric compound has only one formal head and only one semantic head, then the two must coincide. If a compound has one or more formal heads and no semantic head, it will be considered exocentric. If a compound has one or more semantic heads and no formal head, it will also be considered exocentric.

This differs from the approach taken in the present study, which focuses on the semantic head and takes account of metonymy and metaphor when determining whether or not a compound has a head (see §4.2).

Regarding the position of the head, Guevara and Scalise (rightly) reject the righthand head rule of Williams (1981), which defines the head of a morphologically complex word to be “the righthand member of that word”. They also reject the idea that the position of the head is a parameter that is fixed for any given language (Scalise 1994), and instead “prefer to maintain […] that in every language there is a canonical position of the head, which may be disregarded by certain compound-types”. The notion of ‘compound type’ is then defined in terms of four properties: output category (e.g. N), structure (e.g. N+N), classification (e.g. SUB) and position of the head (e.g. Right).

In order to reveal typological universals, the Morbo/Comp database is subdivided into four “genetically related groups”:

- Romance: Catalan, French, Italian, Spanish
- Germanic: Dutch, English, German, Norwegian, Swedish
- Slavic: Bulgarian, Polish, Russian, Serbo-Croatian
- East Asian: Mandarin, Japanese, Korean

Data from these languages, comprising about 3,000 compounds, is analysed in terms of the following five features which the authors express using the generic structure [ X ṭ Y ] Z (p.116):

1 The authors acknowledge that Mandarin, Korean and Japanese are neither genetically related nor typologically homogeneous.
The typology and semantics of binominal lexemes

- a) $Z$ = Output Category
- b) $X$ and $Y$ = Input Categories
- c) $\mathcal{R}$ = Relation between constituents (Classes)
- d) $[X \ Y]$ = Combination of Categories
- e) $[X \ Y] \ Z$ = Headedness

Scales of preference are computed for each of these features and produce the results shown in Figure 7.

Output category: $N > A > V > \text{Adv} > (\ldots)$
Input category: $N > A > V > \text{Adv} > (\ldots)$
Classification: $\text{SUB} > \text{ATT} > \text{CRD}$
Headedness: Right > No Head > Left > Both
Combinations: $[N+N] > [A+N] > [N+A] > [A+A] > [V+N] > [N+V] > [V+V] > (\ldots)$

Figure 7: Scales of preference in compounding (Guevara & Scalise 2009)

These results can be summarized as follows:

- nominal compounds are more common than adjectival compounds, which are more common than verbal compounds (etc.);
- the most common constituents are nouns, followed by adjectives, verbs (etc.);
- subordinate compounds are more common than attributive compounds, which in turn are more common than coordinate compounds;
- right-headed compounds predominate, followed by exocentric compounds, left-headed compounds and coordinate compounds;
- noun-noun combinations are most common, followed by other combinations as shown.

A few more insights can be gleaned from a closer reading of the text, for example, the greater prevalence of coordinate compounds in East Asian languages (32%) compared to the mean (19%). In addition, Guevara and Scalise have enough data to provide empirical evidence against a number of “false universals”, including the aforementioned “right-hand head rule”, the “root compounding parameter” (Snyder 2001), and the notion of the head as “locus inflectionis” (e.g. Zwicky 1985).

All in all, however, considering the size and scope of the Morbo/Comp project, these results constitute somewhat meagre pickings. Moreover, the lack of balance
in the sample, and the apparently unsystematic manner in which the data were collected (see §2.1.3), cast some doubt on the validity of the findings. Also, in view of the still untapped cross-linguistic potential of what appears to be the most widespread form of compounding in the world’s languages, I do not share the authors’ opinion that future work should concentrate on the analysis of compound types other than “endocentric subordinate right-headed [N+N]N compounds”. It is true that there have been many studies of NN compounds in individual languages, and also a handful of comparative studies, e.g. Bauer (1978) on English, Danish and French and Takada (2008) on French and Japanese, but, as far as I am aware, there has not been a single large-scale cross-linguistic study of such compounds.

2.2 Word-formation

2.2.1 Aikhenvald (2007)

Aikhenvald (2007) is a survey of the kinds of word-formation patterns found in the world’s languages, written for a volume whose aim was to give field linguists “a good idea of what to look for” when describing a language (Shopen 2007: xv). There is broad coverage of both general issues (including the nature of the word, morphological typology, structure and iconicity, productivity, lexicalization and grammaticalization), and specific types of word-formation: noun incorporation, compounding and derivation, and the chapter ends with a set of “suggestions for field workers in describing types of word-formation.” Here I focus on the sections devoted to compounding and derivation.

The section on compounding starts by addressing the issue of how to distinguish compounds from phrases. Four kinds of criteria are put forward: phonological, morphological, morphosyntactic and semantic. None of these are claimed to be universal, however, and thus “compounds have to be defined on language-internal criteria” (p. 24). Nominal compounds receive separate treatment from verbal compounds and compounding in other word classes. A number of “parameters of cross-linguistic variation” are listed (and exemplified), including:

- free forms, cranberry morphemes and special forms of free morphemes
- case-marked forms (e.g. nominative, genitive, elative, allative)
- closed classes in compounds (e.g. personal, reflexive and deictic pronouns)
- compounds formed on phrases
- typical non-referentiality of compound constituents
- productivity, sources and position of the head (if any)
However, the discussion of each issue is quite cursory. The flavour of the chapter as a whole is conveyed by the following discussion of head position:

In Germanic, Slavic and Finno-Ugric languages the head usually follows the modifier – e.g. Estonian pea-linn (head-city) ‘capital’, vana-linn (old-town) ‘downtown, old town’, cf. German Haupt-stadt (head-town) ‘capital’ – while in Romance languages the modifier can follow the head, as in Italian caffelatte ‘type of coffee’, or precede it, e.g. Portuguese boa-vida (good-life) ‘a bon vivant’ (cf. noun phrase vida boa (lit.: ‘life good’) ‘good life’). In Tagalog nominal compounds, the head typically precedes the modifier, thus creating the reverse order to that in their English counterparts (Schachter and Otanes 1972: 110), e.g. puno-ng-mangga (tree-LINKER-mango) ‘mango-tree’, tubig-ulan (water-rain) ‘rainwater’.

The further discussion of nominal compounds focuses on two kinds of contrast: the one based on the nature of the compound head (endocentric vs. exocentric vs. coordinate), the other on the word class of its constituents (root vs. synthetic):

- **Endocentric compounds** denote a subclass of items referred to by one of their elements (i.e. the head); exocentric compounds denote something which is different from either of their components;¹ and coordinate compounds “consist of two juxtaposed nouns which refer to a unitary concept”.

- **Root compounds** “do not have a verb base”, while synthetic compounds “consist of a verbal root with its argument” (which may be a direct object, an oblique constituent, or an intransitive subject). The latter are said to overlap with lexical compounding, which is one of the five functional types of noun incorporation distinguished by Aikhenvald.²

Notable for its absence from the discussion of compounding is any mention of the formation of new lexical items with more phrasal origins, such as chemin de fer and železnaya doroga. For Aikhenvald, as for most linguists, this process is not part of word-formation. But, we might ask, if not there, then where does it belong? Word-formation purports to be about the ways in which languages enrich their lexicons, but as long as it is viewed as a branch of morphology, and morphology as distinct from syntax, there is no home for the two above-mentioned formations.

¹ In Aikhenvald’s usage this definition includes both typical Romance ‘verbal nexus’ compounds, such as Por. quebra-cabeça [break-head] ‘puzzle, crossword’, which has an implicit head, and bahu-vrihi compounds, such as Eng. egghead ‘a type of intellectual’, which has a metonymic head.

² Quite what this “overlap” resides in is unclear, since all the examples of synthetic compounds are nominal, while all the examples of “lexical compounding” are verbal.
Aikhvenvald’s discussion of derivation follows the same pattern as compounding and centres around different ways of classifying derivational processes:

- The functional classification distinguishes between category-changing processes and category-preserving processes, depending on whether or not the process in question leads to a change of word class. In addition there are category-defining processes which are typical for a particular word class.
- The formal classification distinguishes between affixes and morphological processes: affixes can be continuous (prefixes, suffixes, infixes) or discontinuous (circumfixes, transfixes); morphological processes include apophony, reduplication, prosodic modification and subtraction, conversion, repetition and metathesis, and also acronyms, clippings, abbreviations and blends.

Aikhvenvald (2007) is clearly not a typological study in the sense of Song (2007), with its five distinct stages (page 17). She does present a number of typological classifications within the broad domain of word-formation, but these are not based on any particular sample. In fairness, it is not her purpose to formulate and explain typological generalizations, but simply to describe the diversity of the domain. Nevertheless, certain generalizations are made. Among these are:

- “Compounding is widespread in isolating languages, while derivation is a property of languages of other types” (p. 9).
- “Numeral classifiers as independent words tend to occur in analytic isolating languages” whereas “in synthetic languages numeral classifiers tend to be affixes” (p. 10-11).
- Compounds “typically have fixed constituent order, which may be distinct from the order in noun phrases” (p. 26).
- Most languages of the world have more suffixes than prefixes. No language has prefixes without having suffixes (p. 45).

From the perspective of the present study, the value of the paper lies in its comprehensive coverage of the features to be found cross-linguistically in the domain of word-formation, rather than in any insights regarding language universals.

---

1 As she writes in Aikhvenvald (2013): “Until the majority of human languages have been thoroughly analysed and documented, it seems most judicious to follow a qualitative approach, leaving quantitative analysis until such time in the future when more data is available and can be assessed.”
2.2.2 Štekauer, Valera & Körtvélyessy (2012)

Štekauer, Valera & Körtvélyessy (2012) is a typological study more along the lines advocated by Song. The phenomenon under investigation is word-formation in all of its breadth; a ‘basic sample’ of 70 languages is employed, along with a more balanced subset of 55 languages called the ‘study sample’ (see Table 67 on page 490). Most of the data comes from questionnaires, but published sources were used for some languages. The creation of a typological classification is a task left to the reader. According to the authors, “various sorts of typological classifications can be inferred” from the data they present. These are

primarily determined by the specific method of analysis, semasiological or onomasiological. In particular, the typological classification pertains to the preferences for formal ways of expression of cognitive categories and for the semantic scope of the individual formal means of expression of genetically, morphologically and/or geographically related languages (p. 8).

Given that the authors do not develop any classification, very few generalizations are possible, and there are therefore correspondingly fewer observations to explain. The study thus takes only the first couple of steps along Song’s path and should be regarded, as the authors state, as “a first, tentative probe” (p. 329).

A useful aspect of the study is the inclusion of the questionnaire as an appendix. The questionnaire starts with a metadata section with fields for the name of the language, and its genetic affiliation, geographic area and “morphological type”. The genetic classification is based primarily on the World Atlas of Language Structures (Haspelmath et al. 2005) and supplemented by informants, but the source of the morphological type is unclear. The only clue as to the typology employed (but not the details, definitions or criteria used in the classification) is a table on p. 11 (Table 12). This is a pity, since the morphological typology plays a major role in the study, its values being specified for every language in every one of the many “language list” tables throughout the book, and appearing as one of the three independent variables in the statistical analyses in Chapter 7.

<table>
<thead>
<tr>
<th>Morphological (Sapirean) classification</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agglutinative languages (of various types)</td>
<td>30</td>
</tr>
<tr>
<td>Fusional (of various types)</td>
<td>12</td>
</tr>
<tr>
<td>Isolating (of various types)</td>
<td>7</td>
</tr>
<tr>
<td>Polysynthetic (of various types)</td>
<td>6</td>
</tr>
</tbody>
</table>

*Table 12: Morphological classification (Štekauer et al. 2012)*
Chapter 2. Earlier work

The body of the questionnaire has three parts. Part I asks (1) which word-formation processes\(^1\) are productive in the language, (2) their level of productivity on a scale of 1 to 5, and (3) more detail regarding the four basic word-formation processes: (A) prefixation, (B) suffixation, (C) compounding and (D) reduplication. The questions relating to compounding (3C) give a feel for the overall enterprise and are therefore reproduced in their entirety in Figure 8. The questions under points (d) and (e) are pertinent to the present study. Part II of the questionnaire concerns the most productive ways of forming various semantic categories of noun and verb, and Part III gives contributors the opportunity to provide additional comments that might complete the picture of productive word-formation processes in the language.

(a) Is compounding recursive?
(b) Are there adjectival (Adjective + Adjective) compounds?
(c) Does the language make productive use of verbal compounds?
(d) Does the language make productive use of noun (Noun + Noun) compounds?
   Which of the following are found:
   1. Stem + Stem compounds?
   2. Stem + Link + Stem compounds (the link being specific to compounding)?
   3. At least one Stem is phonologically modified
(e) If the language makes productive use of compounds both with and without a linking element, which type is more productive?
(f) Are there any copulative compounds?
   Which, if any, of the following are found:
   1. Substantival?
   2. Adjectival?
(g) Are there any exocentric compounds? Which, if any, of the following are found:
   1. words like redskin?
   2. words like French garde-manger?

*Figure 8: Extract from questionnaire (Štekauer et al 2012)*

The book based on the data thus acquired is in two parts. The first consists of two theoretical chapters devoted to the scope of word-formation and the fuzzy nature of the boundary between word-formation and syntax. Spencer (2000: 315) is quoted regarding the separation of compounds from phrases as being “one of the more vexed problems in morphological theory”. Perhaps this is why there is no further mention in the study of “phrasal lexemes” (Masini 2009) like chemin de fer and železnaya doroga: Apparently Štekauer and his colleagues follow Aikhenvald in excluding such constructions from the domain of word-formation.

\(^1\) The 20 processes comprise those listed by Aikhenvald in her formal classification of derivation processes (page 43) and combinations thereof, together with the other two major word-formation types, incorporation and compounding.
The second part of the book contains the cross-linguistic analysis and comprises five chapters. The first three of these adopt a semasiological perspective and investigate three kinds of process: (i) the combination of free morphemes (compounding, reduplication and blending); (ii) processes involving bound morphemes (affixation of various kinds); and (iii) processes that do not involve the addition of derivational material (conversion, segmental alternation, suprasegmental alternation, subtractive processes, including back-formation). The fourth chapter takes an onomasiological approach and examines the formal mechanisms used to express various semantic categories: nominal (agent, patient, instrumental, locative, animate gender); evaluative (augmentatives and diminutives, phonetic iconicity and word-classes); verbal (causative, transitivity, intransitivity and iterativity and/or intensification); and word-class changing (action nouns and abstract nouns). The final chapter contains a statistical analysis and is followed by a short epilogue.

The first four chapters of Part 2 consist mainly of tables of the data collected using the questionnaire. For example, the first table in Chapter 3 lists the 50 languages (out of the basic sample of 55 languages) that make productive use of compounding, along with their genetic affiliation (language family, not genus), morphological type and geographical area. Noun-noun compounds are treated in section 3.1.1.4, which consists of a table listing the 44 languages that exhibit this feature, examples of compounds with and without linking elements, and a cursory discussion of three theoretical issues: the position of the head, semantics and linking elements.

The final chapter is quite different and consists of a set of statistical analyses using the chi-squared test and Multiple Correspondence Analysis (MCA). Both methods are applied to all twenty word-formation processes using the three independent variables: language family, morphological type, and basic word order. With the chi-squared test and a null hypothesis of a 50-50 chance of any language exhibiting each of the twenty processes, the principal result is that “four word-formation processes occur consistently more frequently than expected: prefixation, suffixation, compounding and reduplication”, and that these “occur regardless of the internal classifications used (by language family, morphological type or word order), even if only suffixation does so for all the types within the independent variables.”

MCA is a clustering technique that makes it possible to explore similarities and differences across data sets that involve multiple variables. In a first analysis, which covers every language family, only Afro-Asiatic, Austroasiatic and Indo-European are well-discriminated. By discarding all the other language families, associations are revealed between:
- Austroasiatic and the absence of suffixation
- Indo-European and the absence of tone/pitch and the presence of both prefixation and suffixation
- Afro-Asiatic and the presence of suffixation and reduplication

The same technique is used with a more fine-grained exploration of various subtypes of the four word-formation processes prefixation, suffixation, compounding and reduplication. The analysis of compounding is restricted to five language families (Altaic, Indo-European, Niger-Congo, Sino-Tibetan and Totonacan)\(^1\) and consists of the plot shown in Figure 9.

![Figure 9: MCS analysis of compounding (Štekuater & al. 2012)](image)

The authors explication of these results are quoted here in full:

> Language families have fairly distinct patterns of presence/absence, as can be seen from their separation along the axis for Dimension 1. Thus, Dimension 1 discriminates two groups:

(a) Sino-Tibetan appears to be associated with Indo-European as regards the presence of most types of compounding, and

---

\(^1\) The other 23 families in the study sample were excluded either because “the questionnaire data on compounding do not apply” or because their quality values were below 0.11.
(b) Totonacan appears to be associated with Altaic and Niger-Congo as regards the absence of most types of compounding.

Regarding processes, the picture is less clear but, in general, presences appear to be associated with presences and absences with absences. Thus, for example, the presence of recursive compounding is associated with the presence of adjective + adjective compounding, and the absence of adjective + adjective compounding is associated with the absence of copulative compounding. Some exceptions can be noted: presences of some processes are associated with absences, like the presence of verbal compounding and the absence of phonological change.

Finally, concerning the association between language families and processes, the separation of languages into two groups is paralleled by the separation between absence and presence of types of compounding. The clearest association seems to occur between the language family Totonacan and the absence of noun stem + noun stem compounding, both high along Dimension 2 (pp. 319–320).

The final statistical analysis concerns the onomasiological data and relates to the expression of various semantic categories by different word-formation processes. The chief result obtained is that “suffixation is the process which is used most by the sample languages throughout all the semantic categories used”. Of interest to the present study is that compounding is used rather little for such purposes, in particular for the formation of agent, patient, instrument and action nouns.

In conclusion, Štekauer, Valera & Körtvélyessy (2012) provides plenty of data across a relatively broad sample of languages, but fails to uncover any particularly interesting generalizations. It confirms the rich diversity of word-formation in the world’s languages (despite ignoring phrasal lexemes), but stops short of trying to explain the associations that it has discovered or making any kind of predictions. According to the authors, “such a motivation or prediction would be a matter of speculation rather than of clear linguistic facts” (p. 304).

2.3 Prefiguring binominals

I coined the term binominal lexeme for the present study, but the concept itself – that of constructions that serve the same function as noun-noun compounds – is not without precedent. It figures implicitly in the following three studies: Levi (1978) on ‘complex nominals’, Rainer (2013) on ‘relational adjectives and their competitors’, and Bauer & Tarasova (2013) on ‘adnominal nominal modification’. Taken together they serve to validate binominal lexeme as a comparative concept.
2.3.1 Levi (1978) – Complex nominals

Levi (1978) is one of the most cited studies of compounding in English. Unlike Warren, writing the same year on semantic relations, Levi is concerned with both syntax and semantics, and writes within the framework of Generative Semantics. Of interest in the present context is Levi’s notion of ‘complex nominals’, a term that covers three “partially overlapping sets of expressions”, viz. (a) nominal compounds, (b) nominalizations, and (c) noun phrases with non-predicating adjectives (p. 1). On the basis of the examples provided, these can be defined more precisely as:

(a) compounds consisting of two root nouns (e.g. apple cake, windmill)
(b) deverbal nouns modified by a denominal adjective or a root noun (e.g. musical criticism, metal detection)
(c) denominal adjective plus noun (e.g. musical criticism, electric shock)

There is some overlap between Levi’s notion of complex nominals and the concept of binominals as defined in §1.2.4: all of Levi’s nominal compounds are binominals, since they consist of two thing-roots; none of her nominalizations are, since they include an action-root; and some of her ‘nonpredicate NPs’ are – more precisely, those that involve a base noun representing a nominal concept, but not those that involve a deverbal noun. On the other hand, the present concept of binominals goes considerably beyond Levi’s notion of complex nominals in that it includes:

a) Constructions that involve one or two inflected nominal roots – including genitives (Bezhta kilos hino [iron:GEN way] RAILWAY); pertensives (Malagasy lalam.by [road:PER.iron] RAILWAY); and various other possessive constructions (Galibi Carib emoli sakilali [nose:3SG aperture:3SG] NOSTRIL).

b) Constructions that involve two nominals linked by a grammatical element such as a preposition (Tarifit abrid n mašina [road PREP train] RAILWAY), a postposition (Japanese budō no ki [grape POSTP tree] VINE), or some other kind of particle (Swahili tundu la pua [hole CON nose] NOSTRIL).

c) Constructions that consist of a thing-root and a thing-affix (Slovak želez.n.ica [iron.ADIZ.NMLZ] RAILWAY) or a thing-root and a noun classifier (Bora tůú.heju [nose.CM(hole)] NOSTRIL).

1 Earlier studies, including Mätzner (1860), Jespersen (1942) and Hatcher (1960), are discussed in Chapter 6.

2 These categories are not disjunct, in that deverbal nouns qualified by denominal adjectives (e.g. musical criticism) belong to both types (2) and (3).
Levi’s claim, that “all [complex nominals] must be derived by just two syntactic processes, predicate nominalization and predicate deletion” (p. 6), is not relevant to the present study, with its cognitive-functional orientation, but her typology of nine “recoverably deletable predicates” is relevant to the discussion of semantic relations in Chapter 6 (see §6.1.1 and §6.1.2).

2.3.2 Rainer (2013) – Relational adjectives, etc.

As the title indicates, the principal research question addressed by Rainer (2013) is whether relational adjectives can express any kind of semantic relation. The term ‘relational adjective’ is used to refer to “denominal adjectives whose suffix is said to serve a purely transpositional function, converting a noun into an adjective”, such as solar (< Latin solis ‘sun’).

Rainer concludes that relational adjectives do indeed seem to be able to express “any relation … except for the privative relation and for cases where some specific relation is blocked due to the interference of a rival pattern” (p. 26). In addressing his research question, Rainer devotes considerable space to the topic of relational adjectives and their competitors, the latter comprising (i) genitives, (ii) nominal compounds, (iii) prepositional phrases and attributivizers, and (iv) derivation. The paper takes a cross-linguistic approach, focusing on Latin and Slavic when discussing genitives, on German for compounds, Romance (with passing mention of the Hungarian attributivizer) with regard to prepositional phrases, and Arabic in the discussion of derivation.

However, the main interest of the paper from the perspective of the present study resides in Rainer’s implicit recognition of the cross-linguistic comparative category of binominals. In fact, Rainer comes extremely close to the definition adopted in §1.2.4, when he refers to “the expression of complex concepts consisting of two nominal concepts linked by some relational concept” (p. 27). For him, this is the “core competence” of relational adjectives; for me, it is the defining characteristic of binominal lexemes. Thus, aside from the fact that the present notion of binominals also includes noun classifier constructions, which are not mentioned by Rainer, his concept of “relational adjectives and their competitors” is essentially identical to that of binominal lexemes, and thus serves to validate it as a comparative concept.
2.3.3 Bauer & Tarasova (2013) – Adnominal modification

Bauer & Tarasova (2013) is an investigation into whether the meaning relationships that hold between the constituents of endocentric noun-noun compounds in English are also found in other constructions in which a noun is modified by another noun. For the purpose of their study, the authors take Levi’s nine “recoverably deletable predicates” (introduced above in §2.3.1) as their starting point and show that the very same relations are to be found not only in noun-noun compounds, but also in five other nominal modification constructions.

Levi’s RDPs are CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT. Since the first three of these are reversible, they embody a total of 12 relations which I will describe and exemplify later in §6.1.1. Bauer and Tarasova’s investigation itself is discussed in more detail in §6.1.2; in the present context, it is the six English nominal modification constructions that are of interest. These are listed in (7) and illustrated with the examples of Levi’s USE relation (N₂ USE N₁) given by the authors.

(7) a. noun-noun compounds (steam iron)
   b. associative (i.e. relational) adjective plus noun (manual labour)
   c. prenominal possessives (car’s driver)
   d. postnominal possessives (driver of the car)
   e. neoclassical compounds (hydromancy < water + divination)
   f. blends (paratroops < parachute + troops)

Bauer and Tarasova are able to demonstrate that all 12 of the relations derived from Levi’s RDPs can be utilized by each of the six constructions in (7). The authors therefore conclude that these relations are not specific to compounding and must arise “from the nature of the modification”. This conclusion is of interest in two respects, firstly in the context of my attempt later on to develop a general model of associative relations (see Chapter 9), and secondly, as further evidence for binominal lexeme as a coherent category, at least in English.

It should be noted that Bauer and Tarasova’s study is not limited to naming units, but also includes what Koptjevskaja-Tamm (2004) terms ‘anchoring’ (as opposed to ‘nonanchoring’ relations (cf. §7.2.1). Thus, (7c) car’s driver and (7d) driver of the car do not qualify as binominals because they are not naming units and therefore fall outside the scope of the present study. On the other hand, car driver denotes a generic concept rather than a specific individual and is a naming unit, but it, too, does not qualify as a binominal since it contains an actional element (DRIVE) and is thus OT1, not OT3. However, this does not mean that the constructions in (7c) and (7d) are outside the scope of the present study: Bauer & Tarasova themselves offer
three examples of prenominal possessives that parallel (7c) and clearly are naming units: *dog’s breakfast* (“a confused mess or mixture”),\(^1\) *ladies’ man* (“a man who is fond of, attentive to, and successful with women”),\(^2\) and *wolf’s bane* (“aconite; esp., a tall Eurasian plant (*Aconitum lycoctonum*) with showy, yellow flowers”).\(^1\) None of Bauer & Tarasova’s examples of postnominal possessives (7d) are clearly naming units, presumably because this is a relatively marginal lexeme-formation strategy in English, but a word like *man-of-war* (“a combatant warship of a recognized navy”)\(^1\) shows that it does occur.

Thus all six of Bauer and Tarasova’s constructions fall within the definition of binominal lexeme, and the first four of them actually occur in the database that will be described in the following chapter. Neoclassical compounds and blends do not occur there, presumably because they are not common enough to have been captured using the list of 100 meanings that I used as the basis for my data collection, or because they tend to be less analysable synchronically and may therefore have been coded as monomorphemic by contributors. However, a few examples did turn up in a larger data set (see §3.1.3 regarding the list of 201 meanings). These include the Hindi neoclassical form *par.deśī* [foreign_region.inhabitant] STRANGER, which is “a Sanskrit word built like German *Ausländer*” (Claus-Peter Zoller, p.c.),\(^3\) and the Italian *pomeriggio* AFTERNOON, which is “a sort of blend between *pomeridiano* ‘postmeridian’ and *meriggio* ‘midday (ancient)’” (Francesca Masini, p.c.).

To summarize: Bauer and Tarasova’s category of “nominals that modify other nominals” is very nearly co-extensive with that of binominal lexemes, provided the former is restricted to lexemes that have a naming function. The only difference resides in the fact that the notion of binominal as defined in Chapter 1 includes two types – denominal derivation and classifier constructions – that are not covered by Bauer and Tarasova. The fact that their six constructions share the core definitional property of involving unstated (or underspecified) semantic relations – and, not least, that those relations are demonstrably of the same types – serves, once again, to confirm the validity of binominal lexeme as a coherent category.

---

\(^1\) [https://www.merriam-webster.com/dictionary/](https://www.merriam-webster.com/dictionary/)

\(^2\) [https://www.collinsdictionary.com/](https://www.collinsdictionary.com/)

\(^3\) The status of neoclassical compounds as possible subtypes of derivation is discussed further in §6.1.2.
2.4 Morphological complexity

This section brings together two rather different studies: the World Loanword Database (WOLD) and Matthias Urban’s (2012) dissertation on “lexico-semantic associations”. What they have in common, and what makes them relevant to the present project, is that they both involve large datasets, adopt an onomasiological approach (from meaning to form, cf. §2.1.2), and are concerned (each in their own way) with morphological complexity.

![Figure 10: Pepper as a loanword in WOLD](image)

2.4.1 Haspelmath & Tadmor (2009) – WOLD

The World Loanword Database (Haspelmath & Tadmor 2009a) is the online version of Loanwords in the world’s languages: A comparative handbook (Haspelmath & Tadmor 2009b). It comprises vocabularies (mini-dictionaries of about 1000-2000 entries) of 41 languages from around the world, with comprehensive information about the loanword status of each word. Sources and donor languages are provided for loanwords in each of the 41 languages, and users are able to compare loanwords across languages. WOLD was particularly useful to me because the data set is freely available under a Creative Commons license and could be used as a starting point for the present project. For this reason, the description given here focuses primarily on the structure and contents of the database, rather than the use to which it was put in the original project. The database contains data from 41 languages (see the map in Figure 18 on page 90 for the areal distribution):

Archi, Bezhta, Ceq Wong, Dutch, English, Galibi Carib (Kali’na), Gawwada, Gurinji, Hausa, Hawaiian, Hmong Daw (White Hmong), Hupdë, Imbabura Highland Quechua, Indonesian, Iraqw, Japanese, Kanuri, Kekchi (Q’eqchi’), Ket, Kildin Sami, Lower Sorbian, Malagasy, Manange, Mandarin Chinese, Mapudungun, Old High German, ...

1 http://wold.cld.org/
Regarding the sample, Haspelmath & Tadmor (2009c:3) write:

In selecting languages for inclusion in the project, an effort was made to represent the world’s genealogical, geographical, typological and sociolinguistic diversity. However, the overriding factors were practical. Languages could only be included if a specialist in the language volunteered to invest the considerable amount of time and effort needed to complete the database and to write a book chapter based on the findings. Indeed, no serious and timely offer to contribute a database and book chapter was turned down.

The authors admit that their language sample is not ideal and that “some regions or language families are over- or underrepresented, as are some typological and sociolinguistic types.” They also point out that the inclusion of a number of closely related languages led to some skewing of the statistics, since, for example, a word loaned by a parent language would count as a loanword in each of its descendants, whereas in fact it represents a single borrowing event. While they cannot claim that the sample is representative of the world’s linguistic diversity, the authors emphasize that it is much better than anything that existed before their project and preferable to using just one or two languages:

Our sample includes languages indigenous to all continents and belonging to many language families. Some of the languages are spoken by hundreds of millions while others only by a few thousands or even a few hundred. Some have a history going back millennia, while others are not normally written to this day. Some are official languages of nation states while others are spoken by ethnolinguistic minorities. Typologically, the sample includes highly isolating languages as well as synthetic languages, both more fusional ones and more agglutinative ones.

The data are organized around a set of 1,460 meanings based on the Intercontinental Dictionary Series (Key & Comrie 2015). These meanings are grouped into 24 semantic fields and five broad semantic categories: Noun, Verb, Adjective, Adverb

---

1 See Appendix A for genetic affiliations, language codes, etc. Note that some of the language names used here differ from [those in WOLD] (see Typographical and naming conventions on page xxv).

2 Agriculture and vegetation, Animals, Basic actions and technology, Clothing and grooming, Cognition, Emotions and values, Food and drink, Kinship, Law, Miscellaneous function words, Modern world, Motion, Possession, Quantity, Religion and belief, Sense perception, Social and political relations, Spatial relations, Speech and language, The body, The house, The physical world, Time, Warfare and hunting.
and *Function word*, which correspond to Things and entities, Actions and processes, Properties, Manner and location, and Grammatical meanings (p. 7). Examples of meanings and how they are categorized are given in Table 13.

<table>
<thead>
<tr>
<th>meaning</th>
<th>semantic category</th>
<th>semantic field</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE EYELID</td>
<td>Noun</td>
<td>The body</td>
</tr>
<tr>
<td>THE TRAIN</td>
<td>Noun</td>
<td>Modern world</td>
</tr>
<tr>
<td>TO EAT</td>
<td>Verb</td>
<td>Food and drink</td>
</tr>
<tr>
<td>RIPE</td>
<td>Adjective</td>
<td>Food and drink</td>
</tr>
<tr>
<td>SOUR</td>
<td>Adjective</td>
<td>Sense perception</td>
</tr>
</tbody>
</table>

*Table 13: Examples of WOLD meanings*

Each of the 1,460 meanings is represented by zero or more translation equivalents (‘words’) in each of the 41 languages. Sometimes a language has no counterpart for a particular meaning, while in other cases it has several counterparts. For each word, information is provided about its orthographic form, analysability1 (with a morpheme gloss for analysable words), loanword status, age, etc. For the most part the morpheme glosses follow the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel 2015), some with more detail than others, for example, the Archi word meaning ‘the nostril’ (8).

(8) *muč-li-n klan* [nose(iii)-OBL.SG-GEN hole(IV)SG.NOM]

The inclusion of the morpheme gloss, along with the semantic classification, the number of meanings and the size of the language sample, meant that the WOLD data was able to fulfill two functions in the present project. Firstly, it enabled a starter set of meanings to be extracted in a principled fashion (as described in §3.1.1), and secondly, it provided an initial source of data for the project (see §3.3.1). The question of analysability played an important role in the original loanword project:

In assessing the possible loanword status of a word, the first question was whether the word was analyzable (i.e. morphosyntactically complex) within the language. If this was the case, it was almost certain that it was created by speakers of the language rather than borrowed from some other language. *Such words were not considered loanwords*, even when they contained borrowed elements (p. 12, emphasis added).

---

1 A word is considered analysable only if it is synchronically analysable for lay speakers. If the word was analysable to linguists but not to lay speakers, it is marked ‘semi-analysable’ (p. 12).
In other words, if a word was marked as analysable, the additional information provided by the contributor – including the morpheme gloss – was not utilized in the loanword project. This information, however, was essential to the present project, as will be seen in the following chapter.\(^1\)

2.4.2 Urban (2012)

At first glance Urban’s dissertation appears to overlap considerably with this work: Both are concerned with morphologically complex items in the nominal lexicon and both apply an onomasiological, data-driven approach. However, the research questions they address are quite different. Urban describes the dissertation as a “typological study concerned with formal and semantic patterns in the lexicon with a focus on referring (“nominal”) expressions” (p. 773). That, in a sense, is also topic of the present work. More precisely, though, Urban’s work (as the title suggests) is an investigation into two different, but related aspects of the lexicon: (i) morphological analyzability and (ii) semantic associations. With respect to the former of these, Urban presents the following (abridged) research questions:

- Are there significant differences between the languages of the world with respect to the degree of morphologically complex terms in the lexicon?
- Is the predominance of simplex lexical items in the better-known European languages an atypical phenomenon, attributable to extensive language contact and concomitant lexical borrowing?
- What is the role that the language-specific means of word-formation have to play?
- Is it appropriate to postulate a typological trait for languages with a pronouncedly high degree of complex formations in the lexicon, and how could this trait be reasonably delimited?
- How do differences in the degree of morphological complexity in the lexicon correlate with other typological variables, in particular those relating to structural properties?

In addition to the degree of morphological complexity in the lexicon, Urban is interested in the kinds of semantic association exhibited by both morphologically

\(^1\) As someone who spent two of his previous lives working with international standards (ISO 8879: SGML and ISO 13250: Topic Maps), and issues related to data longevity and reuse, it is very gratifying to have been able to reuse the WOLD data, especially the parts that had been basically discarded by the original project. This was possible because of the project editors’ commitment to the principles of Open Data, for which they are to be commended.
complex terms (9a) and also monomorphemic lexical items that exhibit semantic extension in the form of polysemy (9b). Both examples in (9) involve a “lexico-semantic association” between the concepts SKIN and BARK.

(9) a. Mbum MDD ngàŋ-kpù [skin-tree] ‘bark’
    b. Bezhta KAP beš ‘skin, bark’

Urban assumes, probably correctly, that an “identical or at the very least highly similar” cognitive process is taking place which links the meanings ‘skin’ and ‘bark’ in both of these instances:

From a semantic point of view, this [cognitive process] is the important commonality between the terms in the two languages. The difference lies in the formal realization: in Bezhta no overt sign of the relationship on the level of the signifier is present (beš can refer to both ‘skin’ and ‘bark’), whereas in Mbum, the meaning ‘bark’ is realized by a morphologically complex term. But, to reiterate, the semantic pattern is fundamentally the same (pp. 57–58).

Following Koch & Marzo (2007), Urban adopts the term ‘lexical motivation’ to cover “ties between words and the meanings they convey, regardless whether they are realized formally by word-formation relations or by polysemous or ambiguous conflation in a single monomorphemic lexical item.” In order to study such ties, Urban adopts an onomasiological approach and starts from a set of 160 nominal meanings, most of which are also found in WOLD. (Unfortunately, no explanation is provided of the selection procedure.) Meanings are grouped into four semantic domains: Topological and nature-related terms; Artefacts; Body-part terms; and terms for Phases of the day (together with a small number of miscellanea). Despite the number of meanings per domain varying considerably (71, 26, 52 and 11, respectively), some limited use is made of this parameter in the analysis. Of the 160 meanings, 17 are represented by binominals in English, *viz.* ADAM’S APPLE, AIRPLANE, EYEBALL, EYEBROW, EYELASH, EYELID, FINGERNAIL, HEADLAND, MILKY WAY, NOSTRILS, RAINBOW, RIVER BED, SATURDAY, SUNRISE, SUNSET, WATERFALL, WHIRLPOOL. This gives a rough measure of the frequency of binominal concepts in Urban’s meaning list, i.e. approximately 10%.

A genealogically balanced variety sample of languages was constructed using the method developed by Rijkhoff et al. (1993) and Rijkhoff & Bakker (1998), which involves the calculation of a Diversity Value (see page 87), adapted for use with the most recent genetic classification developed by Dryer (2005). Data was collected primarily from dictionaries and supplemented with information from grammars and grammar sketches.
Turning to the thesis itself, Chapters 1 and 2 introduce the major research questions and brief discuss the history of research in this area. Chapter 3 describes the general design of the study, presents the meaning list and language sample and outlines the analytical framework. It sets about developing a classificatory grid (inspired by Koch & Marzo 2007) to map the various kinds of semantic relation against the ways in which they are formally expressed. Urban’s starting point for his classification of “formal relations”, which make a basic division into analysable and unanalysable forms, is a set of data points that all mean FLAME (10).

(10)  a. Hausa HAU harshe [tongue, flame]
     b. Khoekhoe NAQ #nora.b [flame.3SG:M]
     c. Lenakel TNL nam.nam- [tongue.RED-]
     d. San Mateo del Mar Huave HUV netitit < ne.atitit [NMLZ.to_flame_much]
     e. Toaripi TQO a.uri [fire.tongue]
     f. Kildin Saami SJD tōl.njūxxčem’ [fire:GEN:SG.tongue]
     g. Swahili SWH ulimi wa moto [tongue of fire]
     h. Fijian FIJ yame(.yame) ni buka [tongue.(RED) POSS fire]
     i. Rama RMA abúng ngárkali ~ abúng ngarkalima [fire flame]

(10a) classed as unanalysable because its morphological structure is not transparent for native speakers, but (10b-i) are all classed as analysable. (10b-d) are subclassified as derived and subdivided into alternation (10b), reduplication (10c) and plain, i.e. conversion (10d). (10e-i) are grouped together as lexically analysable. Urban notes that (10e-h) all contain two morphemes with lexical meaning and could in theory be further subclassified as either compound or phrasal lexeme. However, he argues that there are no satisfactory cross-linguistic criteria for making such a distinction and groups them together under the “deliberately ambiguous” cover term ‘lexical analysability’. Within this group (10i) is subclassified as redundant (i.e. pleonastic) because one constituent has the same meaning as the whole. Since the more specific word-formation strategies are not used in the analysis, the classification of morphologically complex forms consists essentially of two categories: derived and lexical. A major consequence of this is that “differences in the presence of additional grammatical material that different languages require to be present are disregarded.” Thus one axis of the classificatory grid (the formal axis) consists of just two types. This represents a major difference from the present project, in which I develop a more fine-grained typological classification (see Chapter 5).
Likewise, Urban’s semantic classification consists of a “very basic and very broad distinction of semantic relations into two types, traditionally called similarity-based or metaphorical and contiguity-based or metonymic” (p. 85). Similarity relations are further subdivided into perceptual or functional, and contiguity relations into spatial, functional, perceptual, provenience, configurational, and temporal. These subdivisions are not operationalized, however, and are merely offered as a “preliminary typological grid that may serve as a starting point for further work on this topic.” Again this contrasts with the present work which develops a much more fine-grained semantic classification (see Chapter 6).

Urban’s Chapter 4 presents a “coarse first step” in which languages are positioned on a continuum from strongly derivational (and hence weakly lexical) to strongly lexical (and hence weakly derivational), depending on the number of derived vs. lexical items they exhibit. Their positions on the continuum are analysed in terms of areal distribution using three established models for dividing the world into geographical areas (Dryer-6, Nichols-11 and Nichols-3). This is followed by a detailed discussion of “data from languages with different typological profiles” and interaction with the subtypes ‘derived type dominant’, ‘lexical type dominant’ and ‘mixed profile’. The conclusion is that the analysis clearly reveals “an association of preferred word-formation techniques with certain linguistic types”, and that languages in which the derived type is dominant cluster in certain areas of the world, “most notably the American Northwest (affixal type of polysynthesis) and the Northwest Amazon region in South America (derivational use of noun classifiers).”

Chapter 5 contains a quantitative analysis of the data. Using typological parameters and values drawn from the World Atlas of Language Structures (Haspelmath et al. 2005), Urban attempts to provide an explanation for the variation in the frequency of analysable lexical items. This turns out, perhaps surprisingly, to correlate with “phonological complexity (in particular complexity in the consonant inventory and in the structure of the syllable), as well as with the canonical structure of the nominal lexical root” (p. 773). The simpler the syllable structure, the smaller the consonant inventory, the shorter the monomorphemic native lexical morpheme, “the more analysable terms the sample languages have in their nominal lexicon” (p. 299). As for correlations regarding the kinds of semantic relation most often exhibited by a language, the most striking finding is that “the dominant word-formation device influences whether the language will favour contiguity- or similarity-based denominations in morphologically complex lexical items.” Urban concludes:
This is a non-trivial finding, since…there is no a priori reason that compounds must be metaphorical and derivatives must be metonymic semantically. It is also a highly interesting finding because, put in other words, one can observe here that languages, depending on the nature of aspects of their grammar (i.e. word-formation), carve up the essentially same or near-same reality, as represented by the meanings on the list which are presently studied, in quite different ways (p. 298).

Chapter 6 is concerned with semantic associations – “links between meanings in certain semantic domains as well as common metaphorical transfers across languages.” It investigates both areal patterns and “globally common strategies to express a particular meaning.” The investigation of semantic associations and their cross-linguistic distribution starts with an adjacency diagram, showing how certain meanings cluster together. There follows an analysis of various semantic fields (e.g. AEROSOLS), the meanings belonging to them (in this case, SMOKE, STEAM, CLOUD and FOG), and the strength of associations between the various meanings. The concern is thus mainly with the frequency of associations between individual meanings and not their semantics.

A short concluding chapter sums up the dissertation’s contribution as follows:

This study hopefully demonstrated two things: first, that the lexicon is not just merely “an appendix of the grammar, a list of basic irregularities” (Bloomfield 1933: 274), a doctrine that is still very much alive in many theoretical approaches to Linguistics, but that its formal structure is systematically determined by complexity of the roots and of the sound system. Neither are semantic structures completely random, but they are both amenable to areal influence in colexifying and analyzable terms, and, with regard to the latter, they co-vary to some extent with the type of word-formation most commonly used in individual languages (p. 369).

Overall, the dissertation is a treasure trove of interesting material, but there is rather little in the way of easily discernible conclusions. The present study shares with Urban’s a concern with the broad domain of complexity in the nominal lexicon and the onomasiological approach, but the research interests of the two are quite different. For Urban, semantic associations are central (and investigated through polysemy as well as morphological complexity), but these are restricted to two very broad types (contiguity and similarity), whereas in the present study, semantic relations are investigated at a greater level of detail. Moreover, while Urban’s formal typology is limited to two types (derived and lexically analysable), the present work develops a more detailed classification of one particular type of complex nominal, binominal lexemes.
2.5 Chapter summary

This chapter has discussed previous work, from cross-linguistic studies of compounding (and word-formation more generally), to studies that prefigure the notion of binominals and onomasiological studies of the lexicon. Each of them has helped shape the present study in one way or another. My project was originally envisaged as a large-scale version of Bauer (2001), but replicating that study (with the focus narrowed to nominal compounding) led to the realization that the comparative concept should be defined in functional-semantic rather than formal terms if interesting progress was to be made. This was confirmed implicitly by the paucity of new insights produced by the Morbo/Comp project and in the work of Štekauer et al. The results of Morbo/Comp also raised awareness of the need for a principled way to collect data and the need for cross-linguistic breadth.

Rainer (2013), supported by Bauer & Tarasova (2013), and to some extent Levi (1978), provided confirmation that there are certain topics – in particular, those involving semantic relations – that can be usefully investigated by adopting the broader, functionally-oriented perspective of binominal lexemes. That this would involve plenty of formal diversity was confirmed by both Aikhenvald (2007) and Štekauer et al. Arnaud (2004b) gave the first hint that an onomasiological approach might be productive, and both Haspelmath & Tadmor (2009b) and Urban (2012) showed that such an approach is workable on a large scale. The former also constituted a source of readily available data, while the latter showed that the range of topics that could potentially be investigated with data of that kind is very great: so great, indeed, that one should beware the danger of trying to cover everything. Beyond this, there is any amount of previous work that is relevant to this project (concerning, for example, choice of comparative concept, sampling techniques, developing a classification, analysing semantic relations, metonymy and more). That work will be presented and discussed in the context of later chapters where it more properly belongs.
3 Meanings, languages and data

This chapter is about methodology and data. It covers the second stage of Song’s recipe for doing typology (cf. page 17): that is, generation of a language sample, which I take to include the collection of data. Since the meaning list constitutes the foundation of the database, I explain in some detail in §3.1 how I constructed it. I describe how the initial set of 159 meanings was generated from WOLD, how it was tested against five languages and extended to a list of 201 meanings, and the principled manner in which it was later reduced to the final set of 100 meanings.

§3.2 is concerned with the language sample. Here I discuss various sampling techniques and present the sample of 106 languages used in the project. In §3.3 I cover the process of data collection, which was based on open data from the WOLD database, questionnaires, dictionaries and grammars, each of which presented its own challenges. Finally, §3.4 offers a short summary of the overall process.

3.1 The meaning list

Having decided upon a data-driven, onomasiological approach (see §1.4.1), it was necessary to construct a meaning list. The primary desiderata for such a list are that it should yield as many binominals as possible (both within each individual language and across the sample of languages as a whole), and that it should yield as many different types of binominal as possible. In other words, it should exemplify as many morphosyntactic strategies as possible for each language.

The most efficient and principled method of arriving at such a list would be to query an existing set of data, and my initial thought was to use the Morbo/Comp database for this purpose. As described in §2.1.3, that database was reported to contain 80,000 compounds from over 20 languages, annotated inter alia for category (N, V, A, P, etc.), structure ([N+N], [N+A], [V+N], etc.) and gloss (i.e. English translation equivalent). It was envisaged that a suitable list of meanings could be extracted from this database by querying it for compounds of category N with the structure [N+N], grouping these by gloss, and ordering the resulting list of glosses by the number of compounds representing each gloss. However the data turned out to be unsuitable for this purpose. This was because fewer than 5% of the 80,000 compounds had in fact been analysed and classified by type, and not all of these
had been glossed. With only 704 glossed [N+N] compounds distributed across 23 languages, there was no basis for selecting a set of meanings in a principled manner.

Fortunately, the WOLD database (described in §2.4.1) turned out to be more suitable for the purpose of extracting a set of meanings, for three reasons: firstly, it contains a large and fully annotated data set (1,460 meanings in 41 languages); secondly, it is constructed around the notion of ‘meanings’, so these do not have to be inferred from translation equivalents; and thirdly, it is freely available online. Unlike the Morbo/Comp database, however, entries in WOLD are not annotated in such a way that it is possible to query for forms consisting of two nouns. As noted on page 55, information is provided about analysability, and analysable forms are given a morpheme gloss: for example, Archi muč-li-n klan NOSTRIL is glossed as [nose(iii)-OBL.SG-GEN hole(IV)SG.NOM]. But such glosses are not generalized into forms that correspond to Morbo/Comp’s [N+N], which in the case of the Archi example might be given as N.GEN N or Mod.GEN Head. In short, WOLD can be queried for complex nominals, but not for binominals. For this reason, a multi-step process was used. In the first step, an initial list of meanings was extracted from WOLD. This consisted of the 159 meanings that are most often represented by complex (i.e. analysable) nominals (§3.1.1). In the next step, this list of meanings was tested using a small number of languages where it was known in advance what kinds of binominal construction to expect (§3.1.2). As a result of this examination, the list was extended to 201 meanings (§3.1.3). Data was then collected and analysed for a total of 50 languages and, on the basis of this data, it became possible to reduce the list to the 100 meanings that were most often represented as binominals (§3.1.4).

3.1.1 Initial extraction of 159 meanings

Binominals almost always denote things and entities. In other words, they denote the kinds of meanings that belong to the semantic category Noun in WOLD; they very rarely denote actions and processes (Verb), properties (Adjective), manner and location (Adverb), or grammatical meanings (Function word). Moreover, they are always analysable (i.e. morphosyntactically complex), since by definition they consist (at a minimum) of two thing-morphs. Building on these facts, and exploiting the metadata available in WOLD, an SQL query (11) extracted the subset of meanings with the semantic category “Noun” that are represented by analysable forms in one or more languages. These were sorted in descending order of the number of languages in which the meaning is represented by one or more analysable forms.
The query returned a three column table with the headings id (identifier), meaning, and langs, with the data sorted in descending order of langs (cf. the boldfaced portions of the query). Those meanings that are most frequently represented by complex (i.e. analysable) words were therefore at the top of the list. The resulting CSV file containing 928 rows was imported into an Excel spreadsheet called meaning (Figure 11). Observe that the meaning NOSTRIL is represented as a complex form in 35 of the 41 languages, MIDDAY in 32, etc.

In its original form, the query counted the number of analysable words for each meaning, rather than the number of languages that had an analysable word for each meaning. As a result, certain meanings received undue weight, which would have led to a smaller yield of binominals; the meaning COUSIN appeared in first place in the original list, with a count of 38, followed by NOSTRIL, with a count of 37. This was because COUSIN has multiple (analysable) counterparts in some languages (for example, Tarifit has six and Mandarin Chinese eight). By querying for languages instead of analysable words, the contribution of these two languages to the meaning COUSIN was reduced from 14 (6+8) to two, its overall value from 38 to 16; and its position on the list from 1st to joint 134th place.
From the list of meanings returned by the revised query, the top 159 meanings were selected. The cut-off point was set at meanings represented by complex words in 16 or more languages. This left space to adjust the sample in the next stage by adding a further 41 meanings (as described below), in order to achieve a total of 200, a target chosen for commensurability with Urban (2012). A higher cut-off point of 17 languages would have resulted in 133 meanings and the need to add a further 67 during balancing; a cut-off point of 15, on the other hand, would have resulted in 191 meanings and space for only a further nine meanings to achieve balance. While any of these three cut-off points could have been chosen, 16 seemed like a reasonable compromise. The result was a “starter set” of 159 meanings, shown in (12), with the number of languages in which each meaning is represented by at least one analysable form (i.e. the value of langs) is given in parentheses.

(12) NOSTRIL (35); MIDDAY (32); EYELID (31); PUPIL, SPIDER WEB, STEPFATHER (29); FISHERMAN, MERCHANT, PARENTS, STEPMOTHER, THUMB (27); DEFENDANT, EARLOBE (26); BREAKFAST, EARTHQUAKE, MURDER, NIPPLE OR TEAT, SKULL, SPINE, WATERFALL (25); EARWAX, POTTER, STEPSON, YOLK (24); CAPTIVE OR PRISONER, DIVORCE, DRINK, EYELASH, PLAINTIFF, RAPE, SHORE, STEPDoughter, TAILOR, TOE (23); ANCESTORS, BEGGAR, FOOD, KID, MARRIED WOMAN, NATIVE COUNTRY, STRANGER, TEACHER, THIEF (22); ANXIETY, BAD LUCK, DAWN, HERDSMAN, QUARREL, SUPPER, WRIST, YOUNG MAN (21); ARSON, BEEHIVE, BEESWAX, BEGINNING, BIRTH CERTIFICATE, BLACKSMITH, BRACELET, DINNER, DOORPOST, EARRING, OLD WOMAN, PALM OF HAND, PTY, SORCERER OR WITCH, WHETSTONE, WHIRLPOOL, WIDOWER (20); ANKLE, DARKNESS, DESCENDANTS, GLOVE, HOSPITAL, HOST, LUNCH, MARRIED MAN, MEAL, PROSTITUTE, REMAINS, SCULPTOR, SHOEMAKER, SHOULDERBLADE, SWELLING, WOMB (19); AFTERNOON, AIRPLANE, BOY, BRUISE, CARPENTER, COCK/ROOSTER, COLLARBONE, COOKHOUSE, DECEIT, DISEASE, GUARD, LICENSE PLATE, MAGIC, MEETING HOUSE, MISTAKE, MOTHER-IN-LAW (OF A MAN); NURSE, PERJURY, PESTLE, ROOF, SERVANT, TOILET, TOOL,
WEDDING, WIDOW (18); CALF, CHIEFTAIN, CROWD, DEFEAT, ENVY OR JEALOUSY, FARMER, FIREPLACE, FISHING LINE, FLAME, FOOTPRINT, FREEMAN, GRIEF, JUDGMENT, LAMB, MARE, NIECE, OLDER SISTER, SCREWDRIVER, SIBLING, SPECTACLES/GLASSES, SUNDAY, TWINS, VEIN OR ARTERY, YOUNGER SISTER (17); ANGER, BABY, COUSIN, EAST, ELECTION, END, FOAL OR COLT, GIRL, HANDKERCHIEF OR RAG, IDEA, INTENTION, ITCH, NEIGHBOUR, POSTCARD, PRAISE, QUEEN, RAINBOW, RAZOR, RIB, SCHOOL, VICTORY, WEAPONS, WEDNESDAY, WEST, YOUNG WOMAN, YOUNGER BROTHER (16).

Two things may be observed from the English words used as labels. Firstly, many of the meanings (but by no means all) are represented as binominals in English – specifically, noun-noun compounds – e.g. eyelid. This inspired a certain degree of confidence that the list of meanings had the potential to yield a substantial crop of binominals cross-linguistically. Secondly, many of the labels (and consequently the English language counterparts) are not binominals. Some are monomorphemic; this shows that, while some meanings (e.g. PUPIL) have analysable counterparts in many languages, they do not do so in all languages. Others are analysable but not binominal: shoemaker contains the actional element MAKE and is thus OT1; and beginning consists of the onomasiological base (the nominalizing affix -ing) and the determined element of the mark (BEGIN), but no determining element, and is thus OT2. Moreover, it can be observed that while SHOEMAKER probably has a rather high potential for binominal expression in some languages – for example through the combination of the two nominal concepts SHOE and AGENT, parallel with potter, and as is the case in Wichí sapatu.wu [shoe.AGT] – the same cannot be said for BEGINNING, which often involves an actional element with the meaning BEGIN.2

3.1.2 Evaluation against five languages

Since a major goal of the project is to document the variety of morphosyntactic strategies that are used to express binominal concepts, it was necessary to test the initial list of meanings against some languages for which both data and the ‘answer key’ (i.e. information regarding the kinds of strategies to anticipate) were readily available. The languages chosen for this purpose were French, Welsh, Turkish,

1 Note that the names of these meanings are all prefixed by the word “the” in WOLD in order to distinguish Nouns from homonymous Verbs (prefixed with “to”). In the present work the article is not used, since only meanings belonging to the semantic category Noun have been selected.

2 This also turns out to be the case with the WOLD data, where none of the words meaning BEGINNING are binominals.
Nizaa and Japanese. French and Japanese are represented in WOLD and thus the required data was readily available. For Welsh, Turkish and Nizaa data was collected from dictionaries and word lists. Each form that denoted one of the 159 meanings was analysed in order to determine the kind of construction it manifested. The results are detailed in the remainder of this section.

French

Based on the literature (Arnaud 2015; Bourque 2016), the following constructions were anticipated for French:

- **Head Mod**
- **Head de Mod** (and also, but marginally, **Head de:DET Mod**)
- **Head à Mod**, (and also, but marginally, **Head à:DET Mod**)
- **Head PREP Mod**, (where PREP stands for prepositions other than de and à).

One additional pattern, not mentioned by either Arnaud or Bourque, was found in the WOLD data: **Head Mod.ADJZ**, e.g. *incendie volontaire* [fire will.ADJZ] ARSON. Two patterns (**Head Mod** and **Head à Mod**) were not encountered. In order to ensure that such constructions were represented, the 20 meanings in (13) and (14) were selected from Bourque’s database and added to the list of meanings in (12).\(^1\)

(13) **Head Mod** (5x)

*timbre-poste* [stamp-post] **POSTAGE STAMP** (8), *oiseau-mouche* [bird-fly] **HUMMINGBIRD**, *mot-clé* [word-key] **KEYWORD**, *nœud papillon* [knot butterfly] **BOW TIE**, *papier toilette* [paper toilet] **TOILET PAPER**

(14) **Head à Mod** (15x)


---

\(^1\) Meanings given in bold are represented in WOLD but were not among the top 159 produced by the SQL query. For these meanings the number of languages in which it is represented by a complex word in WOLD is given in parentheses.
Welsh

A set of data points was created for Welsh using Evans & Thomas (1994) and Hawke, Fychan & Roberts (2014). Based on the literature (Zimmer 2000; Awbery 2004), the following constructions were anticipated:

- **Head.Mod, Head DET Mod, Head PREP (DET) Mod, Mod.Head**

There were no missing patterns, but only three examples of the **Mod.Head** type were found in the 159 meaning sample. The 20 meanings added on the basis of the analysis of French data (above) resulted in a further 10 examples of this type. Adding the two meanings in (15) then yielded three more examples.

(15)  **Mod Head (2x)**  

gyddf.dorch [neck.chain], gyddf.gadwyn [neck.chain] **NECKLACE** (10),  
rheil.ffcrrdd [rail.road] **RAILWAY**

Turkish

Using Akdikmen (2006) a set of data points was created for Turkish. Based on the literature (Kornfilt 1997; van Schaaik 2002), the following constructions were anticipated:

- **Mod Head.3SG, Mod Head**

Both of these were encountered with relatively high frequency. However, two subtypes of **Mod Head** mentioned by van Schaaik (2002: 21) were not encountered. These are (i) **Mod Head** constructions in which baş ‘head’ is the first element; and (ii) **Mod Head** constructions in which the first element denotes a material. In order to have such constructions represented, the meanings in (16) and (17) were added.

(16)  **Modbaş Head**  

baş.kent [head.city] **CAPITAL CITY**

(17)  **ModMATERIAL Head**  

altın yüzk [gold ring] **GOLD RING**, taş köprü [stone bridge] **STONE BRIDGE**

Nizaa

Due to the lack of a comprehensive dictionary it was not possible to create a full set of data points for Nizaa. However it was possible to use Theil’s unpublished wordlist (Theil Endresen nd) to find out whether the constructions predicted in the literature (Pepper 2010b) would be revealed by the set of meanings. Those constructions are

- **Head Mod, Mod Head**
Both patterns were encountered, despite the fact that rather few of the 159 meanings have translation equivalents in the data sources (18,19).

(18) **Head Mod**

*nīi dānnī* [person work] **FARMER** (17); *nīi sīnnī* [person darkness] **CAPTIVE OR PRISONER** (23); *kōw būnī* [bone head] **SKULL** (25)

(19) **Mod Head**

*nīyār yāā* [nose hole] **NOSTRIL** (35); *twaāy lāāa* [ear dirt] **EARWAX** (24); *būnī kōwī* [head bone] **SKULL** (25)

Since the list of meanings covered both left- and right-headed compounds in Nizaa, there was no need for further additions.

**Japanese**

Based on the literature (Hinds 1986; Kageyama 2009; Hasegawa 2015) the following constructions were anticipated:

- **Mod Head, Mod no Head**

55 examples of **Mod Head** were found in the WOLD data but only two of **Mod no Head**. Some of the latter (20) were already present in WOLD but not part of the initial meaning list, and Kageyama (2016: 492ff) proffered further examples (21).

(20) **Mod no Head** (in WOLD)

*waki no shita* [armpit GEN under] **ARMPIT** (11); *ki.no.mi* [tree GEN fruit] **NUT** (10); *e.no.gu* [painting GEN ??] **PAINT** (9); *budō no ki* [grape GEN tree] **VINE** (14); *toshishita no kyōdai* [age:down GEN sibling] **YOUNGER SIBLING** (9); *toshiue no kyōdai* [age:up GEN sibling] **OLDER SIBLING** (9)

(21) **Mod no Head** (additional)

*ko.no.ha* [tree GEN leaf] **LEAF**; *hi.no.de* [sun GEN rise] **SUNRISE**; *kumo.no.su* [spider GEN web] **SPIDER WEB**; *nomi.no.iti* [flea GEN market] **FLEA MARKET**; *ama.no.kawa* [heaven GEN river] **MILKY WAY**; *mago.no.te* [grandchild GEN hand] **BACK SCRATCHER**

In order to increase the representation of **Mod no Head** constructions, ARMPIT, NUT, VINE, FLEA MARKET and MILKY WAY were added to the set of meanings and the word *kumo.no.su* SPIDER WEB was added to the data set, since it does not appear in the Japanese WOLD vocabulary (http://wold.clld.org/meaning/3-819#2/24.3/-4.8), despite being one of the WOLD meanings.

---

1 Theil’s word list contains two entries with the definition *crâne* ‘skull’ (see Pepper 2016: 301).
3.1.3 Expansion to 201 meanings

Based on the results of testing against the five languages discussed in the preceding section, the list of 159 meanings was extended by the 30 meanings in (22), seven of which (shown in bold) were already present in WOLD and 23 were new.

(22) ARMPIT; BACKPACK; BEE; BICYCLE PUMP; BOW TIE; CAPITAL CITY; DAIRY COW; FLEA MARKET; GOLD RING; HAND BRAKE; HANDBAG; HORSESHOE; HUMMINGBIRD; KEYWORD; LIPSTICK; MAIL BOX; MILKY WAY; NECKLACE; NUT; PADDLE WHEEL; POSTAGE STAMP; RAILWAY; STONE BRIDGE; SUGAR CANE; TOILET PAPER; TOOLBOX; TOOTHBRUSH; VINE; WATER PUMP; WINDMILL

In order to bring the total number of meanings to 200, a further eleven meanings were selected from WOLD (23). The selection process was based on three criteria: (i) the meaning should be relatively frequently represented by analysable words in WOLD; (ii) it should have the (subjectively judged) potential to be represented by a binominal, and (iii) it should have the potential to be found in most of the world’s languages. Criterion (i) ensured that the meaning would often be represented by a complex form, even if it didn’t meet the threshold set for the original list of 159 meanings. Criterion (ii) was based on hints implicit in the English form (e.g. EYE + BROW, TREE + TRUNK) or, in the case of simplex English forms, information available elsewhere (such as Urban 2012) indicating that such meanings are often represented by complex forms (e.g. BARK from TREE + SKIN, CAVE from EARTH + HOLE). Criterion (iii) meant avoiding meanings that were overly culture-specific.

(23) ARCTIC LIGHTS; BARK; BICYCLE; CAVE; EYEBROW; SPRING OR WELL; STABLE OR STALL; TEAR; THATCH; TRAIN; TREE TRUNK

Finally, FIREWOOD (not present in WOLD) was added to the list of meanings on something of a whim, because it seemed to be a good candidate for being found in most languages and for being represented by binominals in many of them. This brought the total to 201, 177 of them present in WOLD and 24 new. The complete list of 201 meanings is given in Appendix H (Table 68) on page 491.

3.1.4 Reduction to 100 meanings

Having finalized the list of 201 meanings it was possible to start the process of data collection, which is described in §3.2.3. During that process it became apparent that the list of meanings was far too large and that, given the overall scope of the project, the amount of work required to collect and analyse forms representing 201 meanings in each of the targeted 100 languages would be prohibitive. It therefore became
imperative to reduce the number of meanings drastically. By this time, data had been collected and analysed for 50 languages: French, German, Hindi, Hungarian, Irish, Italian, Lithuanian, Polish and Welsh, plus the 41 WOLD languages (see §2.4.1). In addition to the properties in WOLD (analysability and morpheme gloss), each item had been annotated in such a way that binominals could be identified and counted (see Table 22 on page 102). This showed that some meanings were far more likely to be represented by binominal constructions than others. For example, of the 52 words meaning NOSTRIL, 40 were binominals (most of them consisting of nouns meaning NOSE and HOLE), whereas none of those meaning VICTORY were (many of the latter were derived from verbs with the actional meaning WIN). This indicated that the set of meanings could be pruned without significant loss of variety in terms of binominal constructions. The pruning was accomplished in several steps.

First, three columns were added to the spreadsheet meaning, as shown in Figure 12. Column D (words) shows the total number of words per meaning (including monomorphemic and unanalysable forms) in the expanded database; column E (NN) shows the number of binominals per meaning; and column F (freq) shows the frequency of binominals as a percentage of the number of words (NN / words). The table is sorted in descending order by column F.1

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ID</td>
<td>meaning</td>
<td>langs</td>
<td>words</td>
<td>NN</td>
<td>freq</td>
</tr>
<tr>
<td>2</td>
<td>4-231</td>
<td>nostril</td>
<td>38</td>
<td>52</td>
<td>40</td>
<td>77 %</td>
</tr>
<tr>
<td>3</td>
<td>4-221</td>
<td>earlobe</td>
<td>43</td>
<td>45</td>
<td>32</td>
<td>71 %</td>
</tr>
<tr>
<td>4</td>
<td>4-331</td>
<td>palm of hand</td>
<td>29</td>
<td>31</td>
<td>21</td>
<td>68 %</td>
</tr>
<tr>
<td>5</td>
<td>1-771</td>
<td>arctic lights</td>
<td>24</td>
<td>26</td>
<td>17</td>
<td>65 %</td>
</tr>
<tr>
<td>6</td>
<td>4-222</td>
<td>earwax</td>
<td>45</td>
<td>48</td>
<td>31</td>
<td>65 %</td>
</tr>
<tr>
<td>7</td>
<td>3-819</td>
<td>spider web</td>
<td>48</td>
<td>49</td>
<td>31</td>
<td>63 %</td>
</tr>
</tbody>
</table>

*Figure 12: View of the table 'meaning' showing frequencies*

Observe that NOSTRIL is represented by 52 words, of which 40 are binominals. This gives a value of 77% (40/52) for what might be called the binominal potential, or binominality, of NOSTRIL. Next after NOSTRIL in terms of binominality we find EARLOBE, PALM OF HAND, ARCTIC LIGHTS, etc. The meanings that are least often denoted by binominals (have the lowest binominality) are DIVORCE, ITCH, PRAISE, etc.

---

1 Column C shows the number of languages in which the meaning is represented by an analysable word in WOLD; that information was no longer relevant at this stage since more languages had been added in the meantime. The other values were updated automatically as more data was added.
QUARREL, REMAINS, SWELLING and VICTORY, none of which are represented by as much as a single binominal. Like VICTORY, most of these are derived from action-roots and are thus OT1 or OT2 and not OT3.

The complete ranking of meanings in binominality order is given in Table 69 of Appendix H on page 493. This ranking offers a principled method for trimming the list of meanings. One approach would be to choose a cut-off point at random, for example, at 25, 50, 100 or 120 meanings. However, it is important to know what the consequences of any choice will be, since there is a trade-off to be made: the smaller the size of the meaning sample, the easier it is to collect data, but the greater the risk of missing out on interesting morphosyntactic patterns. In order to arrive at an optimal cut-off, a consequence analysis was performed based on a further stage of the formal analysis, described in Chapter 4. As explained in detail there, every binominal was annotated for its construction, as shown in Figure 13.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>language</td>
<td>word</td>
<td>gloss</td>
<td>construction</td>
</tr>
<tr>
<td>2</td>
<td>Gawwada</td>
<td>sint.itte</td>
<td>[nose.SG:F]</td>
<td>Base.F</td>
</tr>
<tr>
<td>3</td>
<td>German</td>
<td>nase.n.loch</td>
<td>[nose.LE.hole]</td>
<td>Mod.LE.Head</td>
</tr>
<tr>
<td>4</td>
<td>Gurinji</td>
<td>jitji jarriny</td>
<td>[nose hole]</td>
<td>Mod Head</td>
</tr>
<tr>
<td>5</td>
<td>Hmong Daw</td>
<td>qhoq ntswg</td>
<td>[hole nose]</td>
<td>Head Mod</td>
</tr>
<tr>
<td>6</td>
<td>Iraqw</td>
<td>foxár duunga'</td>
<td>[hole:of nose]</td>
<td>Head.CON Mod</td>
</tr>
<tr>
<td>7</td>
<td>Kanuri</td>
<td>súwúlí kánzà.bè</td>
<td>[opening nose.GEN]</td>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>8</td>
<td>Ket</td>
<td>olan.d qúk</td>
<td>[nose.GEN hole]</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>9</td>
<td>Lower Sorbian</td>
<td>nos.owa žérka</td>
<td>[nose.ADJ hole]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>10</td>
<td>Swahili</td>
<td>tundu la pua</td>
<td>[hole CON nose]</td>
<td>Head CON Mod</td>
</tr>
<tr>
<td>11</td>
<td>Takia</td>
<td>ndu.n awa.n</td>
<td>[nose.3SG mouth.3SG]</td>
<td>Mod.3SG Head.3SG</td>
</tr>
<tr>
<td>12</td>
<td>Wichi</td>
<td>to.nhes.pe'</td>
<td>[POSS.nose.LOC]</td>
<td>POSS.Base.LOC</td>
</tr>
<tr>
<td>13</td>
<td>Yakut</td>
<td>murun ḡayāyah.a</td>
<td>[nose hole.3SG]</td>
<td>Mod Head.3SG</td>
</tr>
</tbody>
</table>

*Figure 13: Extract of the table 'word' showing constructions*

To exploit this annotation, a script was written to calculate how many constructions would appear in the data for meaning samples of sizes ranging from 1 to 201. The results of this analysis are shown in Figure 14. The graph shows the number of constructions found in the 50-language sample for various sizes of meaning sample. Note that constructions are language-specific, so Dutch Mod Head and English Mod Head, for example, are counted as different constructions.
With 201 meanings the yield across 50 languages was 276 constructions, an average of 5.52 constructions per language. At the other end of the scale, a single meaning (the most binominal meaning, NOSTRIL) yields 38 constructions, since the data set contains 38 different constructions that express this meaning (24).¹

(24)  
Archi Mod.GEN Head  
Ceq Wong Head Mod  
Dutch Mod Head  
English Mod Head  
Galibi Carib Mod.POSS Head.POSS  
Gawwada BASE.FEM  
German Mod.LE.Head  
Gurinji Mod Head  
Hausa Head.LK Mod  
Hawaiian Head Mod  
Hmong Daw Head Mod  
Hungarian Mod Head  
Hupdē Mod Head  
Imbabura Quechua Mod Head  
Indonesian Head Mod  
Iraqw Head.CON Mod  
Irish Head Mod.GEN  
Japanese Mod Head  
Kanuri Head Mod.GEN  
Kekchi Head Mod  
Ket Mod.GEN Head  
Lower Sorbian Mod.ADJ Head  
Malagasy Head Mod  
Manange Mod Head  
Mandarin Chinese Mod Head  
Mapudungun Head Mod  
Oroqen Mod.APOSS Head.POSS  
Otomí Head Mod  
Saramaccan Mod Head  
Seychelles Creole Head Mod  
Swahili Head CON Mod  
Takia Mod.3SG Head.3SG  
Thai Head Mod  
Vietnamese Head Mod  
Wichí POSS.BASE.LOC  
Yakut Mod.Head.3SG  
Yaqi Mod Head  
Zinacantán Tzotzil Head Mod

Table 14: Binominal constructions yielded by NOSTRIL

---

¹ The discrepancy between the number of binominals that denote NOSTRIL (40) and the number of constructions (38) is due to two languages that have two words with the same construction for this meaning: Indonesian Head Mod lubang hidung [hole nose] and rongga hidung [cavity nose], and Swahili Head CON Mod mwanzi wa pua [bamboo CON nose] and tundu la pua [hole CON nose].
Adding the second most binominal meaning, EARLOBE, which was represented by 32 binominals, increased the number of constructions to 50. Note how the addition of these 32 binominals resulted in the addition of only 12 new constructions (25); the remaining 20 constructions were already represented by words denoting NOSTRIL. Note also that a few languages appear in both (24) and (25), for example German, for which we find two different constructions: Mod.LE.Head and Mod.Head.

(25) French Head PREP Mod Old High German Mod Head
Galibi Carib Mod Head.POSS Polish Head Mod.GEN
German Mod.Head Takia Mod Head.3SG
Ket Mod Head Tarit Head PREP Mod:STC
Kildin Saami Mod Head Welsh Head Mod
Lithuanian Mod.GEN Head Wichí POSS.Mod Head

Table 15: Additional constructions yielded by EARLOBE

As more meanings are added to the list, the number of constructions increased, as shown in Figure 14. With a list of 40 meanings (roughly 20% of the original list of 201), the number of constructions reaches 171: the Pareto principle (or “80/20 rule”) thus does not apply with this data set, since only 62% of the effects (i.e. constructions) comes from 20% of the causes (meanings). In order to achieve 80% coverage (221 constructions out of the maximum of 276), a set of 84 meanings is required. This would correspond to a 58% reduction of the original sample size and require the inclusion of every meaning (from NOSTRIL to KEYWORD) with a binominality of 25% or more (cf. Table 69 on page 493). In this scenario, a total of 55 constructions would be lost; these are listed in their entirety in Table 70 on page 495, along with each individual binominal that instantiates the construction. For convenience, a subset is reproduced here as Table 16. (The significance of the dagger † is explained below.)

The following observations may be made regarding Table 16 (and Table 70):

- Most of the constructions (34 out of 55) are only instantiated by a single binominal, and are therefore probably somewhat peripheral; this is the case for all the types listed for Bezhta and Dutch, for example.
- One construction, Malagasy NMLZ.Base.CIRC, with six instances, is clearly of importance in the language and should therefore not be lost.
- In Zinacantán Tzotzil there are two constructions that have four instances each, Mod.Head and opaque. The latter can obviously be discarded without loss of information, but the former should preferably be retained.
Table 16: Constructions lost with a sample of 84 meanings (extract)
In Yaqui, the construction Mod:PL Head also has four instances. The database was a little inconsistent here, since plural forms of modifiers were otherwise treated in the same way as singular forms, as in French *ruche d’abeilles* [hive of bee_PL] BEEHIVE and *cire d’abeille* [wax of bee] BEESWAX, which both were categorized as Head PREP Mod. Once this inconsistency was resolved, the Yaqui construction no longer existed.

One construction, Wichí Base.AGT, has three instances, all denoting professions, and should be retained.

All of the other types are represented by two instances each.

Altogether there are 50 different meanings represented in Table 70 – that is, among the set of constructions that would be lost if the sample were to be restricted to the 84 meanings with the highest binominality. (Recall that Table 70 is the complete table of which Table 16 is just an extract.) Those 50 meanings are listed in Table 17, in descending order of the number of times that they occur in Table 70 (column A). Column B gives the binominality value. We observe that the meaning CHIEFTAIN is expressed by five of the “lost constructions” and that it has a relatively high binominality value of 21. In other words, the effect of adding CHIEFTAIN as the 85th meaning would be to increase the yield of constructions in the 50 language data set from 221 to 226. Now, at first glance it might seem that the best way to proceed in order not to lose too many constructions is to increase the number of meanings selected from the set of 201 such that the meanings that occur most frequently in Table 70 (i.e., those at the top of Table 17) are included in the meaning list. However, if we test this approach using the top six meanings (CHIEFTAIN, MAGIC, BEE, MARRIED WOMAN, MERCHANT and MARRIED MAN) we can see that it is suboptimal.

<table>
<thead>
<tr>
<th>meaning</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>chieftain</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>magic</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>bee</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>married woman</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>merchant</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>married man</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>cookhouse</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>widower</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>neighbour</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>toilet</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>widow</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>captive</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>meaning</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>wedding</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>stepmother</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>stepdaughter</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>deceit</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>shoemaker</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>herdsman</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>blacksmith</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>sugar cane</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>tool</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>hospital</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>bicycle</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>yolk</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>meaning</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>stepfather</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>meeting house</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>guard</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>younger sister</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>afternoon</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>weapons</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>prostitute</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>spring or well</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>servant</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>beggar</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>airplane</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>intention</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>arson</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>meaning</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>younger brother</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>tailor</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>defendant</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>envy</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>screwdriver</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>murder</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>disease</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>grief</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>election</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>pity</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>beginning</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>anxiety</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 17: Meanings represented by lost constructions
Table 69 on page 493 shows these six meanings ranked as numbers 98, 105, 101, 103, 117 and 135, respectively. For these meanings to be included, the cut-off point in terms of binominality value would have to be increased by 51: from 84 to 135 (the rank of MARRIED MAN). This would result in the inclusion of many meanings (e.g. POSTAGE STAMP, ROOF and COUSIN, ranked 85th to 87th) that would contribute nothing to reducing the number of lost constructions since they do not figure in Table 70 and hence are not listed in Table 17. Moreover, the addition of both MARRIED MAN and MARRIED WOMAN is unnecessary, since they employ the same constructions wherever they occur in Table 70 (i.e., in Kanuri, Malagasy, Mapudungun and Takia). It is sufficient to add just one of them.

From this it was clear that adding more meanings simply on the basis of the number of “lost constructions” that they would bring back into the data set was not the most efficient way of increasing the number of constructions represented in the data set as a whole. A more efficient procedure would be to add meanings based on their binominality value, but only if they occur in Table 17. On this basis the 16 meanings shown in boldface were selected from Table 17 and added to the list of 84 that resulted from applying a cut-off value of 25%. All 16 have a binominality of at least 17% (column B of Table 17) and their addition brings the number of meanings to a total of 100.

The final list of 100 meanings is shown in Table 18. 82 of them are shared with WOLD, and 18 (shown in boldface) are new. This list covers 253 (92%) of the 276 constructions contained in the 50 language data set based on 201 meanings. Those that were lost are marked with a dagger (†) in Table 70 (and Table 16). Some of these losses are clearly unfortunate. For example, it would be interesting to know that Hindi has the Mod. ADJZ Head construction in its binominal arsenal, that Irish exhibits the occasional head-initial compound, and that Seychelles Creole has a Base. AGT agentive derivation. However, it was never the goal of the present study to detect every single binominal construction in every language. On the plus side, the retention rate of 92% is remarkable considering that the size of the meaning list has been more than halved, from 201 to 100.

Table 18 also includes three different classifications. As mentioned in §2.4.1, the 1,460 meanings in WOLD are assigned to a set of 24 semantic fields (cf. page 54) and the same classification is retained in the present study for those 82 meanings that also occur in WOLD. The 18 meanings not found in WOLD were assigned to semantic fields following the same principles. Thus, for example, STONE BRIDGE is assigned to Motion on analogy with the bridge (WOLD code 10.74).
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>#</th>
<th>meaning</th>
<th>semantic field</th>
<th>semantic type</th>
<th>hatcher type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ankle</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>2</td>
<td>arctic lights</td>
<td>The physical world</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>3</td>
<td><strong>backpack</strong></td>
<td>Modern world</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>4</td>
<td>bee</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>5</td>
<td>beehive</td>
<td>Animals</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>6</td>
<td>beeswax</td>
<td>Animals</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>7</td>
<td>bicycle</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>8</td>
<td><strong>bicycle pump</strong></td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>9</td>
<td>blacksmith</td>
<td>Basic actions and technology</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>10</td>
<td>boy</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>11</td>
<td>bracelet</td>
<td>Clothing and grooming</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>12</td>
<td>breakfast</td>
<td>Food and drink</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>13</td>
<td>carpenter</td>
<td>Basic actions and technology</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>14</td>
<td>chieftain</td>
<td>Social and political relations</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>15</td>
<td>cock/rooster</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>16</td>
<td>collarbone</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>17</td>
<td>cookhouse</td>
<td>The house</td>
<td>location</td>
<td>place</td>
</tr>
<tr>
<td>18</td>
<td><strong>dairy cow</strong></td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>19</td>
<td>dinner</td>
<td>Food and drink</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>20</td>
<td>doorpost</td>
<td>The house</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>21</td>
<td>earlobe</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>22</td>
<td>earring</td>
<td>Clothing and grooming</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>23</td>
<td>earwax</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>24</td>
<td>eyebrow</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>25</td>
<td>eyelash</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>26</td>
<td>eyelid</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>27</td>
<td>farmer</td>
<td>Agriculture and vegetation</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>28</td>
<td>fireplace</td>
<td>The house</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>29</td>
<td>fisherman</td>
<td>Warfare and hunting</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>30</td>
<td>fishing line</td>
<td>Warfare and hunting</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>31</td>
<td>flame</td>
<td>The physical world</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>32</td>
<td><strong>flea market</strong></td>
<td>Social and political relations</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>33</td>
<td>foal or colt</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>34</td>
<td>footprint</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
</tbody>
</table>

*Table 18: Final set of 100 meanings*
<table>
<thead>
<tr>
<th>#</th>
<th>meaning</th>
<th>semantic field</th>
<th>semantic type</th>
<th>hatcher type</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>girl</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>36</td>
<td>glove</td>
<td>Clothing and grooming</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>37</td>
<td>gold ring</td>
<td>Clothing and grooming</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>38</td>
<td>hand brake</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>39</td>
<td>handbag</td>
<td>Basic actions and technology</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>40</td>
<td>handkerchief or rag</td>
<td>Clothing and grooming</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>41</td>
<td>herdsman</td>
<td>Animals</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>42</td>
<td>hospital</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>43</td>
<td>host</td>
<td>Social and political relations</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>44</td>
<td>keyword</td>
<td>Modern world</td>
<td>advanced</td>
<td>abstract</td>
</tr>
<tr>
<td>45</td>
<td>kid</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>46</td>
<td>lamb</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>47</td>
<td>license plate</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>48</td>
<td>lunch</td>
<td>Food and drink</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>49</td>
<td>magic</td>
<td>Religion and belief</td>
<td>basic</td>
<td>abstract</td>
</tr>
<tr>
<td>50</td>
<td>mail box</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>51</td>
<td>mare</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
</tr>
<tr>
<td>52</td>
<td>married woman</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>53</td>
<td>midday</td>
<td>Time</td>
<td>location</td>
<td>time</td>
</tr>
<tr>
<td>54</td>
<td>milky way</td>
<td>The physical world</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>55</td>
<td>mother-in-law (of a man)</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>56</td>
<td>native country</td>
<td>Social and political relations</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>57</td>
<td>neighbour</td>
<td>Social and political relations</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>58</td>
<td>niece</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>59</td>
<td>nipple or teat</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>60</td>
<td>nostril</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>61</td>
<td>paddle wheel</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>62</td>
<td>palm of hand</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>63</td>
<td>postcard</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>64</td>
<td>potter</td>
<td>Basic actions and technology</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>65</td>
<td>queen</td>
<td>Social and political relations</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>66</td>
<td>railway</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>67</td>
<td>rainbow</td>
<td>The physical world</td>
<td>natural</td>
<td>concrete</td>
</tr>
</tbody>
</table>

Table 18: Final set of 100 meanings (cont.)
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>#</th>
<th>meaning</th>
<th>semantic field</th>
<th>semantic type</th>
<th>hatcher type</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>rib</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>69</td>
<td>shoemaker</td>
<td>Clothing and grooming</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>70</td>
<td>shore</td>
<td>The physical world</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>71</td>
<td>shoulderblade</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>72</td>
<td>skull</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>73</td>
<td>sorcerer or witch</td>
<td>Religion and belief</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>74</td>
<td>spectacles/glasses</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>75</td>
<td>spider web</td>
<td>Animals</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>76</td>
<td>spine</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>77</td>
<td>stable or stall</td>
<td>Animals</td>
<td>location</td>
<td>place</td>
</tr>
<tr>
<td>78</td>
<td>stone bridge</td>
<td>Motion</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>79</td>
<td>sugar cane</td>
<td>Agriculture and vegetation</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>80</td>
<td>Sunday</td>
<td>Time</td>
<td>location</td>
<td>time</td>
</tr>
<tr>
<td>81</td>
<td>supper</td>
<td>Food and drink</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>82</td>
<td>tear</td>
<td>Emotions and values</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>83</td>
<td>thatch</td>
<td>The house</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>84</td>
<td>thumb</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>85</td>
<td>toe</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>86</td>
<td>toilet</td>
<td>Modern world</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>87</td>
<td>toilet paper</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>88</td>
<td>tool</td>
<td>Basic actions and technology</td>
<td>basic</td>
<td>concrete</td>
</tr>
<tr>
<td>89</td>
<td>toolbox</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>90</td>
<td>toothbrush</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>91</td>
<td>train</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>92</td>
<td>tree trunk</td>
<td>Agriculture and vegetation</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>93</td>
<td>vein or artery</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>94</td>
<td>vine</td>
<td>Agriculture and vegetation</td>
<td>natural</td>
<td>concrete</td>
</tr>
<tr>
<td>95</td>
<td>water pump</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>96</td>
<td>Wednesday</td>
<td>Time</td>
<td>location</td>
<td>time</td>
</tr>
<tr>
<td>97</td>
<td>widower</td>
<td>Kinship</td>
<td>person</td>
<td>person</td>
</tr>
<tr>
<td>98</td>
<td>windmill</td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>99</td>
<td>wrist</td>
<td>The body</td>
<td>body part</td>
<td>concrete</td>
</tr>
<tr>
<td>100</td>
<td>yolk</td>
<td>Food and drink</td>
<td>natural</td>
<td>concrete</td>
</tr>
</tbody>
</table>

*Table 18: Final set of 100 meanings (cont.)*
Meanings can also be classified by semantic type. Many different classifications have been proposed in the literature. One of these is a scheme (labelled hatcher type in Table 18) consisting of the seven types: Person; Animal; Concrete Object, Substance, Condition; Place; Time; Activity; and Miscellaneous Abstract Entities. This was proposed by Hatcher (1960) as a starting point for subcategorizing the system she had devised for classifying the semantic relations of non-appositional compounds. Since I will later adopt Hatcher’s classification of semantic relations (see §6.2.2-6.2.3), it seemed reasonable to test her system of semantic types as well. However, as I show in the next section, this led to a very unbalanced result and therefore an alternative scheme (labelled semantic type in Table 18) was created instead.

3.1.5 Overall evaluation

With the benefit of hindsight, and having collected and analysed data based on the list of 100 meanings, it is possible to enumerate the desiderata for such a list and consider the extent to which the set of meanings that was finally arrived at was optimal in each respect. There are four main desiderata:

1. maximal diversity
2. maximal yield
3. cross-categorial balance
4. cross-linguistic representation

From the description of the methodology given in the preceding four sections it will be apparent that the primary considerations in the present study were those related to maximizing the yield of binominals and the diversity of constructions. In the first stage of the process, the initial extraction of 159 meanings (§3.1.1), two criteria were applied in order to subset the WOLD data (cf. the query on page 64): (i) the words should be analysable (because every binomial must by definition be analysable) and (ii) the meaning expressed should belong to the semantic category Noun (since the combination of noun plus noun almost invariably results in a noun). The result of the query was a list of meanings commonly represented by complex nominals, from which those that occurred in the greatest number of languages were selected as a starting point. Note that, at this stage of the process, the information required in order to find those meanings most commonly represented by binominals was not available.

In the second stage of the process, the focus was on maximizing the diversity of binominal constructions. This was accomplished by testing the initial set of 159
meanings against five languages for which information regarding constructions was readily available in the literature (§3.1.2), and then adding new meanings that were considered likely to ensure increased variety (§3.1.3), up to a total of 201. Then, in the final stage, having identified and analysed binominals in data from 50 languages, the list of 201 meanings was reduced to 100 meanings in such a way as to retain as much constructional diversity as possible (§3.1.4).

The degree to which the final set of meanings fulfilled the goals of maximum yield and diversity is hard to assess objectively. Given the lack of any previous work in this area, it is hard to see how the overall process could have been improved, but with hindsight it is evident that certain minor improvements could have been made. For example, it would have been sufficient to have either DINNER or SUPPER but not both, since most languages do not distinguish the two concepts, and one might also eliminate BREAKFAST and/or LUNCH as well, for the same reason. Moreover, I should have anticipated that these four meanings would be expressed by one and the same construction in many languages, such that the inclusion of just one of them would have sufficed.

It also turned out that some meanings yielded a lower percentage of binominals in the final data set than in the data set based on 50 languages. These are discussed in §4.4.2 under the rubric Data analytics by meaning. In some cases, this may be because insufficient attention was paid to the last of the four desiderata listed above, cross-linguistic representation. Arnaud, whose “toy” example provided the initial impetus for the onomasiological approach adopted in this study, was quoted earlier (§2.1.2) emphasizing the need to “reduce cultural differences to a minimum” and suggesting that the most appropriate concepts might be “body parts, natural species, meteorological phenomena and widely used artefacts”. He notes further that his own choice of concepts

did not result in sufficient coverage: thus, ‘cradle’ and ‘quiver’ are not relevant in the culture corresponding to Bininj Kun-Wok. The cradle is not indigenous in Khmer culture, but it does have a name. [Fr. foyer ‘home’ < fireplace] corresponds to metonymic shifts that are not found in all languages.

Clearly, body parts are an important source of concepts for a study of binominals, but only secondary body parts, such as EARLOBE, EYELASH and NOSTRIL (all of which are in the final list), since they tend to be named in terms of primary body parts such as EAR, EYE and NOSE (none of which are included). On the other hand, body parts cannot be allowed to dominate, since they would tend to use the same construction and thus limit the overall diversity of morphosyntactic structures.
Natural species are less suitable: those that are well-known, such as LION or GOAT, tend to have monomorphemic names, while sub-species, such as MOUNTAIN LION or MOUNTAIN GOAT, have much more restricted habitats and are less likely to have denotations in many languages. Only with domestic species, such as SHEEP and HORSE, is this problem avoided, and then only when specified for age or sex: LAMB and MARE, although monomorphemic in English, are often denoted by complex nominals cross-linguistically (for example, as ‘sheep child’ or ‘woman horse’). These are well-represented in the meaning list and given special attention in §8.4.

Arnaud’s meteorological phenomena (to which might be added natural phenomena in general) figure fairly widely in the present list with meanings such as RAINBOW, MILKY WAY, SHORE, BEESWAX and FLAME, but at least one such phenomenon, ARCTIC LIGHTS, should have been removed on the grounds of limited geographical occurrence.

Finally, widely used artefacts suffer from the same problem as natural species: those that are widely used and/or general tend to be denoted by a single morpheme, such as BAG, BRIDGE or RING. In order to yield binominals, more narrowly defined subtypes are required (thus BACKPACK, STONE BRIDGE and EARRING, all of which are on the list), but this, in turn, tends to reduce the number of languages in which the concept is applicable. One way to mitigate this problem would be to loosen the concepts in some way, such as “bag carried in a particular way (e.g. using the arm, back, hand, head or shoulder)”, “bridge made of a particular material (e.g. stone, iron, wood)”, but that was not done in the present study.

Figure 15: Cross-linguistic representation of meanings

Figure 15 shows which meanings most often have reflexes in the languages of the sample (whether as binominals, other kinds of complex nominal, monomorphemic words or loanwords) and which are least often represented. On average, each
meaning is represented in 77 of the 106 languages, and as expected, body parts dominate the top end, while concepts belonging to the semantic field Modern world dominate the bottom. A bias towards languages spoken in more technologically advanced societies is thus built in to the meaning list and was commented on by a number of contributors. However, there was always a danger that removing this bias might reduce the diversity of structural types. Since the aim of the present study is to explore diversity rather than make predictions, it is arguably better to accept the bias rather than risk removing it.

As for the desideratum of categorial balance, this relates to the possibility of making statistical comparisons across groups of meanings. These are often more reliable if the groups are of roughly the same size and none of them too small. Two kinds of groupings are relevant to the list of meanings: semantic field and semantic type. The distribution of meanings across the set of semantic fields inherited from WOLD is very uneven (Figure 16). As a result, semantic field cannot form the basis of any robust comparisons or typological generalizations (unless these are restricted to, say, The body vs. Modern world).

The distribution across Hatcher’s set of semantic types was also very uneven, as noted above (cf. Figure 17a). (Note that this does not mean that the scheme is unfit for the purpose for which it was intended: non-appositional compounds in general. Since the latter can consist of action- and property-roots as well as thing-roots, types such as Place, Time, Activity and Abstract are more likely to occur than they do in binominals.) In my revised system of semantic types Hatcher’s Place and Time are combined into Location, and Concrete is split into Body part, Natural
phenomenon, Basic technology (or concept), and Advanced technology (or concept). The resulting distribution of the 100 meanings (Figure 17b) is much more balanced and therefore more suitable for statistical analysis.

![Figure 17: Distribution of meanings across semantic types](image)

The 100 meanings shown in Table 18 constitute the *tertium comparationis* of this study. That is, the independent points of comparison that allow data to be compared across languages. Having established this list, data collection could start in earnest. For commensurability with Urban (2012) I had set a goal of 100 languages: the question was, what sample of languages to use, and how to collect the data. These issues are discussed in sections §3.2 and §3.3.

### 3.2 The language sample

#### 3.2.1 Types of sample

The question of language sampling arises from the fact that it is not possible within the scope of a single study to include data from every one of the 7,000 languages that are extant in the world today. This is not just a matter of the limited resources available to typological studies, but also the lack of documentation for most of the world’s languages. Bakker (2011) estimates that for about two thirds of existing languages, no grammar or even grammatical sketch is currently available, and it can be assumed that the situation regarding dictionaries is at least as bad. As a result, typological studies are invariably based on a subset, or sample, of languages. However, as Bakker points out, “there is no such thing as an all-purpose sample. Different kinds of research questions call for different sampling strategies and sample sizes.” There are essentially three kinds of language sample:
The typology and semantics of binominal lexemes

- **probability sample** a (stratified) sample that is areally and genetically as un-biased as possible and that therefore permits the broadest generalizations and predictions
- **variety sample** a sample that is optimized in order to illustrate diversity as fully as possible
- **convenience sample** an (opportunity) sample based primarily on the availability of data

For a **probability sample** the world is divided into a certain number of regions and equal numbers of languages are selected from each region such that none of them belong to the same family or genus. The most well-known and widely-used division into geographical areas is to be found in Dryer (1992) and consists of six macro-areas: Africa, Eurasia (excluding Southeast Asia), Southeast Asia & Oceania, Australia-New Guinea, North America and South America. A revision of this scheme by Hammarström & Donohue (2014), based on purely geographical criteria, was adopted for the 2013 version of WALS (Dryer & Haspelmath 2013) and consists of Africa, Australia, Eurasia, Papunesia, North America and South America. Other schemes include Tomlin’s (1986: 28, 301) set of 26 (mostly) “non-controversial linguistic or cultural areas,” Nichols’ (1992: 25) 10 “large areas which are areally coherent and areally discrete from one another”, and the revision of the latter by Nichols, Witzlack-Makarevich & Bickel (2013) into 10 macro-areas and 24 smaller scale areas.

The division into genetic groupings is complicated by the fact that our knowledge of family relationships among languages is far from complete; by the availability of different, competing classifications; and by the need to select a time-depth that is appropriate to the specific research question. Based on the 625 languages in his database at the time, Dryer (1992) listed 253 genetic groups that are “roughly comparable in time depth to the subfamilies of Indo-European”. These distribute across his six macro-areas as follows: Africa (47), Eurasia (36), Southeast Asia & Oceania (21), Australia-New Guinea (30), North America (71), South America (48). A more extensive and up-to-date genealogical language list is available from WALS and consists of 543 genera divided across 258 top-level language families (“the highest level widely accepted by specialists”) and containing a total of 2,679 languages.

The term **variety sample** appears to have been first used by Rijkhoff et al. (1993), who describe its purpose and significance as follows:
[If] one tries to account for all possible realizations of a certain meaning or structure across languages, like definiteness or relative clause, then the sample should display the greatest possible diversity. This approach is particularly relevant in the greater context of a theory of grammar. In a variety sample (as opposed to a probability sample) it is very important to have cases of the rarest type, since “exceptional types test the theory” (Perkins 1988: 367). If a general theory of grammar is to be universally valid it has to provide for the grammars of all languages, whatever their genetic origin, linguistic type, or geographical location (p. 171).

A sampling technique called Diversity Value (DV) was developed by Rijkhoff and his colleagues for generating variety samples of any predetermined size, based on a language classification chosen by the user (see also Rijkhoff & Bakker 1998). An alternative technique, called the Genus-Macroarea (GM) method, was developed by Miestamo (2003; 2005) and the two are compared in Miestamo, Bakker & Arppe (2016). The details of the two methods need not concern us here, since neither was used in the present study. However, it is worth noting that while DV (unlike GM) involves no areal stratification, the results achieved by the two techniques are quite similar.

The two types of sample discussed above are ideals. In practice, most typological studies make use of a convenience sample. This is because “there may be practical circumstances which force a researcher to just grab the data which happen to be available and sufficiently reliable” (Bakker 2011). As will be seen below, this is largely the case in the present study. However, as Song (2001: 20) points out:

a good number of ground-breaking typological works are based on such convenience samples (e.g. Greenberg 1963; Comrie 1976; Keenan & Comrie 1977; Nichols 1986 *inter alia*). The obvious shortcomings in their samples notwithstanding they did not only provide much insight into the nature of human language, which continues to play an important role in typological research. But, more often than not, they also gave impetus to subsequent large-scale research.

In other words, a convenience sample may still produce valuable results, provided Song’s injunction is borne in mind:

Needless to say, any generalizations or inferences based on such convenience samples should only be taken as what they are – suggestions or preliminary findings concerning cross-linguistic patterns, or language universals – and they should naturally undergo further empirical verification, or revision on the basis of more languages, or more adequately constructed language samples.

Of the typological investigations of compounding and word-formation covered in Chapter 2, only Bauer’s study (§2.1.1) is based on a probability sample; it follows
Dryer’s model almost to the letter (see Table 2 on page 22).¹ Urban’s (2012) study (§2.4.2) is based on a variety sample of 109 languages, constructed using Rijkhoff and Bakker’s technique, from which various subsets are extracted, including:

- a “core sample” of 94 languages, each of which contained at least 65 per cent of the items on his list of 160 concepts; and
- a “statistics sample” of 78 languages, which is a genetically balanced subset of the core sample since it is restricted to one language per family.

In selecting languages for inclusion in WOLD (§2.4.1), the editors attempted to represent the world’s genealogical, geographical, typological, and sociolinguistic diversity, but the overriding factors were practical and, as a result, their language sample is “not ideal”, for a number of reasons (Haspelmath & Tadmor 2009b: 3). However, the authors contend that their sample is preferable to using just one or two languages, or to relying on intuition. (For a map showing the geographical distribution of WOLD languages, see Figure 18 on page 90).

The sample used in Arnaud (2004a) is very much one of convenience, consisting as it does simply of those languages that are represented in Arnaud (2004b). Still, all of Dryer’s six geographical areas are represented (see Table 2 on page 22). The Morbo/Comp database, as noted on page 34, is not balanced (all of its 25 languages are spoken in Eurasia), but this was to some degree inevitable, given the kind of data that was collected (see §2.1.3). Finally, the sample used by Štekauer, Valera & Körtvélyessy (2012), although also clearly one of convenience, is somewhat more areally balanced, and a subset of the “basic sample” provides better genetic balance (see Table 67 on page 490).

3.2.2 The present sample

As far as the present study is concerned, one might assume that a variety sample would be most appropriate. As Bakker (2011) points out, the need for a variety sample “arises when linguistic variables are explored about which not much is known”, and that is exactly the case when it comes to binominal lexemes. However, it quickly became apparent that a variety sample constructed according to the DV and GM methods would not reveal some of the variety known to exist. To give an example, the tables found in Miestamo, Bakker & Arppe (2016: 272–273) show that a 100-language sample constructed using the DV method with the Ethnologue

¹ To be precise, Bauer follows Dryer in using his six geographical areas and selecting six unrelated languages from each area. However, he deviates from Dryer’s method in that his data points are based on individual languages (e.g. Danish) rather than genera (e.g. Germanic).
15 classification would only contain 11 languages from Eurasia; the same method used with the Glottolog classification would have just nine Eurasian languages; and the GM method with the WALS classification would be restricted to a mere eight Eurasian languages. Given that there are 36 genera in Eurasia (according to Dryer 1992), of which 10 are Indo-European, this means that Indo-European would be represented with two or at most three languages, which means that the diversity discussed in §5.6.2 would not be observable.

If a variety sample of the type found in the literature was not appropriate, neither was a probability sample, which would reveal even less diversity. As a result of such considerations, and not least due to resource constraints, a policy of ‘managed opportunism’ was adopted. The 41 WOLD languages (see page 53) constituted the foundation of the sample, and to these were added four of the five languages used for testing the meaning list (§3.1.2).¹ Thereafter, languages were added based on questionnaires completed by language experts and, to echo the words of the WOLD editors, no serious and timely offer to contribute to the database was turned down.

In order to compensate a little for the resulting lack of genealogical, areal and typological balance, a few languages (e.g. Central Yupik, Navajo and Puyuma) were added using dictionaries, but it was not a goal to achieve a fully balanced sample. The resulting convenience sample is known to represent a greater diversity than would have been achieved by following an established sampling method to the letter. Moreover, the inclusion of multiple languages from a single genus permits comparisons at the micro-typological level, such as the investigation in §5.6.2 into how binominal fingerprints correlate with genus. Consequently, I seek solace in Saulwick’s observation that there is “value in providing a micro-typology of interfamilial differences, because these differences aid the typologist in developing a more nuanced appreciation of (cross-)linguistic facts” (2007: 107).

The final sample consists of the 106 languages listed in Table 19 and shown in Figure 18. (The map includes language codes which can be read by zooming in the electronic version of this document.) A more detailed listing, ordered by ISO code, and including the genetic affiliation and geographical location of every language, is to be found in Appendix A. (Genetic affiliation includes genus, as defined by Dryer 1992, and language family, the topmost grouping in Glottolog 2.7, which in some cases are one and the same.) Sources for each language are listed in Appendix B. In the next section I describe and evaluate the sample in more detail.

¹ All except Nizaa, for which too little data was available.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Language</th>
<th>Language</th>
<th>Language</th>
<th>Language</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Äiwoo</td>
<td>Akkadian</td>
<td>Amharic</td>
<td>Anindilyakwa</td>
<td>Archi</td>
</tr>
<tr>
<td>Assamese</td>
<td>Baa</td>
<td>Bambara</td>
<td>Bandial</td>
<td>Barain</td>
</tr>
<tr>
<td>Basque</td>
<td>Bezhta</td>
<td>Cabécar</td>
<td>Caijia</td>
<td>Central Yupik</td>
</tr>
<tr>
<td>Ceq Wong</td>
<td>Chakali</td>
<td>Croatian</td>
<td>Czech</td>
<td>Datooga</td>
</tr>
<tr>
<td>Dutch</td>
<td>English</td>
<td>Estonian</td>
<td>Finnish</td>
<td>French</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>Gawwada</td>
<td>German</td>
<td>Greek</td>
<td>Gurindji</td>
</tr>
<tr>
<td>Harakmbut</td>
<td>Hausa</td>
<td>Hawaiian</td>
<td>Hebrew</td>
<td>Hindi</td>
</tr>
<tr>
<td>Hmong Daw</td>
<td>Ho-Chunk</td>
<td>Hungarian</td>
<td>Hupdê</td>
<td>Imbabura Quechua</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Iraqw</td>
<td>Irish</td>
<td>Italian</td>
<td>Japanese</td>
</tr>
<tr>
<td>Kalamang</td>
<td>Kam</td>
<td>Kambaata</td>
<td>Kanuri</td>
<td>Kekchí</td>
</tr>
<tr>
<td>Ket</td>
<td>Kildin Sami</td>
<td>Korean</td>
<td>Kupsabiny</td>
<td>Latvian</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>Lower Sorbian</td>
<td>Malagasy</td>
<td>Malayalam</td>
<td>Maltese</td>
</tr>
<tr>
<td>Mamara Senoufo</td>
<td>Manange</td>
<td>Mandarin Chinese</td>
<td>Mapudungun</td>
<td>Mbyá Guaraní</td>
</tr>
<tr>
<td>Murui Huitoto</td>
<td>Navajo</td>
<td>Nepali</td>
<td>Norwegian</td>
<td>Old High German</td>
</tr>
<tr>
<td>Oroqen</td>
<td>Polish</td>
<td>Puyuma</td>
<td>Querétaro Otomi</td>
<td>Romanian</td>
</tr>
<tr>
<td>Russian</td>
<td>Saramaccan</td>
<td>Selice Romani</td>
<td>Seri</td>
<td>Seychelles Creole</td>
</tr>
<tr>
<td>Sidamo</td>
<td>Slovak</td>
<td>Somali</td>
<td>Srenge</td>
<td>Swahili</td>
</tr>
<tr>
<td>Tagalog</td>
<td>Takia</td>
<td>Tarifit</td>
<td>Thai</td>
<td>Ticuna</td>
</tr>
<tr>
<td>Trinitario</td>
<td>Turkish</td>
<td>Tuwari</td>
<td>Vietnamese</td>
<td>Walman</td>
</tr>
<tr>
<td>Warta Thuntau</td>
<td>Wawa</td>
<td>Welsh</td>
<td>Western Farsi</td>
<td>Western Mari</td>
</tr>
<tr>
<td>Wichi</td>
<td>Wik-Mungkan</td>
<td>Wolof</td>
<td>Yakut</td>
<td>Yaqui</td>
</tr>
<tr>
<td>Zinacantán Tzotzil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 19: Complete list of languages in the sample

Figure 18: Areal distribution of language sample
3.2.3 Overall evaluation

The sample consists of 106 languages distributed across 42 families and 72 genera. On average there are 1.47 languages per genus, 2.52 languages per family, and 1.7 genera per family, but the distribution is uneven (Table 20). Two families, Indo-European and Afro-Asiatic, account for 34% of the languages in the sample and 17% of the genera, and 28 families are represented by just one language. Of the 72 genera, 18 are represented by more than one language (26).

(26) Slavic (6), Cushitic (5), Germanic (5), Indo-Aryan (4), Semitic (4), Oceanic (3), Romance (3), Baltic (2), Bantoid (2), Celtic (2), Chadic (2), Core Mayan (2), Finnic (2), Gur (2), Nilotic (2), North-Central Atlantic (2), Nuclear Torricelli (2), Turkic (2) family languages genera

<table>
<thead>
<tr>
<th>family</th>
<th>languages</th>
<th>genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Indo-European</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>2. Afro-Asiatic</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>3. Atlantic-Congo</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>4. Austronesian</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>5. Uralic</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6. Altaic</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>7. Nilo-Saharan</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8. Sino-Tibetan</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9. Austro-Asiatic</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10. Mayan</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11. Nakh-Daghestanian</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12. Nuclear Torricelli</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13. Pama-Nyungan</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14. Pidgins &amp; Creoles</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15. Other families</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 20: Genetic distribution of languages

The genetic balance of the sample could be restored by selecting one language from each of the genera in (26). This would result in a sample of 72 languages, none of which belong to the same genus. In order to fulfil the requirements of a probability sample, that sample would then have to be further reduced to account for geographical distribution since, as Table 21 shows, the sample is unbalanced in that respect too.
A quarter of the sample is accounted for by Africa and over a third by Eurasia. (Pidgins and Creoles are ignored in this context since their genetic relationships cross-cut areal boundaries: Saramaccan and Seychelles Creole, which have Eurasian lexifiers, are found in South America and off the coast of Africa, respectively.) Of the six large geographical areas, NG/Australia is least well represented, with eight languages in all, and thus a balanced subset of the current sample could contain no more than \(6 \times 7 = 42\) languages.

Defining such a subset has not been a priority in the present study since the focus has been on exploring diversity rather than making predictions. In future research, efforts will be made to expand the representation of NG/Australia, North America and South America. Data from 11 more languages (five from NG/Australia, four from North America and two from South America) would suffice for a balanced sample of 72 languages (12 per area), but this would involve discarding data from 27 Eurasian, 12 African and two Oceanian/SE Asian languages that are already in the database and might, as noted above, lead to some loss of variety.

### 3.3 Data collection

This section describes the principal sources of data: the WOLD database (§3.3.1), questionnaires (§3.3.2) and dictionaries (§3.3.3). Additional data and information concerning attributive possession that forms the basis of the analysis in §7.2 was mostly collected from grammatical descriptions, as documented in Appendix D. §3.3.4 describes the kind of clean-up that was performed on the data. In the spirit of the “re-doing typology” debate in *Linguistic Typology* 10(1), this section contains some technical detail which is intended to aid the further reuse of the WOLD data and enable replication of the present study.
3.3.1 Online database

The WOLD database, described in §2.4.1, was used not only to generate the initial list of 159 meanings (as described in §3.1.1), but also to furnish the data for 82 of the 100 meanings covering the 41 languages listed on page 53. There are a number of sources for the WOLD data. At the outset of this project (2015) two datasets could be downloaded from the WOLD website at http://wold.clld.org: a CLDF file and a two-file dataset in RDF format. However, neither of these were complete, and both lacked the all-important field containing the gloss, which I needed in order to be able to analyse complex forms. The GitHub site at http://github.com/clld/wold2 contained a file called data.zip which held the complete data set in the form of 48 CSV and JSON files (Haspelmath, Tadmor & Forkel 2010). These included the glossing information, but it was buried inside a JSON field and inaccessible to most tools. That problem could be solved by importing the data into a relational database that supports JSON, but doing so from a set of 48 files would be time-consuming and error-prone. Fortunately, the data was available in the form of a PostgreSQL dump in another project at http://github.com/clld/wold-data. This file, once downloaded and uncompressed, could be loaded directly into my database system.1

1 It was now possible to extract an initial data set from WOLD using the SQL query shown on the next page (28). This finds every translation equivalent (word) for each of the 100 meanings present in WOLD, and for each such word, the query selects the identifier, language, meaning, word and gloss. The result of the query, a table containing 3,326 rows, was written to a CSV file of 3,422 lines.

The discrepancy between the number of rows in the query result and the number of lines in the CSV file was due to the fact that some cells in the table contained a new line (→), for example the gloss for the Hausa word meaning NOSTRIL (27). Unfortunately, Microsoft Excel erroneously interprets a new line that occurs inside a quoted string as a record end, so this file would not import cleanly. The CSV file was therefore edited using regular expressions to replace such new lines with spaces.2

2 The search string was `"[^\r\n]+"` and the replace string `\1\2` (where \1 denotes a word space). The search expression picks out unclosed quoted strings at the end of a line and replaces the new line with a space. The \r was necessary to avoid picking up end-of-line cells that contained a final semicolon. This works because rows are delimited by the two-byte sequence CR-LF (i.e. \r\n), whereas new lines inside cells are LF only (i.e. \n).

---

1 I used PostgreSQL 9.4 in order to ensure maximum compatibility with the data format.
2 The search string was `/\A[^\r\n]+\n/` and the replace string `/\1/` (where \1 denotes a word space). The search expression picks out unclosed quoted strings at the end of a line and replaces the new line with a space. The \r was necessary to avoid picking up end-of-line cells that contained a final semicolon. This works because rows are delimited by the two-byte sequence CR-LF (i.e. \r\n), whereas new lines inside cells are LF only (i.e. \n).
The typology and semantics of binominal lexemes

(28) select
    u.id as id,
    l.name as language,
    p.name as meaning,
    u.name as word,
    u.jsondata::json->'gloss' as gloss
from
    meaning as m,
    word as w,
    unit as u,
    counterpart as cp,
    value as v,
    valueset as vs,
    parameter as p,
    language as l
where
    w.pk = u.pk
and m.pk = p.pk
and cp.word_pk = w.pk
and cp.pk = v.pk
and v.valueset_pk = vs.pk
and vs.language_pk = l.pk
and vs.parameter_pk = p.pk
and m.semantic_category = 'Noun'
and (  
    p.name = 'the ankle' or
    p.name = 'the arctic lights' or
    [ ... 95 lines omitted ... ]
    p.name = 'the windmill' or
    p.name = 'the wrist' or
    p.name = 'the yolk'
)
order by language, meaning, word

Once this issue had been rectified, the file could be imported into Excel as a four-column table (Figure 19).

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>language</td>
<td>meaning</td>
<td>word</td>
</tr>
<tr>
<td>2</td>
<td>Hausa</td>
<td>the nostril</td>
<td>kàfář háncìi</td>
<td>kàfá-r háncii [orifice-GEN nose]</td>
</tr>
<tr>
<td>3</td>
<td>Swahili</td>
<td>the nostril</td>
<td>mwanzi wa pua</td>
<td>bamboo of nose</td>
</tr>
<tr>
<td>4</td>
<td>Swahili</td>
<td>the nostril</td>
<td>tundu la pua</td>
<td>hole of nose</td>
</tr>
<tr>
<td>5</td>
<td>Tarifiyt Berber</td>
<td>the nostril</td>
<td>tìnzā</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Thai</td>
<td>the nostril</td>
<td>ruucamùuk</td>
<td>ruu-camùuk [hole-nose]</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 19: Extract of the data after import

The WOLD database furnished translation equivalents for 82 of the 100 meanings selected for this project. To collect data from the 41 WOLD languages for the 18 meanings not found in WOLD, I used two methods. Dictionaries were consulted where available (and where I had knowledge of the script); this was the case for 10
of the 41 languages, viz. Dutch, English, Hausa, Hawaiian, Indonesian, Iraqw, Japanese, Kanuri, Romanian and Swahili. For the remaining 31 languages the original WOLD contributors were contacted by email. 15 of these responded and were willing to help. This low response rate (bearing in mind that contributors were only being asked to provide translations for 18 meanings) gave a strong indication that cold calling potential contributors by email would not be a generally viable method for collecting large amounts of data.

One important observation that was made during this process was that the number of meanings for which translation equivalents existed varied considerably from one language to another. In the case of Kanuri only two of the 18 were found in the source (reké SUGAR CANE and fómfom WATER PUMP) and neither of these were analysable. For Iraqw, only three meanings were found in the dictionary (miwa SUGAR CANE, fura TOOTHBRUSH and yaaér doori MILKY WAY), and of these only the latter could be analysed [river:CON sky].

On the other hand, all 18 of the additional meanings had counterparts in Dutch and all of them were analysable. Once again, this highlighted a bias in the meaning list due to the inclusion of concepts that are only familiar in industrialized societies. At this stage of the project, however, there was nothing that could be done to rectify this.

3.3.2 Questionnaires

The experience of using dictionaries to supplement the WOLD data (§3.3.1), and even more that of adding complete data sets from Welsh and Turkish dictionaries in order to test the meaning list (§3.1.2), made it clear that I would not be able to collect data for a further 50 languages by my own efforts in the time available (the reasons are discussed in the next section). It was therefore decided to solicit the help of other linguists. A questionnaire was devised that consisted of two worksheets in an Excel workbook, the first containing a page of instructions and the second the list of meanings, along with French, Spanish and Russian translation equivalents, for use when the contact language is other than English (Appendix F). Contributors were asked to provide the most canonical translation equivalent for each meaning, using the Latin script (or IPA) and the native script (where applicable), together with a morpheme gloss for morphologically complex translation equivalents, using a modified version of the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel
It was also possible to add comments regarding the source of loans and calques.

On the basis of the four languages that I had added myself, I estimated that it would take a minimum of eight hours to complete the original 201 meaning questionnaire, depending on how well the contributor knew the language and its morphological complexity. As mentioned above, it was clear from the experience of contacting colleagues by email to supplement the WOLD data with an additional 18 meanings that the response rate from cold calling for a questionnaire of this size would be very low indeed. I therefore recruited contributors through networking at various events (summer schools, conferences, etc.) during the course of the project.

The first version of the questionnaire, containing all 201 meanings, was completed by 25 contributors before being replaced by the version containing 100 meanings, which was completed by a further 40 colleagues. I owe these linguists an enormous debt. Without them the present project would not have been possible.

Of course, this part of the data collection was not all plain sailing. A number of people who agreed to supply data failed to deliver, and others had to be prodded several times. A very small amount of the data turned out to be unusable because the contributor lacked the expertise to perform the morphological glossing, but all in all the quality of the data received was sufficient for the purpose to hand.

3.3.3 Dictionaries

Using bilingual dictionaries to collect data for a project of this kind poses a variety of challenges relating to (i) the availability of such a dictionary, (ii) the familiarity of the researcher with the script it employs, (iii) its size, and (iv) the amount of morphological detail it contains.

The first point is obvious: If no dictionary or wordlist exists for a language, the only way to collect data is from a speaker, usually by means of a questionnaire. Moreover, the dictionary must be two-way (or else there must exist two one-way dictionaries that complement each other). This is because using a dictionary to collect and analyse data is a two-step process: a meaning (e.g. RAILWAY) is first looked up in the source-to-target section (e.g. English-Malagasy); then the translation equivalent (here, lalamby) must be looked up in the target-to-source section (e.g. Malagasy-English), in order to understand whether it is monomorphemic or polymorphemic and, if the latter, what its constituents are.

---

1 See Typographical and naming conventions on page xxv for the modified glossing rules.
Knowledge of the script is also an absolute prerequisite for being able to utilize a dictionary, again for obvious reasons. Thus, for the present researcher it was not possible to supplement the WOLD data for languages like Thai and Chinese, despite the availability of good dictionaries, because of insufficient familiarity with the scripts. It was also not possible to collect new data for languages such as Arabic, Hebrew and Persian, or languages where the dictionary used an unfamiliar contact language, e.g. Russian. For such languages, questionnaires were the only solution.

Size matters when using a dictionary because it is important that the data be as complete as possible, in order to reveal the full variety of constructions. As a case in point, only one of the 18 additional meanings used to supplement the WOLD data (shown in bold in Table 18 on page 78) was found in the dictionary used for Malagasy (Vaovao 1969), viz. lalamby RAILWAY (which was actually found under ‘railroad’). Whether this is because the language truly does not have words for the other 17 meanings, or because a dictionary of 118 pages was too small to include them, is hard to say. At any rate, such uncertainty is less likely to arise when a more comprehensive dictionary is available.

Finally, the more morphological information contained in the dictionary, the easier it is to determine the structure of complex words. To illustrate this point, take once more the case of lalamby: Looking up the word in a Malagasy-English dictionary (Richardson 1885) confirms that it does indeed mean ‘railway’, but no additional information is provided. Close by one finds lalana, meaning ‘a road, a way, a path’, which is a typical constituent in many words for railway, but there is no entry for *by, which might otherwise be a candidate for the second constituent. A conjecture that the second constituent might mean ‘iron’ is confirmed by looking up the latter in the English-Malagasy dictionary, where it is translated as vy. The difference is not an issue, since it can be assumed that assimilation is occurring. A more serious problem is that we are none the wiser regarding the actual process that takes place when lalana and vy combine to produce lalamby. Fortunately, the answer was to be found in the only grammar available to the present researcher (Parker 2014) – a 66-page reprint from 1883 – which mentions “euphonic changes” (including replacing v with b), which occur “when n or m is inserted between two words as the sign of an indefinite possessive or ablative case”, and the fact that “final syllables -na, -ka, and -tra are contracted sometimes by rejection of the final syllable” (pp. 8-9, see also p. 33). This process occurs in several of the Malagasy words found in WOLD, e.g. tràno.N. hàla [house.GEN.spider] SPIDER WEB. Applying the same pattern, the word could be glossed as [road.GEN.iron]. The point being made is that
arriving at the correct solution would have required much less effort if the dictionary had included morphological information of the kind illustrated in (29).

(29) railroad N lalamby < lalana ‘road, way, path’ + -N- (POSS) + vy ‘iron’

The need for such detailed information becomes even more acute with languages such as Central Yupik and Navajo, that have highly complex morphology.

3.3.4 Data cleansing

Before the analysis could proceed some limited data cleansing was required. The WOLD data was mostly accepted in the form found in the database, but obvious typos and coding errors were corrected and consistency enforced. In addition, all glosses were revised to conform to the conventions used throughout this project: that is, the use of standard abbreviations, and the use of periods instead of hyphens to mark morpheme boundaries (see Typographical and naming conventions on page xxv). Thus, for example, AG.N, which is used to denote nomen agentis in the Iraqw vocabulary, was changed to AGT (‘agentive’), while Takia yu sa-n byouŋ [war possession-3SG things] WEAPONS was changed to yu sa.n byouŋ [war POSS.3SG things].

In many cases the glossing did not follow the guidelines set down by WOLD, which was to provide “a morpheme-by-morpheme gloss, i.e. a hyphenation and a gloss in square brackets” (Haspelmath & Tadmor 2009b: 12) and to use standard abbreviations given in the Leipzig Glossing Rules and/or Croft (2003: xix). For example, the two items in (30a) were amended as shown in (30b), both for consistency and to enable automated processing of the database.

(30) a. Eng. nostril [NOSE + archaic THIRL ‘hole’]
   Otomí ‘yoda [’yo = ?; da = eye]

b. Eng. nos.tril [nose.hole]
   Otomí ‘yo.da [??.eye]

Sometimes the glossing was not even internally consistent. For example, in the Takia data the word sa-n occurs four times, glossed twice as [its] and twice as [POSS-3SG]. For consistency, all four occurrences were amended to [POSS.3SG]. Furthermore, the glossing was not always accurate, as demonstrated by the Malagasy words discussed in the preceding section, where the abbreviation GEN (genitive) was used instead of PER (pertensive) or CON (construct) to gloss a suffix that marks the possessee rather than the possessor. These were also corrected manually. In at least one case the wrong semantic category had been assigned to a meaning.
Thus **TO ULULATE** is misclassified as Noun throughout the original database.¹ Some mono-morphemic words were classified as analysable by the contributor, e.g. Vie. *mí EYELID*. In such cases the classification was simply changed to unanalysable. Moreover, some words were classified as analysable rather than semi-analysable or unanalysable even though they are not analysable to present-day lay speakers, which was the criterion that should have been applied (see Haspelmath and Tadmor 2009: 12). A case in point is the *nostril* example given in (30) above. While native speakers have an intuition that this word is related formally (as well as semantically) to *nose*, none would see it as the combination of words meaning NOSE and HOLE. It should therefore not have been classified as analysable in the database. Despite this, such instances were retained on the grounds that it would not have been possible for me to double-check every single gloss: I needed to trust my contributors. Finally, certain words were classified as (fully) analysable even though the gloss shows that one or more components is opaque, as was the case with the Otomo example in (30). These were reclassified as semi-analysable. Data provided via the questionnaire required similar treatment, despite the detailed instructions that had been provided to the contributors (see Appendix F on page 485).

### 3.4 Chapter summary

In this chapter I have documented how I arrived at the list of meanings, presented the language sample and described how the data were collected and cleaned.

The meaning list plays a crucial role in a project such as this, since it basically determines what kind of data will be collected. I therefore went to considerable pains to ensure that it was constructed in a principled manner and that it was maximally fit for purpose. Even so, with hindsight it can be seen that the list could have been improved in a number of ways. I hope that, by providing detailed documentation of the process and an evaluation of the results, I have made it easier for researchers who wish to use the onomasiological approach to avoid some of my mistakes.

The other factor that is crucial in determining the shape of the data is the language sample. My sample was neither genetically nor areally balanced, despite covering 72 genera spread across the whole world; nor was it a variety sample in the sense of Rijkhoff, Bakker and Miestamo. But this, I argued, was not a problem, given that my research questions did not involve making predictions, and was in some

¹ See [http://wold.cld.org/meaning/18-99912#6/5.500/56.000](http://wold.cld.org/meaning/18-99912#6/5.500/56.000). This did not affect the present project since the meaning did not appear in either the original set of 159 WOLD meanings or the additional set of 18 meanings.
sense even an advantage, since it allowed me to make some microtypological comparis-sons that would not otherwise have been possible.

Also in this chapter I described some of the pros and cons of various data collection methods (viz. open databases, dictionaries and questionnaires). None of these are perfect, they each present their own challenges, but together they provided me with a set of over 10,000 data points which I make available via TROLLing, the Tromsø Repository of Language and Linguistics. In the next chapter I describe how the data were annotated and engage with some of the theoretical issues that arose.
Chapter 4. Data annotation

4 Data annotation

The annotation (or coding) of the data consisted of three parts: (i) distinguishing binominals from other words (§4.1), (ii) determining the head of each binominal (§4.2), and (iii) establishing the type of construction (§4.3). All three tasks were first performed on an early version of the database consisting of 201 meanings and data from 50 languages. (Recall that annotation of the first 50 languages provided the basis for reducing the number of meanings from 201 to 100 in a principled manner, as I documented in the preceding chapter.) After reducing the meaning list from 201 to 100, the same tasks were performed on data from a further 56 languages.

In this chapter I describe each task in the context of theoretical issues that arose. The latter include: distinguishing roots from affixes; distinguishing thing-roots from action-roots and property-roots; distinguishing thing-affixes from action-affixes and property-affixes (and the differential treatment of the three); the notion of ‘head’ in the context of binominals consisting of two thing-roots and binominals consisting of a thing-root and a thing-affix; and how to generalize the structures exhibited by binominals into more schematic (language-specific) constructions. To conclude, I present analytics regarding vocabulary size, morphological complexity and binominal frequency, both cross-linguistically and across meanings (§4.4).

4.1 Identifying binominals

The first annotation task consisted in a rough categorization of each word based on its onomasiological type (cf. page 11). Words consisting primarily of two thing-morphs (cf. §1.2.4) were assigned the code nn. Words consisting of three or more thing-morphs can usually be decomposed into binary constituents and tend to share the same structure as binominals, but they are not binominals *sensu stricto* and were therefore assigned the codes 3n or 4n in order to enable closer examination at a later stage. Binominals whose constituents are in a coordinate relation were assigned the code co, despite consisting of two thing-morphs, since coordinate compounds had been ruled out of scope in this study (see page 9). Most analysable words containing an actional element were flagged v, but those consisting of one
action-root and two thing-roots, e.g. Vie. bữ ăn sáng [meal eat morning] BREAKFAST, were coded separately as NVN, in order that they could be revisited later. In addition, certain other codes were applied, as shown in Table 22.

<table>
<thead>
<tr>
<th>code</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>words consisting of two thing-roots (OT3) – binominal lexemes</td>
</tr>
<tr>
<td>3n</td>
<td>words consisting of three thing-roots</td>
</tr>
<tr>
<td>4n</td>
<td>words consisting of four thing-roots</td>
</tr>
<tr>
<td>CO</td>
<td>coordinate constructions</td>
</tr>
<tr>
<td>V</td>
<td>form containing a verbal (OT1 or OT2)</td>
</tr>
<tr>
<td>NVN</td>
<td>words consisting of two lexical nominals and a verbal (OT1)</td>
</tr>
<tr>
<td>XX</td>
<td>words that require closer analysis</td>
</tr>
<tr>
<td>x</td>
<td>other analysable word (not relevant for this study, many OT4)</td>
</tr>
<tr>
<td>sa</td>
<td>semi-analysable word</td>
</tr>
<tr>
<td>un</td>
<td>unanalysable word</td>
</tr>
</tbody>
</table>

*Table 22: Preliminary classification of structural types*

The coding is illustrated below in Figure 20, which shows a subset of the Polish data. Some rows and columns have been omitted and the contents of the columns *word* and *gloss* have been simplified. (In the database, *word* does not contain morpheme breaks; these are instead shown in *gloss* (thus, for example, cells F5897 and H5897 actually read "kolej żelazna" and "kolej żelazna [course iron.ADZ]").

Columns c1 and c2 show the first and second (major) constituent, respectively.

The reason for adding structural codes to the data was to identify binominals. As explained in §1.2.4, these are defined as lexical items that consist primarily of two thing-morphs. The latter are in turn defined as morphs that profile (or denote) things, prototypically physical objects (animate or inanimate). The task therefore consists in identifying thing-morphs and differentiating them from other kinds of morph, in particular action-morphs – that is, morphs that profile actions.

In the discussion that follows I consider first thing-roots and thereafter thing-affixes. Croft (2000) points out that the division between root and affix, has been assumed to be clear-cut: “In fact it is not, and the chief reason for this fact is the diachronic process of grammaticalization, which causes root morphemes bearing lexical meaning to acquire grammatical meaning, generally becoming affixes in the process.”

---

1 There is also a note (not shown here) indicating that this form is dated. The modern word is kolej.
Table 1: Subset of the Polish data

<table>
<thead>
<tr>
<th>word</th>
<th>gloss</th>
<th>meaning</th>
<th>str</th>
<th>c1</th>
<th>c2</th>
<th>pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>zorza polar.na</td>
<td>[dawn pole.ADJZ]</td>
<td>arctic lights</td>
<td>NN</td>
<td>dawn</td>
<td>pole</td>
<td>L</td>
</tr>
<tr>
<td>past.uch</td>
<td>-</td>
<td>beehive</td>
<td>un</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>śniad.anie</td>
<td>[eat_breakfast.NMLZ]</td>
<td>breakfast</td>
<td>V</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>platek uch.a</td>
<td>[lobe ear.GEN]</td>
<td>earlobe</td>
<td>NN</td>
<td>lobe</td>
<td>ear</td>
<td>L</td>
</tr>
<tr>
<td>wosk.o.wina</td>
<td>[wax.LE.NMLZ]</td>
<td>earwax</td>
<td>NN</td>
<td>wax</td>
<td>NMLZ</td>
<td>R</td>
</tr>
<tr>
<td>dziewczyn.ka</td>
<td>[girl.DIM]</td>
<td>girl</td>
<td>NN</td>
<td>girl</td>
<td>DIM</td>
<td>R</td>
</tr>
<tr>
<td>złot.y pierścionek</td>
<td>[gold.ADJZ ring]</td>
<td>gold ring</td>
<td>NN</td>
<td>gold</td>
<td>ring</td>
<td>R</td>
</tr>
<tr>
<td>pastuch</td>
<td>[mind/graze.AGT]</td>
<td>herdsman</td>
<td>V</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ryb.ak</td>
<td>[fish.AGT]</td>
<td>fisherman</td>
<td>NN</td>
<td>fish</td>
<td>AGT</td>
<td>R</td>
</tr>
<tr>
<td>poł.u.dnie</td>
<td>[half.LE.day]</td>
<td>midday</td>
<td>NN</td>
<td>half</td>
<td>day</td>
<td>R</td>
</tr>
<tr>
<td>droga mlecza.na</td>
<td>[road milk.ADJZ]</td>
<td>milky way</td>
<td>NN</td>
<td>road</td>
<td>milk</td>
<td>L</td>
</tr>
<tr>
<td>sutek</td>
<td>-</td>
<td>nipple or teat</td>
<td>un</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>nozdrze</td>
<td>-</td>
<td>nostril</td>
<td>un</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>poczt.ów.ka</td>
<td>[post.ADJZ.DIM]</td>
<td>postcard</td>
<td>NN</td>
<td>post</td>
<td>DIM</td>
<td>R</td>
</tr>
<tr>
<td>garnc.arz</td>
<td>[pot.AGT]</td>
<td>potter</td>
<td>NN</td>
<td>pot</td>
<td>AGT</td>
<td>R</td>
</tr>
<tr>
<td>król.owa</td>
<td>[king.F]</td>
<td>queen</td>
<td>NN</td>
<td>king</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td>kolej żelaz.na</td>
<td>[course iron.ADJZ]</td>
<td>railway</td>
<td>NN</td>
<td>course</td>
<td>iron</td>
<td>L</td>
</tr>
<tr>
<td>kolej</td>
<td>-</td>
<td>railway</td>
<td>un</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>pajęcz.yna</td>
<td>[spider.F]</td>
<td>spider web</td>
<td>NN</td>
<td>spider</td>
<td>F</td>
<td>R</td>
</tr>
<tr>
<td>kamień.ny most</td>
<td>[stone.ADJZ bridge]</td>
<td>stone bridge</td>
<td>NN</td>
<td>stone</td>
<td>bridge</td>
<td>R</td>
</tr>
<tr>
<td>palec u nogi</td>
<td>finger PREP leg</td>
<td>toe</td>
<td>NN</td>
<td>finger</td>
<td>leg</td>
<td>L</td>
</tr>
<tr>
<td>pociąg</td>
<td>-</td>
<td>train</td>
<td>un</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>wino.rosł</td>
<td>[wine.grow:NMLZ]</td>
<td>vine</td>
<td>V</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>wiatr.ak</td>
<td>[wind.NMLZ]</td>
<td>windmill</td>
<td>NN</td>
<td>wind</td>
<td>NMLZ</td>
<td>R</td>
</tr>
<tr>
<td>nad.garst.ek</td>
<td>[over.handful.DIM]</td>
<td>wrist</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>żółt.ko</td>
<td>[yellow.DIM]</td>
<td>yolk</td>
<td>NN</td>
<td>yellow</td>
<td>DIM</td>
<td>R</td>
</tr>
</tbody>
</table>

Figure 20: Subset of the Polish data

Having despaired of the task of distinguishing between ‘words’ and ‘affixes’ in his 2011 paper on the indeterminacy of word segmentation, Haspelmath (2019) offers the follow definition of ‘affix’:

(31) **affix**: a non-promiscuous bound form which is not a root

This of course presupposes a definition of ‘root’, which Haspelmath (2012: 123 fn. 9) defines in passing as ‘morphs that denote things, actions, or properties’. This
makes the assumption that only roots can denote things, actions and properties, which conflicts with the position taken in the present study that nominal affixes denote things, verbal affixes denote actions, and adjectivizers denote properties (cf. §1.2.4). Haspelmath’s definition of affix thus breaks down even before we attempt to interpret the notions of ‘promiscuity’ and ‘boundness’.

Croft (2001) discusses the distinction between root and affix in the context of “heads” and roots in morphology, and argues that the semantic notion of primary information bearing unit (PIBU), but not that of profile equivalent (roughly, “head”), is “relevant to morphological organization at the word level with respect to the root-affix distinction” (p.268); however, he stops short of providing a definition of affix. There are thus no widely accepted cross-linguistic definitions for roots and affixes, but for present purposes the kind of standard definitions given in introductory textbooks like Haspelmath & Sims (2010) will suffice (32). We will have to deal with certain in-between phenomena, such as affixoids (see page 164), but these are not very frequent and will not affect the overall analysis to any great extent.

(32)   affix: a morpheme that must attach to a base and cannot occur by itself
   root: a base that cannot be analysed further – i.e. a base that consists of a single morpheme
   base: the base of a morphologically complex word is the element to which a morphological operation applies

4.1.1 Identifying thing-roots
Most of the time it is fairly straightforward to identify roots that profile things and distinguish them from roots that profile actions or properties. This can be demonstrated using the data in Figure 20. Referring to the English glosses in column G, we observe that all the following denote things (some less prototypically than others):

    pole, (breakfast), lobe, ear, wax, girl, gold, ring, (day), road, milk, pot, king, course, iron, spider, stone, bridge, finger, leg, wine, wind, handful

Similarly, the following denote actions (in a broad sense that includes states):

    eat, mind/graze, grow

Some of the other glosses are less clear-cut. For example, fish can denote either a thing or an action – as can dawn and post; yellow can denote a property or a thing (as in ‘the yellow of an egg’) – as can half. The problem of distinguishing things

---

1 This concept is discussed in §4.2.
from actions and properties is analogous to that of determining the direction of conversion from one word class to another, for which Marchand (1964) developed a set of eight criteria. Of these, the most important is that of semantic dependence, as this is “as often as not sufficient in itself to solve the question” (p. 10).\(^1\)

According to the criterion of semantic dependence, “the word that for its analysis is dependent on the content of the other pair member is necessarily the derivative”. Several illustrations are provided:

The verb saw must be derived from the substantive saw. Saw sb is satisfactorily defined as ‘a cutting instrument with a blade, having a continuous series of teeth on the edge’. That the instrument may be used for the action of sawing need not be included in the definition. On the other hand, the content analysis of the verb must necessarily include the semantic features of the substantive saw: saw vb ‘use a saw, cut with a saw’. — The verb knife is naturally analysable as ‘wound with a knife’ whereas the substantive knife does not lean on any content features of the verb knife, which does not exist in the vocabulary of many speakers who commonly use the noun. — A parallel case we have in telephone vb and telephone sb. For its analysis, the verb relies on the semantic features of the substantive telephone. — Though seemingly parallel to telephone and saw the case of whistle sb (the name of the instrument) with regard to whistle vb is the reverse. The analysis of the verb does not call for any semantic features of whistle sb (the instrument). Whistling is aptly described by ‘forcing the breath through the teeth or compressed lips’ whereas the instrument whistle has for its explanation recourse to the content features of the verb: whistle ‘instrument used for whistling’ (pp. 12-13).

In the absence of other evidence, the criterion of semantic dependence can be used to distinguish Things from other kinds of denotation. Thus fish as Action is dependent on fish as Thing (it can only be defined through reference to fish as Thing) and therefore the latter is primary. In the same way, yellow as Thing is dependent on yellow as Property (it can only be defined through reference to its colour), which is thus primary.

Sometimes, though, other evidence compelled a different diagnosis to that suggested by semantic dependence. The most important was the test used to determine the semantic head (described below). When yellow combines with egg to denote yolk, as is the case in, for example, Welsh melyn.\(\ddot{w}\)y [yellow.egg], it is clear that yellow is the head (since a yolk ISA yellow something and not an egg). This fact suggests that, in this particular case, yellow should be regarded as denoting a Thing (“the

\(^1\) The other criteria are restriction of usage, semantic range, semantic pattern, phonetic shape, morphologic type and criterion of stress.
yellow bit”) and not a Property (even if, as a thing, it is largely defined in terms of a property). Turning briefly from semantic types to word classes, what this in effect means is that adjectives – and not only relational adjectives, but also qualitative adjectives – are somehow slightly more acceptable as denoters of things than verbs. This can be explained through reference to Givón’s (2001: 54) scale of temporal stability (Figure 21).

<table>
<thead>
<tr>
<th></th>
<th>most stable</th>
<th></th>
<th>least stable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tree, green</td>
<td>sad, know</td>
<td>work, shoot</td>
</tr>
<tr>
<td>noun</td>
<td>adj</td>
<td>adj, verb</td>
<td>verb, verb</td>
</tr>
</tbody>
</table>

*Figure 21: Givón’s scale of temporal stability*

Givón’s claim is that

Nouns, verbs and adjectives may be placed on the scale of time-stability of coherently bundled experience… The prototype of the class noun occupies the most time-stable end of the scale… Prototype verbs occupy the other end of the time-stability scale…

While prototype nouns code bundles of experienced features (‘horse’, ‘chair’, ‘woman’, ‘tree’), the cognitive status of adjectives is a bit more murky… Many languages do not code durable single properties of nouns as adjectives, but rather as verbs (see examples further below) and occasionally even nouns. But as Dixon (1982) has noted, if a language has the lexical category adjective at all, it tends to include at least the most durable physical properties of prototype nouns: size, shape, color, consistency, texture, weight, smell, taste. This supports our view (and Bertrand Russell’s) that prototype adjectival concepts are abstracted from the direct experience of prototype noun-coded entities. **It also explains why prototype adjectives occupy the same extreme time-stable end of our temporal stability scale as prototype nouns** (pp. 50-54; emphasis added).

This provides theoretical justification for regarding certain prototypical adjectives, such as colours, as thing-roots, at least when they behave like nouns as the head of a nominal construction. Additional confirmation comes from the scale of property concepts presented by Croft (forthc.) which shows colour terms to be among those most likely to recruit object modification constructions, cf. (145) on page 333.

Another kind of evidence comes from the word class of the item in question. This cannot always be determined, but sometimes the morphological shape of the item (or some other element of the word) provides an indication of the word class. In the
Äiwoo word *me.ki.tei* [person.IPFV.fish] FISHERMAN, the presence of the imperfective affix *-ki* provides a fairly good indication that *tei* refers to the Action of fishing rather than the Thing involved in that action.

However, care must be taken in such cases. For instance, it cannot be simply assumed that combination with a nominalizing affix indicates an action-root. In many languages “nominalizing” affixes can also combine with thing-roots (e.g. Eng. *potter* < *potN* + *-er*). In case of doubt, internal evidence can sometimes be of help. Thus Mapudungun *challwa.fe* [fish.NMLZ1] FISHERMAN was accepted as a binominal lexeme based on internal evidence from *ruka.fe* [house.NMLZ1] CARPENTER.1 Without the presence in the database of this form, where *house* is clearly a thing-root, *challwa* might have been rejected on the internal counterevidence of *wizü.fe* [give_shape.NMLZ1] POTTER.2

A similar situation is encountered with Kambaata *hoga’.aan.chu* [plough.AGT.ACC] FARMER and *tum.aan.chu* [pound/forge.AGT.ACC] BLACKSMITH, in which the status of *plough* and *pound/forge* is unclear. In this case, evidence for the presence of action-roots in such constructions is available in the form of the words for HERDSMAN and SORCERER OR WITCH, which, to judge by the gloss contain elements that are clearly actional: *herding* and *perform magic*, respectively (for the actual data see Appendix E). Admitting the former two words would result in a new type of binominal construction for which no other internal evidence exists. In this case, however, the criterion of semantic dependence argues in favour of inclusion.

In general, in cases of doubt an inclusive approach was adopted, on the grounds that it is better to include some dubious data than to risk missing out on potentially interesting phenomena. This was, however, tempered by alertness to the fact that any anomalies that turn up in the later analysis should be regarded with suspicion if they can be associated with such data.

I will conclude this discussion of issues involved in identifying thing-roots with the case of *over* in Pol. *nad.garst.ek* [over.handful.DIM] WRIST. In the Polish, *nad-* is a locative prefix, whereas in the English gloss it is a locative preposition with an inherently relational, un-thing-like character. However, elsewhere in the database where *over* appears in the gloss, the situation is less clear-cut, for example in Mapudungun *wente nge* [over eye] EYELID, where *wente* seems to be considerably more thing-like and to denote a location (that is, ‘the area above something’) rather

---

1 See the next section regarding the contribution of the affix *-fe* to the word’s status as a binominal.
2 Smeets (2008: 500) provides additional confirmation that *challwa* is a noun (and *challwa-* a verb).
than a relation as such.¹ The same applies to Kalamang kelkam elao.un [ear under.3POSS] EARLOBE, where the presence of the possessive marker strongly suggests a thing-like interpretation of under (‘the underneath part of something’). Thus, in its relational sense as a preposition, over seems to fall outside the thing-action-property trichotomy applied here,² but as a noun denoting a location it can be interpreted as a kind of place, and thus as a thing-morph.

4.1.2 Identifying thing-affixes

Whereas the presence of any root other than a thing-root was enough to disqualify a word from being regarded as a binominal, affixes are treated slightly differently, as an examination of the Polish affixes in Figure 20 will reveal.

In accordance with the onomasiological perspective outlined in §1.2.3, affixes that derive nouns are regarded as profiling Things, while those that derive verbs are considered to profile Actions and those that derive adjectives profile Properties. Examples of thing-affixes in Figure 20 are the very general nominalizer NMLZ and the relatively specific agentive (AGT), diminutive (DIM) and feminine (F) affixes.³ These are all considered to profile Things at some level of generality. When such an affix combines with a thing-root, the result is a binominal lexeme (labelled NN in Figure 20) according to the definition adopted in this project.

The Polish data does not contain any action-affixes, but the Romanian data does: the -i suffix in vrăj.i.tor [magic.VBLZ.AGT] SORCERER OR WITCH. Constructions like these are treated analogously to Vietnamese bûa ăn sáng [meal eat morning] BREAKFAST; that is, the verbalizer is regarded as an actional element, the form is interpreted as Onomasiological Type 1, and the code NVN is assigned. The presence of an action-affix is thus sufficient to disqualify a word from being considered a binominal.

That is not the case, however, with property-affixes like ADJZ. Whereas a non-derived adjective like Pol. chory ‘ill’ in combination with a thing-root or thing-affix is classified as Onomasiological Type 4 (cf. blackbird, page 11), a denominal adjective like Pol. mlecz.na [milk.ADIZ] ‘milky’ in combination with a thing-root or thing-affix is classified as Type 3, i.e. as a binominal.⁴ In other words, when it comes to determining binominal status, thing-roots and thing-affixes are treated

¹ Smeets (2008: 572) translates wente as “(N, Adv) top, on top of”.
² In the onomasiological model it would fall within the fourth category, concomitant circumstance.
³ Note that the abbreviation F denotes a feminine derivational affix and not a gender marker.
⁴ It is unclear whether this accords 100% with Štekauer’s (1998) system.
uniformly (both are acceptable nominal constituents), as are action-roots and action-affixes (both are equally unacceptable). But while a property-root like *chory* is unacceptable, a property-affix like *-na* is simply disregarded (for the purpose of determining whether or not the item is a binominal). The reason for this is that both thing- and action-affixes are substantive: they signal the presence of a new thing or action in the expression. Property-affixes, on the other hand are non-substantive: a denominal adjectivizer merely signals the presence of some kind of relation between its base and the qualified noun, and therefore has a purely grammatical function.

Like property-affixes, inflectional affixes such as *GEN* (genitive) and *PL* (plural), and linking elements (LE) such as the *-o*- in Pol. *wosk.o.wina* [wax.LE.NMLZ] EAR-WAX are also ignored for the purpose of determining binominal status. So too are function words, including prepositions, determiners and pronouns. All of these, however (with the exception of number markers) are taken into consideration when determining the kind of construction (see §4.3 below), and even more so when it come to determining the morphosyntactic strategy (see Chapter 5).

4.2 Determining the head

Once binominals had been identified, the English gloss of the primary constituents was added in columns M and N (labelled c1 and c2, respectively). This would prove useful in ensuring consistency in later stages of the analysis. In addition, the position of the head was indicated in column U, as either L (left-headed) or R (right-headed), cf. Figure 20.1

As noted above, coordinate constructions were assigned their own code (*CO*). As a result, meanings that tend to be represented by coordinate constructions (e.g. *COUSIN*, *PARENTS* and *SIBLING*) were not retained when the list of meanings was reduced from 201 to 100 (cf. §3.1.4). The consequence of this was that in the final version of the database the vast majority of binominals (over 90%) exhibit a clear head-modifier structure.2

There are two reasons why it is a desideratum to know which of the primary constituents is the head. First, the position of the head affects the type of construction.

---

1 The traditional terms left-headed and right-headed can be confusing in languages that are written from right to left. They should not be interpreted literally, but as synonyms for head-initial and head-final, respectively.

2 Most of the exceptions involve age or gender and are discussed in §8.4.
Thus, a head-initial compound **Head Mod** is considered to be a different construction from a head-final compound **Mod Head**. The second is that the nature of the semantic relation between the two main constituents of a binominal cannot be determined without knowing which is the head and which is the modifier. For example, in terms of Bourque’s 2014 classification used in Chapter 6, the relation in *houseboat* is FUNCTION (“a boat that serves as a house”), while that in *boathouse* is PURPOSE (“a house intended for boats”). If one did not know which of the two constituents *house* and *boat* was the head, it would not be possible to determine the precise relation exhibited by each binominal.

In the following discussion of how the head was identified separate treatment is given to binominals consisting of two thing-roots and binominals consisting of a thing-root and a thing-affix.

### 4.2.1 Binominals consisting of two thing-roots

The question of how to identify the head in compounds (the prototypical type of binominal lexeme) has been addressed by Scalise & Fábregas (2010), who suggest (p. 124) recognizing three distinct types of head:

- **semantic head**: defines the semantic class of the whole word
- **categorial head**: defines the lexical category of the word
- **morphological head**: defines the formal properties of the compound as a lexical item (e.g., its gender and its inflectional class).

For various reasons, only the semantic head is relevant to the present project. First, the primary constituents of most binominals are nouns; therefore, since binominals are in practice also nouns, there is no way of knowing which of the constituents is the categorial head. Second, not all languages exhibit gender or inflectional class, and even when they do, the relevant information is often not easily available, so morphological criteria cannot be used consistently. Third, the theoretical framework underlying the present research, Cognitive Linguistics, gives primacy to meaning, and thus favours the semantic head (or ‘profile determinant’, to use Langacker’s 1987 term for what is essentially the same thing). And finally, only the semantic head can be used to determine the nature of the semantic relation (see Chapter 6).

Croft (1996) finds Langacker’s concept of headhood to be correct “as far as it goes” but leads to the “unsatisfactory conclusion” that noun phrases with a determiner and clauses with an auxiliary are two-headed. In order to address this problem, Croft introduces the notion of a ‘primary information-bearing unit’, or PIBU, and proposes that the following definition:
"A (semantic) head is the profile equivalent that is the primary information-bearing unit, that is, the most contentful item that most closely denotes the same kind of thing that the whole constituent denotes" (Croft 2001: 259).

The concept of PIBU is required for disambiguation when a construction contains two elements at different levels of abstraction (e.g. noun and determiner or main verb and auxiliary), and both can be regarded as profile determinants (or equivalents, to use Croft’s preferred term). This is never the case with binominals: most have only one profile determinant, and where there are two (e.g. in some coordinate compounds), they are at the same level of abstraction (as in Scalise and Bisetto’s example poeta-pintor). As a result the same results are obtained with Langacker’s profile determinant approach and Croft’s combination of profile equivalent and PIBU. (As we shall see in the next section, however, a different situation obtains with binominals that consist of a thing-root and a thing-affix.)

The semantic head is thus determined via the ISA condition. The formulaic statement of this condition, given by Scalise and Fábregas in (33), makes the (unwarranted) assumption that the head is the second (right-hand) constituent.

(33) In a compound [ [ ]x [ ]y ], z “IS A” Y

Their more general prose formulation states that “whatever concept the whole compound expresses, it is a subclass of the concept that its head denotes. In other words...the whole compound must be a hyponym of its head” (Scalise & Fábregas 2010: 111). Applying this as a diagnostic to the Polish examples in the shaded cells in Figure 20 (refer to page 103 for the glosses) produces the results in (34):

(34) The head of zorza polarna is zorza because arctic lights ISA dawn, not a pole
    The head of platek ucha is platek because an earlobe ISA lobe, not an ear
    The head of złoty pierścienek is pierścienek because a gold ring ISA ring, not a gold
    ?? The head of południe is pol because midday ISA half, not a day
    The head of droga mleczna is droga because the Milky Way ISA road, not a milk
    The head of kolej żelazna is kolej because a railway ISA course, not an iron
    The head of kamieńny most is most because a stone bridge ISA bridge, not a stone
    The head of palec u nogi is palec u nogi because a toe ISA finger, not a leg

1 Croft (to appear) defines the head (of a construction) as “essentially the most contentful word that most closely denotes the same function as the phrase (or clause) as a whole.”
2 The minor semantic shift of dawn from ‘sunrise’ to ‘brightening of the sky’, in order to construe the arctic lights as a kind of dawn, seems to be acceptable.
3 A toe can be construed (metaphorically) as a finger. In some languages, toe and finger are colexified and have the meaning digit, but the word might still be glossed as ‘finger’.
Most of these results are uncontroversial, but *pol.u.dnie* [half.LE.day] MIDDAY requires comment. In the database, almost all analysable words meaning MIDDAY have a constituent glossed as ‘day’ (or ‘sun’, which often co-lexifies with ‘day’, cf. Urban 2012: 703). When this is the case, the other constituent is always ‘middle’ (or occasionally – and metaphorically – ‘heart’) or ‘half’. The first of these clearly denotes a superclass (or hypernym) of MIDDAY (the middle of the day ISA middle, not a day). As to the second, it seems reasonable to interpret ‘half’ in this context as denoting the ‘halfway point of a day’ rather than ‘half of a day’ (the latter would apply if the meaning denoted by the construction were something like ‘before noon’ and/or ‘afternoon’); clearly, the middle of the day ISA halfway point, not a day, and also not a half.

As these Polish examples show, in most cases, determining the head of a binominal is rather straightforward, notwithstanding the three situations identified by Scalise and Fábregas in which compounds do not follow the ISA rule (2010: 111ff). In the first place, coordinate constructions occur only very rarely in the data, for the reason cited above; when they do, the constituents are synonyms, as in Mandarin Chinese 火焰 huo3.yan4 [fire/flame.flame/fire] FLAME.

Secondly, the two types of compound that Scalise and Fábregas call “exocentric” either do not occur, or else are given alternative interpretations. Compounds of the Romance V+N type (e.g. It. portalettere [carry.letters] POSTMAN, cf. Eng. pickpocket) are not considered to be binominals because they contain an actional element. As for constructions such as Eng. skinhead and It. testa rasata [head shaven] ‘skinhead’, where the referent is not a direct hypernym of either constituent, these are regarded in the constructionist framework adopted here as metonymical head-modifier structures. The English example would be regarded as a binominal, since it consists of two thing-roots, whereas the Italian counterpart would not, since it contains the action-root rasare.

The third situation identified by Scalise and Fábregas as being problematic for the ISA rule concerns words in which the meaning cannot be derived from the meanings of their constituents, for example Sp. pati-difuso [leg-distributed] ‘puzzled’. This can be because the origin is lost in the mists of time, or due to unfamiliar cultural traditions. An example of the latter is Nizaa tâá gûr [father leg] YOUNGEST CHILD. According to Rolf Theil (p.c.), in Nizaa culture it is traditional for the youngest child to act as his father’s metaphorical ‘leg’ and to support him physically in his old age. Thus, in this word, gûr ‘leg’ is clearly the head. A further example is Takia tamol sos [man Derris_root] WIDOWER. While the motivation for this word
cannot be determined, a widower is far more likely to be conceived as a kind of man than as a kind of root. Cases such as these are, however, extremely rare in my data.

4.2.2 Binominals consisting of a thing-root and a thing-affix

Turning now to binominals that consist of a thing-root and a thing-affix, while the notion of head is relatively uncontroversial in the case of noun-noun compounds, and other, more “phrasal” binominals consisting of two thing-roots, there is much more contention when it comes to derivation and inflection. Some idea of this disagreement can be gleaned from Bauer (1990), who cites a number of major contributions to the discussion (including Marchand 1969; Williams 1981; and Zwicky 1985), reviews the criteria used to determine the head in syntax, and then attempts to apply the latter to English morphology. He concludes:

The obvious, though not necessarily the correct, conclusion to draw is that heads have no place in morphology. Certainly, if they have a role to play, this role needs to be defined much more carefully than has been the case up until now… Given the things that ‘head’ is supposed to do at the moment, we would not be much worse off without our heads (p. 29-39).

Bauer thus hedges his overall conclusion, but he is considerably more categorical in his discussion of derivational affixes:

Processes such as nominalization or adjectivalization, it is suggested, have only grammatical meaning: that is, something like ‘turn the base into a noun’ is the meaning of nominalization affixes. Other affixes, such as the prefix un- have only lexical meaning: something like ‘negative’. Others, such as -er, it is suggested, have both: the lexical meaning is something like ‘person or object which is typically the subject of the verb used as the base’, while the grammatical meaning is ‘noun’. [This suggests] that nominalization and adjectivalization markers have no meaning. If that is the case, they cannot take part in any hyponymous relationship (p. 5).

Štekauer (2000) takes up the discussion and starts by briefly sketching the views of Marchand, Williams, Anderson, Lieber, Selkirk and Zwicky before presenting his own proposal, which he then proceeds to test against Bauer’s list of criteria. Štekauer’s proposal is to identify the head with the onomasiological base of the naming unit:

Rather than identifying head either positionally or morphologically […] the proposed approach shifts the criterion of headedness to the extralinguistic level, in particular, to the conceptual level of generation of naming units. By implication, [the] head can be a suffix, a prefix, or a word-formation base (p. 341).
Štekauer tests six of Bauer’s criteria with four Onomasiological Types. The first, and most important, criterion is that of hyponymy. All four types pass this test easily, “which follows naturally from the […] definition of the [onomasiological base] as that constituent of the [onomasiological structure] which stands for the most general concept”. Štekauer’s examples are given in (35), with the onomasiological base (i.e. the head) emphasised in boldface.

(35) OT1: truckdriver
     OT2: writer
     OT3: honeybee
     OT4: restart

Štekauer assumes that the first three examples are uncontroversial (Bauer allows that the agentive -er suffix can have lexical as well as grammatical meaning, as seen in the quote given above), but he recognizes that the OT4 example might raise objections:

Nonetheless, I believe that the reader will agree that the meaning of affixes is more general than any lexical meaning. Thus, the meaning of re- (REPETITION of an Action) is more general than the specified Action. In other words, any particular Action can be repeated or returned to the original state. Analogically, Bauer’s ambiguous disinter is disambiguated in this way: dis- (OPPOSITE) is head and inter is the [onomasiological mark].

Bauer’s counterexamples are explained as follows:

the [base] of dialectal is -al because its meaning (RELATED TO) is superordinate to the more specific meanings of what is “related to” (dialect, education, continent). In the same way, -ish in greenish is the [base] because its meaning is much more general (APPROXIMATION) than that of green. Similar considerations apply to diminutives, such as duckling. -ling (DIMINUTIVE) is more general than duck.

Štekauer does not explicitly mention nominalization and adjectivization, which Bauer suggested have only grammatical meaning, as general processes, but it is quite clear from the more specific instances discussed that affixes having these functions would also be regarded as constituting the onomasiological base. This means that every derivational affix found in binominals (words with the code NN in Figure 20 on page 103) would be regarded as the onomasiological base, with the

---

1 It will be recalled from §1.2.3 that OT5 represents the absence of onomasiological structure since the mark is not expressed. This type therefore “does not admit discussion of headedness,” according to Štekauer. It covers instances of conversion, as in time_VERB < TIME_NOUN.
exception of affixes glossed ADJZ, which only occur in words where there is another, more likely candidate for the role of onomasiological base (e.g. *kolej* in *kolej żelaz.na* [course iron. ADJZ] RAILWAY).

The onomasiological position accords with that of Cognitive Grammar. Langacker has not specifically addressed the issue of derivational morphemes as profile determinants in a general way, but simply taken this for granted (p.c.) based on particular sorts of examples prevalent in the early CG literature. Cases include, for instance, the Cora nominalizer ‘*a* in *ne.wes. ’a* [my.plant.NMLZ] ‘my plant/what I planted’, which is “both dependent (being elaborated by another component [i.e. wes-]) and the profile determinant” (Langacker 1988: 118), and the English agentive suffix ‘*er*’, which is considered to be the profile determinant in the construction [V.er] (Langacker 2002: 129). Langacker sums up the position of Cognitive Grammar as follows:

From the outset, derivational morphemes that change grammatical category have been analyzed in Cognitive Grammar as profile determinants (heads) in the constructions effecting their combination with a stem. They are conceptually dependent on the stem, making schematic reference to it, and this functions as the construction’s elaboration site. The derivational affix imposes its conceptual organization on the content provided by the stem. In particular, it imposes its profile, and since an expression’s profile determines its grammatical category, the derived form undergoes a “change” in category, becoming a specific instance of the one represented by the affix (p.c.).

**Haspelmath (1992)** invokes grammaticalization theory and the origin of derivational affixes in earlier free lexical items to support **Williams’ (1981)** affixal head theory, but excludes evaluative (diminutive and augmentative) affixes on the grounds that they typically originate as sound-symbolic extensions of the stem, rather than as independent lexemes.

Croft’s position is different; he questions the role of the head in morphology:

The head-dependent distinction, to the extent that it is applicable to morphemes, does not account for the root-affix structure; and that is because PIBU status, not profile equivalence, is the most relevant semantic property for this structure. The properties

---

1 In fact, the position extends beyond category-changing derivational morphemes to include inflectional morphemes, such as the plural ‘*s* in *pins*: “Thus, [PL] functions as the profile determinant in the [[PIN]-[PL]] construction, and the composite structure [PIN-PL] has the profile of a mass object instead of a discrete object” (Langacker 2002: 123).
of profile equivalence and PIBU part with each other in morphology in many derivational forms. For example, in agent nominalizations, the root (verbal or nominal) is the PIBU, but the agent-nominalizing affix is the profile determinant (2001: 269).

The theoretical frameworks that the present research builds on thus diverge in some important respects, but provided one equates head with onomasiological base and profile determinant/equivalent, they lead to the same results – except in the case of evaluative morphology, where I choose to follow Štekauer rather than Haspelmath. This has the added advantage of consistency, which is reflected in the codes given in Figure 20. As can be seen, all the Polish derivational binominals use suffixes and are therefore deemed to be right-headed.

4.3 Defining the construction

Once the head of each binominal had been determined, it was a simple matter to infer the underlying schemas (i.e. constructions) by generalizing each primary constituent (thing-morph) to either Head or Mod. Any additional constituents were retained in approximately the same form as the original gloss, except for certain simplifications; the latter consisted of removing some of the details from certain glosses, where these were considered unnecessary for the purpose of the present study and in order to improve readability. Thus, for example, the original gloss for Archi mučlin klan NOSTRIL was simplified from its form in WOLD – muč-li-n klan [nose(III)-OBL.SG-GEN hole(IV)SG.NOM] – to the form used in the database – muč.li.n klan [nose.OBL.GEN hole] – on the assumption that the inflection class would not be relevant for this study. Then, in defining the type of construction, information about the presence of an oblique suffix was omitted on the same grounds and the construction was defined as Mod.GEN Head.

Binominals consisting of a thing-root and a thing-affix (i.e. derivations) were treated differently, for the following reason. If every affix were to be replaced by Head and its base by Mod, every derivation would end up having exactly the same schema (Mod Head for suffixation and Head Mod for prefixation). This would result in the loss of potentially interesting information regarding the amount of variety exhibited by different languages in the domain of derivation. For binominals of this kind the construction was generalized by representing the non-head as Base and retaining the gloss for the affix, as in (36).
(36)  Base.AGT  garnc.arz [pot.AGT] POTTER
       Base.DIM  żółt.ko [yellow.DIM] YOLK
       Base.F    pajęcz.yna [spider.F] SPIDER WEB
       Base.NMLZ wiatr.ak [wind.INS] WINDMILL
       Base.ADJZ.DIM poczt.ów.ka [post.ADJZ.DIM] POSTCARD
Figure 23: Subset of the Polish binominal data showing constructions
The result of this analysis is illustrated in Figure 23, which shows the binominals (and only the binominals) from the Polish data in Figure 20, together with their constructions. Rows in this table are ordered by construction, with right-headed constructions (Mod.LE.Head and Mod.ADJZ Head) appearing first, followed by left-headed constructions (Head PREP Mod, Head Mod.GEN and Head Mod.ADJZ), and finally derivations (Base.NMLZ, Base.F, Base.DIM, Base.AGT and Base.ADJZ.DIM). Recall that this is just a subset of the Polish binominal data. The full set consists of 57 binominals spread across 10 constructions. A brief summary of those constructions, along with the number of instances and an example of each can be found for Polish – and every other language in the sample – in Appendix D. The complete data set can be found in Appendix E.

4.4 Data analytics

This section provides a general description of the data based on the coding performed during the first phase of the study, that is, before developing the typological and semantic classifications. As described in the preceding sections, that annotation includes the following properties:

1. morphemic gloss (analysable words only)
2. structural type (binominals vs. non-binominals)
3. head position (L or R, binominals only)
4. construction type

Table 23 gives the overall database statistics and Figure 23 shows the distribution by language and meaning.

<table>
<thead>
<tr>
<th></th>
<th>Totals</th>
<th>By language (n=106)</th>
<th>By meaning (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>sd</td>
</tr>
<tr>
<td>Words (S)</td>
<td>8,899</td>
<td>84.0</td>
<td>25.02</td>
</tr>
<tr>
<td>Analysable words (A)</td>
<td>5,346</td>
<td>50.4</td>
<td>21.93</td>
</tr>
<tr>
<td>Binominals (NN)</td>
<td>3,738</td>
<td>35.3</td>
<td>17.97</td>
</tr>
</tbody>
</table>

*Table 23: Overall database statistics*}

The codes S, A and NN are used below in simple formulae that denote ratios or percentages. For example, A/S denotes the number of Analysable words as a percentage of the total number of words (or vocabulary Size) for any given language or meaning and thus serves as a rough indicator of the degree of morphological complexity in that context. Similarly, NN/S denotes the number of binominals as a percentage of the total vocabulary and gives an indication of the degree of “binominality”, i.e. the degree to which meanings are expressed as binominals within or across languages.
In the following sections, these numbers are broken down, first by language (§4.4.1) and thereafter by meaning (§4.4.2).

### 4.4.1 Data analytics by language

Figure 24, in which a single, very long bar chart has been chopped into three pieces, gives a breakdown of the numbers in Table 23 by language. For each language the darker, left-hand column shows the number of words in the sub-database ($S$); the lighter column in the middle shows the number of analysable words ($A$), which is always less than or equal to $S$; and the right-hand column shows the number of binominals ($NN$), which is always less than or equal to $A$. Languages are listed in ascending order of vocabulary size ($S$), which is not the same as the orders obtained if the number of analysable words or the number of binominals are used instead.

---

1 Here and elsewhere the use of colours in a plot has no significance if there is no legend.
Figure 24: Basic data summary (by language)
A number of observations can be made on the basis of this diagram. The first is that the vocabulary size ($S$), varies considerably across languages. Furthermore, there is substantial variation in the proportion of analysable words, $A/S$ (compare the heights of the green and blue columns). Thus, around half of the words in Tuwari and Wawa are analysable, almost all of those in Datooga are, and the great majority of those in Murui Huitoto are. Moreover, the number of binominals as a proportion of either the total number of words ($NN/S$), or the number of analysable words ($NN/A$), also exhibits great variety (compare the height of the orange column with the other two columns). Thus, almost all of the analysable words in Wawa, Ceq Wong and Srenge are binominals, whereas almost none of those in Datooga are. Each of these observations is discussed in more detail in the following sections.

**Vocabulary size**

The size of each sub-database (i.e. the number of words in the database for a given language) varies considerably, from 21 (Tuwari), at the left-hand edge of Figure 24, to 144 (Korean) at the right-hand edge. This variation can be accounted for by four factors. Firstly, some contributors provided multiple translation equivalents for the same meaning, whereas others provided only the most common or precise translation equivalent. With the benefit of hindsight, it can be seen that more careful instructions should have been given in this regard. Secondly, loanwords were included by some contributors but not by others. Again, more precise instructions could have been given; however, this would not have solved the problem completely, since it is often difficult to draw the line between loanwords that have been absorbed into the language (and should therefore be included, for commensurability with the WOLD data), and instances of single-word code-switching, which should not (cf. the discussion in Haspelmath & Tadmor 2009: 12). Thirdly, as discussed in §3.1.5, the meaning list used as the basis for data collection has a certain bias towards languages spoken in technologically advanced societies. As a result, certain meanings tend to lack translation equivalents in some languages. Finally, the data sources available for some languages were incomplete. (It was noted in §3.3.1 that some of the original WOLD contributors did not respond to requests for additional data for the 18 meanings that are not found in WOLD.) The combined effect of the last two factors is that minority and extinct languages (like Akkadian and Old High German) tend to appear towards the (top) left-hand end of Figure 24, whereas national languages of large and developed states tend to appear towards the (bottom) right-hand end. (Exceptions to this tendency are mostly due to the effects of the first two factors.)
Morphological complexity

Saussure (1983: 132 [183]) suggested that languages can be divided into *langues lexicologiques*, which are characterized by a low degree of motivation, and *langues grammaticales*, with a much higher degree of motivation. According to Ullmann (1966: 222), words can be motivated in three different ways: phonetically (e.g. *swish*, *sizzle*), morphologically (e.g. *arm-chair*, *thinker*) and semantically (e.g. ‘the bonnet of a car’, ‘the pivot on which a question turns’). Based on the present data, a rough measure of morphological motivation in the nominal lexicon can be adduced from the ratio of analysable (i.e. motivated) to unanalysable (unmotivated) items in each sub-vocabulary.

A rough measure can be found by simply dividing the number of analysable words (A) by the total number of words (S). However, certain caveats are in order. The first is that words denoting technologically advanced concepts, which have a higher tendency to be complex, are less likely to occur in certain languages, such as those of hunter-gatherer communities; as a result, languages of this type will appear to have a lower degree of morphological motivation. In addition, some contributors may have interpreted analysability as synchronically analysable for linguists as opposed to for lay speakers;¹ this will tend to inflate the figure for morphological motivation. It should also be noted that the actual figures obtained using the present data will be much higher than those for each language as a whole, because the set of meanings from which the database is constructed was chosen in order to maximise the yield of analysable items (cf. §3.1).

Figure 25: Morphological complexity (A/S) by language

¹ This is the case with the English vocabulary; cf. the discussion of the gloss provided for NOSTRIL on page 97.
For the database as a whole the degree of morphological complexity is roughly 60%, but the figure varies considerably across languages, from a high of 98% for Baa to a low of 16% for Selice Romani. Figure 25 shows languages at the extreme ends of the spectrum. Mandarin, often regarded as a language of compound words (Arcodia 2007), is a special case, but the results for Baa call for further investigation. The position of Swahili and Murui Huitoto at the high end is due to the presence of noun classes and noun classifiers in these languages, and the high value obtained for Datooga is not indicative of morphological motivation, since it is mainly due to purely grammatical suffixes (the primary suffix -óo/-ée(n), the singulative suffix -(C)éan, and/or the singular suffix -ta/-da). On the other hand, the languages at the other end of the spectrum do seem to be genuine langues lexicologiques.

**Binominal frequency**

The languages in the sample can also be compared in terms of the extent to which they use binominal word-formation strategies. One measure of this is the number of binominals as a proportion of the total vocabulary size \((NN/S)\), a figure that varies from a high of 77% (Mandarin) to a low of 3% (Datooga), with an average value of 40.8%. Figure 26 shows the ten languages at each extreme of the scale.

![Figure 26: Binominal frequency (NN/S) by language](image)

The case of Mandarin and the compound-rich structure of its lexicon was noted above, and the same trait accounts in large part for the high ranking of Japanese and Vietnamese. Explanations for the other languages at the high end of the scale await further research which is beyond the scope of the present project. At the other end of the scale, both Datooga and Puyuma are sparsely represented in the database (29 and 54 words, respectively). The data for Datooga are mostly mono-morphemic (if one disregards the grammatical affixes mentioned above), while complex words in Puyuma are more likely to contain an action-root than a thing-root: for example,
the words denoting SPECTACLES/ GLASSES, GLOVE and BRACELET all contain the verbal prefix *pu*- ‘put’, combined with words meaning ‘eyes’, ‘hand’ and ‘arm, respectively.

4.4.2 Data analytics by meaning

The same data can be described in terms of meanings, as shown in Figure 27. As in Figure 24, the leftmost column in each group of three shows the total number of words in the database for each meaning (*S*); the middle column shows the number of these that are analysable (*A*); and the rightmost column shows the number of binominals (*NN*). Meanings are arranged in ascending order of vocabulary size (*S*). Similar observations can be made of this diagram: the size of the sub-vocabularies (*S*) varies considerably across meanings; there is great variation in the proportion of analysable words (*A/S*); and the number of binominals as a proportion of either the total number of words (*NN/S*) or the number of analysable words (*NN/A*) also exhibits great variety. Each of these observations is discussed further below.

**Vocabulary size**

The size of each sub-database (i.e. the number of words in the database for a given meaning) also varies considerably, from 35 (PADDLE WHEEL), at the left-hand edge of Figure 27, to 126 (BEE) at the right-hand edge.

Figure 28 shows how the size of the meaning vocabularies varies across different semantic fields (see Table 18). The best coverage is found for meanings belonging to field of Religion and belief. There are two of these, MAGIC and SORCERER OR WITCH, represented by 101 and 121 words respectively (on average, 111). Other meanings with extensive coverage include Kinship (six meanings: BOY, GIRL, MARRIED WOMAN, MOTHER-IN-LAW (OF A MAN), NIECE, WIDOWER), The body (19 meanings) and Emotions and values (one meaning: TEAR). At the other end of the scale are the semantic domains of Motion (one meaning: STONE BRIDGE), Modern world (19 meanings), and Time (three meanings: WEDNESDAY, SUNDAY, MID-DAY), where the coverage averages 80 words or fewer.
Figure 27: Basic data summary (by meaning)
Note that these figures represent the total number of words for each meaning in the database, not the number of languages that have one or more words for that meaning. Thus, one reason why SORCERER OR WITCH has such a high score is because some contributors provided translation equivalents for both SORCERER and WITCH (e.g. Mandarin Chinese wūshī and wūpō). Moreover, some languages have many words for this particular meaning (the database contains five each from Querétaro Otomi and Seychelles Creole, and four from Zinacantán Tzotzil). Thus, SORCERER OR WITCH is actually only found in 94 of the 106 languages, despite being represented by 114 words in the database.

**Morphological complexity**

The same procedure that was used above to measure and compare the degree of morphological complexity in languages (i.e. the number of analysable words as a proportion of the total number of words) can be used – with the same caveats – to measure and compare the degree of morphological complexity associated with individual meanings. For the database as a whole, the average measure is roughly 62.5%, but as before, the figure varies considerably, from a high of 97% for GOLD RING to a low of 27% for BEE. Figure 29 shows meanings that are at the extreme ends of the scale. It is perhaps not surprising that GOLD RING and STONE BRIDGE are the meanings that are most frequently represented by analysable words; both are subtypes of basic level concepts (RING and BRIDGE) that are differentiated by the material they are made of. A natural way of naming such entities is to combine the name of the parent concept with the name of the material, which results in an analysable form. The few instances in the database of non-analysable words associated with these meanings, such as Mamara Senoufo pɔ (< Fr. pont), actually denote the parent concept. The extent to which items denoting concepts like STONE
BRIDGE and GOLD RING are lexicalised is difficult to determine, especially across a large number of languages, but these two are clearly at the less lexicalized end of the continuum of nonanchoring possessives. The issue of lexicalisation, and how to distinguish lexical items from phrases, was avoided in the present study by the selection of meanings that tend to be found in dictionaries. With the benefit of hindsight, we might conclude that STONE BRIDGE and GOLD RING should not have been on the meaning list; however, their inclusion does not appear to have skewed the results of the study in any major way.

Otherwise, it is striking that almost all the other eight meanings to the left in Figure 29 (WINDMILL, KEYWORD, TOOLBOX, etc.) all belong to the semantic field Modern world. This confirms the intuition that the recency and inherent complexity of technologically advanced concepts tend to be reflected in the complexity and degree of motivation of the names that are coined for them.

![Figure 29: Morphological complexity (A/S) by meaning](image)

At the other extreme, to the right in Figure 29, are the ten meanings that are least often denoted by analysable words. The case of BEE is worth commenting on, since it was not among the 159 meanings originally extracted from WOLD (see §3.1.1). This was because it is represented by an analysable form in only 11 of WOLD’s 41 languages. It was added to the set of meanings on the strength of Fr. abeille à miel in order to increase the representation of Head à Mod constructions (see §3.1.2). This was clearly a mistake, since abeille à miel has the more specific meaning HONEY BEE; BEE itself is simply abeille. Of the other meanings that tend to be expressed by unanalysable words, some have a tendency to be borrowings of what might originally have been an analysable form (e.g. Kanuri báskùr < Eng. bicycle, Central Yupik nuussnik TOILET < Rus. núžnik ‘latrine’), while others are more-or-less basic level concepts.
Binominal frequency

The degree to which the 100 meanings are expressed by binominals as opposed to other kinds of lexeme ranges from 86% in the case of STONE BRIDGE to 11% in the case of MAGIC, with an average of 44.6%. Figure 30 shows the ten meanings at each end of the scale; they are identical to those in Figure 29, except that RAILWAY appears instead of KEYWORD. At the other end of the scale, borrowing and basic level concepts are once again factors, but in addition the concept in question often involves an action. This is the case with HERDSMAN, which is often denoted by a word containing an actional element, be it Onomasiological Type 1, e.g. Assamese go.rokh.ia [cow.watch.AGT], or Onomasiological Type 2, e.g. Malagasy mpi.àndry [AGT.watch], rather than a binominal.

![Image: Figure 30: Binominal frequency (NN/S) by meaning]

Another useful measure of the frequency of binominals by meaning is shown in Figure 31, in which the number of binominals (NN) is divided by the number of analysable words (A) rather than by the total number of words (S). Essentially the diagram addresses the question, if a meaning is expressed by an analysable word, how likely is that word to be a binominal? The meanings to the left all score very highly in this regard (91% or more); their modifying concepts tend to be materials, body parts, and the like. Those to the right all tend to involve an action (e.g. MAKE, COOK, CATCH) or a property (e.g. SICK) and are thus expressed by OT1 and OT2 forms to an even higher degree.
Chapter summary

This chapter described the coding of the data in some detail, in order to highlight certain theoretical and practical issues relating to three major tasks: distinguishing binominals from other words, determining the head of the construction, and defining the schematic construction instantiated by each binominal. In order to identify binominals roots must be distinguished from affixes, and thing-morphs must be distinguished from action- and property-morphs. In general, two things-morphs are required, and the presence of an action-root, an action-affix or a property-root is usually sufficient to disqualify a form from binominalhood. However, the case of ‘yellow’ in combination with ‘egg’ to denote YOLK drew attention to the fact that the border between property-roots and thing-roots can be gradient. The particular status of properties was further underlined by the differential treatment accorded to property-affixes as opposed to (most) property-roots.

Determining the head of the construction turned out to be rather straightforward in the case of binominals consisting of two thing-roots but required an analytical choice to be made in the case of binominals consisting of a thing-root and a thing-affix. For consistency, evaluative morphology was treated in the same way as derivational morphology in general with the affix accorded the status of head.

I also provided a detailed description of the binominal data by language and by meaning, focusing on vocabulary size, morphological complexity and binominal frequency. This showed considerable variation across all three parameters and revealed a number of issues that call for further investigation. The next chapter discusses the typological classification that was developed on the basis of the annotated data.
5 Typological classification

Chapters 1 and 2 of this work covered the first stage of Song’s recipe for “doing typology” (see page 17), namely identification of a phenomenon to be investigated. Chapters 3 and 4 then described the second stage, generation of a language sample. The third stage is the creation of a typological classification, and that is the focus of both this chapter and the next. Stassen (2002: 766) makes the point that it is quite possible for one and the same database to give rise to several different typologies, none of which has to be intrinsically “better” than the others. This is because classifications are simply tools for answering research questions and can vary depending on the kind of criteria on which they are based. In this chapter I present a two-level typological classification based on the morphosyntactic features of binominals. In the next chapter I develop another two-level classification based on the semantic relation between the two major constituents of a binominal.

In §5.1 I review two classifications that serve as points of departure for developing a typological classification of binominals: Koptjevskaja-Tamm’s (2002) classification of possessive noun phrases (PNPs) in Europe (§5.1.1), and Croft’s (2003) use of possessive constructions to illustrate a “cross-linguistically valid description of morphosyntactic structures” (§5.1.2). This section also highlights the problems inherent in representing a classification as a hierarchy.

In §5.2 a two-level classification of binominals is developed, together with an alternative, non-hierarchical representation that avoids the problems highlighted in the preceding section. The nine-way classification is based primarily on the degree of marking and the locus of marking; the more detailed, 18-way classification adds constituent order to these parameters. A detailed description of each of the nine basic types follows in §5.3, along with numerous examples from the data set to illustrate the range of constructions that they each cover.

Two “gaps” in the classification, revealed by the non-hierarchical representation, are discussed in §5.4 and explanations are put forward to account for them. §5.5 then exemplifies the kinds of gradient phenomena found in the data and shows how they are handled by the proposed representation; §5.6 presents some data analytics; and §5.7 provides a conclusion.
5.1 Theoretical prerequisites

5.1.1 Koptjevskaja-Tamm (2002; 2003)

The classification that stands out in the literature as providing a possible starting point for the present study is Koptjevskaja-Tamm’s (2002) study of possessive noun phrases (PNP) in Europe (Figure 32). The typology defines three major types: synthetic, juxtaposition and analytic. In the juxtaposition type, the possessor and possessum are simply placed next to each other, in either head-initial or head-final order, with no additional grammatical material. In the synthetic type, affixes are attached to either the possessum (head), the possessor (dependent), or both; and in the analytic type, the possessor and possessum are supplemented by an additional free morpheme, which may be either a preposition or a linking pronoun. The types are organized in a hierarchy with the leaf nodes labelled juxtaposition, dependent-marking, double-marking, head-marking, prepositions and linking pronouns. A seventh type, possessive compounding, is not included in the typology since it is “mainly restricted to Northern Swedish”.

![Figure 32: Major structural types of PNPs in Europe](after Koptjevskaja-Tamm 2002: 144)

The classification accommodates most of Koptjevskaja-Tamm’s data, but as she points out, it is only a “rough and simplified approximation to the great structural diversity within European PNPs”. This highlights one of the disadvantages of hierarchical models: they tend to imply the existence of discrete, non-overlapping categories, and to obscure the existence of gradient phenomena that do not fit neatly into one or another category. Another disadvantage of such models is that they force the analyst to give precedence to one way of grouping categories over another, which may be equally valid. Thus, the principal subgrouping in Figure 32 is based on degree of fusion, with an opposition between synthetic and analytic and with juxtaposition in between. The subdivisions of synthetic and analytic are based on
locus of marking and functional criteria, respectively. This is not only inconsistent, it also obscures important facts about relationships between types, such as that prepositions (also) mark the dependent, and that case affixes arise from adpositions that become attached to the noun (Croft 1990: 34). This, however, is a criticism of the visual organization, not the substance of Koptjevskaja-Tamm’s classification, which provides a useful starting point for the present project. Consider the four examples presented on page 1:

(37) a. Ger. *eisen.bahn* [iron.way] – appears to exemplify juxtaposition
b. Fr. *chemin de fer* [way of iron] – is clearly of the prepositional type
d. Rus. *želez.naja doroga* [iron.ADIZ road] – is dependent-marked

In addition, one other PNP type is to be found in the list of binominals on page 14:

(38) e. Western Farsi *surāx.e bin.i* [hole.EZ nose.ADIZ] – uses double marking

Clearly, there is significant commonality between PNPs and binominals. Only the linking pronoun type is not encountered amongst binominals. However, binominals exhibit a variety of strategies that goes beyond Koptjevskaja-Tamm’s set of six types. To start with, some languages outside Europe (the area investigated by Koptjevskaja-Tamm) employ postpositions instead of prepositions, the Japanese *no* construction in *kumo no su* [spider GEN web] being a classic case. We must therefore generalize the category prepositions to adpositions. We also need to add two more types, as follows:

- **derivations**, such as the Czech form *pavuč.ina* [spider.NMLZ] SPIDER WEB, in which the concept WEB can be regarded as having been generalized to THING and expressed by a nominalizing affix (cf. the discussion of ‘head replacement’ on page 167);
- **classifier constructions**, as in Murui Huitoto *ui.tirá' [eye.CL(hair)] EYELASH, in which the constituent that carries the meaning HAIR is neither an affix nor an independent lexeme, but rather belongs to a special class of nominal called a noun classifier.

A further difference between possessive noun phrases and binominals is that the category (synthetic) dependent-marking covers two distinct types of binominal: those with an adjectival suffix, e.g. Rus. *želez.naja doroga* [iron.ADIZ road], and those with a genitive suffix, e.g. Bezhta *kil.os hino* [iron.GEN way] RAILWAY. Distinguishing these as adjectival and genitival introduces a functional element into an otherwise formal typology, based on locus of marking and degree of fusion, but
this can be justified on two grounds. Firstly, the two types involve different morphological processes: the one, transpositional (word-class changing) derivation, the other non-transpositional (word-class preserving) inflection. Secondly, these two types can be encountered as competing strategies in one and the same language, as in Lithuanian gelež.in.kelis [iron.ADJZ.way] RAILWAY and aus.ų vaškas [ear.GEN wax] EARWAX: the data therefore require us to make this distinction.

The preceding discussion suggests that a classification of binominals will consist of the eight strategies (or types) listed in Table 24. Five of Koptjevskaja-Tamm’s six PNP types are retained, but with the following modifications: her juxtaposition is subdivided into compounding, derivation and classifier (constructions), and her dependent-marking into genitival and adjectival. In addition, prepositions becomes adpositions and head-marking is renamed construct, following Creissels (2017).

<table>
<thead>
<tr>
<th>PNP types</th>
<th>binominal types</th>
</tr>
</thead>
<tbody>
<tr>
<td>juxtaposition</td>
<td>compound</td>
</tr>
<tr>
<td></td>
<td>classifier</td>
</tr>
<tr>
<td></td>
<td>derivation</td>
</tr>
<tr>
<td>dependent-marking</td>
<td>genitival</td>
</tr>
<tr>
<td></td>
<td>adjectival</td>
</tr>
<tr>
<td>double-marking</td>
<td>double</td>
</tr>
<tr>
<td>head-marking</td>
<td>construct</td>
</tr>
<tr>
<td>prepositions</td>
<td>adpositions</td>
</tr>
<tr>
<td>linking pronouns</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Table 24: Types of PNP and binominal*

Having identified a set of core classes, the question arises how best to arrange them into a system. Because of the shortcomings described above, I would like to avoid hierarchical representations. Inspiration for an alternative representation comes from Croft, who uses adnominal possession in order to illustrate the classification of cross-linguistic structural types, and whose approach emphasises the function of marking rather than its locus.

5.1.2 Croft (2003)

The development of Croft’s classification over a period of thirty years is described in Croft (to appear). An initial enumeration of types in Croft (1990) lists fusion, special form, affixation or compounding, juxtaposition, case, adposition, agreement, indexation, classification and linker, followed by the comment that
A more thorough analysis of the genitive construction types would involve the following morphosyntactic features:

1. **Word order of G and N**: GN, NG
2. **Degree of morphological fusion of G and N**: none (juxtaposition), compounding/affixation, fusion
3. **Additional morpheme**:
   a. **Existence**: absent, present
   b. **Type**: case/adposition, agreement, index, classifier, linker, combinations thereof
   c. **Constituent of**: G, N, neither (i.e. ternary branching structure)
   d. **Word order**: precedes, follows, between (if not a constituent of either)
   e. **Degree of morphological fusion to G or N**: juxtaposition, affixation, suppletion (p.37).

The same classification (but with the term concord instead of agreement) is found in Croft (1995), along with examples and a diagram showing diachronic relations between the different strategies. A major overhaul is then presented in Croft (2003) with the introduction of a three-way grouping into **simple**, **relational** and **indexical** strategies, the distinction between the latter two harking back to a contrast between relational and ‘deictic’ in Croft (1988).

The defining property of simple strategies is that they “do not involve an additional morpheme, beyond those used to express the possessor and the possessum.” They are of three types: **juxtaposition**, **concatenation** and **fusion**. The last of these is not relevant in the present context, since once two nominals have fused, they cease to constitute a binominal in the sense used in this study; to cite a classic example, English *lord*, though originally a compound consisting of Old Eng. *hlāf* (‘bread, loaf’) and *weard* (‘ward, guardian, keeper’), is not a binominal because it is no longer analysable for present-day lay speakers.

The distinction between juxtaposition and concatenation is essentially that between juxtaposition and possessive compounding in Koptjevskaja-Tamm’s typology, except that Croft’s concatenation covers both affixation and compounding. As I already indicated in Table 24, I intend to keep the latter two separate. The border between juxtaposition and compounding, however, as Koptjevskaja-Tamm (2004) points out, is notoriously difficult to draw, and much more research is needed for determining to what degree this distinction makes sense cross-linguistically. Until then in many cases we have to rely on the local tradition (p. 175).
Croft himself notes that “affixation and compounding are historical developments from juxtaposition: the juxtaposed elements become morphologically bound.” But morphological boundness is a matter of degree, as a result of which we find various phenomena that fall in between the prototypical cases of juxtaposition (at one end of the gradient) and concatenation (at the other). The boundary between juxtaposition and compounding is particularly hard to maintain in the domain of binominals, since the criteria for determining the boundary between the two vary from one language to another:

In English, some compounds are distinguished from syntactic phrases by stress (contrast a 'black board and a 'blackboard, for instance). In other languages there may be special morphophonemic processes which apply between the elements of compounds, there may be tone sandhi patterns or particular tonal patterns which apply to compounds, there may be some phonological merger between the elements of the compound (Dakota, Hebrew, see section 7), and so on (Bauer 2001: 695).

Even within a single language, such criteria may not be applied consistently, as demonstrated by the English pair 'apple cake and apple 'pie, neither of which seems to be more or less deserving of the appellation ‘compound’ than the other, despite their differing stress patterns. Nor are orthographic criteria to be trusted, as witness the variable spelling of a word like flowerpot ~ flower-pot ~ flower pot and the various ways of transliterating a Japanese binominal like 蜘蛛の巣 SPIDER WEB (as either kumo no su or kumonosu).

Given the absence of clear, cross-linguistic criteria for distinguishing juxtaposition from compounding it would be reasonable to combine them into a single type, but there are several compelling reasons for keeping them separate. Firstly, it would be inconsistent to ignore word boundaries in simple strategies and to respect them in, say, relational strategies. As we know, affixal case markers develop historically from adpositions; here, too, the juxtaposed elements become morphologically bound and here, too, the border can be difficult to draw (cf. the case of Indic languages on pages 167 and 178). But some languages, such as Polish (cf. Figure 23), exhibit both types, and so we want to keep adpositions and case affixes distinct. Consistency requires us to do the same with juxtaposition and compounding.

Secondly, it will turn out that a combined juxtaposition/compounding category will be almost an order of magnitude larger than all the others and render any statistical comparisons less robust. More importantly, there is reason to believe that the two strategies might have different properties and, if so, we want to capture these. For the purpose of the present study, I have therefore chosen to maintain the distinction
and, like Koptjevskaja-Tamm, rely on the “local tradition”, which basically means trusting the instincts of the documenters; the proof of the pudding will be in the eating, as we will see in §7.1.1. As a result of this decision, juxtaposition is to be added to the eight types in the right-hand column of Table 24.

The boundary between juxtaposition and affixation, on the other hand, is easier to maintain: juxtaposition involves two thing-roots, whereas affixation involves a thing-root and a thing-affix; in order to be able to maintain a boundary between the two we need to be able to distinguish roots from affixes. While tricky (cf. §4.1), this can usually be done, provided we are willing to accept occasional in-between phenomena such as affixoids. Croft’s juxtaposition and affixation thus correspond to the compound and derivation types already found among our eight binominal types. The key property that compounding and derivation have in common, as Croft’s analysis makes clear, is that they “do not involve an additional morpheme, beyond those used to express the possessor [modifier] and the possessum [head]” (p. 32). This insight will form the basis of the alternative, non-hierarchical system presented in the next section.

In contrast to simple strategies, the relational and indexical strategies identified by Croft do involve additional morphemes, the purpose of which is to “encode the relation between the possessor and the possessum”. Relational morphemes are case markers (case affixes or adpositions), whereas indexes are agreement markers. Croft points out that “a cross-linguistic definition of case/adposition and indexation on a structural basis is difficult” and he therefore provides a semantic definition:

A more suitable definition would be that a case marker/adposition is relational, that is, a morpheme that denotes the semantic relation that holds between the noun phrase and the verb, while agreement is indexical, that is, a morpheme that denotes the argument itself (Croft 1988; §2.1). This definition is essentially a semantic one (p. 16–17).

This definition is framed in terms of the relations between a verb and its arguments, but it is clearly intended to be more generally applicable and I therefore propose to reformulate it more succinctly as follows:

(39) A relational morpheme (or relator) denotes a semantic relation, whereas an indexical morpheme (or index) refers to a participant in such a relation.

Croft’s distinction between relational and indexical strategies corresponds fairly closely to that between flags and person indexes made by Haspelmath (to appear).¹

¹ At least for the paradigm cases (see Croft to appear for a detailed comparison).
flag
A flag is a bound form that occurs on a nominal and that indicates the semantic or syntactic role of the nominal with respect to a verb (in a clause) or with respect to a possessed noun (in a complex nominal).

person index
A person index is a bound form denoting a speech role or a highly accessible third person referent that occurs on a verb (or in second position) to indicate a verb’s argument, or on a noun to indicate its possessor.

Both dichotomies were developed as alternatives to the head/dependent-marking distinction proposed by Nichols (1986; 1992) and adopted by Koptjevskaja-Tamm in her classification of PNPs. In both cases a primary motivation was to avoid problematic issues with the broad notions of head and dependent (see Croft 2001 Ch. 7; Haspelmath 2019a §6). More importantly, from a theoretical perspective, Croft’s concepts are based solely on the function of marking rather than the locus of marking; Haspelmath’s are based on both; and Nichols’ are based solely on the latter.

Logically, these parameters are orthogonal to one another: in theory, a relator could appear on either the head or the dependent, and so could an index, and this is indeed the case in predicate-argument relations; however, in referent-modifier relations, which is what are involved in binominals, “it seems to be true that … case markers (relators) occur on the modifier and [person] indexes on the head” (Croft, p.c. 2018-12-14). As a consequence, the three approaches lead to almost exactly the same results: indexes occur on heads (in Haspelmath by definition, and in Croft by default), while flags/relators occur on the possessor, i.e. the dependent. Croft (2007: 360) suggests that the tendency for indexation to correlate with head-marking is because it “primarily functions to particularize a relational concept type [and is thus] most useful when it is grammatically more closely associated with the relational concept”, i.e., in this case, the head. Conversely, head-marking consists of indexes (and linkers, discussed below), while dependent-marking consists (almost exclusively) of flags or relators. In practice, therefore, it makes little difference whether one uses function or locus as a parameter.¹ This might be one reason to adopt the more widely used head/dependent distinction, especially as the terms head and dependent are relatively uncontroversial in the context of binominals. Another reason, as we will see, is that it is easier to apply consistently in practice.

¹ That is not to say that there is no difference at all. Croft suggests (p.c.) that non-person indexes may occur on dependents, but no clear-cut examples of this have been found in the data.
In addition to simple, relational and indexical strategies, Croft’s system includes ‘more grammaticalized strategies’: linkers and special forms (Figure 33). Linkers are found in contexts where the grammaticalization process has proceeded “to the point that it makes progressively less sense to classify a marker as indexical or relational, because there is just one invariant morpheme that is used to code the dependency” (p. 38); they are overt forms that encode a relation but do not form a paradigmatic contrast with other forms. Hence one cannot define a linker as either indexical (since it does not contrast with other forms in person, number and/or gender) or relational (since it does not contrast with other forms to distinguish semantic relations) (Croft 2019).

The first point to note about this is that it is often unclear how to apply this concept to binominals. For example, in a binominal the dependent is never pronominal, its referent is always a third person, so there cannot be a contrast of person. As a result, a morpheme that would be regarded as a person index in a possesive construction, such as the 2/3SG prefix t- in (40), which contrasts with 1SG n- , 1PL q- and 2/3PL ky- in Mam, would presumably have to be regarded as a linker in the domain of binominals. Or would it be a nonperson index, if it (very occasionally) contrasts in number with ky-?

\[ t\_kamb'\_meeb'a \] [3SG.prize orphan] ‘[the] orphan’s prize’

A similar issue arises with relators. Take the case of French de and Japanese no, which both occur (frequently and less frequently, respectively) in the binominal data (see Appendix E for examples). Both of these participate contrastively in the
adpositional systems of their respective languages, but the situation with respect to binominals is different: *de* contrasts with *à* (and other prepositions) whereas *no* does not contrast with anything. Is the latter then a linker rather than a relator, or are they both relators? The argument could go either way (which of course is not surprising given the diachronic relationship between adpositions, case affixes and linkers). The question is whether or not it is useful to treat the two differently.

The other point to note is that such issues do not arise if the classification is based on the locus of marking. Provided it can be ascertained whether the marker attaches to the head or the dependent, there is no need for an additional category such as linker. The example usually cited by Croft is the Persian *ezāfe*, which arose from a relative pronoun. This is generally deemed to be a constituent of the head, as in *zanbur.e asal* [bee.EZ honey] HONEY BEE (Tehranisa 1987: 56), and therefore falls fairly uncontroversially into Koptjevskaja-Tamm’s head-marking type. For these reasons I have chosen not to include linker as a separate binominal type.

The category ‘special forms’ includes segmental morphemes and base modification. The most well-known case of the latter is the Semitic construct state, illustrated by the Hebrew example *mesila.t barzel* [track.STC iron] RAILWAY (5f). Here, the base modification clearly applies to the head. In Berber languages, on the other hand, it is the dependent that is modified in constructions that involve what some Berber linguists call the construct state, as seen in the Tarifit example *ŧisi u.fus* [bottom STC hand] PALM OF HAND (60j). While it might make “progressively less sense” to classify these as relational or indexical, there is no problem determining the locus of marking. Once again, adopting locus as a parameter in the classification obviates the need for the additional category ‘special form’.

Finally, Croft also discusses the “somewhat problematic” case of classifiers, but comes to the conclusion, on the basis of diachronic evidence (i.e. their nominal origins) that they are fundamentally indexical in nature (p. 38). In *Croft (to appear)*, on the other hand, classifiers are “split as to whether they functioned indexically – referring to the modified referent – or relationally – denoting the semantic relation between the two concepts”. But neither of these approaches cover the kind of binominals that involve classifiers. A case in point is Murui Huitoto *ui.tiraï* [eye.CL(hair)] EYELASH, where the classifier with the meaning HAIR has the same function as the head constituent of a compound such as Thai *khôn.taa* [hair.eye]. For this reason – and because they are clearly separate from compounds and derivations – I choose to treat them as a fourth type of simple strategy (i.e., one that involves no additional morpheme).
Chapter 5. Typological classification

The great advantage of Croft’s approach is that the various strategies are embedded in a theoretically motivated framework based on the grammaticalization paths that link them together. The principal disadvantage is that it is hard to apply in practice, at least in the domain of binominals, especially in a large cross-linguistic study, given the lack of detail available in most grammars. Another disadvantage is that it results in even more types than the locus approach. It is, of course, an empirical question which approach is most fruitful – the proof of the pudding is in the eating – and ideally one should apply both approaches and compare the results. However, that is beyond the scope of the present study. I have therefore opted to take the more practical route and focus on the locus of marking. The degree to which this bears fruit, and whether the alternative would prove more productive, will be the topic of Chapter 7, *Typological generalizations*.

5.2 A non-hierarchical alternative

Croft (2003: 40) summarizes the basic properties for describing morphosyntactic structures as follows:

(i) additional morpheme: none, relational, indexical, linker, (special form);
(ii) degree of fusion of elements: none, concatenation, fusion;
(iii) order of elements.

The first two properties, modified in accordance with the preceding discussion, suggest an alternative way of representing the eight types of binominals arrived at in Table 24. (The third property is discussed on page 144 below.) We start by adding juxtaposition and reordering the contents of Table 24 as Table 25, supplemented with data regarding the number of additional morphs (over and above the two required thing-morphs), examples of each type, and three-letter mnemonic labels. These labels are used extensively in the remainder of the present work, so it is important that they be internalized by the reader. Even more important, though, is to understand that these labels are simply mnemonics intended to evoke prototypical exemplars of each type: they are not descriptive labels! Thus, for example, prp labels the analytic dependent-marking type that includes not just prepositional constructions (e.g. French *chemin de fer*), but also postpositional constructions (e.g. Japanese *kumo no su*); similarly, gen labels a synthetic dependent-marking type that allows for any kind of case affix, not just the genitive. (The label adp could, of course, have been used instead of prp for greater transparency, but it would be easily confused with adj, and the problem of naming gen would still exist.) More
precise descriptions of each type, along with many more examples, will follow in the next section (§5.3).

<table>
<thead>
<tr>
<th>markers</th>
<th>strategy</th>
<th>code</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>juxtaposition</td>
<td>jxt</td>
<td>VIE đường sắt [road] RAILWAY</td>
</tr>
<tr>
<td></td>
<td>compounding</td>
<td>cmp</td>
<td>DEU eisen.bahn [iron] RAILWAY</td>
</tr>
<tr>
<td></td>
<td>classifier</td>
<td>cls</td>
<td>HUU ui.tiraí [eye.CL] EYELASH</td>
</tr>
<tr>
<td></td>
<td>derivational</td>
<td>der</td>
<td>CES pavuč.ina [spider] SPIDER WEB</td>
</tr>
<tr>
<td>1</td>
<td>adjectival</td>
<td>adj</td>
<td>RUS želez.naja doroga [iron] RAILWAY</td>
</tr>
<tr>
<td></td>
<td>genitival</td>
<td>gen</td>
<td>KAP kil.os hino [iron] RAILWAY</td>
</tr>
<tr>
<td></td>
<td>prepositional</td>
<td>prp</td>
<td>FRA chemin de fer [way] RAILWAY</td>
</tr>
<tr>
<td></td>
<td>construct</td>
<td>con</td>
<td>PLT lala.m. by [road] RAILWAY</td>
</tr>
<tr>
<td>2</td>
<td>double</td>
<td>dbl</td>
<td>TBC -emo.li sakila.li [nose.POSS] NOSTRIL</td>
</tr>
</tbody>
</table>

*Table 25: Reordered binominal types*

In order not to complicate the typology unnecessarily, two or more consecutive morphs attached to either the modifier or the head are counted as a single morph. For example Bezhta *kil.os hino* [iron.OBL GEN way] RAILWAY is simplified to *kil.os hino* [iron GEN way] and treated as having just one additional morph. In this I follow Nichols (1992), who found such a simplification necessary “because the precise amount of multiple case marking in the constructions I am surveying is generally not made clear in grammars, so no consistent count could be made” (p. 62). A further reason is that there are too few examples of this phenomenon in my data to justify defining separate types to cater for them.

It is possible to arrange these nine binominal types (or strategies) hierarchically, but to do so would require choosing between a grouping based on degree of fusion (i.e. analytical vs. synthetic, as in Figure 32), or one based on locus of marking (i.e. head, dependent, both or none), both of which would obscure important facts. Instead, I have chosen to arrange the types on a two-dimensional grid (Figure 34) which incorporates three different parameters (number of markers, locus of marking and degree of fusion), and allows for the addition of a third dimension that captures the order of elements.
The vertical axis represents the number of markers (levels 0, 1 and 2, respectively), and caters for binominals consisting of:

- (level 0): two elements (i.e. with no additional morpheme beyond the head and the modifier),
- (level 1): three elements (i.e. with one additional morpheme attached to either the head or the modifier), and
- (level 2): four elements (i.e. with additional morphemes attached to both the head and the modifier).

The types jxt (juxtaposed nouns), cmp (noun-noun compounds), der (denominal derivations) and cls (noun-classifier constructions) are situated on level 0; adj (constructions involving relational adjectives), gen (constructions with an affix on the modifier), prp (constructions involving adpositions) and con (constructions with an affix on the head) are situated on level 1; and dbl (constructions in which both the head and the modifier are marked) are situated on level 2.

The horizontal axis on level 1 represents the degree of fusion. The latter, illustrated in Figure 35, is a continuum that ranges, in Bybee’s (1985: 12) words from “the most highly fused means of expression, lexical expression”, to “the most loosely joined means of expression, syntactic or periphrastic expression”. The “derivational-inflectional-free grammatical” section of Bybee’s scale is replicated in the left-hand (dependent-marking) side of level 1. Degree of fusion is also relevant to the vertical organisation of level 0, since compounds often evolve from juxtaposition and are in turn the source of classifier constructions and derivations, but it is obviously not relevant to level 2, which only contains a single type.
The typology and semantics of binominal lexemes

Level 1 has been partitioned in order to capture the *locus of marking*. To the left are various forms of dependent-marking (i.e., where some kind of marker attaches to the modifier) and to the right, the one attested form of head-marking. The types **gen** and **con** both involve non-transpositional affixes and are therefore situated opposite each other on level 1. (The parallel between these two types is underscored by the fact that the term ‘genitive’ is sometimes applied to possessive markers that attach to the head, cf. the discussion of the Malagasy data on page 98.) Dixon (2010b) has advocated using the term ‘pertensive’ when the morphological marking is on the possessum in order to avoid confusion, but I have chosen to follow Creissels (2017) and I use the term construct form as the basis for the mnemonic **con**.

Since no binominals have been found representing types that mirror **adj** and **prp**, gaps are shown in those positions. These lacunae, labelled (**prn**) and (**nml**), represent the potential existence of two further types, as yet unattested: an analytic form of head-marking (which would correspond to **prp**), and a type in which the head bears a transpositional affix (which would correspond to **adj**). These “missing types” are discussed in §5.4.

The use of ovals with gradient fill rather than the boxes with sharp outlines found in more traditional representations (including Koptjevskaja-Tamm’s) is intended to convey the fact that the types are not clearly defined categories but rather points within a multidimensional space at which phenomena tend to cluster. This enables the representation of grammaticalization paths and various “in between” or gradient phenomena, as will be seen in §5.5.

One parameter not shown in the diagram is Croft’s (2003: 40) third grammatical property: the order of elements or, as it is usually called in studies of compounding such as Scalise & Fábregas (2010), the position of the head. Such a parameter is required in order to cater for situations in which a language exhibits both left- and right-headed binominals of one and the same type, such as **Mod Head** and **Head Mod** (both **jxt**) in Vietnamese (41a); and **Mod.ADJZ Head** and **Head Mod.ADJZ** (both **adj**) in Polish (41b).
5.3 Binomial types

The classification of binominals developed in the preceding section consists of nine types (jxt, cmp, der, cls; prp, gen, adj; con; dbl). These are arranged on a grid that captures the degree of marking, the locus of marking and the degree of fusion. An additional parameter, head position, can be represented by adding a second plane (or third dimension). The resulting model reveals two “gaps”, points in the two-dimensional space at which one might expect to find binominals but where none are attested in the data. The latter are discussed below in §5.4 but first, definitions are provided for the nine attested types, along with examples and more detailed discussion. Intermediate types are alluded to where applicable in this section and discussed more fully in §5.5 under the rubric Gradience.

5.3.1 No additional marker: jxt, cmp, der and cls

Level 0 of the model contains four types, jxt, cmp, der and cls, each of which has exactly two components – the two thing-morphs that are the primary constituents of a binominal – and there is no additional grammatical material.

jxt

The prototype of jxt (“juxtaposition”) is comprised of a head and a modifier, both of which are independent lexemes, or thing-roots. There is no additional grammatical material and little or no fusion between the two constituents, which are treated as separate words.

As noted above (page 136), there is no accepted cross-linguistic definition of the notion of word. Nevertheless, every language appears to have such a notion; or, as
Bauer (2000: 255) puts it, “all languages have a unit which falls between the minimal sign and the phrase”. As explained above, I follow Croft in making a distinction between juxtaposition and compounding and, in the absence of more robust criteria, I adopt Koptjevskaja-Tamm’s policy of relying on the “local tradition”. Given the nature of my data, I employ the orthographic heuristic that a word space or hyphen signals juxtaposition (jxt), whereas the lack of either signals compounding (cmp). We will see in §7.1 that this is, in fact, sufficiently robust to reveal at least one interesting universal.

Examples of the jxt strategy (all with the meaning RAILWAY) are given in (42):

(42) Vietnamese đường sắt [road iron]  Kildin Sami rūvv’t čuekas [iron road]
     Saramaccan talán fūtu [train foot]  Ho-Chunk mqas nqagq [metal road]
     Western Farsi rāh āhan [way iron]  Yakut timir suol [iron path]
     Cabécar kóbâkâ ŋlg [train road]  Western Mari kòrtni korná [iron road]

The jxt type is found in 22% of the binominals in the database (§5.6.1). It occurs in 76 of the 106 languages and is a significant word-formation type (accounting for at least 10% of binominals) in 53 of these. It accounts for the majority of binominals in 22 languages and is the only binominal word-formation strategy in Ceq Wong, Datooga, Imbabura Quechua, Seychelles Creole, Vietnamese and Walman.

cmp

The prototype of cmp (“compound”)¹ is comprised of a head and a modifier, both of which are independent lexemes, or thing-roots. There is no additional grammatical material, but a high level of fusion between the two constituents, such that the binominal constitutes a single word.

The paradigm case of this type is the noun-noun compound, in which two nouns are simply concatenated. The cmp type is the most frequent type in the data set of this study, accounting for almost 30% of all the binominals in the database. It occurs in 67 languages and is a significant word-formation type (accounting for at least 10% of binominals) in 48 of these (§5.6.1). Furthermore, it accounts for the majority of binominals in 24 languages, and is the only binominal word-formation strategy in Caijia and Tuwari (cf. Figure 41 on page 173). Examples of the cmp type (again, all with the meaning RAILWAY) are given in (43).

---

¹ Recall that the labels used for types are not intended to be taken literally. They are mnemonics whose purpose is to bring to mind prototypical instances of the type.
Chapter 5. Typological classification

(43) Baa krà.kisà [road.train]  Bambara teren.sira [train.road]
Hawaiian ala.hao [path.iron]  Hungarian vas.út [iron.road]
Irish iarn.ród [iron.road]  Japanese tetsu.dō [iron.road]
Korean chel.kil [iron.road]  Mapudungun tren.rüpü [train.way]
Thai thaang.rótfay [way.train]  Welsh rheil.ffordd [rail.road]

Less prototypical examples of the cmp type are compounds that contain a linking element; these are discussed on page 165 in the context of gradient phenomena.

der

The prototype of der (“derivation”) comprises a nominal base (thing-root) and a nominalizing affix (thing-affix). Less prototypically, and not attested in the present database, a neoclassical compound may be considered to consist of two thing-affixes and thus belongs to this type (see §6.1.2 on neoclassical compounds). Some typical examples of the der type are listed in (44).¹

(44) a. Polish wiatr.ak [wind.NMLZ] WINDMILL
   b. Mapudungun kütral.we [fire.LOC] FIREPLACE
   c. Malagasy mpam.intana [AGT.fishhook] FISHERMAN
   d. Czech kůz.le [goat.DIM] KID
   e. French brac.elet [arm.DIM] BRACELET
   f. Hausa sàráu.niyáa [king.F] QUEEN
   g. Gawwada kaank.itte [horse.SG:F] MARE
   h. Gawwada sint.itte [nose.SG:F] NOSTRIL
   i. Sidamo lukk.iffo [hen.SGLT_M] COCK/ROOSTER

Only affixes that contribute some tangible semantic content are considered in scope. The meaning contribution may be very general (THING, 44a) or it may be more specific (INSTRUMENT 44a, LOCATION 44b, AGENT 44c, etc.). Note that the gloss provides only a rough indication of the meaning contribution of the affix and is not claimed to be consistent, for two reasons: firstly, the exact meaning of many derivational affixes is hard to pin down and may exhibit considerable variation; and secondly, some contributors have taken pains to provide specific glosses while others have supplied more general glosses. As a case in point, Nagórko (2016) highlights the instrumental nature of the Polish suffix -ak, whereas it is glossed more generally as NMLZ in the database (44a).

¹ The glosses given here and elsewhere are based on those of the original contributors and reflect not only functional diversity but also terminological variety across different traditions of linguistics.
Diminutives are deemed to bear the meaning contribution SMALL THING (cf. the discussion in §4.2.2). They can denote a small version of the entity denoted by the base (44d) or something small that is in some way related to the base entity (44e). Gender-denoting affixes are included only when they mark a clear semantic alternation. In (44f–h) QUEEN alternates with KING, MARE with HORSE and NOSE. On the other hand, and to judge by the gloss provided by the contributor, Gawwada xarrap.atte [spider_web.SG:F] SPIDER WEB cannot be regarded as a binominal of any kind since the suffix appears not to derive a new meaning through gender alternation.

In §4.2 the position was advanced that the affix should be regarded as the head of derivational binominals on semantic grounds (see the discussion on page 113 ff). The affix usually has a very general meaning and thus represents a hypernym of the referent. Thus a windmill (44a) is a kind of thing (or instrument), a fireplace is a location (44b), a fisherman is an agent (44c), etc. Furthermore, the affix often takes the place of what would have been the head element in a parallel construction of type cmp. This is again illustrated by (44a): the Polish suffix -ak combines with the word meaning WIND to denote the same concept as the combination of WIND and MILL in the German equivalent, wind.mühle [wind.mill]. If a Windmühle is a Mühle (mill) that is powered by Wind, then a wiatrak must be some THING (or INSTRUMENT) that is powered by wiatr (wind). Similarly, the concept FIREPLACE is often denoted by a combination of forms meaning FIRE and PLACE (as in English, and also in Welsh lle tân [place fire]. In Mapudungun, it is denoted by suffixing the word for FIRE with a locative, which can therefore be regarded as the semantic head of the construction, just as lle ‘place’ is in the Welsh word. The same applies to French bracelet, in which the diminutive suffix denotes a SMALL THING that is located on the bras (arm), cf. Japanese ude.wa [arm.ring].

cls

The prototype of cls (“classifier”) is comprised solely of a nominal base and a noun classifier, where the denotatum of the binominal is different from that of the base, such that the classifier is used to derive a new meaning rather than for classification.

This is the least satisfying of the types in the classification, not least because it is the only type that is not defined solely in terms of structural criteria. Its inclusion in the typology is motivated by the existence of forms such those in (45) and (46) (examples from Urban 2012: 126–127), which clearly qualify as binominals.
(45) Arabela
   a. *quitia.aca* [breast.teat.CL(liquid)] MILK
   b. *namiji.aca* [eye.CL(liquid)] TEAR

(46) Bora
   a. *íñu.héju* [earth.CL(hole)] CAVE
   b. *túú.heju* [nose.CL(hole)] NOSTRIL

The classifier morphemes in these examples (-aca and -héju) have exactly the same function as the corresponding head constituents of, say, Thai, *náam.taa* [water.eye] TEAR and *ruu.camùuk* [hole.nose] NOSTRIL. However, they cannot be used in isolation, so they are not thing-roots, and thus these binominals do not belong in the cmp type described above. The classifiers also constitute a closed class, which sets them off from the typical nominal constituents of the cmp type. In both respects they are more like thing-affixes, so these binominals could be classified under der. But classifiers differ markedly from affixes in having very precise semantics. This does not, in and of itself, constitute sufficient reason to separate them off from the der type, but the matter does not end there, as we shall see.

Aikhenvald (2000; 2017), citing criteria articulated earlier by Allan (1977), defines classifiers as “morphemes which occur in surface structures under specifiable conditions, denoting some salient semantic characteristics of the entity to which an associated noun refers”. The examples from Arabela (45) and Bora (46) belong to one of seven subtypes, which Aikhenvald calls noun classifiers,¹ characterized by the fact that they “occur with a noun independently of any other constituent of a noun phrase or a clause”. They can be affixes to nouns, as above, but they are also often “independent words with generic semantics”, as in (47).

(47) Minangkabau (Aikhenvald 2000)
   a. *batang limau* [CL(tree) lemon] LEMON-TREE
   b. *buah limau* [CL(fruit) lemon] LEMON-FRUIT

If the Arabela and Bora examples are classified as binominals of type der, then these must be classified as binominals of type jxt, and noun classifiers as a group would then be split across two binominal types. That is not necessarily a problem, but it suggests that a better solution, one that would make it possible to investigate the classifier phenomenon more closely, is to define a separate subtype, cls.

¹ The seven subtypes identified by Aikhenvald are: (i) genders and noun classes, (ii) noun classifiers, (iii) numeral classifiers, (iv) classifiers in possessive constructions, (v) verbal classifiers, (vi) locative classifiers and (vii) deictic classifiers.
However, the matter does not end there either. Examples to parallel those from Minangkabau are also found in Atlantic-Congo languages. In (48) pairs of singular and plural noun class prefixes, *m/-mi* - and *Ø/-ma*-, distinguish trees from fruits, in just the same way as the Minangkabau classifiers *batang* and *buah*. If the Minangkabau words qualify as binominals, so too should the Swahili forms.

(48) Swahili (Russell 2003)
   a. *m.limau / mi.limau* [CL.3:lemon / CL.4:lemon] LEMON TREE(S)
   b. *limau / ma.limau* [CL.5:lemon / CL.6:lemon] LEMON FRUIT(S)

And so should the Bandial examples in (49), where the noun class prefixes serve, among much else, to distinguish between animals and their offspring (cf. the Czech diminutive suffix in 44d, above).

(49) Bandial (Watson 2015)
   a. *ji.jamen* [CL.(ji):goat] KID
   b. *e.jamen* [CL.(e):goat] GOAT

Such noun class prefixes are not noun classifiers in Aikhenvald’s typology, instead they are classified under subtype (i) ‘genders and noun classes’. One of the major differences between these two subtypes, according to Aikhenvald, is that in noun class languages every noun belongs to a noun class, whereas in noun classifier languages, every noun does not have to take a classifier.¹ Consequently, there would be a very substantial cost to admitting words like (49a, b): almost every word in the Swahili and Bandial sub-vocabularies would qualify as a binominal of type *cls* and, as a result, the data from noun class languages would swamp those from noun classifier languages and give a distorted overall impression of the *cls* type. That problem may not be insurmountable provided we remain aware of it, but unfortunately the matter does not end here either. Consider (50).

(50) Gawwada
   a. *piʔ atte* [kid.SG:F] KID
   b. *xarrap atte* [spider_web.SG:F] SPIDER WEB

If the Bandial examples are regarded as binominals, why not also the Gawwada? The only real difference between a two- or three-gender system and a noun class system of the Atlantic-Congo type is the size of the system; Aikhenvald groups them under the same subtype. And yet, the Gawwada examples cannot by any

¹ In addition, agreement is a necessary feature of noun class/gender systems but not of noun classifier systems. However, this does not impinge on the present discussion.
stretch of the imagination be regarded as the functional equivalents of noun-noun compounds. Moreover, admitting them into the pantheon of binominals would lead to the kinds of construction we are interested in being completely drowned out. Somewhere on this slippery slope a line has to be drawn.

That line could be drawn between Aikhenvald’s two subtypes; this would amount to defining noun classifiers, but not noun class markers, as thing-morphs. (45)–(47) would then be categorized as binominals, while (48)–(50) would not. This would have the unfortunate result that (47) and (48), which really are parallel in every way, would be accorded different treatments. The line could also be drawn by contriving a distinction between noun class languages and gender languages based on the size of the system: say, more than three for noun class languages and two to three for gender languages. The line would then go between (49) and (50). Not only would this be somewhat arbitrary, it would also result in the aforementioned imbalance between noun class languages and noun classifier languages.

The solution adopted here is to draw the line instead between (49a) and (49b). The basis for making such a distinction is that in (49a) the denotatum of the whole (KID) is different from that of the base (GOAT), whereas in (49b) they are the same (GOAT and GOAT). In (49a) the noun classifier contributes a meaning component that changes the denotatum, i.e. its function is derivational. In (49b) this is not the case; nor is it in the two examples from Gawwada (50). This analytical choice results in a much smaller harvest of cls binominals in the data set than the second solution outlined above: only 37 out of the total of 3,738 binominals in the database are of this type, and they are distributed across just six languages: Äiwoo, Bandial, Harakmbut, Murui Huitoto, Swahili and Trinitario with 3, 3, 4, 11, 4 and 12 examples respectively. This is few enough that every one can be presented and discussed here.

(51) Äiwoo
   (a) kio mi.sigiläi [hen BN:GNL.male] COCK/ROOSTER
   (b) dâbu mi.polee [day BN:GNL.seven] SUNDAY
   (c) dâbu mi.eve [day BN:GNL.three] WEDNESDAY

The items attached to the general bound noun mi- in the Äiwoo examples (51b) and (51c) are borderline thing-morphs since they denote numbers. However, they employ the same construction Head BN:GNL Mod as (51a), in which sigiläi ‘man’ is “a perfectly good noun in its own right” (Åshild Næss, p.c.). All three were admitted as binominals, albeit non-prototypical, in order to retain maximum variety, but it now seems that the latter two are more properly regarded as action-roots.
Given the definition arrived at above, only three words in the Bandial vocabulary qualify as binominals and they share a single construction, **CL.Base** (52). The same construction is found in over 30 other words in the data set, but these are excluded on the grounds that the denotatum of the whole is the same as the denotatum of the base. Examples include *e.uyat* [CL(e).beehive] BEEHIVE, *sa.mbul* [CL(sa).flame] FLAME and *mu.fu* [CL(mu).tear] TEAR. In the first two examples in (52) a base that would normally belong in the *e-/si-* paradigm,¹ where many animals are to be found, is placed in the *ji-/mu-* paradigm which carries the semantics of diminutiveness (Watson 2015). This process parallels that of other word-formation processes in which the meaning GOAT is combined with the meaning DIMINUTIVENESS to denote the meaning KID (53).

(53)  

| der | Czech | kûz. le [goat. DIM] |
| cmp | Hawaiian | kao keiki [goat child] |
| con | Hausa | yâ.r àkwîyàa [daughter.LK goat] |
| prp | Tagalog | bata.ng kambing [child.LK goat] |

In the third example of (52), the root denoting SORCERY is transferred to the *a-/u*-paradigm where the majority of nouns denoting humans are formed in order to denote a person who practices sorcery — a SORCERER OR WITCH. In short, what distinguishes the three Bandial words admitted as binominals is that they have been formed through a process of noun class *alternation*. The same can be said of the Swahili examples in (54). For example, in *uchawi* [CL14:witch] MAGIC the root has been transferred from the person noun class 1/2, where *mchawi* denotes WITCH, to noun class 14, which is typically used for abstract concepts, in order to denote the meaning MAGIC.

(54)  

| Swahili |  |
| u.chawi | [CL14:witch] MAGIC |
| fumo | [CL9:spear] CHIEFTAIN |
| m.zabibu | [CL3:grape] VINE |

¹ According to Watson, the defining function of this paradigm is “to form nouns denoting entities that are bounded and individuated” (p. 275). In addition to animals it contains nouns denoting body parts and artefacts and is the default paradigm for loanwords.
Chapter 5. Typological classification

The 11 Murui Huitoto binominals listed in (55) differ from the Swahili binominals in (54) in two respects: firstly, the classifying morphemes are suffixed rather than prefixed, and secondly, their semantics are much more specific. Whereas a typical Bantu language has between 10 and 20 noun classes, Murui Huitoto has more than 100 classifiers, the meanings of which, as the examples show, can be as specific as ‘cover’, ‘hair’, ‘cavity’ and ‘stem’.

(55) Murui Huitoto

\begin{itemize}
  \item \textit{taizí.ko.ño} [heel.CL(cover).CL(fem)] ANKLE\textit{ui.tíraї} [eye.CL(hair)] EYELASH
  \item \textit{jitaї.ño} [adolescent.CL(fem)] GIRL
  \item \textit{jїfai.ño} [in_law.CL(fem)] MOTHER-IN-LAW (OF A MAN)
  \item \textit{enaize.ño} [grandson/nephew.CL(fem)] NIECE
  \item \textit{defo} [nose:CL(cavity)] NOSTRIL
  \item \textit{moo.rui} [father.CL(day)] SUNDAY
  \item \textit{riїno.kaiї} [woman.CL(stem)] THUMB
  \item \textit{eї.kaiї} [foot.CL(stem)] TOE
  \item \textit{ra.o} [thing.CL:flex] VINE
  \item \textit{onoyї.kї} [hand.CL(cluster)] WRIST
\end{itemize}

Once again, each of the examples in (55) consists of a stem whose denotation is quite distinct from that of the word as a whole, the difference being attributable to the meaning contribution of the classifier. By contrast, the Murui Huitoto words in (56) all denote the same concept as the stem to which they are attached, and are therefore not classified as binominals.

(56) Murui Huitoto

\begin{itemize}
  \item \textit{enї.e} [land.CL(gen)] NATIVE COUNTRY
  \item \textit{feka.be} [shoulderblade.CL(leaf)] SHOULDERBLADE
  \item \textit{nїkui.o} [vein.CL:flex] VEIN OR ARTERY
  \item \textit{kiїfo} [hive:CL(cavity)] BEEHIVE
  \item \textit{iyaї.ma} [chief.CL(masc)] CHIEFTAIN
  \item \textit{fekї.ma} [widower.CL(masc)] WIDOWER
\end{itemize}

The examples from Harakmbut (57) and Trinitario (58) exhibit similar patterns to those from Murui Huitoto but also greater complexity, sometimes involving additional morphemes over and above those denoting the base concept and the classifier. Due to the paucity of the data, it is not possible here to disentangle the various subtypes of strategy being employed here.
The typology and semantics of binominal lexemes

(57) Harakmbut

\begin{itemize}
\item aymörö.po [honey.CLF(round)] BEE
\item sërä.po [honey.CLF(round)] BEE
\item wa.kmbere.ku.to.pa [NPOS.\text{forehead}.CLF(\text{edge?}).SPAT(\text{down}).CLF(\text{rod})] EYEBROW
\item wa.kpo.ku.ti.mba [NPOS.\text{eye}.CLF(\text{edge?}).SPAT(\text{on}).CLF(\text{hand})] EYELID
\end{itemize}

(58) Trinitario

\begin{itemize}
\item mopó.si [\text{bee\_related}.CLF(round)] BEE
\item mopó.ji [\text{bee\_related}.CLF(\text{shapeless})] BEESWAX
\item -ugi.mo [\text{eye}.CLF(\text{fabric})] EYELID
\item (e)san(e).ti.m(o)re [field.npos.clf(fan)] FARMER
\item ta.pu.gi [3NH.point(?).CLF(trunk)] FLAME
\item -mut.pewo’u [\text{clothes}.CLF(\text{hand})] GLOVE
\item -jek.pew’u [\text{interior}.CLF(\text{hand})] PALM OF HAND
\item -mits.gi [?.\text{CLF(trunk)}] SPINE
\item -it(i).ne.re.pi [\text{blood}.PSD.NZR?.CLF(snake)] VEIN OR ARTERY
\item tapaj.ro.gi [\text{door}.\text{DER}.CLF(trunk)] DOORPOST
\item -iypé.re.ku [\text{foot}.\text{DER}.CLF(\text{inside})] FOOTPRINT
\item -pow.ro.cho [\text{wing}.\text{DER}.CLF(\text{board})] SHOULDERBLADE
\end{itemize}

5.3.2 One additional marker: prp, gen, adj and con

Level 1 of the taxonomy contains four types. In three of these (prp, gen and adj) the additional marker forms a constituent with the modifier, and in the fourth (con) it forms a constituent with the head. What they all have in common is that they contain an additional (grammatical) morpheme, over and above the two basic constituents, the head and the modifier, both of which are lexical items.

\textbf{prp}

The prototype for prp (“prepositional”) consists of a head and a modifier, both of them independent lexemes, and another independent lexeme that forms a constituent with the modifier.

In the most common case (59a–e), the additional lexeme is a preposition (hence the choice of mnemonic for this type), but it may also be a postposition (59f) or a particle named according to a language-specific descriptive category, such as a connector (59g), linker (59h), relator (59i), possessive classifier (59j), article (59k) or determiner (59l).
(59) a. French *chemin de fer* [road of iron] RAILWAY
   b. French *moulin à vent* [mill to wind] WINDMILL
   c. French *arc-en-ciel* [bow-in-sky] RAINBOW
   d. Lower Sorbian *lapka na woku* [flap on eye] EYELID
   e. Welsh *papur lle chwech* [paper for toilet] TOILET PAPER
   f. Hindi *dāṃt kā braś* [tooth POSTP brush] TOOTHBRUSH
   g. Swahili *tundu la pua* [hole CON nose] NOSTRIL
   h. Tagalog *daa.ng-lakal* [road.LK-iron] RAILWAY
   i. Kupsabiny *loleet nyēē cēēkēēnē* [bag REL:SG back] BACKPACK
   j. Takia *graian sa.n anay* [evening POSS.3SG food] SUPPER
   k. Maltese *xatt il-bahar* [shore DEF-sea] SHORE
   l. Welsh *asgwrn y cefn* [bone DET back] SPINE
   m. Barain *assi ge ṣ kee* [bone REL:3SG:M PREP head] SKULL

(59m) is remarkable in that the “additional morpheme” actually consists of two items: a relator (*ge*) and a preposition (*ŋ*), both of which are associated with the modifier. Following the convention established earlier, these are counted as a single marker.

While prepositions are fairly common, postpositions are rare and can be problematic, in that they can often be analysed as case affixes rather than adpositions. The Hindi (59f) example is a case in point. It is glossed by the contributor using the abbreviation GEN, but *ka* is also commonly regarded as a postposition. In this study the Hindi construction has been classified as prp. The issue is discussed in more detail under the rubric Gradience on page 166ff.

The most commonly used adpositions are those whose function includes marking possession, such as the French *de* (59a) and the Hindi *ka* (59f), but some languages permit other prepositions to be used as well, as illustrated by the various locatives (59b-d) and the purposive (59e). In other languages, the particle has a more general, associative meaning that is used for a wide variety of relations and not solely for possession. Examples include the Swahili connective *-a* (59g), the Tagalog linker *ng* (59h) and the Kupsabiny relational marker *nyēē* (59i).

**gen**

The prototype for gen (“genitival”) consists of a head and a modifier, both of them independent lexemes, with an additional, non-transpositional affix or segmental marker attached to the modifier.
Typically the marking indicates the genitive case (60a–c) or possessive function (60d), but other cases occur as well, including dative (60e), locative (60f), lative (60g) and proprietive (60h). Case-marking may be indicated segmentally, as in (60a) Irish gaoithe [wind:GEN] < gaot [wind:NOM] and (60b) Kildin Sami töl [fire:GEN] < tōll [fire:NOM].

Atypical the marking indicates the genitive case (60a–c) or possessive function (60d), but other cases occur as well, including dative (60e), locative (60f), lative (60g) and proprietive (60h). Case-marking may be indicated segmentally, as in (60a) Irish gaoithe [wind:GEN] < gaot [wind:NOM] and (60b) Kildin Sami töl [fire:GEN] < tōll [fire:NOM].

The Tarifit example (60j) exemplifies the kind of confusion that can arise if one assumes that descriptive categories are the same across languages. Here the modifier, ‘hand’, normally fus, is in what some Berber linguists call the “construct state” (hence the gloss).\(^1\) This term is used in Semitic linguistics to describe a special form of the head in possessive constructions. Consequently, Berber words glossed with STC belong to the type gen (dependent-marked, relational) whereas Semitic words glossed with STC belong to the type con (head-marked, indexical), cf. (65c) on page 159.

adj

The prototype for adj (“adjectival”) consists of a head and a modifier (both of them independent lexemes), with an additional, transpositional morpheme attached to the modifier.

1 Other terms used are annexed state (‘état d’annexion’, Kossmann 2000) and dependent form (‘avhengig form’, Endresen 1990).
The database contains 196 instances of this type, the great majority from European languages, either Indo-European or Uralic (61). The six Slavic languages (Croatian, Czech, Lower Sorbian, Polish, Russian and Slovak) account for 130 of them alone. Whether this is because adjective as a productive lexical category is more frequent in Europe than elsewhere is a matter for further research.

The most common descriptive category for the additional morpheme is adjectivizer, but the terms attributivizer and proprietive are also found. Polish and Hungarian are notable for having two distinct constructions of this type. In Polish the same construction can be either head initial (61f) or head-final (61g). In Hungarian there are two different adjectival suffixes: -i (labelled ADJZ) and -s (labelled PROP), both of which can be attached to a wide variety of nouns (Kiefer 2009).

The remaining instances (17 in all) are found scattered across the globe in Africa (62a-b), the Caucasus (62c-e), Asia (62f-i), New Guinea (62j) and Central America (62k). The examples from Archi (62c-d) are somewhat borderline as binominals since the modifying element is more adverbial than nominal. They are included in order to exemplify the use of relational adjectives in Archi, which can be derived from nouns as well as adverbs (Kibrik 1994).

The examples from Archi (62c-d) are somewhat borderline as binominals since the modifying element is more adverbial than nominal. They are included in order to exemplify the use of relational adjectives in Archi, which can be derived from nouns as well as adverbs (Kibrik 1994).
The typology and semantics of binominal lexemes

**con**

The prototype for **con** (“construct”) consists of a head and a modifier (both of them independent lexemes), with an additional, non-transpositional affix or segmental marker morpheme attached to the head.

The term ‘construct’ is traditionally used in Semitic linguistics but is extended by Creissels (2017; forthc.) to cover any obligatory marking on a noun that fulfils the role of head in noun-modifier constructions, provided it does not cross-reference features of the modifier that condition its use. Without that proviso (the reasons for which are not stated), Creissel’s term ‘construct’ would cover every instance of the type **con**. Binominals of this type are glossed in a variety of ways (63-65). Labels used in traditions other than Semitic include linker, possessive, genitive and pertensive. The latter term, proposed by Dixon (2010b), is restricted to possessive constructions (unlike Creissels’ construct), but does permit cross-referencing.

As described in §5.2, and for the reasons given there, the type **con** covers Croft’s linkers and special forms (to the extent that these occur on the head), as well as his indexical strategies. Croft subdivides the latter according to whether or not they encode the category of person, i.e., **person indexation** and **nonperson indexation**, respectively. Maria Koptjevskaja-Tamm (2003: 645) makes what appears to be a similar distinction, albeit using different terminology, between two subtypes of head-marking:

1. **relators**, whereby the form of the head signals the presence of the dependent in the same NP, without, however, specifying its features;
2. **indexers**, whereby the form of the head varies according to the properties of the dependent.

Her relators correspond to Croft’s linkers and special forms, since they do not exhibit contrast, whereas her indexers may involve either person indexation or nonperson indexation. Examples of the latter are found in Barain (63a-b) and Hausa (63c-d), where the markers are *-jil-(g)eti* or *-rl-n*, respectively, depending on the gender of the dependent.

(63)  
b. Barain *nokuno non.ji* [goat child.POSS:3SG:M] KID  
c. Hausa *kàfá.r hàncii* [orifice.LK nose] NOSTRIL  
d. Hausa *dóoki.n kàrfée* [horse.LK metal] BICYCLE

---

1 In addition, Haspelmath (2009) proposes the term “anti-genitive”.
On the other hand, the glossing of (64a-d), all of which make reference to the third person, suggests that these are indexers in MK-T’s terms and examples of person indexation in Croft’s.

(64)  
a. Kalamang kāŋgir pul.un [eye skin.3POSS] EYELID  
b. Kekchí x.na’aj xam [3ERG.place fire] FIREPLACE  
c. Takia su mala.n [breast eye.3SG] NIPPLE OR TEAT  
d. Yakut χarax u.ta [eye water.3SG] TEAR

Lastly, the invariant possessive affixes in Kupsabiny (65a) and Malagasy (65b) are relators for MK-T but linkers for Croft. So too are the Galbi Carib possessive suffix -li and its allomorph -yi (65c-d), since their distributions are phonological and not conditioned by features of either the head or the modifier. The form of the Hebrew construct case (65e) is determined by the gender of the head, so it is an MK-T relator, whereas it is a special form for Croft. Despite this variation, all of the preceding examples are binominals of type con.

(65)  
a. Kupsabiny karīt.aap maata [car.POSS fire] TRAIN  
b. Malagasy lala.m.by [road.PER.iron] RAILWAY  
c. Galbi Carib manati poti.li [breast tip.POSS] NIPPLE OR TEAT  
d. Galbi Carib upupo kuwai.y [head calabash.POSS] SKULL  
e. Hebrew mesila.t barzel [track.CON iron] RAILWAY

5.3.3 Two additional markers: dbl

Level 2 of the model comprises a single type, dbl, which consists of a head and a modifier (both of them independent lexemes), with additional morphemes attached to both.

(66)  
b. Takia patu.n kəpəbo.gan [egg.3SG yellow.3SG] YOLK  
c. Seri i.to i.naail [3:POSS.eye 3:POSS.skin] EYELID  
d. Oroqen dalay.yi u.kə:n [sea.GEN bank.3SG:POSS] SHORE  
e. Central Yupik i.nai̯mi im ce̯ni.i [sea.REL shore.POSS:ABS] SHORE  
f. Somali bam.ka biyo.ha [pump.DEF water.DEF] WATER PUMP  
g. Maltese l-isfar tal-bajda [DEF-yellow of:DEF-egg] YOLK  
h. Akkadian bīt habūb.āti [house:STC bees.OBL] BEEHIVE  

The examples demonstrate considerable variation in terms of the kinds of markers (case, definiteness, possession, construct, etc.) and the ways in which they are combined. Sometimes it is the same affix that attaches to both major constituents (66a-
c). In some languages the markers appear to cross-reference each other (66b-c), in others the affix on the head cross-references the modifier (66d-e). Somali exhibits two definiteness markers (66f) and Maltese a combination of definite marker and definite preposition (66g). Finally, Akkadian (66h) exhibits an older form of the Semitic construct state with the modifier in the oblique case.

The variety encountered here suggests that a more fine-grained classification might be possible. However, the database contains only 71 instances of dbl, spread over 14 languages, which is deemed too few for this to be feasible on the basis of the present data. With more data these could be analysed in terms of Croft’s distinction between relators, indexes and linkers.

5.4 Unattested strategies

The nine types of binominal discussed in the preceding section were situated on a two-dimensional grid in Figure 34 (page 143). It will be recalled that the three levels on the vertical axis represent the number of components: level 0 contains four types (jxt, cmp, der and cls); level 1, four types (prp, gen, adj and con); and level 2, one type (dbl). There is no significance to the horizontal positioning on levels 0 and 2 other than aesthetic, but the positioning on level 1 does have import. Firstly, level 1 is divided into two sections, with dependent-marking strategies to the left and head-marking strategies to the right. The three dependent-marking strategies (prp, gen and adj) are situated from right to left, in that order, such as to reflect Bybee’s (1985) scale based on degree of fusion: prp approximates to Bybee’s ‘free grammatical’ stage, gen to her ‘inflectional’ stage, and adj to her ‘derivational’ stage (see Figure 35 on page 144).¹ The single head-marking strategy (con) is situated in the middle of the right-hand section, not for aesthetic reasons, but in order to highlight its symmetrical relation to gen: whereas gen is a non-transpositional affixing strategy with the marking on the dependent, con is a non-transpositional affixing strategy with the marking on the head. Once the nine types are laid out in this manner, two apparent gaps are revealed, labelled (prn) and (nml) in Figure 34. These are the head-marking correlates of prp and adj, respectively. In this section I discuss possible explanations for these lacunae.

¹ Note, once again, that the definitions of adj and gen are not based on the notions of derivation and inflection, but rather on the distinction between transpositional and non-transpositional affixation. However, the two dichotomies do tend to correlate with one another: Inflection is often (but not always) non-transpositional, while derivation is often (but not always) transpositional, especially in the context of binominals.
5.4.1 Head-marking correlate of prp (prn)

The first missing type is the head-marking correlate of prp, which I have labelled prn for reasons that will become apparent. If such a type exists, it must consist of a head, a modifier and another independent lexeme that forms a co-constituent with the head, e.g. Mod {X Head}. So what kind of item might be a candidate for the role of X? One way to approach this question is to look for a relation prn ↔ con that is isomorphic with the relation prp ↔ gen. It is well-established that adpositions (prp) are a common source of case markers (gen): “Diachronically, case affixes arise from adpositions that become affixed to the noun” (Croft 1990: 34). The missing type prn could thus be whatever is the source of con.

According to Croft, such “bound indexation markers” develop out of pronouns (in the case of person indexation) and articles (in the case of nonperson indexation). An example of the former is the Hausa suffix -n (singular masculine or plural) or -r (singular feminine), which attaches to the head in possessive constructions (67a, c). This suffix also occurs in the Head.LK Mod construction responsible for 40 of the 43 Hausa binominals in the database used in the present study (cf. example 63c on page 158). According to Creissels (2009) this suffix results from the cliticization of a pronoun nalta that is co-referent with the head noun in the synonymous construction illustrated by (67b, d).¹

(67) Hausa (cf. kàree ‘dog’, saaniyaa ‘cow’)
   a.  kàre.n Daudà [dog.CSTR.SG.M Dauda] ‘Dauda’s dog’
   b.  kàree na Daudà [dog that one (SG.M) of Dauda] ‘Dauda’s dog’
   c.  saaniya.r̂ Daudà [cow.CSTR.SG.F Dauda] ‘Dauda’s cow’
   d.  saaniyaa ta Daudà [cow that one (SG.F) of Dauda] ‘Dauda’s cow’

In other words, the source of Head.LK Mod is Head, PRON Mod. The latter construction would be considered an instance of the missing type prn if the pronoun formed a constituent with the head (i.e., {Head PRON} Mod), but that is not the case. Instead, the pronoun forms a constituent with the modifier (Head {PRON Mod}), which means that (67b) and (67d) – if they were binominals (which they are not, because they do not have a naming function) – would be instances of prp, not prn. This is an example of reanalysis, in which an element preposed to the modifier in a head-initial construction is reinterpreted as a postposed marker on the head (68).

(68)  kàree {na Daudà} Head, PRON Mod ➔ {kàre.n} Daudà  Head.LK Mod

¹ Newman (2000: 300) calls na / ta a (free) (genitive) linker. It can combine with personal pronouns, but is not itself a pronoun, according to him.
Clearly, constituency must be taken into consideration when looking for examples of the missing type prn. There are six logical possibilities (69). The component X might be a pronoun or an article, but it must form a constituent with the head. This means that constructions (69c) and (69d) are highly unlikely: they could only occur in a non-configurational language.

(69)  

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>{X Head} Mod</td>
<td>{Head X} Mod</td>
</tr>
<tr>
<td>a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Head Mod X</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>X Mod Head</td>
<td>Mod X Head</td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td></td>
<td>Mod Head X</td>
</tr>
</tbody>
</table>

An example of (e), in which a pronoun copy of the dependent (ha ‘he’) is preposed to the head, is provided by Nichols (1992: 79) from Atakapa (AQP), together with a schematic English rendition (70a,b).

(70)  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>yukhiti icak kau ha tal</td>
</tr>
<tr>
<td></td>
<td>Indian man dead he skin</td>
</tr>
<tr>
<td></td>
<td>‘the skin of a dead Atakapa’</td>
</tr>
<tr>
<td>b.</td>
<td>[ [ the man] [ he skin ] ] ‘the man’s skin’</td>
</tr>
</tbody>
</table>

(70a) is, of course, a possessive construction, not a binominal, so it does not count as an instance of the missing binominal type prn. However, since binominals often recruit their morphosyntactic strategies from possessive constructions (as we will see in §7.2), it is perfectly possible that the type does exist. Were it to be found, it would correspond to Koptjevskaja-Tamm’s “linking pronoun”, the one type in her PNP classification (see Figure 32 on page 132) that I did not find in my binominal data. While the linking pronoun type of PNP is rare in European languages, its status across the world’s language is unknown and it seems eminently possible that the binominal type prn could exist somewhere. Finding it, however, must remain a topic for further research.

5.4.2 Head-marking correlate of adj (nml)

The second missing type is the head-marking correlate of adj, labelled nml. If such a type exists, it must consist of a head, a modifier and a transpositional (word class changing) morpheme attached to the head, just as adj consists of a head, a modifier and a transpositional morpheme attached to the modifier, cf. the Russian example želez.naja doroga [iron.ADJZ road] RAILWAY. There are two logical possibilities: either the additional morpheme is a nominalizer, as in Mod Head NMLZ, in which case the “head” element would not be a thing-root; or it derives some other word class, as in Mod Head ADJZ, in which case the resulting construction would not denote an
entity. In neither case would the form in question be regarded as a binominal. In other words, nml as a type of binominal is a logical impossibility, at least as long as one thinks in terms of major word classes; it is not found in the data for a good reason.

5.5 Gradience

The two-dimensional representation of the typology of binominal lexemes was developed in order to account for gradient phenomena. This section discusses instances of constructions that fall in between the nine major types. It is based primarily on the data collected but includes some examples from other sources. Each subsection refers to one of the numbered items in Figure 36.

Univerbation cmp ➔ simplex

Univerbation is the term given to the historical process by which an (analysable) item consisting of two or more morphemes develops into an (unanalysable) item consisting of a single morpheme. Examples of such simplex forms would belong below level 0 in Figure 36, but since by definition they would not be binominals, such a level is not required for the classification. However, since univerbation is a gradual process, it is only to be expected that there will be partially analysable items that are intermediate between the types on level 0 (der, jxt, cmp, cls) and

---

1 The latter is found in Geisheim’s (1839: 75) adverbial use of the word Eisenbahnisch [railway:ADJZ] in his poem Kopf oben: „Eisenbahnisch sich bewegend“ (‘move like a locomotive’, lit. ‘railway-ish’).

2 Marie-Elaine Van Egmond (p.c.) has suggested a candidate for the nml type in Anindilyakwa, but it appears to involve a change of noun class rather than a change of word class.
that lower level. Strictly speaking, no such items should be present in the database of this study, since in order to qualify as analysable in WOLD, words had to be so for lay speakers (Haspelmath and Tadmor 2009: 12) and the same criterion was applied for data collected specifically for this project. Inevitably, however, some intermediate forms did make their way in and were retained on the grounds that it would not be possible to remove them all. A prime example is (71a), which was originally a compound but which is no longer identifiable as such, despite the first constituent still being recognisable as ‘nose’. In (71b), on the other hand, the process of univerbation has reached its end-point: What started out as a prototypical binominal of type cmp is today completely opaque to lay speakers.

(71) a. Eng. nostril < nose + thirl (‘hole’)
    b. Ger. messer ‘knife’ < Proto-Ger. *matiz ‘food’ + *sahsq ‘knife, dagger’

2 Affixoids cmp ➔ der

The difference between cmp and der is that the former consists of two thing-roots whereas the latter consists of a thing-root and a thing-affix. But the distinction between root and affix is not clear-cut; the two exist as end-points on a continuum that can be defined in terms of autonomy versus dependence (Tuggy 1992). Between these two end-points one finds phenomena, called affixoids, that are neither fully autonomous nor fully dependent. Booij (2010) gives a number of examples from Dutch in which a noun acquires a specialized meaning when used as the head of a compound (72).

(72)  baron ‘baron’ > rich dealer: afval-baron [trash-baron] ‘rich dealer in trash’
      boer ‘farmer’ > seller: sigaren-boer [cigar-farmer] ‘cigar seller’
      man ‘man’ > seller: bladen-man [magazine-man] ‘magazine seller’
      marathon ‘marathon’ > long-session: jazz-marathon ‘jazz marathon’

In the absence of more detailed information from the contributors it is not clear how often this kind of phenomenon occurs in the database, but one fairly clear example is Eng. herd.s.man [herd.POSS.man], in which phonological reduction of the second element -man from /mæn/ to /mən/ indicates a status intermediate between root and affix, even though it may not have “broken away from MAN, becoming a lexical formative on its own” (Matthews 1991: 94).
Chapter 5. Typological classification

**Bound nouns** \( \text{cmp} \rightarrow \text{cls} \)

As was pointed out in §5.3.1, the type \( \text{cls} \) (“classifier”) is the least well-defined and the most poorly represented in the database. If terminology is anything to go by, it consists of a number of somewhat disparate phenomena, as witnessed by the “seven way for nouns to meet nouns” in Àiwoo (Næss 2017; 2018; forthc.). One of the types found in this language is called a “bound noun” (see 51a on page 151), a term suggestive of something intermediate between a noun and a classifier, which would in turn give rise to binominals mid-way between \( \text{cmp} \) and \( \text{cls} \). The same term is used by Van Linden (2016) in describing Harakmbut, a language in which common nouns “divide into two morphologically defined classes: potentially free vs. obligatorily bound nouns.” An example of the latter is the bound root -\( \text{mba} \) in (73). Again, this suggests the possibility of an intermediate type, but further research is required.

(73) Harakmbut \( \text{tare’}. \text{mba’} \) [manioc.hand] MANIOC LEAF

**Linking elements** \( \text{gen} \rightarrow \text{cmp} \)

In many languages noun-noun compounds involve linking elements. Almost all the examples in the database are from Indo-European languages (74a-d), the only exceptions being from Korean (74e) and occurring in what Yeon & Brown (2011) describe as “compounds in which the two elements are linked together by the addition of the so-called ‘genitive s’ (p. 31). The latter, which causes tensing (or reinforcement) on the following plain consonant, is best regarded as a linking element in the modern language. Bauer (2001) cites an example from Cambodian (74f) and mentions Yoruba as having a “purely phonological” linking element that involves prolongation of the final vowel, and W. Bauer (1993) mentions a type of compounding involving a linking element -\( \text{aa} \)- “which is being increasingly used at present” in Maori (74g).

Many elements of this kind have their origin in case and/or number suffixes that have become semantically bleached and now often conflict with the grammar. For example, in the German \( \text{regierung.s.chef} \) [government.LE.head] ‘head of government’ the linking element -\( \text{s} \)-, a reflex of the masculine genitive, is here attached to a feminine noun. The Greek linking element -\( \text{o} \)- (74b) originates in an ancient thematic vowel but today functions solely as a compounding marker (Ralli 2013). Binominals such as these can be said to occupy the space between the types \( \text{cmp} \) and \( \text{gen} \) but are arguably closer to the former than the latter. Other linking elements, like those in Yoruba and Cambodian, may only ever have had a phonological role. Binominals with linking elements thus present a challenge when coding the data:
classifying them consistently as either cmp or gen could obscure important distinctions in Germanic and Greek respectively. A solution would be to adopt Croft’s notion of linkers, but this alternative was rejected for the reasons given in §5.1.2. A decision was therefore taken to classify them in such a way as to bring out any contrasts that might be relevant in each individual language. Thus, for example, binominals with linking elements are coded as gen in Germanic (to contrast with the cmp strategy otherwise found in those languages) and as cmp in Greek (to contrast with the “true” gen strategy).

(74) a. German nase.n.loch [nose.LE.hole] NOSTRIL
b. Greek siōir.o.ðromos [iron.LE.road] RAILWAY
c. Lithuanian vor.a.tinklis [spider.LE.web] SPIDER WEB
d. Russian golen.o.stop [shank.LE.foot] ANKLE
e. Korean kho.s.kwumeng [nose.GEN.hole] NOSTRIL
f. Cambodian yian.ǝ.thaan [vehicle.LK.place] GARAGE
g. Maori waiata-aa-ringea [song.LK.hand] ACTION SONG

Adpositions or case affixes? prp ➔ gen

As noted above (§5.4.1), case affixes often arise from adpositions becoming affixed to the noun. As a result, the status of some binominals as either prp or gen can be hard to determine. A classic example is the Japanese no construction which some linguists analyse as a genitive suffix (75a) and others as a postposition (75b). The orthography offers no clue since the particle no is written in Hiragana (の) while the other words are written in Kanji (蜘蛛の巣). In order to facilitate comparison with Korean, in which the equivalent possessive particle 의 (-uy) is always written as a suffix, a decision was taken to classify the Japanese forms as gen rather than prp. In the event, however, none of the Korean binominals used this construction.

(75) Japanese
a. kumo.no.su [spider.GEN.web] SPIDER WEB
b. kumo no su [spider POSTP web] SPIDER WEB
Maltese
c. mithna tar-rih [mill OF:DEF-wind] WINDMILL

The orthography used in Maltese, on the other hand (75c), suggests that the combination of the preposition ta ‘ and the definite article il-, which occurs commonly in binominals, is neither a separate word nor a prefix, but rather a clitic. This, again, lies somewhere between prp and gen.
Sometimes grammatical descriptions analyse equivalent constructions in closely related languages in rather different ways. This applies to possessive constructions in Hindi and Nepali. Whereas in Hindi (76a) the possessive morpheme is written, transliterated and referred to as a postposition, in Nepali (76b) it is treated as a suffix. In the present work, these differences are reflected in the assignment of the binominals in question to different types (prp and gen, respectively), which results in different ‘binomial fingerprints’ (cf. §5.6.3). This underscores the importance of the gradient approach to classification adopted in §5.2.

(76) a. Hindi मकड़ी का जाला makṛī kā jālā [spider POSTP web] SPIDER WEB

b. Nepali माकुराको जालो mākurā.ko jālo [spider.GEN web] SPIDER WEB

Inflection or derivation? gen ~ adj
It was noted above (footnote 1 on page 160) that the definitions of gen and adj do not make reference to the notions of inflection and derivation, but rather to the distinction between transpositional (word-class changing) and non-transpositional affixation. The reason for this is that the traditional distinction between inflection and derivation, whereby derivational affixes change the word-class of their base, while inflectional affixes do not, has been shown to be too simplistic. Haspelmath (1996) uses the example of Slavic possessive adjectives to show that the difference between inflection and derivation is one of degree, with Upper Sorbian being at the inflectional end of the scale and Russian more towards the derivational end. Defining gen and adj in terms of inflection and derivation would thus result in intermediate forms. Defining them in terms of transposition, on the other hand, results in a more clear-cut distinction.

Head replacement adj ➔ der
The type adj belongs to level 1 in the classification whereas der belongs on level 0; the former has three components whereas the latter has just two. An intermediate between these two is represented by the Slovak word železnica (77a). The structure of this word parallels that of the Russian železnaja doroga (77b) precisely, except for the use of the nominalizing suffix -ica instead of a lexical head.

(77) a. Slovak želez.n.ica [iron.ADJZ,NMLZ] RAILWAY

b. Russian želez.naja doroga [iron.ADJZ road] RAILWAY

Thus, in one sense the word belongs on level 3 under adj. On the other hand, as a derived word it has more in common with other derivations and, indeed, Slovak
linguists recognize an alternative synchronic analysis, železnica [iron.NMLZ], an undoubted instance of the **der** type:

There are two possible starting points for the analysis of the word železnica:

1. It is derived from železo (iron) and can be paraphrased as follows: “the object which is related to iron” (which moves on iron)

2. It is derived from železný (iron**ADJ** as univerbization from železná dráha (‘railway’) (Martina Ivanova, p.c. via Lívia Körtvélyessy).

This form can thus be seen as intermediate between **adj** and **der** and represents a type that occurs rather often in certain Slavic languages, in which the head element of an adjectival binominal is replaced by a more general nominalizing suffix.

### Morpheme loss dbl ➔ con

The final example of intermediate (gradient) phenomena is that of morpheme loss. Citing data from Hungarian, Kirmandji, Arbore and Maltese, Koptjevskaja-Tamm (2003) shows that “the step between double-marking and head-marking [in PNPs] is not necessarily big”:

Head-marked PNPs in Maltese, similarly to head-marked PNPs in Kirmandji, have developed from earlier double-marked PNPs, partly due to the breakdown of the case system of modern Arabic dialects compared to Classical Arabic, in which the posses- sor regularly appeared in the genitive case (p. 647).

The same appears to be the case with binominals, and not just between **dbl** and **con** (the example shown in Figure 36), but also between **dbl** and **gen**, between **prp** and **gen**, and between **gen** and **con** on the one hand and **cmp** on the other. Or more generally, between any strategy involving \( n \) additional morphemes and one including \( n-1 \) morphemes. One particularly striking example is Welsh, in which the dominant type at an earlier stage of the language was **gen** (as it still is in Irish), but following the loss of case marking is today **jxt** and **cmp**. Elsewhere in the database there are indications that this process is at work in Galibi Carib, Tarifit and Swahili.

In the case of Galibi Carib (78a-c) are **dbl**, **gen** and **con**, respectively. The double-marked pattern (a) may represent an earlier construction from which the others have developed.

(78) Galibi Carib

a. *emo.li sakila.li* [nose.POSS aperture.poss] NOSTRIL

b. *pana.li weti* [ear.POSS dirtiness] EARWAX

c. *manati poti.li* [breast tip.POSS] NIPPLE OR TEAT
(79) is one of three words in which the preposition \( n \) is given as optional. With the preposition the construction is considered an instance of \textit{prp}; without it, it is an instance of \textit{gen}; but both are dependent-marked since the modifier is in the (Berber) construct state (cf. page 156).

(79) Tarifit \textit{tisi (n) ufus} [bottom (of) hand:STC] PALM OF HAND

There is also an example in Swahili (80) of a construction in which the associative marker is given as optional. Since it is the only occurrence, it is classified as \textit{prp} along with other words that exhibit this marker, but it may also indicate gradience.

(80) Swahili \textit{gari (la) moshi} [car (CON) smoke] TRAIN

5.6 Data analytics

In §4.4 the binominal data were described in terms of very basic properties, viz. vocabulary size, analysability and frequency, by language and by meaning. In this section the data are presented in terms of the morphosyntactic strategy they exhibit, in other words, binominal type. In the next section I describe the distribution of binominals by type, raise the issue of compounding as a universal, and introduce the notion of binominal preference. I then discuss intralingual competition (§5.6.2) before returning to the universality of compounding (§5.6.3). Finally, in §5.6.4, I discuss “binominal fingerprints” in the context of genetic and areal patterning.

5.6.1 Distribution of binominals by type

\textbf{Overall frequencies}

Figure 37 shows the number of binominals by type, which gives the following overall ranking:

(81) Overall ranking: \textit{cmp} >> \textit{jxt} >> \textit{gen} > \textit{der} > \textit{con} > \textit{prp} > \textit{adj} > \textit{dbl} > \textit{cls}

The relative frequency of the \textit{cmp} and \textit{jxt} types (1,922 out of 3,738, or 51.4\%) is striking and confirms the widely-held belief that compounding (defined here as the use of either the \textit{jxt} or \textit{cmp} strategy) is the most common method of word formation in the world’s languages (e.g. Dressler 2006; Gagné & Spalding 2006; Booij 2007; Bauer 2009).
Lest it be thought that this overwhelming bias in favour of compounding is due to areal bias in the language sample (cf. §3.2.3), Figure 38 provides a breakdown by geographical area in terms of percentages. It shows that the areas that are most widely represented in the sample (Africa and Eurasia, with 24 and 37 languages, respectively) actually have lower than average proportions of the jxt and cmp types combined (25.9% and 44.1% respectively); this suggests that Figure 37, if anything, underestimates the predominance of these types across the world. But is compounding a universal, as has been suggested? I return to that question in §5.6.3, following a discussion of intralingual competition.
Further insights to be gleaned from the two figures include the fact that gen, the next most common strategy after jxt and cmp, is especially frequent in Africa and Eurasia, and almost unknown in Oceania/SE Asia. The prp type is found mainly in Africa and Eurasia, but also in Oceania/SE Asia, whereas the adj type is mostly limited to Eurasia. The der strategy is found everywhere (but particularly in Eurasia and North America), as is the con strategy (except among pidgins and creoles). Finally, dbl and cls occur too infrequently in the data for any reliable conclusions to be drawn.

**Language preferences**

Almost every language in the sample shows a preference for one or another type of binominal: 37 of the 47 binominals in Amharic are of type gen, 15 out of 22 in Iraqw are con, 26 of 43 in Lower Sorbian are adj, etc. Sometimes the preference is very marked; for example, all 51 binominals in Vietnamese are jxt. Sometimes it is less so: of the 47 binominals in Baa, 26 are cmp and 20 are gen. Applying Dryer’s (2013) relative frequency criterion for dominance, whereby a dominant form should be at least twice as frequent as any other, 70 out of the 106 languages in the sample can be said to have a dominant binominal type. In a few languages there is no one type that occurs more frequently than the others: five of the 14 Æiwoo binominals are jxt and five are prp, Galibi Carib has six each of con and jxt out of a total of 20, and of Selice Romani’s eight, four are gen and four are der; such languages are described as ‘mixed’.

Table 26 shows preferences for each language, and Figure 39 summarizes the data in the form of a bar chart, in order to enable comparison with the overall distribution of binominal types shown in Figure 37 (see page 170).

While the shape of the two diagrams is fairly similar, we observe that the rankings differ in one crucial respect: the overall ranking of der is higher than its ranking in terms of preference:

(82) Overall ranking: cmp >> jxt >> gen > [der] > con > prp > adj > dbl > cls

Preference ranking: cmp >> jxt >> gen > con > prp > [der] > dbl > adj > cls

This indicates that while derivational word-formation is rather widespread, it is seldom the principal word-formation strategy, at least in the language sample of this study: it is the preferred type in seven of the 106 languages (cf. Table 26) and the dominant type in just two: Central Yupik and Puyuma.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>type</th>
<th>count</th>
<th>languages (n=106)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmp</td>
<td>28</td>
<td>Baa, Bambara, Basque, Caijia, Chakali, Dutch, English, Finnish, German, Harakmbut, Hindi, Hungarian, Japanese, Kam, Ket, Korean, Malagasy, Malayalam, Mamara Senoufo, Mandarin Chinese, Navajo, Norwegian, Old High German, Querétaro Otomi, Saramaccan, Thai, Ticuna, Tuwari</td>
</tr>
<tr>
<td>jxt</td>
<td>26</td>
<td>Bandial, Cabécar, Ceq Wong, Datoga, Gurindji, Hawaiian, Hmong Daw, Ho-Chunk, Hupdë, Imbabura Quechua, Indonesian, Kalamang, Kildin Sami, Manange, Mapudungun, Mbyá Guaraní, Seychelles Creole, Srenge, Vietnamese, Walman, Warta Thuntai, Welsh, Western Mari, Wichi, Wik-Mungkan, Yaqui</td>
</tr>
<tr>
<td>gen</td>
<td>15</td>
<td>Amharic, Archi, Assamese, Bezhta, Estonian, Gawwada, Greek, Irish, Kambaata, Kanuri, Latvian, Nepali, Sidamo, Wawa, Zinacantán Tzotzil</td>
</tr>
<tr>
<td>con</td>
<td>10</td>
<td>Anindilyakwa, Hausa, Hebrew, Iraqw, Kekché, Kupsabiny, Turkish, Western Farsi, Wolof, Yakut</td>
</tr>
<tr>
<td>prp</td>
<td>8</td>
<td>Barain, French, Italian, Maltese, Romanian, Swahili, Tagalog, Tarifit</td>
</tr>
<tr>
<td>der</td>
<td>7</td>
<td>Central Yupik, Croatian, Lithuanian, Oroqen, Polish, Puyuma, Slovak</td>
</tr>
<tr>
<td>dbl</td>
<td>4</td>
<td>Akkadian, Seri, Somali, Takia</td>
</tr>
<tr>
<td>adj</td>
<td>3</td>
<td>Czech, Lower Sorbian, Russian</td>
</tr>
<tr>
<td>cls</td>
<td>2</td>
<td>Murui Huitoto, Trinitario</td>
</tr>
<tr>
<td>mixed</td>
<td>3</td>
<td>Äiwoo, Galibi Carib, Selice Romani</td>
</tr>
</tbody>
</table>

Table 26: Preferred binominal types by language

![Diagram showing preferred binominal types by language](image)

Figure 39: Preferred binominal types by number of languages

The geographical distribution of preferred binominal types is shown in Figure 40.
Intralingual competition

One striking difference between the sampled languages is the number of binominal strategies employed. Some, like Caijia, use only one, while others use as many as six different strategies. This is illustrated in Figure 41, which gives the number and ISO codes of every language grouped by the number of strategies it exhibits, based on the nine-way classification which does not take constituent order into account. As a result, facts about languages such as Vietnamese and Polish, which exhibit mixed order within one and the same type (cf. page 261), are not captured. Thus, Polish (POL) exhibits five different binominal types (cmp, prp, gen, adj, and der), of which two (cmp and adj) can occur as either left-headed or right-headed.
As the figure shows, there are nine languages in the sample that only employ a single morphosyntactic strategy: Akkadian, Caijia, Ceq Wong, Datooga, Imbabura Quechua, Seychelles Creole, Tuwari, Vietnamese and Walman. When a language only exhibits a single type, it is almost invariably either jxt or cmp (in other words, one of the two compounding strategies); the only exception in the present data is Akkadian, which only exhibits the dbl type. This underlines the flexibility of compounding cross-linguistically. However, as the figure shows, users of most languages have at least three different strategies to choose between when forming a new binominal. Understanding the various factors that mediate between competing strategies in such cases may prove to be a fruitful area of further research.

As a first step, we can investigate the correlations between different binominal types using the R function cor, and plot the result using a function from the library corrplot (Figure 42). Positive and negative correlations are denoted by colour (blue and red, respectively), and the strength of the correlation by both the size of the circle (and colour intensity) in the upper right triangular field and the value in the lower left field. An explanation for the relatively strong positive correlation between der and adj (with a value of 0.47) awaits further investigation. The strongest negative correlations suggest that cmp tends to be incompatible with dbl (and to some extent con), gen with con and jxt, and jxt with both der and gen. However, none of these correlations are particularly strong and they will therefore not be pursued here.

![Figure 42: Correlations between binominal types](image-url)
5.6.3 Compounding as a universal

Having introduced the notion of intralingual competition and noted that the overwhelming majority of languages have multiple strategies to choose between when forming binominals, we can now consider the question of compounding as a universal from a new and more profitable perspective. Before doing so, however, we should reiterate the point made earlier about the paucity of cross-linguistic studies of compounding: Greenberg does not mention the topic and it is completely absent from the Konstanz Universals Archive. Bauer (2001), in what is still the only study of compounding based on a genealogically and areally balanced sample, states that “it is not clear to what extent compounding can be regarded as a universal”.

Despite this, the literature abounds with statements such as “compounding might be considered to be the universally fundamental word formation process” (Libben 2006), “compounding is a common word formation process in all languages” (Gagné & Spalding 2006) and “compounds are present in all languages of the world (as far as described in grammars)” (Dressler 2006).

There are a number of problems with this mainstream view. The first, as Bauer (2017) points out, is that “there is no overall agreement on such basic issues as the definition of a compound.” Definitions vary from one language to another; the term is (at best) a language-specific descriptive category based on formal criteria and is thus not suitable for cross-linguistic comparison. Secondly, even within a single language, different types of compound exhibit very different properties: English Determinative Compounds involve an unstated semantic relation that allows for relatively free interpretation, whereas Synthetic Compounds contain an actional element that invokes the argument structure of the verb and puts great constraints on the interpretation; Mandarin Nominal Compounds are right-headed whereas Verbal Compounds are left-headed; etc. As a result, the very notion of compound invites us to compare apples and oranges. Thirdly, there is the question of what counts as evidence for the claim that a particular language “has” compounds: are the “few sporadic forms that can be considered compounds” in West Greenlandic (Sadock 2003) sufficient to justify an affirmative answer? Finally, one might ask whether the question of compounding as a universal is even particularly interesting. If the answer is yes, so what? And if no, again, so what?

A more profitable approach may be to start with a comparative concept rather than a descriptive category; to limit the question to a single type or function; and to frame it in terms of tendencies and degrees of preference rather than absolutes. The comparative concept chosen for the present study – binominal lexemes – does not
allow us to address the question of compounds in general, since it restricts the field of enquiry to noun-noun compounds and their functional equivalents, or more precisely, lexical items formed through the information packaging function of object modification (Croft forthcoming). However, since noun-noun compounds are generally believed to be the most widespread form of compounding (cf. Figure 7 on page 40), they constitute the best available proxy for compounding in general. The fact that the comparative concept of binominal lexeme covers multiple – and sometimes competing – strategies, in addition to compounding, then allows us to approach the matter in terms of preferences for one strategy over and above another. Thus, instead of asking ‘is compounding a universal’, I will reframe the question as:

- To what extent are simple strategies the preferred way to combine two thing-roots to create a new lexical item?

The use of the term ‘thing-root’ restricts the strategies to jxt and cmp, and excludes der and cls, the other two strategies that do not involve an additional morpheme.

Ideally, the notion of extent should be defined not only in terms of frequency but also productivity. Productivity cannot be addressed on the basis of the present data, but frequency can.

For the purpose of this discussion I define ‘attested’ to mean that there is at least one example of the type in the data, ‘common’ to mean that at least 10% of the binominals are of the given type, and ‘dominant’ to mean that 50% or more of the binominals are of this type.

Figure 43 shows how widespread the various strategies are across the languages in the database. The darker, leftmost column in each group of three shows the number of languages in which the type is attested (even if only by a single binominal); the
lighter column shows the number of languages in which the type is common; and the rightmost column shows the number of languages in which it constitutes over half of the total. Thus, \textit{jxt} is attested in 76 of 106 languages, common in 53 and dominant in 22; \textit{cmp} is attested in 68, common in 49 and dominant in 24; etc.

The corresponding numbers for languages in which \textit{either jxt or cmp} (or both) are attested, common and dominant are 98, 80 and 53, respectively. In other words, compounding, as defined here, is attested (in our data) in 98 of 106 languages and thus unattested in eight (Akkadian, Czech, Central Yupik, Kupsabiny, Puyuma, Selice Romani and Slovak); it is common in 80 and dominant in 53. Of course, one cannot claim on the basis of evidence from just 100 meanings that compounds never occur in the languages in question, but such occurrences are either exceptions that “prove the rule” that these languages do not have noun-noun compounds (just as left-headed English compounds like \textit{attorney-general} are the exceptions that “prove the rule” that English compounds are right-headed), or else represent a relatively marginal strategy. The numbers also show that noun-noun compounding is uncommon in 28 languages, in which they account for fewer than 10% of binominals, and that they are only the dominant strategy in 53 languages.

The following conclusions can be drawn regarding noun-noun compounds on the basis of the present data:

- NN compounding (at least) is \textbf{NOT} an absolute universal. It is non-existent or marginal in 7.5% of languages, uncommon in 24.5% of languages, and dominant in just 50%.
- However, NN compounding is by far the most widespread form of binominal word-formation, accounting for roughly half of all binominals.
- Moreover, if a language exhibits just one binominal strategy, that strategy is overwhelmingly likely to be one of compounding (i.e. \textit{cmp} or \textit{jxt}).

To the extent that NN compounding is considered a valid proxy for compounding in general, the same conclusions can be extended to the latter.

5.6.4 Genealogical and areal patterning

Building further on the notion of intralingual competition, every language can be said to have a “binominal fingerprint”, which captures its preferences for various types of binominal. Thus, languages like Imbabura Quechua and Vietnamese have a simple fingerprint that can be expressed as \textbf{[100\textit{jxt}]}. At the other end of the scale, Polish has a complex fingerprint consisting of a mixture of five different types in varying proportions, viz. \textbf{[4\textit{cmp}7\textit{prp}5\textit{gen}34\textit{adj}50\textit{der}]} . The question arises whether these
fingerprints show any kind of genealogical or areal patterning. In other words, can the binominal fingerprint tell us what family a language belongs to, or where in the world it is spoken? Here I focus on genealogical affiliation. A prerequisite for determining whether the binominal fingerprint is a good indicator of genus is that the database contains data from multiple languages from the same genus. When such data exist, genera can be compared genetically (within the same language family) or areally (within the same geographical area).

Figure 44: The Indo-European spectrogram

Figure 44 shows the situation in Indo-European, for which the data covers multiple languages in six genera. A visual inspection of the chart suggests that Germanic, Romance and Slavic exhibit quite distinctive and rather consistent patterns, that the patterns are less consistent across Baltic and Indo-Aryan, and not at all consistent in Celtic. Not surprisingly, the exception in Indo-Aryan is Selice Romani, which like all varieties of Romani has adopted the derivational word-formation strategy (der) of the Greek, Romance and Slavic languages it has been in contact with since separating from its Indo-Aryan siblings. The difference between Hindi on the one hand and Assamese and Nepali on the other is less dramatic than might appear; the typically Indic inflecting postpositions still found in Hindi (prp) have simply been reduced (through grammaticalization) to case-marking suffixes (gen) in Assamese and Nepali. As for Celtic, it is noticeable that Welsh patterns more closely with

---

1 Note that this is an example of the kind of question that cannot be addressed using a “balanced” sample.
English than with Irish, a result, no doubt, of both language contact and the loss of the case system in Welsh (the latter, to some extent, perhaps, due to the former).

A more robust way of exploring genetic patterning is to use a statistical clustering method such as Principal Components Analysis (PCA), in which data distributed across a multi-dimensional space is reduced to fewer dimensions. The binominal data occupies a nine-dimensional space (one for each type of binominal), which is impossible to visualize, of course. In Figure 45 each language is instead represented as a point in a three-dimensional space, the axes of which are the three principal components, PC1, PC2 and PC3.\footnote{The following biplots provide keys to the three top left quadrants (R2C1, R1C1 and R1C2) of Figure 45, respectively. They are best read by zooming in the electronic version.}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure45}
\caption{The Indo-European scatter plot}
\end{figure}

The six quadrants containing points provide views onto the data as if through the six faces of a cube. The quadrant in the first column of the second row (R2C1), with the axes PC1 and PC2, provides the clearest view of the data, since these two principal components account for almost 80% of the variance in the data. We can...
observe that the Germanic languages tend to cluster in the bottom left of this quadrant (the outliers are ENG English and GOH Old High German), while Slavic clusters in the bottom right. Romance and Indo-Aryan appear to cluster with each other, but here the two-dimensional view onto the three-dimensional space misleads us. Although the two groups both occupy the same top-right corner, they are actually situated at different depths: their X and Y values are similar, as it were, but their Z values are different. Viewing the data through another face of the cube, either R1C1 (PC1×PC3) or R1C2 (PC2×PC3), reveals that there is actually quite a clear separation between these two genera. There is a marked difference between the two Celtic languages in all three quadrants. In R2C1 Irish (Gle) appears to be closer to Germanic than Welsh (Cym), but the depths are different, as can be seen from the other two quadrants, where Welsh is closer to English. In fact, Irish is much closer to the Baltic languages, in particular Latvian (Lav).

Taken together these results indicate that the binominal fingerprint can often be a rather good indicator of genus, at least in the case of Indo-European, the only family for which sufficient data is available at present.

5.7 Chapter summary

In this chapter, I developed a typological classification of binominals based on the morphosyntactic strategies employed. Nine binominal types were identified:

- four types consist solely of two thing-morphs with no additional material:
  - jxt (juxtaposition)
  - cmp (compounding)
  - der (denominal derivation)
  - cls (derivation classifier constructions)
- three types have an additional marker attached to the modifier:
  - prp (with a marker that is an independent lexeme, typically an adposition)
  - gen (with an affixal case marker, often genitive)
  - adj (with a transpositional morpheme, usually adjectival or attributive)
- one has a non-transpositional affix or segmental marker attached to the head:
  - con (“construct”)
- one has markers attached to both the head and the modifier:
  - dbl (double-marked)

Each of these can occur as a head-initial or head-final construction.
In order to avoid issues inherent in traditional, hierarchical representations, these nine types were presented in a two-dimensional grid whose primary parameters are the number of markers and the locus of marking; parts of the grid also express the degree of fusion. Each type was described in detail together with examples.

The non-hierarchical grid representation revealed two apparent “gaps” in the data for which explanations were offered: whereas the type prn (“pronominal”), the head-marking equivalent of prp, can be expected to occur, albeit relatively seldom, the type nml (“nominalized”), the head-marking equivalent of transpositional adj, is a logical impossibility, since the result would not be a naming unit.

I also discussed a number of gradient phenomena, many of which can be explained in terms of grammaticalization.

The analysis of the distribution of binominals by type, revealed the predominance of the jxt and cmp strategies, both overall and in terms of the preferences shown by individual languages. This also showed how the degree of intralingual competition between different strategies varies considerably from one language to another. While I could confirm the prevalence of noun-noun compounding, its status as a universal was disconfirmed, and one can probably extrapolate this to compounding in general.

Finally, I proposed the novel concept of a language’s binominal fingerprint, which I investigated for Indo-European, the only language family for which the database contains enough data across sufficient genera. The result indicates a potential for binominal fingerprints to contribute interesting insights into both genetic affiliation and the history of language contact.

Other aspects of binominal typology will be examined in Chapter 7. These include constituent order, the relationship between binominals and possessives (or more precisely, anchoring nominal modifier constructions), and the interaction between morphosyntactic strategies and semantics. Before that, however, we must shift our attention from morphosyntax to the nature of the semantic relation that is one of the defining features of binominal lexemes.
6 Semantic relations

In the previous chapter I developed a classification of binominals based on the morphosyntactic strategy employed. Here I develop a second classification, this time based on the kind of semantic relation that pertains between the two major constituents of a binominal. This unstated (or underspecified) semantic relation is a defining feature of both binominal lexemes and the canonical subtype, noun-noun compounds. Guevara & Scalise (2009) suggest that the “inner essence” of a compound can be captured (in the prototypical case) with a “rough schema” (83), where X, Y and Z represent “major lexical categories” (in the case of binominals, these are, of course, nouns or other thing-morphs), and \( \Re \) “represents an implicit relationship between the constituents (a relationship not spelled out by any lexical item)” (p. 107).

(83)  \[ X_N \Re Y_N \] Z_N

Jackendoff (2016) provides a nice set of examples (84) to show that the kind of semantic relation can be “hugely varied”, even across compounds that share a common hypernym, such as cake.

(84)    chocolate cake ‘a cake made with chocolate in it’
    birthday cake ‘a cake to be eaten as part of celebrating a birthday’
    coffee cake ‘a cake made to be eaten along with coffee and the like’
    marble cake ‘a cake that resembles marble’
    layer cake ‘a cake formed in multiple layers’
    cupcake ‘a little cake made in a cup’
    urinal cake ‘a (nonedible) cake to be placed in a urinal’

In N+N compounds (i.e. binominals of type jxt or cmp) there is no indication of the nature of the semantic relation, and the same applies to the types der and cls that also belong to Level 0 in the formal typology. In types that belong to Levels 1 and 2 the semantic relation is underspecified rather than unstated: the presence of one or more additional morphemes indicates the presence of some kind of relation. However, the meaning inherent in those relational morphemes is extremely schematic. The fact that \( \Re \) is not stated explicitly gives rise to at least two questions (Bauer 2017):
1. How can these semantic relationships best be classified?

2. What mechanism allows compounds to be generated (and understood) when the semantics is so variable?

The second of these questions is the focus of Hacken (2016), in which the process of compounding is examined in the context of three theoretical frameworks, viz. Jackendoff’s Parallel Architecture, Lieber’s Lexical Semantics and Štekauer’s Onomasiological theory. The present chapter concentrates on the first of the questions posed above and investigates the kinds of semantic relation found in binominals. To that end, §6.1 provides a brief overview of previous work on this topic: while most of this relates to noun-noun compounds, there is reason to believe that the same approach can be applied to other types of binominal (§6.1.2), including denominal derivations (§6.1.3). This overview is followed in §6.2 and §6.2.2 by detailed descriptions of the two classification schemes on which the present study is based: the one, rather granular, developed by Yves Bourque, the other, more schematic, by Anna Hatcher. The topic of §6.2.3 is Pierre Arnaud’s innovative mapping of his low-level classification to Hatcher’s high-level classification. The actual process of classifying binominals in the present study is discussed in §6.3. This is followed by data analytics in §6.4 and a summary in §6.5.

6.1 Theoretical prerequisites

The history of research into semantic relations in compounding is often rehearsed by researchers entering the field (e.g. Ryder 1994; Pepper 2010b; Szubert 2012; Bourque 2014; Eiesland 2016; Toquero 2018). The reader is referred to these and similar works for more historical details. The aim of the present section is simply to present a brief overview and prepare the ground for a more detailed survey of work that is particularly relevant to my own work.

6.1.1 Background

Interest in semantic relations can be traced at back least as far as Grimm (1826) on compounding in German, and Mätzner (1860) on English, but the starting point for modern studies of the topic is usually taken to be Jespersen’s (1942) treatment in Volume 6 of his Modern English Grammar on Historical Principles. This was followed by a large number of influential and widely cited studies, notably Hatcher

---

1 The cognitive approach (or more precisely, that of Cognitive Grammar) to word-formation is described by Tuggy (2005), and psycholinguistic aspects are explored by Libben & Jarema (2006).

2 Other early work worthy of mention includes Bergsten (1911) and Carr (1939).
Chapter 6. Semantic relations


**Transformational** approaches attempt to derive semantic relations from underlying syntactic structures such as relative clauses; Søgaard cites Rhyne’s (1976) example *ignition key*, which is claimed to come from *key which causes ignition*. The most well-known studies of this type are Lees (1960; 1970) and Levi (1978), which are situated within the frameworks of early generative theory and Generative Semantics, respectively. This line of research was abandoned after the 1970s (Hacken 2009), but while the underlying theories are no longer regarded as viable, the actual results achieved, especially by Levi, have had a lasting effect (see below). Furthermore, Levi’s insight that the kinds of semantic relation found in compounds tend also to appear at the clause level is worth retaining; I will return to it in Chapter 8.

In **slot-filler** theories of compounding constituents are seen as bundles of features with the modifying constituent supplying a value to one of the features of the head. Søgaard refers in particular to Johnston & Busa’s (1999) work within the framework of the Generative Lexicon (Pustejovsky 1995), but the slot-filler approach applies equally to Allen’s (1978) work in the lexicalist framework and that of Rochelle Lieber in lexical semantic analysis (Lieber 2009; 2016). It also applies (in some degree) to the Cognitive Grammar approach exemplified by Tuggy (2005), in which “elaboration sites” equate to slots.¹

Most work on semantic relations, however, has followed what Søgaard calls the **reductionist** approach, in which the researcher attempts to enumerate a limited number of “primitive relationships”. These vary in number from four (in the case of Hatcher 1960) to well over 100, depending on whether one includes subtypes, i.e. the degree of granularity of the analysis. Bourque (2014: 167) provides a table listing 16 studies of this type and the number of relations that they posit. The following list offers a representative selection of such studies within a variety of theoretical frameworks:

¹ For example, Langacker (2008) describes *lid* in *jar lid* as having an elaboration site that is elaborated (specified in finer detail) by *jar*, and points out that “a serious analysis of this or any other construction requires that each structure and each relationship be described in explicit detail” (p. 162).
• **Jespersen (1942):** Jespersen (p. 142ff) identifies eight types of NN compound AB in which B is modified by A: Subject and Object (*sunrise, sun-worship*); Place, including source and goal (*land-breeze, side-glance*); Time, including duration (*nightmare, day-fly*); Purpose (*beehive*); Means (*handwriting*); Characterizing Feature (*sandpaper*); Similarity (*needle-fish*) and Material (*gold ring*).\(^1\) He asserts that the number of possible logical relations between the two elements is “endless” and lists several examples from “the large residue of compounds which do not fit in anywhere”, including *sunflower, sun-dial, weathercock* and *rainbow*. See §6.2.2 for further discussion.

• **Hatcher (1960)** (logic-based): Hatcher reduces seven of Jespersen’s types and two of Mätzner’s to four “logical” types: “A is contained in B” (*seed orange*); “B is contained in A” (*orange seed*); “A is the source of B” (*cane sugar*); and “A is the destination of B” (*sugar cane*). See the discussion in §6.2.2.

• **Levi (1978)** (Generative Semantics): Levi defines nine “recoverably deletable predicates” (RDPs), of which the first three are reversible:\(^2\) *CAUSE* (*tear gas, drug deaths*); *HAVE* (*picture book, lemon peel*); *MAKE* (*honeybee, daisy chains*); *USE* (*steam iron*); *BE* (*soldier ant*); *IN* (*field mouse*); *FOR* (*horse doctor*); *FROM* (*test-tube baby*); *ABOUT* (*tax law*). In addition, she defines four “nominalization types”: *Act* (*parental refusal*); *Product* (*clerical errors*); *Agent* (*city planner*); and *Patient* (*student inventions*). Levi’s assumption is that complex nominals are derived via transformations (including predicate deletion) from underlying sentences, and her deletable predicates are declared to be “recoverable” in order to abide by Chomsky’s (1965) condition that only “recoverable deletions are permitted”.\(^3\) See further discussion in §6.1.2.

• **Warren (1978)** (functionalism): Warren delineates 12 “semantic classes”, most of them named according to the role played by the constituents in the relation: *Source-Result* (*student group*); *Whole-Part* (*spoon handle*); *Part-Whole* (*armchair*); *Size-Whole* (*3-day affair*); *Goal-Obj* (*moon rocket*); *Place-Obj* (*sea port*); *Time-Obj* (*Sunday paper*); *Origin-Obj* (*hay fever*); *Activity-Actor* (*cowboy*); *Copula* (*girl friend*); *Resemblance* (*clubfoot*); and *Purpose* (*ball bat*) (cf. Warren’s summary, p. 229). These are further divided into 70 subclasses.

---

\(^1\) The labels given here are mine. Jespersen employs descriptive phrases.

\(^2\) A reversible relation is one in which the roles can be reversed, as in Whole-Part vs. Part-Whole (cf. the examples from Warren 1978 in the next bullet point).

\(^3\) Hacken (2009) expresses some doubt as to whether this manoeuvre conforms to the “letter and spirit” of Chomsky’s condition.
• **Ryder (1994)** (Cognitive Grammar): Ryder identifies 50 templates representing compound schemas that emerged from her psycholinguistic experiment. These are listed in her Appendix II and include CON [X contains Y] (bag lunch); its inverse, CONIN [X is contained in Y] (teapot); SELL [Y sells/delivers X] (fishwife); and TEND [Y tends/raises/Trains X] (horseman).¹

• **Jackendoff (2016)** (Conceptual Semantics): Jackendoff presents a list of 13 “basic functions”,² six of them reversible: CLASSIFY (beta cell); BE (boy king); SIMILAR (piggy bank); KIND (puppy dog, bear cub); BE AT/IN/ON (sunspot); COMP (tinfoil, sheet metal); MADE FROM (apple juice, sugar beet); PART (apple core, cheesecake), CAUSE (sunburn), MAKE (ant hill, honey bee); SERVES AS (guard dog); HAVE (AIDS baby, writer’s cramp); PROTECT (FROM) (lifeboat; mothball).

When Jespersen declared the set of possible semantic relations to be open-ended, he was echoing sentiments expressed earlier by Carr:

> Although an attempt may be made to classify the compounds from a semantic point of view, it would be impossible to state all the relationships which do occur, and to assign each compound to a particular class (Carr 1939: 319–320).

At the other end of the spectrum we find the fourth kind of theory of compounding identified by Søgaard, which he calls **pragmatic** theories. These claim that there is really only one, very general relationship between the constituents of a compound, and that the compound’s meaning is derived solely from pragmatic knowledge about the world. This position was advanced by Bauer (1979), couched in terms of the predicate deletion approach then current (cf. Levi 1978):

> I suggest that only one ‘verb’ (or more accurately ‘pro-verb’ since it is an abstract unit) should be deleted in the generation of compounds. If only one verb is to account for the range of semantic relations that exist between the two elements of compounds it will have to be very abstract and have a vague meaning. I suggest a gloss something like ‘there is a connection between’.

Bauer never entirely abandons this position. 35 years later he and Tarasova write: “We might prefer to say that rather than there being a set of adnominal relationships, there is just one, the adnominal nominal relationship itself. Such a relationship would have to be described in semantically very imprecise terms, since it covers such a wide range of territory…” (Bauer & Tarasova 2013).

---

¹ The latter two appear to be incorrectly paraphrased by Ryder in Appendix II as [X VERBS Y].

² These are the “most prominent” basic functions. The labels used here are slightly simplified.
All of the above-mentioned works focus exclusively on compounding in English, with a couple of exceptions that investigate German (or early Germanic) and Chinese. In recent years, however, the topic of semantic relations has been explored in other languages, including Nizaa (Pepper 2010b), Danish (Szubert 2012), French (Arnaud 2003; 2016; Bourque 2014), Norwegian (Eiesland 2016) and Spanish (Toquero 2018). It has also received a lot of attention in computational and corpus linguistics (e.g. Vanderwende 1994; Moldovan et al. 2004; Girju et al. 2005; Ó Séaghdha 2008; Tratz & Hovy 2010; Nakov 2013; Schäfer 2018) and was even the focus of an NAACL-HLT Workshop on Semantic Evaluations task on “the interpretation of noun compounds using paraphrasing verbs and prepositions” (Butnariu et al. 2009).

The position taken in the present study is that the number of relations one identifies will be a function of the degree of granularity one requires them to express. It can therefore be anything the researcher desires, from one (as suggested by Bauer) to unlimited (as opined by Carr and Jespersen). Furthermore, even if the total number turns out to be unbounded, “the vast majority”, as Tratz & Hovy (2010) point out, will fit within “a relatively small set of categories.” As we shall see, in the present study I actually apply two classifications: the one fairly granular, consisting of 25 relations, the other much more abstract and consisting of just four relations.

What the studies mentioned in this section have in common is that they all focus on NN compounds, i.e. binominals of type cmp (§5.3.1), or in Levi’s case, cmp and adj (her “noun phrases with non-predicating adjectives”). However, Jackendoff does note (2016: 30) that the meanings of the compound construction overlap to an extent with those of N of NP and continues: “This suggests that we are dealing with a common stock of rather primitive semantic relations that can be expressed through various (morpho)syntactic frames, compounding among them.” We will return to this important point in §8.3, but for now the task before us is to examine evidence that the same kinds of semantic relation can be applied not just to NN compounds, but to binominals in general.

### 6.1.2 Nominal modification (Bauer & Tarasova 2013)

The work of Bauer and Tarasova was introduced in §2.3.3. The focus there was on the fact that their study of adnominal nominal modification in English covers six different constructions in which one noun modifies another, and thus prefigures the concept of binominal lexemes. In this section I am interested in their investigation into the semantic relation (or “meaning link”, as they call it), but I will start out by showing how their six constructions fit into the typology developed in the previous
chapter. The list of constructions and examples given in (7) on page 51 is repeated
below (85) with each of Bauer and Tarasova’s six types mapped to one of the nine
binominal types. This is followed by comments on each construction.

(85)  
(a) **jxt**  noun-noun compounds (steam iron)
(b) **adj**  associative (i.e. relational) adjective plus noun (manual labour)
(c) **gen**  prenominal possessives (car’s driver)
(d) **prp**  postnominal possessives (driver of the car)
(e) **der**  neoclassical compounds (hydromancy < water + divination)
(f) **cmp**  blends (paratroops < parachute + troops)

(a) Steam iron differs from eisen.bahn [iron.way] RAILWAY in that the elements
are not fused and is thus an instance of jxt rather than cmp.

(b) Manual labour is clearly an instance of adj, parallel to Rus. želez.naja doroga
[iron.ADJZ road]. The source of the derivation (Lat. manus ‘hand’) is not quite
so transparent, but the connection between manual and the hand is fairly clear
to most native speakers.

(c) Although car’s driver is not a binominal, the examples given in §2.3.3 (dog’s
breakfast, ladies’ man and wolf’s bane) are all parallel to Bezhta kilo.s hino
[iron.GEN way] and thus of type gen.

(d) Again, while driver of car is not a binominal (see §2.3.3), man-of-war most
surely is; it parallels Fr. chemin de fer [way PREP iron] and is thus of type prp.

(e) Neoclassical compounds like hydro.mancy present a more interesting case.
The traditional term for such constructions notwithstanding, they do not fit the
category cmp since hydro- and -mancy are thing-affixes, not thing-roots.1 (They
might also be regarded as intermediate forms that are more affix-like than root-
like, since they cannot occur as independent words.) Such forms can be re-
garded as a non-prototypical subtype of der and closer to the two morpheme
analysis of Slovak želez.nica [iron.NMLZ] (page 168) than to cmp.

(f) Finally, paratroops is best regarded as a non-prototypical subtype of cmp, as
it consists of two thing-roots, one of them a truncation (of parachute). To the
extent that para- (with this etymology2) is established as a prefix, paratroops
constitutes another gradient, affixoid-like phenomenon between cmp and der
(see page 164).

---

1 Wiktionary lists 349 English words containing the prefix hydro- and 98 words containing the suffix

2 Wiktionary lists 17 such words, including paraglider and parajump. Words formed from Greek παρά
or French/Italian para- do not count, even though parachute itself belongs to the latter type.
In conclusion, each of the six constructions considered by Bauer and Tarasova in their investigation of adnominal nominal modification in English fits very neatly into the formal classification of binominals.

I turn now to Bauer and Tarasova’s principal research question, which concerns the “meaning link” between the nominal constituents in each of their constructions. They present ample evidence for the fact that “the meaning relationships which can be found in [endocentric noun-noun] compounds … can also be found in a range of other constructions in which a noun modifies another noun.” As noted earlier (§2.3.3), they opt to use Levi’s set of nine recoverably deletable predicates (listed on page 186), explaining their choice as follows:

We adopt Levi’s list because it is relatively well-known, because it has been shown to provide good coverage of the data (see, for instance, the evaluation in Kunter 2011: 153), and because it provides an independent list of semantic relationships for us to work with. We should emphasize, however, that our use of this set of categories does not indicate any commitment on our part to the particular set that Levi provides; Levi’s set of categories is merely a convenient list and our decision to use this classification over the others is dictated by operational needs (p. 4).

A number of drawbacks are identified in Levi’s scheme, but these need not concern us here. The important point is that Bauer and Tarasova are able to find examples of every one of Levi’s 12 relations for each of their six constructions. Table 27 provides a complete listing.¹

The important conclusion to be drawn from Bauer and Tarasova’s work is that the kinds of relation that occur in noun-noun compounds also occur in other types of binominal, at least in English. Four of the eight types of binominal described in §5.3 (cmp, adj, gen and prp) are covered by their work, which also provides evidence concerning a subtype of der (neoclassical compounds).

¹ Notes on Table 27: Labels in parentheses and small caps in the Relation column are alternative names for Levi’s relations based on the role played by the modifier (N₁ in the case of English compounds). POSSESSOR-POSSESSUM is treated by Levi as a subtype of the PART-WHOLE (HAVE) relation. See Tarasova (2013: 43) for a naming scheme that uses the numbers 1 and 2 to indicate the direction of the relation (e.g. CAUSE₁ and CAUSE₂ for my CAUSE and RESULT); her terms are CAUSE₁, CAUSE₂, POSSESSION₁, POSSESSION₂, COMPOSITION₁, COMPOSITION₂, INSTRUMENT₂, ESSIVE₂, LOCATION₂, PURPOSE₂, SOURCE₁ and TOPIC₂.
<table>
<thead>
<tr>
<th>Relation</th>
<th>N N</th>
<th>AREL N</th>
<th>N’s N</th>
<th>N of N</th>
<th>Neo-classical</th>
<th>Blend</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1 CAUSE N2</td>
<td>sex scandal, withdrawal symptom</td>
<td>viral infection</td>
<td>nature’s bounty</td>
<td>smell of bourbon</td>
<td>hydathrosis, necrophobia</td>
<td>contrail, parascending</td>
</tr>
<tr>
<td>N2 CAUSE N1</td>
<td>tear gas, shock news</td>
<td>malarial mosquitoes</td>
<td>Israel’s creation</td>
<td>creation of Israel</td>
<td>cinematograph, oncogene</td>
<td>slimnastics</td>
</tr>
<tr>
<td>N1 HAVE N2</td>
<td>lemon peel, school gate</td>
<td>feminine intuition</td>
<td>dog’s breakfast</td>
<td>cost of the flight</td>
<td>neuroglia, opthalmia</td>
<td>channel, parawing</td>
</tr>
<tr>
<td>N2 HAVE N1</td>
<td>camera phone, picture book</td>
<td>industrial area</td>
<td>ladies’ man</td>
<td>owner of the cafe</td>
<td>liriodendron, odontocete</td>
<td>cineplex, dinosuem</td>
</tr>
<tr>
<td>N1 MAKE N2</td>
<td>court order, snowball</td>
<td>molecular chain</td>
<td>Kellogg’s cornflakes</td>
<td>Odyssey of Homer</td>
<td>electromagnetism</td>
<td>cremains, glassphalt</td>
</tr>
<tr>
<td>N2 MAKE N1</td>
<td>computer industry, silk worm</td>
<td>musical clock</td>
<td>letter’s author</td>
<td>writer of thrillers</td>
<td>phonorganon</td>
<td>Motown</td>
</tr>
<tr>
<td>N2 USE N1</td>
<td>steam iron, wind farm</td>
<td>manual labour</td>
<td>car’s driver</td>
<td>driver of the car</td>
<td>electrocardiograph, hydromancy</td>
<td>jazzercise, paratroops</td>
</tr>
<tr>
<td>N2 BE N1</td>
<td>island state, soldier ant</td>
<td>professorial friends</td>
<td>Dublin’s fair city</td>
<td>sign of the cross</td>
<td>cryptonym</td>
<td>thugson, foolosopher</td>
</tr>
<tr>
<td>N2 IN N1</td>
<td>field mouse, letter bomb</td>
<td>autumnal rain</td>
<td>Thursday’s lunch</td>
<td>people of the forest</td>
<td>nephralgia</td>
<td>Californication, airmada</td>
</tr>
<tr>
<td>N2 FOR N1</td>
<td>arms budget, steak knife</td>
<td>avian sanctuary</td>
<td>wolf’s bane</td>
<td>day of rest</td>
<td>speedometer</td>
<td>Identikit, palimony</td>
</tr>
<tr>
<td>N2 FROM N1</td>
<td>business profit, olive oil</td>
<td>solar energy</td>
<td>New Zealand’s wines</td>
<td>heat of the sun</td>
<td>otorrhoea</td>
<td>Chicagorilla, anecdata</td>
</tr>
<tr>
<td>N2 ABOUT N1</td>
<td>tax law, love letter</td>
<td>criminal policy</td>
<td>university’s statutes</td>
<td>Book of British Birds</td>
<td>ethnography, hydrograph</td>
<td>exploitation, snotlem</td>
</tr>
</tbody>
</table>

Table 27: Levi’s (1978) classification applied to six binominal types
The other three types (con, dbl and cls) are not covered, but there is other evidence to suggest that these, too, exhibit the same kinds of semantic relation. We have already encountered (in the discussion of binominal types in §5.3) examples of these types that parallel binominals of type cmp (and others), both in terms of the meanings of the constituents and the meaning of the whole. This is shown in (86)-(88), where the (a) is an instance of one of these three types, and (b) is an instance of cmp. In each pair of examples the same two meanings combine to produce the same target meaning. This strongly suggests that the same cognitive process, involving the same semantic relation, is at play in each pair. In terms of Levi’s classification, (86) would be instances of N₁ MAKE N₂ (cf. snowball in Table 27) while (87) and (88) would be instances of N₂ HAVE N₁ (cf. camera phone).

(86) IRON + TRACK = RAILWAY
(a) con Hebrew mesila.t barzel [track.CON iron] RAILWAY
(b) cmp German eisen.bahn [iron.track] RAILWAY

(87) EGG + YELLOW = YOLK
(a) dbl Takia patu.n kdabog.an [egg.3SG yellow.3SG] YOLK
(b) cmp Welsh melyn.ŵy [yellow.egg] YOLK

(88) EYE + HAIR = EYELASH
(a) cls Murui Huitoto ui.tїraї [eye.CL(hair)] EYELASH
(b) cmp Thai khǒn.taa [hair.eye] EYELASH

Evidence such as this suggests that an exploration of semantic relations across the whole 100 language sample might bear fruit. Similar encouragement is to be drawn from the affirmative answer reached by Rainer (2013) to the question posed in his paper: Can relational adjectives express any relation? (see §2.3.2). As for whether the same applies to binominals of type der, which Bauer and Tarasova did not explore: should the reader feel that the argument advanced above on the basis of neoclassical compounds is too tenuous, more evidence of the type just put forward in (86)-(88) can be found in (89).

(89) WIND + MILL or INSTRUMENT = WINDMILL
(a) der Polish wiatr.ak [wind.INST] WINDMILL
(b) cmp Finnish tuuli.mylly [wind.mill] WINDMILL

In this case the meanings of the two head constituents is not identical, but there exists a hypernymy relation between them: a mill is a kind of instrument. It therefore seems reasonable to assume that the semantic relation is the same in each case: a tuulimylly is a mill that is POWERED BY wind, while a waitrak is an instrument
Chapter 6. Semantic relations

that is powered by wind. In Levi’s terms these are both instances of $N_2$ USE $N_1$ (cf. steam iron). Further, and perhaps even more compelling, evidence that der binominals entail the same kinds of relation, is provided by Janda (2011), to which we now turn.

6.1.3 Metonymy and derivation (Janda 2011)

On the face of it, Janda is not concerned with semantic relations at all, but rather the role of metonymy in word-formation (or more precisely, derivational word-formation, since she does not deal with inflection). However, as Janda points out (p. 363), citing Jakobson & Halle (1956) and Peirsman & Geeraerts (2006), one of the three main strategies for classifying metonymy is the Aristotelian concept of contiguity and, furthermore, the approach adopted by Peirsman and Geeraerts’ is compatible with both of the other two strategies: Frames / ICMs and Domains / Dominions. Contiguity, characterized by Blank (2003) as relations that are based on “spatial, temporal and logical” connections between concepts, is thus fundamental to all modern accounts of metonymy.

Janda frames her argument primarily in terms of metonymy rather than contiguity and was taken severely to task for this by Brdar & Brdar-Szabó (2014). However, the latter’s objections boil down to a terminological disagreement. Essentially, Janda extends the definition of metonymy to cover contiguity relations in general, while her critics, as “metonymy-people”, are concerned to keep their turf pure. None of this, however, detracts from the insightful observations Janda makes about what we will call contiguity relations in derivation, starting with her very first examples (90), (91).

(90) PART FOR WHOLE
   a. We need a good head for this project.
   b. Russian brjuxan (lit. ‘belly’-an) ‘person with a large belly’
   c. Czech břicháč (lit. ‘belly’-áč) ‘person with a large belly’

(91) CONTAINED FOR CONTAINER
   a. The milk tipped over. (cf. Peirsman and Geeraerts 2006: 281)
   b. Russian saxarnica (lit. ‘sugar’-nica) ‘sugar-bowl’
   c. Czech květináč (lit. ‘flower’-áč) ‘flower-pot’

1 Aristotle defines three types of associative relation: contiguity, similarity and contrast. We will return to these later in a number of contexts. For an excellent account of the relationship between metonymy, contiguity, frames, domains, prototypes and gestalts, see Koch (1999).
The typology and semantics of binominal lexemes

Janda’s claim is that the English examples (a) illustrate lexical metonymy, while the Russian (b) and Czech (c) examples illustrate parallel examples of metonymy in word-formation. If we reformulate this to state that the same contiguity relations are exhibited in (90a-c) and in (91a-c), there can be no doubt that Janda is correct: in (90) it is the PART-WHOLE relation and in (91) it is the CONTAINER-CONTENTS relation. Now, these are exactly the kinds of semantic relation that one finds in binominals. This is immediately obvious in the glosses of (91b) and (91c), i.e. Eng. sugar bowl and flower-pot, both of which are binominals of type cmp (and instances of Levi’s N₂ HAVE N₁ relation, cf. picture book in Table 27).

Exact parallels for the Russian and Czech words in (90) are harder to find, for the simple reason that belly tends to default to the human belly, so there is little need for a binominal lexeme that combines the concepts of PERSON and BELLY. But the same semantic relation is evident in words denoting the bellies of other animates, such as Ger. schwein.e.bauch [pig.LE.belly] ‘pork belly’. PART-WHOLE is in fact by far the most common semantic relation in our data, as we shall see in the next chapter, and the database is replete with body-part binominals based on that very relation. The examples in (92) are just the tip of a very large iceberg.

(92) a. gen Navajo ’a.ké.ts’iin [3SG.foot.bone] ANKLE
   b. dbl Hebrew tenux ha.'ozen [lobe:STC DEF.ear] EARLOBE
   c. cls Trinitario ugi-mo [eye-CL(fabric)] EYELID
   d. con Hausa bàaki.n màamá [mouth.LK breast] NIPPLE OR TEAT
   e. cmp Querétaro Otomi oky.xiñu [hole.nose] NOSTRIL
   f. der Central Yupik cu.araq [person.AR(AQ)] TOE¹
   g. adj Lower Sorbian ruc.ne zgibk [hand.ADIZ joint] WRIST

All of these exhibit the PART-WHOLE relation which, like CONTAINER-CONTENTS, is another subtype of the N₂ HAVE N₁ relation in Levi’s scheme. Thus, the two contiguity relations underlying Janda’s initial examples of derivation are seen to be present in binominals of all types. Janda provides a plethora of other examples, not only in her paper but also in the data set that she has made publicly available (Janda 2014). I return to her work when considering metonymy in the larger context of associative relations in §9.1.

¹ The postbase (suffix) -araq is described as a “little piece of N”, where N refers to the base, here cu-‘person’ (Jacobson 1995: 741). A toe in Central Yupik is thus literally a little piece of a person.
6.2 An integrated approach to classification

6.2.1 A low-level classification

It is striking how researchers working in the field of semantic relations tend to avoid reusing existing schemes. I am as guilty of this as anyone (Pepper 2010b). This urge to reinvent the wheel could be due to the not-invented-here syndrome, but it could also relate to the slippery nature of semantic relations: they are notoriously hard to pin down and sometimes quite subjective. As a result it can be difficult to adopt a system developed by someone else. Sometimes it is just easier to start from scratch than struggle to put oneself into the mind of the other person. Be that as it may, for this study I decided that I would not reinvent the wheel. I would select a pre-existing classification, apply it to my own data and then make up my mind about its suitability and need for extension or revision. In the event I ended up selecting two pre-existing classifications, those of Bourque (2014) and Hatcher (1960), which contrast nicely with one another, Bourque’s being rather granular and Hatcher’s extremely schematic. In this section I describe and justify the selection of Bourque’s system, and in the next I do the same for Hatcher.

Bourque’s dissertation is entitled Toward a typology of semantic transparency: The case of French compounds. The core of it is an examination of the semantic relations found in 1,048 compounds of the type NN and N à N, based on a classification developed by Bourque himself. There were a number of reasons for selecting this classification for the present study:

1. The work is based on a detailed and comprehensive survey covering some 20 earlier schemes from which he rather explicitly synthesizes his own, referring to them as “retained relations”. Bourque thus stands on the shoulders of giants, instead of simply starting from scratch, as I did in Pepper (2010b).

2. Unlike some other researchers, Bourque recognizes that some relations work in two directions, e.g. part-whole (wheelchair) and whole-part (table leg); these are termed reversible.

3. The scheme consists of a moderate number of classes (15 + 10 inverse = 25), a size which was deemed appropriate for the present study. It is more granular than the 12-14 relations of Levi, Warren and Downing (not to mention Hatcher’s four); this is important, because it is much more challenging to classify the data correctly and consistently using a very abstract system. On the other hand, it is not as fine-grained as Ryder’s system of 50 relations; applying the latter to over 3,000 binominals would have been prohibitive.
4. Each relation is accompanied by a full **description**, the presentation extending over 40 pages. This makes it easier for another researcher to understand just what is intended by each relation.

5. The metalanguage is English while the object language is French. As a result, the **terminology** Bourque employs is less Anglocentric than that encountered in studies based only on English. Levi’s use of English verbs and prepositions (e.g. HAVE, BE and FROM) does not work well cross-linguistically because of the way prepositions fail to match up across languages, and the fact that there is no HAVE-verb in many languages. In addition, Bourque uses NON-HEAD (or modifier) and HEAD instead of N₁ and N₂ (cf. Levi, Jackendoff and others), or A and B (cf. Hatcher), which imply head-final order.

6. English and French **examples** are provided for each relation. This makes the dissertation accessible to researchers who are not fluent in French.

7. Each relation is accompanied by a **test frame** or “template” consisting of both an English and a French paraphrase. The ensures much greater accuracy in the assignment of relation types to individual binominals (cf. §6.3.1).

8. Bourque includes discussion of **overlaps** between different relations. This is important because semantic relations, at least at Bourque’s level of granularity, are prototypical rather than Aristotelian categories. Because of the gradience they exhibit, some binominals could be seen as instances of multiple relations. While this is in the nature of the phenomenon, it is helpful to receive a third party’s confirmation that certain cases will be slightly ambiguous.

9. Finally, Bourque has made his data available on his website (Bourque 2016), thus facilitating further exploration of his classification.

In sum, these nine aspects of Bourque’s work provide a compelling argument for adopting his classification scheme.

Table 28 lists Bourque’s 15 semantic relations, 10 of which are reversible, for a total of 25 types. Each type is accompanied by its paraphrase (template) and the English example given by Bourque (with the minor corrections noted in footnotes), and each relation is supplied with linking material, some of which is used in the template and some of which is offered as a supplement. The scheme is reproduced exactly as given in Bourque’s dissertation and includes certain infelicities (and at least one error: sunburn as Basic CAUSE, i.e. “a burn that causes sun”). In §6.3.2 I revise and extend the scheme, correcting errors in the process.
<table>
<thead>
<tr>
<th>Label</th>
<th>Type</th>
<th>Template</th>
<th>Linking material</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernymy</td>
<td>Basic</td>
<td>an H of kind M</td>
<td>kind of, type of</td>
<td>oak tree</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M is a kind of</td>
<td></td>
<td>bear cub</td>
</tr>
<tr>
<td>Coordination</td>
<td></td>
<td>a C is an H and an M</td>
<td>is also, is both / and</td>
<td>boy king</td>
</tr>
<tr>
<td>Similarity</td>
<td></td>
<td>an H that is similar to M</td>
<td>similar to, like</td>
<td>ant lion</td>
</tr>
<tr>
<td>Function</td>
<td></td>
<td>an H that serves as M</td>
<td>functions, serves as</td>
<td>buffer state</td>
</tr>
<tr>
<td>Possession</td>
<td>Basic</td>
<td>an H that possesses M</td>
<td>possess (have / of)</td>
<td>career girl</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M possesses</td>
<td></td>
<td>family estate</td>
</tr>
<tr>
<td>Part</td>
<td>Basic</td>
<td>an H that is part of M</td>
<td>part of (have / of)</td>
<td>table leg</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M is part of</td>
<td></td>
<td>wheelchair</td>
</tr>
<tr>
<td>Location</td>
<td>Basic</td>
<td>an H located at/near/in M</td>
<td>at, near, in, etc.</td>
<td>window seat</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M is located at/near/in</td>
<td></td>
<td>bedroom</td>
</tr>
<tr>
<td>Composition</td>
<td>Basic</td>
<td>an H made of M</td>
<td>composed/made of</td>
<td>sugar cube</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M is made of</td>
<td></td>
<td>sheet metal</td>
</tr>
<tr>
<td>Source</td>
<td>Basic</td>
<td>an H (made) from M</td>
<td>(made) from</td>
<td>cane sugar</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M is (made) from</td>
<td></td>
<td>sugar cane</td>
</tr>
<tr>
<td>Cause</td>
<td>Basic</td>
<td>an H that causes M</td>
<td>causes</td>
<td>sunburn</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M causes</td>
<td></td>
<td>motion sickness</td>
</tr>
<tr>
<td>Production</td>
<td>Basic</td>
<td>an H that makes M</td>
<td>makes, produces</td>
<td>honey bee</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M makes</td>
<td></td>
<td>beeswax</td>
</tr>
<tr>
<td>Topic</td>
<td></td>
<td>an H about M</td>
<td>about</td>
<td>history conference</td>
</tr>
<tr>
<td>Time</td>
<td>Basic</td>
<td>an H that occurs at/during M</td>
<td>during, at, in, before, etc.</td>
<td>summer job</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H at/during which M occurs</td>
<td></td>
<td>golf season</td>
</tr>
<tr>
<td>Use</td>
<td>Basic</td>
<td>an H that uses M</td>
<td>use / with, by</td>
<td>steamboat</td>
</tr>
<tr>
<td></td>
<td>Rev.</td>
<td>an H that M uses</td>
<td></td>
<td>hand brake</td>
</tr>
<tr>
<td>Purpose and Proper Function</td>
<td></td>
<td>an H intended for M</td>
<td>for</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

*Table 28: Bourque's original system of semantic relations*  
(cf. revised Hatcher-Bourque classification, page 238)
We observe that Bourque’s classification is not fundamentally different from those sketched briefly on pages 186ff and indeed, the same kinds of relations recur again and again in the literature. In his Appendix B, Bourque provides a very helpful comparison table showing how 29 of these relations are named and used across 13 different schemes (in addition to his own). From this we can read that Levi’s FOR is termed ‘Purpose’ by most researchers (but ‘What for?’ by Vanderwende); that Jackendoff is alone with his PROTECT (FROM) function, apart from Adams, who has a subtype of Instrumental (B1, B2), and Arnaud, whose cryptic types AJ and AK cover exactly the same two relations;¹ and that Levi’s HAVE covers both part-whole and possession. Taken as a whole, Bourque’s work is the most comprehensive, systematic and useful study of semantic relations that I have encountered, and it is a pity that it is not more widely known. Much more could be said about it, but the foregoing will suffice for present purposes. We will return to it later.

I turn now to another impressive and undervalued piece of research, one which occupies the other end of the granularity scale from Bourque, namely Hatcher’s (1960) development of a four-way classification of determinative, non-appositional noun-noun compounds in English.

6.2.2 A high-level classification

The starting point for Hatcher’s paper is a trenchant critique of Jespersen’s (1942) attempt to classify semantic relations, mentioned above on page 186. As noted earlier, Jespersen admitted that his analysis was incomplete and that there were many compounds which “do not fit in anywhere”. He claims that his failure is due simply to the inherent unclassifiability of his material: “the number of possible logical relations between the two elements is endless”; “the analysis of the possible sense-relations can never be exhaustive”. But, says Hatcher:

it all too often happens that scholars in linguistics proclaim a given problem to be insoluble, when they themselves have not worked out the categories necessary for its solution; we should, then, examine the outline offered by Jespersen to see if some of the difficulty he encountered may not be explained by his method of classification. For example, was his set of categories constructed with logical rigor: and, before surrendering to the “difficult” types that he mentions, had he been able, at least, to account for all the “easy” compounds, subdividing these as carefully as his patience and his talent permitted? The subdivision of the obvious may lead to greater understanding of the less obvious, if one is guided by logically consistent criteria (p. 356).

¹ aj: “N2 est ce contre quoi N1 est fait, conçu, mis sur pied (alarme incendie)”; ak: “N2 est ce contre l’effet de quoi N1 est fait, conçu, mis sur pied (minimum vieillesse)” (Arnaud 2003: 74).
Thereupon, Hatcher sets about tearing Jespersen’s system to pieces (her words, see p. 365). She starts by listing seven major divisions (93), omitting one of Jespersen’s original eight (Similarity *needle-fish*) on the grounds that it more properly belongs to “apposition”, which she wants to keep separate.\(^1\)

(93)  
1. Action or agent noun as Subject (*sunrise*) or Object (*sun-worship*)  
2. Location in/from/to (*headache, land-breeze, side-glance*)  
3. Time at/duration (*nightmare, day-fly*)  
4. Purpose (*beehive*)  
5. Means (*handwriting*)  
6. Characterizing Feature (*sandpaper*)  
7. Material (*gold ring*)

Examining each of these in turn, Hatcher notes a lack of careful subdivisions, an absence of any principle of symmetry, and mixing of two basic criteria, Reference (e.g. action) and Relation (e.g. subject), exclaiming:

Little wonder that to Jespersen the difficulty of classifying our compounds was insuperable. But how could he fail to see the inadequacy of his categories? How could any linguist today construct a system of classification based on two (or three) kinds of main criteria? (p. 361).

Jespersen’s blindness to the flaws of his outline is put down to his “consummate lack of interest in the problem”; as to the second question, the answer is that what Jespersen actually offers is no more than a simplification of the classification suggested by Mätzner in 1860. “It is sad,” writes Hatcher, “that a linguist of today cannot go beyond a linguist of the nineteenth century; it is sadder still if his work is inferior to his predecessor’s. And this is the case, here”.

Hatcher now sets about rearranging Jespersen’s scheme as depicted in Figure 46. For consistency she chooses to base her new scheme on Relation only and to avoid Reference, so she starts by separating the first three of Jespersen’s types (1)-(3),\(^2\) all of which are either based on reference or mixed, from the rest (4)-(7), which are all relational. The former are set to one side, and to the latter she adds the two relational types that Jespersen for some reason failed to adopt from Mätzner (*α* *broomstick* and *β* *castor oil*). She then proceeds to reorganize these six relational types as follows:

---

\(^1\) See the discussion of appositional compounds on pages 206ff. I will reinstate Similarity later since, unlike Hatcher in her 1960 paper, I am also interested in determinative appositional binominals.  
\(^2\) Numbers (and Greek letters) in bold refer to the nine divisions in Figure 46.
Figure 46: Hatcher’s reworking of Jespersen’s classification
The first step is to combine (6) Characterizing Feature sandpaper and (7) Material gold ring into a more general category “A is somehow, to some extent, contained, comprehended in B”. In the second step, (5) Means handwriting and (β) B-obtained-from-A castor oil are combined into the category “A is somehow the source of B”, the concept of source being “expanded, spiritualized”, such as to “transcend concrete limitations”. In the third step, (α) B-is-a-part-of-A broomstick is used to establish a category “B is contained in A”, which is the inverse of the first new category (“A is contained in B”). Hatcher symbolizes these two “mutually exclusive” categories by ⬅️ (A is contained in B) and ➡️ (B is contained in A); I will, however, henceforth symbolize them as $A \subset B$ and $A \supset B$, respectively.\(^1\) Finally, in the fourth step, Hatcher establishes that “all the various compounds” of (4) Purpose beehive “present A as the destination, end of B”, which is the opposite of the second type, Source. In contrast to the “static contrast” of $A \subset B$ and $A \supset B$, this pair exhibit a “dynamic contrast”, which she indicates using the symbols $A \rightarrow B$ (“A is the source”) and $A \leftarrow B$ (“A is the destination”).

\[
\begin{align*}
(94) & \quad (a) \ A \subset B \quad \text{“A is contained in B”} \\
& \quad (b) \ A \supset B \quad \text{“B is contained in A”} \\
& \quad (c) \ A \rightarrow B \quad \text{“A is the source of B”} \\
& \quad (d) \ A \leftarrow B \quad \text{“A is the destination of B”}
\end{align*}
\]

Having reduced the six relational categories of Jespersen/Mätzner to two pairs of mutually exclusive concepts (94), Hatcher now turns her attention to the referential types in order to see how they might be accommodated in her new scheme. She starts with Place (2), Time (3) and their subdivisions. The subtype “A=Place in which B is found” (seagull), labelled in/at in Figure 46, fits clearly into $A \supset B$, i.e. “A contains B, as a place ‘contains’”. Similarly, in “A=Time in which B is found” (spring frost), also labelled in/at, A contains B “in the mysterious way in which Time contains”. Continuing:

As for “A= Place from which .. . “ (land breeze; cf. also spring water; kitchen fats; store clothes), we have to do with $A \rightarrow B$; here also belongs the less frequent type I have added: “A=Time from which...”: Ming bronze, 18th-century snuffbox. The third division of Place (“A= Place to which...”) belongs obviously to $A \leftarrow B$: not only side glance but New York express, Tokyo flight (as well as the type knee-pants discussed earlier) (p. 365).

\(^1\) I do so based on three considerations: precision ($\subset$ means subset of, a form of containment); æsthetics ($A \subset B$ nicely balances $A \rightarrow B$); and convenience ($\subset$ is available in more fonts).
This caters for both of the to and from subtypes of Place and Time, except for the fact that neither Jespersen nor Hatcher had any examples of Time-to. Regarding the latter, Hatcher notes: “If Jespersen had looked for a Time-to-which as a parallel to Place-to-which, he probably would not have found it. This may be partly explained by the fact that Place-to-which itself is not very frequent (and Time has the habit of imitating Place)” (p. 358). All that remains of (2) Place and (3) Time is extent, which falls neatly into place into the fourth type not yet tapped: A ⊂ B, which subsumes Jespersen’s (6) Characterizing Feature. Hatcher advises us to forget about Jespersen’s example life annuity, which belongs rather to Purpose, and to consider 2-hour discussion and 3-foot pole, entities characterized by their extent, which they “contain”.

Finally, the two verbal types (1) Subject and (2) Object are “easy”:

Sunshine and sun worship, these perfect opposites, fall under A→B and A←B, respectively. Surely the subject is the “source” of its own activity (in putting sunshine under A→B, we are merely adding Agent to Agency); and in sun-worship (A←B), the sun is that toward which the worship is directed.

Thus we see that both the referential and the relational types of Matzner-Jespersen can be included in our two pairs of relational criteria: the static Ⓚ and Ⓛ, and the dynamic A→B and A←B (p. 365).

Hatcher concludes this part of her analysis by pointing out that the scheme she has developed has two advantages over the one she has just “torn to pieces”. Firstly, it is logically conceived, and therefore neater and more pleasing aesthetically; and secondly, it is far more comprehensive, and thus may “be able to account for all possibilities of determinative, non-appositional compounding in the English language,” which she suggests are surely not “endless”. At the same time she expresses the hope that her work represents not a “result”, but rather a beginning, and that it will offer “a more spacious framework” within which research dedicated to the proposition that “all compounds are endowed by their creators with the right to belong somewhere” may proceed more profitably and hopefully than before.

But Hatcher isn’t done yet. She recognizes that her framework leaves much to be desired: the four main categories are “comfortably vague and elastic”, but there is also a need for a more fine-grained set of sub-categories. Hatcher herself does not develop these, but she does suggest a way forward, this time based on Reference rather than Relation. To that end, she suggests a scheme based upon conventional criteria, utilizing a sevenfold classification of the noun (95). I will not pursue her idea here, fascinating though it is, because it is beyond the scope of my study.
(95) 1 Person  
2 Animal  
3 Concrete Object, Substance, Condition  
4 Place  
5 Time  
6 Activity  
7 Miscellaneous Abstract Entities

Hatcher’s work is often cited (Citeseer, accessed 2018-05-30, counts 328 citations), but usually dismissed, often on less than scientific grounds. For example, Søgaard writes:

such an account is by definition both arbitrary (Bauer 1978; van Santen 1979) and incomplete because of the infinite set of compounding relationships. For illustration, try to place a compound such as car thief in [Hatcher’s] four-way typology. Is a car thief a ‘car in a thief’, a ‘thief in a car’, a ‘thief as the goal of a car’ or a ‘thief as the source of a car’? (Søgaard 2005: 320).

Unfortunately for Søgaard the last two paraphrases are incorrect: He has muddled up the order of A and B. The head of the construction (B) is thief, not car, so these two paraphrases should read: a ‘car as the goal of a thief’ and a ‘car as the source of the thief’. With the correct paraphrase, it is obvious that the car is indeed the goal of the thief (i.e. A ← B). Søgaard’s objection must therefore be rejected.

Two authors that have taken Hatcher seriously are Noailly (1990) and Arnaud (2016). Noailly is interested in French “substantifs épithètes” (attributive nouns), i.e. nouns which occupy the position of N₂ in a noun phrase of the type Art N₁ N₂, where the two nouns N₁ and N₂ follow one another directly without an intervening preposition or pause (although they may be hyphenated). Since French binominals of this type (cmp) are left-headed, N₂ refers to the modifier, not the head: a oiseau-mouche [bird-fly] ‘hummingbird’ is a kind of bird, not a kind of fly. Noailly groups such nouns into four types based on the function they perform, viz. Qualification, Coordination, Complementation and Identification.

The type Qualification (e.g. homme-grenouille [man-frog] ‘frogman’) is defined as an N₁ “qui est un N₂”, and is thus an appositional compound in Hatcher’s terms. In Coordination (e.g. moules-frites [mussel-fries] “a popular main dish of mussels and fries originating in Belgium”) the two nouns are “ordonnés au même rang” and thus constitute a coordinate compound. Regarding Complementation, Noailly suggests that it covers most determinative compounds (e.g. oiseau-mouche), except those that are grouped separately under Identification. This is the case when
the modifier is a proper noun (e.g. *le président Pompidou*), and when the modifier (e.g. *chien*) denotes a subclass of the head (e.g. *espèce*), as in *l’espèce chien* [DET species dog] ‘*Canis familiaris*’. I won’t pursue this point here since our interest is in the category of Complementation, which Noailly subdivides by the four types of relation elaborated by Hatcher. No new insights emerge, but the very fact that Noailly, *contra* Søgaard, does not appear to encounter any problems in applying Hatcher’s scheme to her data is very encouraging for the present work. I am indebted to Arnaud (2016) for bringing Noailly’s work to my attention. But my debt to him goes far beyond that, to the extent that he deserves a section of his own.

### 6.2.3 Extending Hatcher’s classification

Arnaud’s paper on categorizing the modification relations in French relational subordinative **NN** compounds is full of interesting observations, examples and discussion, and I cannot do it justice here. The interested reader is encouraged to consult it directly. In the present context it is mainly noteworthy for the fact that Arnaud first develops his own highly granular classification, and then maps it onto Hatcher’s scheme, which is what I will do with Bourque’s classification (§6.3.3).

Arnaud’s classification is based on a database of 949 French binominals of type **cmp**, which he dubs “*les composés timbre-poste*” (‘postage stamp compounds’, Arnaud 2003). As none of the then-existing taxonomies of semantic relations seemed satisfactory, he decided to start from the data up, applying the principles of cognitive linguistics, “in particular the idea that relations are emergent phenomena which gain psychological existence” (p. 71). After the first stage of work based on 809 compounds, he ended up with 54 categories, each with a definition of the type shown in (96).

\[(96)\]

a. **NON-HEAD** (concrete, discrete) is one of the parts of **HEAD**
   
   *tailleur pantalon* [suit trouser] ‘trouser suit’

b. **NON-HEAD** is the container of **HEAD**
   
   *bière bouteille* [beer bottle] ‘bottled beer’

c. **NON-HEAD** is the origin (general) of **HEAD**
   
   *arrêt maladie* [stop sickness] ‘sick leave’

d. **NON-HEAD** is the goal/the purpose/the object (general) of **HEAD**
   
   *carte réponse* [card answer] ‘answer (post)card’

In the second stage, in which he analysed a further 140 compounds, Arnaud was obliged to adapt some of his categories slightly and to add four new ones, which
confirmed for him “the frequently expressed opinion that a categorization of compounds cannot be exhaustive” (p. 82). However, he also notes that we are “clearly in a situation of diminishing returns, since most of the units in the second dataset were accounted for by already identified relations.”

Next Arnaud applies his taxonomy to a random sample of 200 compounds drawn from a dataset of 3000 lexicalized English NN compounds. Once more he was obliged to modify a few categories and add some new ones, but no more than six. This is remarkably few considering that compounds are considerably less numerous in French than in English, and that French has an entrenched competing N PREP N construction that is preferred for some types of relation (for example, habitat and part-whole), as Arnaud himself points out (p. 89). Again, this suggests that the number of relations required at any given level of granularity will tend to flatten out and not increase indefinitely.

Of most interest to us is that Arnaud now proceeds to map his set of 58 empirically derived (low-level) relations to Hatcher’s set of four logically derived (high-level) relations. For the most part, this is plain sailing, as seen in (97).

(97) a. NON-HEAD (concrete, discrete) is one of the parts of HEAD
tailleur pantalon [suit trouser] ‘trouser suit’
   \[ \Rightarrow A \subset B \] (trouser \( \subset \) suit)
b. NON-HEAD is the container of HEAD
bière bouteille [beer bottle] ‘bottled beer’
   \[ \Rightarrow A \supset B \] (bottle \( \supset \) beer)
c. NON-HEAD is the origin (general) of HEAD
arrêt maladie [stop sickness] ‘sick leave’
   \[ \Rightarrow A \rightarrow B \] (sickness \( \rightarrow \) leave)
d. NON-HEAD is the goal/the purpose/the object (general) of HEAD
carte réponse [card answer] ‘answer (post)card’
   \[ \Rightarrow A \leftarrow B \] (answer \( \leftarrow \) card)

1. Two responses come to mind: first, a situation of diminishing returns might suggest something like a cumulative Pareto curve (left), which flattens out at some point; second, as noted above on page 186, the number of relations one posits will always be a function of the degree of granularity one aspires to and can range anywhere from one (IS RELATED TO) to however many one has the patience to enumerate.

2 He also examines how Jackendoff’s model (see page 185) can be applied to French compounds.
In each of the four examples the first line gives Arnaud’s label for his low-level relation, the second provides an illustration, and the third shows the mapping to Hatcher’s high-level relation. At this point the reader is advised to recalibrate her brain and think of A and B as denoting, not the first and second constituent, but the Attribute (modifier) and the Base (head). This is because Arnaud’s (French) examples are all head-initial, whereas Hatcher worked with head-final (English) data. While A and B work well for modifier and head in English, the order can be confusing in French. Alternatively, the reader can continue to think in Hatcher’s terms and focus on the English translation rather than the French word or gloss.

We see that Arnaud’s bottom-up deduction melds neatly with Hatcher’s top-down induction. Or at least, it almost does. Arnaud did have problems with some units that he feels do not correspond to one of the four abstract categories. One example is régime jockey [diet jockey] ‘jockey diet’, which denotes a diet that is typical of jockeys. To cater for these cases, Arnaud feels obliged to create a fifth category, ANALOG (not to be confused, he insists, with the attributive relation). In all, four “supplementary abstract categories” were necessary (98) (further examples of each are given in the sections to follow).

(98) a. ANALOG (e.g. régime jockey)
   b. BE (e.g. opéra rock)
   c. NON-HEAD SYMBOLIZES HEAD (e.g. franc or)
   d. NON-HEAD SYMBOLIZED BY HEAD (e.g. pierre papier)

Arnaud himself notes that all four of these categories are “marginal” compared with Hatcher’s initial four, but all the same, they show, in his opinion, that Noailly and Hatcher “erred on the side of abstraction”. But what does it mean to err on the side of abstraction? If anything, surely Hatcher’s categories were not abstract enough to accommodate his difficiliora. But is it true that they cannot do so? In order to find out, it will be very much worth our while to examine in turn each of the 12 low-level relations that seemed to Arnaud to justify the creation of his four new high-level relations.

**ANALOG**

Arnaud’s first new high-level category, ANALOG (98a), is exemplified, as we have seen, by régime jockey (99a), a diet that is typical of jockeys. But when A typifies B, it characterizes B and, as we have seen, Hatcher subsumes Characterizing Feature under $A \subseteq B$, which is where régime jockey properly belongs; no new high-level category is required in order to accommodate it. In Arnaud’s low-level scheme it
is classified under “NON-HEAD has or includes HEAD (of a kind that is specific to it) / ‘NON-HEAD type HEAD’”, along with 36 other examples which he does not list here.\(^1\) It seems likely that most, if not all, examples of this particular subtype of ANALOG can be placed with régime jockey under A ⊂ B, such that this subtype, at least, does not demand the creation of the new high-level category. Three other low-level relations are used to justify the creation of ANALOG (99b-d). Let us examine each of them to see if they really do. (It will turn out that one of them does.)

(99) ANALOG (four subtypes, 62 instances)

a. NON-HEAD has or includes HEAD (of a kind that is specific to it) / ‘NON-HEAD type HEAD’
   régime jockey [diet jockey] ‘jockey’s diet’

b. NON-HEAD is an entity which analogically identifies HEAD (abstract, discrete)
   style nouille [style noodle] ‘noodle style’

c. NON-HEAD is that on which HEAD was identified
   facteur rhesus [factor rhesus] ‘rhesus factor’

d. NON-HEAD names analogically a perceptual characteristic of HEAD
   brasse papillon [breast_stroke butterfly] ‘butterfly stroke’

Turning to (99b), Arnaud has 11 instances of this low-level relation in his database. If they are all of the same nature as style nouille (the example he gives), they will all fit nicely under A ⊂ B (together with régime jockey), for what are noodles in this context but an (A)ttribute that characterizes a style, the (B)ase?

The low-level relation in (99c) has but one exemplar in the database. I consider it a little overzealous to define a whole low-level category for the “X identified on Y” relation, but it is not forbidden: as noted above, the researcher gets to decide the level of granularity she requires for the purpose of her investigation. But how this category is mapped to Hatcher’s scheme is not subjective; it has to be rigorous. Now, it is clearly a Characterizing Feature of the rhesus factor that it was identified on rhesus monkeys. In that case, A ⊂ B is once again the appropriate home.

The fourth low-level relation used to justify ANALOG (99d) presents a different and more interesting case. Here there can be no doubt that some kind of analogy is at work. But is brasse papillon really a non-appositional compound? If, instead, it is appositional, then it falls outside Hatcher’s scope (recall that her 1960 paper looks

---

\(^1\) Most of them can be found in his earlier work (Arnaud 2003). Due to time constraints I have not been able to examine all of them.
only at non-appositional compounds). This, of course, raises the question of what we mean by appositional, and who better to turn to than Hatcher herself and her 1952 paper on *Modern appositional compounds of inanimate reference*. Here, again, I must quote at length. According to Hatcher (p. 4):

With any appositional compound $AB$ we are offered two names for the same object, the first name, obviously, representing a modification, a further classification, of the second. There are, basically, only two types of modification possible: an object belonging to class $B$ (*stone*) may be assigned to a subdivision of that class (*pumice*), in which case we have to do with the relationship between the species and genus (*pumice stone*); or again, an object seen as belonging primarily to class $B$ (*oil*) may also be assigned to a different, but complementary, class $A$ (*fuel*): with *fuel oil* we might speak of ‘cross-classification’. In both cases alike, the creation of the compound $AB$ suggests the predication: ‘$B$ is (also) $A$’. With pumice stone, however, this is a necessary relationship, valid for all examples of $A$ and $B$: $A$ is, by definition, a [subdivision of the class] $B$. This is not true of *fuel oil*, which rests on no generic assumption, but suggests only the possibility that a certain type of fuel and a certain type of oil might coincide.\(^1\)

On the face of it, this definition would seem to exclude Arnaud’s *brasse papillon* (99d) from the category of appositional compounds, since the butterfly stroke is not a kind of butterfly; it is only a kind of (breast) stroke. Therefore it cannot, by this definition, be an instance of either the species-genus subtype or the cross-classification subtype of appositional compounds. However, Hatcher goes on to subdivide the latter subtype as follows:

By means of the device of cross-classification, it is possible to apply to a given entity a second name denoting: (1) the function thereof; (2) the status; (3) a characteristic descriptive feature (p. 4).

For (1) she gives the example *fighter plane*\(^2\) (which she incidentally makes a point of distinguishing from *fighter-bomber*, considering the latter to be a copulative compound, since the two classes to which it is assigned are mutually exclusive). Among Hatcher’s examples of type (2) are *mother country*, “in which we find a reference to status (‘$B$ is [has the rank, value of] $A$’)”, and *biography sensation*. Hatcher then observes that

\(^1\) The distinction between species-genus and cross-classification is related to Jackendoff’s functions, BE and KIND, about which he comments: “Note the distinction between [witch doctor, BE] and [puppy dog, KIND]. A witch doctor is both a witch and a doctor, and a puppy is both a puppy and a dog. But a puppy is, more specifically, a kind of dog, whereas a witch is not a kind of doctor” (p. 27-28).

\(^2\) A disconcerting number of Hatcher’s examples are military, reflecting the sad state of the post-war world at the time she was writing.
in all the examples of the type biography sensation, and in many single examples of the other types, we have had to do with hyperbole or metaphor. When we come to the third type of compound of cross-classification, in which an object is defined according to its most picturesque feature, a metaphor will always be involved” (p. 7, emphasis added).

Clearly, mother country is an example of exactly this, for it is a subclass of mother only in a metaphorical sense. When we turn to type (3), in which an entity is given a second name that denotes a characteristic descriptive feature, we find amongst the examples the word butterfly table, about which Hatcher has this to say:

The fact is that when the object of comparison is a material entity, the whole simile is apt to rest upon only one detail: in a butterfly table shape alone is involved and none of the iridescent, airy, fragile essence of the butterfly is suggested (p. 8).

The case of brasse papillon is exactly parallel. The definition originally offered for the cross-classifying type of appositional compound (“an object seen as belonging primarily to class B may also be assigned to a different, but complementary, class A”) is not as Aristotelian as one might have thought: it permits metaphorical extension of the prototype. The conclusion must therefore be that brasse papillon is to be regarded as an appositional compound, parallel to butterfly table, at least in Hatcher’s scheme, which of course is the one Arnaud is applying. Consequently it should come as no surprise to Arnaud that the four-way classification she devised for non-appositional compounds does not cover this (appositional) case.

But acquitting Hatcher for “erring on the side of abstraction” does not allow us to disregard the case of brasse papillon: we still need to account for it somehow, as Arnaud does by introducing the category ANALOG (which I have reduced from 62 to 13 exemplars, having reassigned three of his four subtypes (99a-c) elsewhere). We clearly need a new high-level category for appositional compounds like brasse papillon; the only question is whether ‘analog’ is the best label for it. Compared to Hatcher’s logically defined pair of reversible relations, both of them based on Contiguity, ‘analog’ seems less than satisfactory. It is time to return to Aristotle.

In my discussion of Janda’s work on “metonymy” in word-formation (§6.1.3) I made the point that the concept underlying both metonymy and the phenomenon discussed by Janda in derivation is Contiguity. I noted (fn. 1 on page 193) that Contiguity is one of three associative relations identified by Aristotle, the others being Similarity and Contrast. Here I suggest that the associative relation underlying both types of appositional compound (species-genus and cross-classification), as well as Arnaud’s brasse papillon (and incidentally also coordinative compounds),
is Similarity. Similarity is, of course, gradient: one thing can be more or less similar to another, the limiting case being that of Identity, when two things are so similar that we deem them to be identical.\(^1\) When we perceive things as being similar, we tend to group them into categories. Conversely, when we perceive them as being slightly dissimilar, we form subcategories – of the species-genus kind: a beech tree and an oak tree are sufficiently similar that we group them both under the more general concept of tree, but sufficiently dissimilar that biologists assign them to their own genera (\textit{Fagus} and \textit{Quercus}, respectively) under the family Fagaceae within the Plant kingdom.\(^2\)

The notion of Similarity thus accounts for both species-genus compounds of the type \textit{fuel oil} and cross-classification compounds of the type \textit{pumice stone}. Moreover, Similarity is what underlies analogy: we make analogies between things when we perceive them as being in some sense similar. In other words, just as Contiguity accounts for Hatcher’s four relations and for metonymy (as well as much else, as I shall claim in Chapter 8), Similarity accounts for both analogy and metaphor. It is important to note that the term Similarity is used here in a rather abstract sense, one that includes both identity and the taxonomic (species-genus) relation as well as similarity proper. Clearly it can be subdivided (I have just done so), but so can Hatcher’s high-level relations: Arnaud’s almost successful attempt to reduce his 58 relations to Hatcher’s four is simply the inverse of subdividing Hatcher’s four into his 58, or Bourque’s 25. Thus it seems that Similarity as a high-level category is at about the right level of generality or abstraction as those of Hatcher.

I conclude, therefore, that Hatcher’s scheme was not incomplete, since it was never intended to cover appositional compounds. However, it can easily be extended to do so by adding a third basic relation, Similarity, to supplement Hatcher’s two pairs, Containment (\(A \subset B\), \(A \supset B\)) and Source/goal (\(A \rightarrow B\), \(A \leftarrow B\)). The new relation is symbolized by the mathematical operator for “almost equal or equal to”, \(A \equiv B\). For the limiting case of Identity, we may use \(A \equiv B\); for the subtype Similarity (proper), \(A \approx B\) (“almost equal to”); and for the species-genus subtype involving

\(^1\) An example of (near-)identity would be Mandarin 毛发 \textit{máo,fà} [hair,hair] ‘hair on human body’ (Ceccagno & Scalise 2006).

\(^2\) Compounds of the species-genus kind, sometimes called pleonastic, epexegetic or subsumptive compounds, are somewhat peculiar in that the head is essentially redundant. Is \textit{tree} in \textit{oak tree} a kind of classifier and, if so, what purpose does it serve? Vennemann (1996: 118; 2003: 318) offers the following explanation: “The addition of a clarifying generic head-noun to an old name usually shows that the old name standing by itself was becoming obfuscate to the language users.”
Taxonomy I suggest the “Up Tack” $A \perp B$.\(^1\) This relation is reversible (cf. Bourque’s hypernymy relation in Table 28, with the basic example *oak tree* and the reversed example *bear cub*, and also the discussion of the latter in §6.3.2); we will therefore also need a genus-species subtype: $ATB$.

**BE**

Having now replaced Arnaud’s ANALOG with Similarity ($A \equiv B$), let us briefly apply ourselves to the other *difficilia* that led him to define new categories, viz. BE; NON-HEAD SYMB HEAD; and HEAD SYMB NON-HEAD (100)–(102).

(100) **BE** (six subtypes, 23 instances)

a. NON-HEAD is the degree of quality of HEAD
   *kirsch fantaisie* [cherry(-brandy) fantasy] ‘fancy kirschwasser’
b. NON-HEAD is the style of HEAD
   *opéra rock* [opera rock] ‘rock opera’
c. NON-HEAD is the physical state of HEAD
   *morphine base* [morphine base] ‘morphine base’
d. NON-HEAD is the duration of the action/preparation of HEAD
   *pulsar millisecond* [pulsar millisecond] ‘millisecond pulsar’
e. NON-HEAD is the status (time) of HEAD
   *match retard* [match delay] ‘late match’
f. NON-HEAD is the status (space) of HEAD
   *match retour* [match return] ‘return match’

Arnaud groups six of his low-level relations into the new high-level category BE (100a-f). Like *brasse papillon, kirsch fantasie* (100a) is an instance of Hatcher’s third (characteristic descriptive) subtype of (appositional) cross-classification: that in which an object is defined according to its most picturesque feature, and therefore $A \equiv B$. In *opéra rock* (100b), rock denotes, as Arnaud says, the style of opera and is thus a Characterizing Feature, which, as we recall, Hatcher has under $A \subset B$. The example in (100c), *morphine base*, is slightly more tricky and therefore worth examining in detail. Base refers to the form of the morphine, so *morphine base* is, in a sense, both a type of base and a type of morphine: a fairly clear example of cross-classification, it might seem, perhaps of type 2, in which the modifier denotes the status of the head. But Hatcher (1952) has no parallel examples; instead, she has

\[^1\] This symbol, also known as the falsum, is sometimes used to denote the “bottom type” in type theory, i.e. the type that has no values. While this has nothing to do with species-genus, the symbol is suggestive of *oak* and *beech* pointing up towards their co-hypernym, *tree* (right).
an extended discussion in her 1960 paper of the “great” type exemplified by mountain range. This is worth quoting in full, because it touches on the difficulty (“in a few cases”) of distinguishing between the four main relational categories:

Can one always recognize Ⓐ, Ⓑ, A→B and A←B? I should say that if we understand
the meaning of our compounds and if we have a decent corpus of examples, we will,
mainly, not have to hesitate. And in the few cases where choice appears difficult, we
may find that the difficult types come in great blocks. Consider mountain-range, and
the many other compounds in which B refers to shape, mass, extension: land mass,
sand dune; sandpile (ragpile etc.), haystack; shell mound, dung hill; melon balls, snow-
ball, snowflake (soap-flakes), snowdrop, tear drop, chocolate drop, butter pat, sugar-
loaf and many, many others (a type never before isolated, to my knowledge). Shall we
assign these to Ⓐ, as representing subtle variants of Material? Compare smoke ring
and gold ring; snowball and rubber ball; sand bar and iron bars. Or perhaps they
belong under Ⓑ; perhaps in snowball the real entity is snow, and ball refers only to an
aspect of this entity, to its shape; must not an entity contain its own shape and not the
reverse? And if sugarloaf, for example, is placed under Ⓑ, it will offer a perfect con-
trast to loaf sugar (print butter, etc.), which can only be interpreted as Ⓐ.

But is it really true that in snowball the second element represents only an aspect of the
first? Is it not more natural to think of ‘a “ball” containing snow’ than of the reverse?
Perhaps, after all, snowball and sugarloaf belong under Ⓑ. This would mean, of course,
that Ⓑ must then accommodate both sugarloaf and loaf sugar – an apparent absurdity.
How, how, shall we classify sugarloaf and snowball, these reminders of the philosophi-
ical problem of Entity and Aspect? Whatever the answer, I know at least that we can
find hundreds of examples of the easily recognizable, snowball type: it is a “great” type.
Thus we should not worry overmuch; the great types abide our classification; our first
task is to isolate them (footnote 22, p. 372).

It is not my intention to adjudicate on this matter here. The point is that, whatever
the difficulties involved in analysing mountain range, snowball – and morphine
base, which I claim is of the same type – these all clearly belong to one of Hatcher’s
four categories (either A ⊂ B or A ⊃ B), and therefore do not justify the creation
of Arnaud’s new category, ANALOG.

As for the other subtypes of BE, in (100d) millisecond is the period of the pulsar,
an extent (in time), and thus belongs in A ⊂ B (cf. Figure 46 on page 200); in (100e)
the delay characterizes the match, hence also A ⊂ B, and the same applies to return
in (100f). Consequently I contend that all six subtypes of BE belong elsewhere and
that BE can be dispensed with altogether.
I turn now to Arnaud’s last two innovations, NON-HEAD SYMB HEAD (102) and HEAD SYMB NON-HEAD (101), each of which has just one subtype, with five and 24 exemplars, respectively. I will deal with the second of these first.

(101)  **HEAD SYMB NON-HEAD** (one subtype, 5 instances)

  NON-HEAD is represented/symbolized by HEAD

  *franc or* [franc gold] ‘gold franc’

According to Wikipedia (accessed 2018-05-24) *franc or* received its original name because it contained a certain amount of gold (290.034 mg, to be precise), and not because it symbolized gold in any way. It is therefore exactly parallel to *gold ring*, the paradigm example of Jespersen’s Material type, which Hatcher merged with Characterizing Feature to constitute her $A \subset B$ (cf. Figure 46). That the *franc or* later came to symbolize gold is hardly relevant to the relation between its constituents *at the time it was coined* (excuse the pun). Since this is the only subtype of HEAD SYMB NON-HEAD, and since it fits neatly into a pre-existing Hatcher category, there is no need for a new one.

(102)  **NON-HEAD SYMB HEAD** (one subtype, 24 instances)

  NON-HEAD represents/symbolizes HEAD

  *pierre papier* [stone paper] ‘shares in real estate’

The exemplar given for the last new category, *pierre papier* (102), is something of an oddity. We have first to ask, what is the head? Is a share in real estate a kind of stone or a kind of paper? The latter seems far more plausible to me, given that paper is often used to “symbolize” an abstract value metaphorically. But in that case, *pierre papier*, unlike every other (French) compound that we have so far discussed, is right-headed, not left-headed. This is indeed strange! The fact that this new category is the inverse of the previous one also alerts us to the strangeness of this form, for it is easy to imagine how B (the head) might symbolize A, but how can A (the non-head) symbolize B? Why would B *need* to be symbolized if it is the head? I have come to the provisional conclusion that this must be a word-formational *rarum* and that the explanation must be sought in diachrony. According to the French Wikipedia (accessed 2018-05-24):

---

1 In fact, it started out in 1803 as the *Franc germinal* (named after the month *Germinal* in the French revolutionary calendar). I have been unable to ascertain when the name changed to *franc or* (or *franc-or*).

2 Recall the ISA condition (page 109).
It is difficult to say exactly when this neologism appeared; presumably, it had to be born in the 1970s before being used more regularly in the 80s and 90s. Its author or authors were inspired by the concept of gold-paper \([d’or-papier]\) used to designate all the forms of investments in gold other than the purchase of physical gold (gold mines, certificates, trackers, etc.) [my translation].

The word \(d’or\) papier corresponds very closely to the English \textit{gold certificate}, a well-behaved Germanic right-headed compound. The French word, on the other hand, is decidedly unruly. A well-behaved French prepositional compound is an instantiation of the construction Head PREP Mod, like our old friend, \textit{chemin de fer}. Here, however, we seem to have PREP Mod Head, indicating that papier is the head. (The alternative, PREP Head Mod, is simply too weird to countenance.) This would appear to indicate that \(d’or\) papier was calqued from \textit{gold certificate}, copying the very un-French head-final word order, but without abandoning the traditional French prepositional word-formation strategy altogether. Later on, perhaps in the 1970s, an age of greatly increased English influence, \textit{pierre papier} was coined on analogy with \(d’or\) papier (*stone paper does not exist in English), but streamlined and anglicized by the removal of the preposition. Much of this is speculation, but it is the most plausible explanation that I have come up with. If it is correct, then papier ‘paper’ is the head and pierre ‘stone’ is yet again a feature which characterizes the head, used metonymically to denote land, or real estate. Consequently, the word belongs in \(A \subset B\). This becomes even more clear if we consider a further comment Hatcher makes regarding this category:

\(\sqsubset\) means “\(A\) is somehow comprehended in \(B\); \(B\) somehow comprehends \(A\), involves \(A\), \textit{concerns} \(A\), is ‘about’ \(A\)” (p. 366, emphasis added).

What does the paper in \textit{pierre papier} concern? What is it ‘about’? Answer: stone, i.e. real estate.\(^1\) Given that the only exemplar of the only subtype of the new high-level category NON-HEAD SYMB HEAD belongs elsewhere, this category, too, can be dispensed with.

**Summary**

I have now shown that all but one of the 12 low-level categories used to support the creation of Arnaud’s four new high-level categories can be accommodated elsewhere. This indicates that Hatcher’s system holds even better than Arnaud himself believes: in point of fact, it is 100% water-tight, provided appositional

---

\(^1\) Incidentally, Hatcher (1960:366, fn. 16) has an nice comment regarding the etymology of ‘about’: “Note that OE \textit{abutan} < \textit{on butan} meant ‘on the outside of’, i.e. ‘containing’.”
compounds are not brought into the equation. And when they are, a single new high-level category suffices to cater for them, that of Similarity, $A \approx B$.

To summarize: While there is much to admire in Arnaud’s paper, and while I am indebted to him for the idea of mapping from a low- to a high-level classification (an idea which I adopt myself in §6.3.3), I strongly dispute his contention that Hatcher’s four types do not cover all of his non-appositional compounds. I do recognize the need for a new high-level category to accommodate one of his low-level types, but I propose we name it Similarity rather than analog, in order not to deviate unnecessarily from Aristotle, and to employ the notation $A \approx B$. In taking this measure, Arnaud and I, between us, have upgraded Hatcher to version 2.0; we have catered for appositional compounds while staying loyal to the spirit of her enterprise. The new system is summarized in (103) and (104), in which I have taken the further step of replacing Hatcher’s $A$ and $B$ with $M$ (for modifier) and $H$ (for head), in order to make it more suitable for cross-linguistic comparison. Hatcher drew attention to the fact that her four relations comprised two pairs, which she characterized as “static” and “dynamic” (see page 201). I have recast them in terms of two superordinate relations, CONTAINMENT and CAUSATION (or source/goal). The four-letter codes in italics are the codes used in the database.

(103) **Contiguity-based**

CONTAINMENT

(a) $M \supset H$ \hspace{1em} $\text{HinM}$ “$H$ is contained in $M$” (*orange seed*)

(b) $M \subset H$ \hspace{1em} $\text{MinH}$ “$M$ is contained in $H$” (*seed orange*)

CAUSATION (source/goal)

(c) $M \leftarrow H$ \hspace{1em} $\text{HtoM}$ “$M$ is the destination of $H$” (*sugar cane*)

(d) $M \rightarrow H$ \hspace{1em} $\text{MtoH}$ “$M$ is the source of $H$” (*cane sugar*)

(104) **Similarity-based**

SIMILARITY

(e) $M \approx H$ \hspace{1em} $\text{MisH}$ “$H$ is similar or identical to $M$”

Similarity has three subtypes (105), one of which is reversible:

(105) **Subtypes of Similarity**

(i) $M \approx H$ “$H$ is similar to $M$” (*butterfly table*) cf. Bourque’s SIMILARITY

(ii) $M \& H$ “$H$ is also an $M$” (*fighter-bomber*) cf. COORDINATION

(iii) $M \perp H$ “$M$ is a sub-type of $H$” (*oak tree*) cf. HYPERNYMY (Basic)

$M \triangleright H$ “$H$ is a sub-type of $M$” (*bear cub*) cf. HYPERNYMY (Reversed)

In addition we might posit a third high-level category (106), based on the third of Aristotle’s associative relations, Contrast, which could be regarded as the opposite
of Similarity, just as $A \supset B$ is the opposite of $A \subset B$, and $A \leftarrow B$ is the opposite of $A \rightarrow B$. The relation of Contrast is not required for the categorization of binominals in my data, but it belongs here in order to complete the picture, since we know that the combination of two thing-roots denoting opposites does occur, as in the case of Chinese 东西 $dōngxī$ [east.west] THING (Ceccagno & Scalise 2006: 238).¹

(106) **Contrast-based**

$$\text{CONTRAST}$$

(f) $M<>H$ **$MnotH$** “H is the opposite of M”

(Chinese 东西 $dōngxī$ [east.west] ‘thing’)

Hatcher’s system of two pairs of relations is thereby turned into a system of three pairs of relations. It is still, in Hatcher’s words, “logically conceived, and therefore neater and more pleasing aesthetically” (I would claim) than Arnaud’s extension. It is also “far more comprehensive”, in that it now also includes appositional and co-ordinative compounds, in addition to non-appositional.

### 6.3 Classifying binominals

In preceding section I provided a brief overview of previous studies on semantic relations in compounding; showed how the work of Bauer/Tarasova and Janda provides grounds for optimism that the same kinds of relations apply to binominal lexemes in general; described in some detail the systems developed by Bourque and Hatcher that I have harnessed in my own work; and, with the help of Arnaud, shown how Hatcher’s system can be extended to cater for coordinate compounds and both types of determinative compound: appositional and non-appositional. In this section I describe the application of both Arnaud’s and Hatcher’s systems to the binominals in my database. But before doing so, I need to rant a bit.

**<rant>**

Researchers working on semantic relations have a tendency to reinvent the wheel by developing their own systems of classification. I did the same in my study of nominal compounding in Nizaa (Pepper 2010b), so I am as guilty as anyone. As far as I am aware, none of the work mentioned earlier in this chapter was conducted exclusively on the basis of another researcher’s system, with the single exception of Bauer and Tarasova (2013), who were not overly interested with the actual set of relations, but only whether the same relations are found across different types of

---

¹ The phenomenon is more common when two property-roots are combined: 高矮 $gāoǎi$ [high.low] HEIGHT, 大小 $dàxiǎo$ [large.small] SIZE and 长短 $cháng.duǎn$ [long.short] LENGTH.
English Adnominal Modification constructions. (They chose Levi’s system simply because it is widely known and of a manageable size.)

This is extremely unfortunate, for how is science to advance if we always go back to square one? It would be preferable if the community of researchers were to build on each other’s work, making minor improvements along the way, with the goal of progressing towards consensus on a small number of well-tested and interlocking classifications, suitable for a variety of purposes. Bourque, Ryder and Arnaud are excused: Bourque because his system was synthesized from earlier ones; Ryder, because her work was based around a psycholinguistic experiment in which the whole point was to allow the relations to emerge from users’ responses; and Arnaud because he tested, reused and developed Hatcher’s system (albeit after constructing his own classification).

What is the reason for this unwillingness to build on the work of other researchers? (Of course, everyone *reviews* previous work, because that is what we are supposed to do, and no doubt some insights get reused when studying the kinds of relations that others have identified; in that sense one is building on the work of one’s predecessors. But that is not the same as working towards the common goal of a robust, flexible solution that has been tested against different kinds of data from a large range of languages.) The first reason is that the material we are dealing with is notoriously slippery; semantics is hard to pin down because the only place it exists is in our minds: getting inside someone else’s head is not easy, and it is made more difficult by the fact that many systems are rather poorly documented. How, then, are we supposed to know what the researcher intended? The second reason, I think, is that judgements regarding the nature of a semantic relation are subjective and dependent on the level of granularity one needs: some might regard Levi as being too vague and high-level, others (Hatcher, no doubt) would find her too low-level (and too unsystematic). The third reason is that no system is perfect (with the exception of Hatcher’s); it is easy to spot inconsistencies and errors in others’ work; we see the errors and think: This is no good, I can do better!

At least, that is what I did in 2010. I rejected Levi, for her use of schematic English verbs and prepositions (HAVE, BE, FOR, etc.), as too Anglocentric for the French translation equivalents in my Nizaa data, and likewise Jackendoff’s eclectic mix of the schematic (CLASSIFY) and the specific (PROTECT (FROM)). I very nearly adopted Warren’s scheme, which appealed to me for its role-based naming system, (e.g. WHOLE-PART vs. PART-WHOLE), to which I return in more detail in §8.3 and
§9.3, but in the end I opted to “let the data speak for themselves” and to allow the relations to “emerge” during the analysis. In other words, I started from square one.

In my defence, I was slightly brain-damaged at the time, having spent the last ten years of my life constructing topic maps, for fun and for clients. Each time this involved getting acquainted with – and analysing – a new domain, which could be anything from Italian opera to pharmaceuticals to Pokemon to WALS. The analysis involved identifying topic types and association types for the domain in question; as explained in §1.1.1, association types express semantic relations, albeit often of a more specific nature (e.g. COMPOSED-BY) than the rather more general relations (e.g. PRODUCED-BY) found in most classification schemes. So I thought I was an expert and could do better…

</rant>

This time my approach is more in keeping with the spirit of reuse fostered by international standards such as SGML and Topic Maps, to each of which I devoted ten years of my life. And this time I am reusing not just one, but two pre-existing classifications: those of Bourque and Hatcher.

6.3.1 The Bourquifier

The reasons for adopting Bourque’s scheme were detailed above on pages 195ff, so I won’t rehearse them here. Instead, I begin by showing the presentation format Bourque uses for each of his 15 basic relations, illustrated for POSSESSION in Figure 47. Each such summary is followed by up to four pages of detailed description. Note the test frames (templates), in both English and French; the examples in both languages; and the use of terminology (Head and Modifier) in the templates that is independent of constituent order.

<table>
<thead>
<tr>
<th>PRODUCTION</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation Type</td>
<td>Structure Template</td>
<td>Examples</td>
<td>Linking Material</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>an H that possesses M un T qui possède M</td>
<td>career girl punk à chien</td>
<td>possess (have / of) possède (a / de)</td>
<td></td>
</tr>
<tr>
<td>Reversed</td>
<td>an H that M possesses un T que M possède</td>
<td>family estate droit d’auteur</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 47: Bourque’s template for POSSESSION

Once the head and the modifier of the compound have been identified, it is a simple matter to plug in them in to the templates to see whether the relations hold or not (107).
But while it helps immensely to have such test frames, it takes some time to run through 25 of them for every one of over 3,700 binominals, and this increases the risk of errors creeping in. In order to alleviate this problem, I created an application in Excel, which I call the Bourquifier; shown in Figure 48 in its original version. To use it, one simply types in the binominal, modifier and head in the appropriate cells and all the paraphrases are automatically populated at the same time. I have illustrated this with Bourque’s example, family estate. The two POSSESSION relations (Basic and Reversed) can be found via the left-hand column and are labelled using the codes I originally devised for my database (POSS and POSS2). The latter is highlighted because it is the one I selected, after examining all 25 in the course of a few seconds, as being the most appropriate.1

<table>
<thead>
<tr>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>family estate</td>
<td>family estate</td>
<td>family estate</td>
<td>POSS2</td>
</tr>
<tr>
<td>relation</td>
<td>code</td>
<td>example</td>
<td>code</td>
</tr>
<tr>
<td>HYPERNYMY</td>
<td>HYP</td>
<td>a family is a kind of estate*</td>
<td>HYP2</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COOR</td>
<td>a family estate is a estate and a family</td>
<td>boy king</td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>a estate that is similar to family</td>
<td>ant lion</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>a estate that serves as a family</td>
<td>buffer state</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POSS</td>
<td>a estate that possesses a family</td>
<td>career girl</td>
</tr>
<tr>
<td>PART</td>
<td>PART</td>
<td>a estate that is part of a family</td>
<td>table leg</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>a estate located at/near/in a family</td>
<td>window seat</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>a estate made of family</td>
<td>sugar cube</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>a estate (made) from family</td>
<td>cane sugar</td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>a estate that causes family</td>
<td>tear gas</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>a estate that makes family</td>
<td>honey bee</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>a estate that is about family</td>
<td>history conference</td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>a estate that occurs at/during family</td>
<td>summer job</td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>a estate that uses family</td>
<td>steamboat</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURPOSE</td>
<td>a estate intended for family</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

![Figure 48: The Bourquifier – FAMILY ESTATE](image)

A few more comments regarding the Bourquifier are in order. Blank areas on the right-hand side obviously indicate that the relation (e.g. COORDINATION) is not reversible. Also, the user has to figure out which articles to use: the Bourquifier is not able to figure out the count/mass noun status of the head and modifier, and in the version shown it here does not check for words beginning with a vowel. The

1 Of course, this does not happen automatically. The Bourquifier is not that smart. Yet.
templates for HYPERNYMY are marked with an asterisk because I felt a need to revise Bourque’s original (cf. Table 28 on page 197), for reasons that will be explained shortly. The Bourquifier was extremely easy to create, so everyone can have their own: all that is required is a fairly simple formula in each of the template cells, illustrated for the basic (a) and reversed (b) possession templates in (108).

(108)  
(a) =CONCATENATE("a ";$J$2;" that possesses a ";$I$2)  
(b) =CONCATENATE("a ";$J$2;" that ";$I$2;" possesses")

Thanks to the Bourquifier the task of assigning semantic relations to individual binominals became much faster and, I like to think, much more reliable. If the reader has any doubt on this matter, I invite her to turn back to the set of relations listed on page 197 and consider how she would handle the word beeswax; glance over the relations and decide which exact one (Basic or Reversed) is most appropriate. Think about the length of time the operation is taking, and consider having to repeat this for over 3,500 binominals. The correct solution is shown, along with all 24 incorrect ones, in Figure 49.

<table>
<thead>
<tr>
<th>The Bourquifier</th>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>relation</td>
<td>code</td>
<td>basic template</td>
<td>example</td>
<td>reversed template</td>
</tr>
<tr>
<td>HYPERNYMY</td>
<td>HYP</td>
<td>a bees is a kind of wax*</td>
<td>oak tree</td>
<td>a wax is a kind of bees*</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COOR</td>
<td>a beeswax is a wax and a bees</td>
<td>boy king</td>
<td></td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>a wax that is similar to bees</td>
<td>ant lion</td>
<td></td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>a wax that serves as a bees</td>
<td>sugar state</td>
<td></td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POS</td>
<td>a wax that possesses a bees</td>
<td>career girl</td>
<td>a wax that bees possesses</td>
</tr>
<tr>
<td>PART</td>
<td>PART</td>
<td>a wax that is part of a bees</td>
<td>table leg</td>
<td>a wax that bees is part of</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>a wax located at/near/in a bees</td>
<td>window seat</td>
<td>a wax that bees is located at/near/in</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>a wax made of bees</td>
<td>sugar cube</td>
<td>a wax that bees is made of*</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>a wax (made) from bees</td>
<td>cane sugar</td>
<td>a wax that bees is (made) from</td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>a wax that causes bees</td>
<td>tear gas</td>
<td>a wax that bees causes</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>a wax that makes bees</td>
<td>honey bee</td>
<td>a wax that bees makes</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>a wax that is about bees</td>
<td>history conference</td>
<td></td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>a wax that occurs at/during bees</td>
<td>summer job</td>
<td>a wax at/during which bees occurs</td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>a wax that uses bees</td>
<td>steamboat</td>
<td>a wax that bees uses</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>a wax intended for bees</td>
<td>animal doctor</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 49: The Bourquifier – BEESWAX**

Of course, it is not always plain sailing, as the quote from Hatcher on page 212 indicates; there are a few cases in which “the choice appears difficult”, but mainly, as she says, “we will not have to hesitate”. And even when we do have cause to hesitate, the difficulty is usually less extreme than in the example used by Hatcher. Mostly, the choice is between two, or possibly three, relations, all of which tend to

1 A later version of the Bourquifier is included in the data set available from the Tromsø Repository of Language and Linguistics.
be subtypes of one and the same high-level relation. For example, PROD2, Bourque’s PRODUCTION (Reversed),\(^1\) overlaps to some extent with CAUS2; many researchers (including Downing 1977, Warren 1978 and Vanderwende 1994) have reduced them to a single relation, as Bourque points out in his discussion of the matter (pp. 204-205). Jackendoff (2010) keeps them separate, as CAUSE(X, Y) and MAKE(X, Y), both of which are reversible, and notes that it is “sometimes hard to distinguish MAKE from CAUSE. Perhaps MAKE (X,Y) decomposes as CAUSE (X, (COME INTO EXISTENCE (Y)))” (p. 441). He then proceeds to classify suntan under the one and sunburn under the other, which hardly seems consistent.\(^2\)

In the case of beeswax, the paraphrases shown in Figure 49 make it rather clear that PROD2 is to be preferred to CAUS2. However, we should not be overly worried about such cases: gradience, as we know, permeates the whole of language, and this is just another example. Where we should be watchful is when the two alternative solutions are subtypes of different high-level types, but that is not the case here: both PROD2 and CAUS2 fit neatly into Hatcher’s A→B, as we shall see in §6.3.3.

It proved to be surprisingly easy to accommodate my 3,738 binominals within the Bourque scheme, but around 100 resisted classification. For the most part this was because the original motivation underlying the combination of concepts could not be ascertained without further information. In some cases this was simply due to the presence of a cranberry morpheme, as in Bezhta häš c ’ic’ [eye:GEN ?] EYELASH. While this is undoubtedly a binominal that conforms to the Bezhta Mod.GEN Head schema, the semantic relation cannot be determined, for obvious reasons. In other cases the difficulty can be ascribed to unfamiliar beliefs or cultural practices, as may be the case with Takia tamol sos [man Derris_root] WIDOWER (mentioned earlier on page 112). All such situations were labelled opaque.

For a handful of binominals (four in number), the motivation was clear, but the relation did not seem to fit any of the existing categories. All such cases involved words containing numerals used to denote days of the week, such as Iraqw deelór tám [day:OF three] WEDNESDAY and Äiwoo dâbu mi.polee [day BN:GNL.seven]. It could be argued that such words do not belong in the database; they are, at best,\(^2\)

\(^1\) For the remainder of this section I will use codes like PROD, PROD2 (in both text and screenshots) for reasons of space, instead of labels like Production (Reversed). In the next section (§6.3.2) I present a revised version of Bourque’s classification, along with a revised set of codes. It is the latter that are used in the database and in the remainder of this work.

\(^2\) “Here, surely, Jackendoff is napping,” if I may be permitted to borrow Hatcher’s comment on Jespersen (1960: 357, footnote 3). Perhaps Jackendoff didn’t have a Bourquifier to help him! Kudos to Bourque (p. 205) for spotting this inconsistency.
extremely peripheral instances of binominal lexemes. Rather than introduce a new type of semantic relation to cater for them, they were assigned the code xrel and disregarded in the semantic analysis.

The code error was used to indicate data points where the word supplied by the contributor did not seem to denote the intended meaning. Examples include Yaqui muumu jo’ara [bee house] and Japanese kaji.ya [smith.shop], which purport to denote BEESWAX and BLACKSMITH, respectively. Since it combines the concepts ‘bee’ and ‘house’, the Yaqui word can be assumed to denote BEEHIVE, not BEESWAX; the likelihood that it colexifies BEEHIVE and BEESWAX is low, since the language has another word denoting the latter concept (sito’ori). As for Japanese kajiya, it denotes a place (the smithy) where the profession is carried out, not the profession itself. So, while both of these words qualify as binominals, encoding their semantic relations as POSS2 and LOC2, respectively, would muddle the later analysis.

Two sets of binominals initially posed a challenge. The first was a group of words that combine ‘man’ with a concept like ‘death’, ‘funeral’ or ‘misery’ in order to denote WIDOWER. None of Bourque’s paraphrases seemed satisfactory. However, thinking back to Jespersen and Hatcher, it became clear that these were covered by Jespersen’s Characterizing Feature which corresponds most closely to Bourque’s TOPIC.

The other set of meanings that at first seemed to indicate a gap in Bourque’s system is the professions: BLACKSMITH, CARPENTER, FARMER, FISHERMAN, HERDSMAN, POTTER and SHOEMAKER. These are not always denoted by binominals; in fact they exhibit a range of word-formation types, including OT 1–4, exocentric constructions and monomorphemic forms (Table 29).

<table>
<thead>
<tr>
<th>Type</th>
<th>English</th>
<th>BLACKSMITH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomorphemic</td>
<td>carpenter</td>
<td>Finnish seppä</td>
</tr>
<tr>
<td>Onomasiological Type 1</td>
<td>shoe.make.r</td>
<td>Ket č.ted.s [iron.hit.NMLZ]</td>
</tr>
<tr>
<td>Onomasiological Type 2</td>
<td>farm.er</td>
<td>Navajo ‘atsid.ii [he_pounds.NFE]</td>
</tr>
<tr>
<td>Onomasiological Type 3</td>
<td>pott.er</td>
<td>Mandarin Chinese duan4.gong1 [forge.worker]</td>
</tr>
<tr>
<td>Onomasiological Type 4</td>
<td>black.smith</td>
<td>black.smith</td>
</tr>
<tr>
<td>Exocentric</td>
<td>NA</td>
<td>Hawaiian ku’i hao [strike iron]</td>
</tr>
</tbody>
</table>

Table 29: Variety of word-formation types – professions

In most binominals denoting professions, the head is either a thing-root denoting a person (usually glossed as man, artisan, professional, doer, worker or guard), or an
agentive affix, like the English -er. The modifier denotes entities such as hoof, iron, hammer, anvil or forge (BLACKSMITH); wood, table or house (CARPENTER); earth, field, garden, farm, land or plough (FARMER); fish, bait or fishhook (FISHERMAN); animal, livestock, cattle, cow or pasture (HERDSMAN); pot, clay or ceramics (POTTER); and boot, shoe or embroidery (SHOEMAKER). These appear to involve a vast range of relations, some of which fit nicely into Bourque’s scheme: USE for a blacksmith with her hammer and anvil, and also her iron (albeit as a material rather than a tool); PROD for a carpenter making a table, a potter making ceramics or a shoemaker making shoes. But then, a fisherman catches fish, a farmer cultivates her field or garden, and a herdsman tends her animals; they neither use nor produce them. Where do they fit into the scheme? What is the relation between a blacksmith and a hoof? And is there not a sense in which we would like all of these to fall under a single relation: “has as profession or occupation”, or simply “occupied with”? Let us see how some previous researchers addressed the problem.

In her corpus of novel compounds in Norwegian Eiesland (2016) has a number of words in which the second constituent mann ‘man’ is typical of compounds that denote professions: bensin.mann [petrol.man] “man who prefers petrol-driven cars to electric”, stigemann [ladder.man] “man seen climbing a ladder at the site of a fire”, and ekspedisjon.s.mann [expedition.LE.man] “expeditioner, man who goes on expeditions.”¹ The first two are classified under her relation “characteristic part” and the last under “part of a whole”. None of them have much to do with profession, with the possible exception of the latter; but while “part of a whole” (Bourque’s PART: “a man that is part of an expedition”) is a reasonable reading, this example serves primarily to add yet another possible relation to the mix of professions.

Jackendoff has violinist and street singer among his examples. To handle the former (as profession, rather than occasional activity or ability), he employs the function OCC (occupation) which appears as if out of a hat (109a), combining it with PLAY (also not a basic function). Two non-basic functions are thus required to cater for this particular profession. For street singer, three functions are employed: OCC (or CHAR), SING and BE (109b). Jackendoff’s functionally based system is much

¹ None of these are found in dictionaries, so I have inferred the meanings from their use on websites: expedisjonsmann: https://www.dagbladet.no/nyheter/vennene-minnes-erne/65939753; bensinmann: http://www.agderposten.no/kjop-tilgang?aId=1.1575557; stigemann: https://www.adressa.no/nyheter/trondheim/bybrannene/prinsen_brannen/article69130.ece (all accessed 2018-05-31).
more powerful than the usual fixed set of categories, but in this study I have committed to the latter, so Jackendoff is not of much help, apart from the fact that he recognizes profession (or occupation) as a significant relation.

(109)  
(a) violin\textsubscript{1}ist\textsubscript{2} = PERSON\textsubscript{2}; [OCC (PLAY (α, VIOLIN\textsubscript{1}))]
(b) street\textsubscript{1} singer\textsubscript{2}\textsubscript{3} = [PERSON\textsubscript{2}; [OCC/CHAR ([SING\textsubscript{3}\textsubscript{β}(α); [BE (β, IN STREET\textsubscript{1})]})]]

Once again Hatcher comes to the rescue, pointing out that Mätzner’s fifth type, “A=that with which a Person B concerns himself or works” (ale-wife, blacksmith etc.) “is superfluous, being a subdivision of Purpose” (fn. 12, p. 362). Jespersen’s Purpose, as we recall, was subsumed by Hatcher under A←B (see Figure 46 on page 200). She is followed by Arnaud, who assigns his low-level relation ‘NON-HEAD is the object of the typical activity of HEAD’, systems engineer to the same high-level category. Thus, the answer to the problem of profession is clear: it belongs under PURP, which I will map to A←B in §6.3.3.

To sum up, the task of classifying my 3,500+ binominals according to Bourque’s system was surprisingly unproblematic, thus testifying to the comprehensiveness of the system and the quality of the documentation, as well as the usefulness of the Bourquifier. However, the system could be further improved with a few tweaks. These are discussed in the next section.

6.3.2 Bourque2

There were some minor errors in Bourque’s classification (as is only to be expected given its complexity), and also room for improvement in a few areas. Here I deal in turn with relations, reversibility, codes, templates and examples.

**Relations**

Most of the names used by Bourque denote conceptual relations (or one of the roles of a conceptual relation) with one exception: HYPERNYMY denotes a lexical relation. For consistency I propose to use TAXONOMY instead.

For clarity it is also desirable to avoid using the same name for a relation and one of its roles, especially when there is a need to distinguish between the relation in general and one of its directions (Basic or Reversed). I therefore propose to rename the relation PART, one of whose roles is itself ‘part’ (the other being ‘whole’); the label PART-WHOLE would do, but there exists a widely established term for this
kind of relation, namely MERONOMY, which I prefer.¹ For the same reason I opt to use CAUSATION instead of Bourque’s CAUSE; TEMPORALITY instead of TIME; and USAGE instead of USE. In some cases it is difficult to find a suitable alternative; this applies to FUNCTION, PURPOSE and TOPIC. However, all of these are non-reversible and so the problem of distinguishing between the general relation and one of its directions does not arise. One such issue remains unsolved and that is the SOURCE relation, where the roles are ‘source’ and ‘result’. Finding no suitable alternative, and deeming SOURCE-RESULT to be too unwieldy, I leave the name of this relation unchanged.

In addition I propose to shorten Bourque’s PURPOSE AND PROPER FUNCTION to just PURPOSE, since the element of intentionality expressed by ‘proper function’ seems to be inherent in the latter. This also avoids confusion with another relation called simply FUNCTION, which does not involve intentionality.

One might question whether one needs both relations. Could they not be subsumed under something more general? The answer, is yes, they could, and later on I will claim that any two relations can be combined into a more general relation, and that any relation can be subdivided into more specific relations. The crucial question is not whether categories can be lumped together or split apart, but what purpose this serves and what costs are involved. One cost of lumping is that paraphrases become much more vague and harder to apply; a cost of splitting is that the boundaries between different categories become much more fuzzy and more data is required in order to populate each category with sufficient data for statistical analysis. My experience applying Bourque’s system is that the level of granularity he has chosen achieves a good balance for the practical purpose of classification, and I therefore opt not to tinker with it more than necessary. It will turn out later that Bourque’s system is in some respects too fine-grained for the purpose of statistical analysis, but the problems this presents are overcome by the mapping to Hatcher’s high-level system.

Similar considerations apply to the decision not to combine PURPOSE and USAGE, as might seem reasonable. To take Bourque’s example, it is true that a hand brake (“(a) brake that (a) hand uses” = USE2) is also “(a) brake intended for (a) hand” (= PURP). However, the inverse does not work: an animal doctor (“(a) doctor intended for (an) animal” = PURP) can hardly be described as “(a) doctor that (an) animal

¹ This term is more general than meronymy, which is specific to linguistics.
uses” (= \textsc{use}2). Thus combining the two would require a fundamental rephrasing of the templates.

The Hawaiian binominal \textit{pahu meli} [box honey] \textsc{beehive} raises another issue. Is a beehive “(a) box that (a) honey is part of” (\textsc{part}2) or that “honey is located at/near/in” (\textsc{loc}2)? Of course, location is involved, and we could also, at a pinch, say that honey is part of the beehive, but what I, at least, would like to say is that the box \textit{contains} honey. So how does \textsc{containment} fit into Bourque’s scheme? It is not one of his 25 relations, but Bourque does not ignore the matter. He discusses it in depth in the context of the overlap between \textsc{part} and \textsc{location}, using the example of \textit{toolbox}. I quote at some length in order to convey the detail of his discussions in general:

Another issue to consider is that some compounds might be analysed as either \textsc{part} or \textsc{location}. This dual analysis is related to the fact that \textsc{location} may subsume \textsc{part}: if something is a part of something else, then it is located at/on/in that thing (cf. Baron and Herslund 2001). One possible solution is to reserve location for only those compounds that actually involve a locative noun, as does Adams (1973). The problem, of course, is that one must treat combinations such as \textit{toolbox} or \textit{treehouse} using some other relation, as they do not, in the strictest sense, involve places. The key distinction that will be used here is one that views the \textsc{part} relation as a reference to an integral component of the whole, without which it would either be incomplete, defective, or non-functional. Thus, a negation test may be used to determine whether the modifier denotes an essential part of the compound. The formulation in (105) below shows how such a test might apply to compounds in which the head denotes the whole (cf. 104 above):

\begin{align}
\text{(105)} & \quad \text{a. a C without an M is still a C} \\
& \quad \text{b. un C sans M est toujours un C}
\end{align}

A positive response to the above sentence would indicate that the modifying noun is not an essential component of the object denoted by the compound, but instead a distinguishing feature. Thus, a toolbox without tools is still a toolbox, which indicates that tools is connected to box via \textbf{some other relationship} (i.e. \textsc{container-contained}). This result is the same for the French \textit{boîte à outils} (i.e. \textit{une boîte à outils sans outils est toujours une boîte à outils}). When applied to compounds that denote a part-whole association, the test produces defective or incomplete readings (pp. 196–197, emphasis added).

The case of “honey box” (\textit{beehive}) is parallel to \textit{toolbox}: a beehive without honey is indubitably still a beehive. The distinction Bourque makes is useful, but I beg to differ regarding his conclusion to treat \textit{toolbox} (and thus also “honey box”) as (mere) location. I suggest that it might be better to bite the bullet and add the relation
CONTAINMENT (which even Bourque recognizes as “some other relationship”) to his system, on the grounds that the ability to perceive containment is a fundamental part of our cognitive endowment. Such a relation would be reversible (110), and it would constitute the prototypical (low-level) subtype of Hatcher’s high-level \( A \subset B \) and \( A \supset B \). As examples, I will adopt Hatcher’s orange seed and seed orange.

(110) CONTAINMENT
(a) \( \text{CONT: “(an) H that is contained in (an) M” orange seed } A \supset B \)
(b) \( \text{CONT-R: “(an) H that contains (an) M” seed orange } A \subset B \)

The other addition made to Bourque’s set of relations was motivated not by any of the binominals in my database, but by one of Jespersen’s examples: sun worship (see page 186). Strictly speaking this is not a binominal since worship denotes an action, not an object. However, the scope of Bourque’s classification is broader than adnominal modification and therefore it should be able to accommodate sun worship. Since the French equivalent, adoration du soleil, is not in Bourque’s database, we are obliged to try and classify it ourselves, and it turns out that none of Bourque’s relations are appropriate. Clearly the notion of the sun as some kind of goal is involved, so one might think that Bourque’s SOURCE would do the job, but no amount of tweaking of either the basic or the reversed template produces a paraphrase that is acceptable for both sun worship and cane sugar (or sun worship and sugar cane). This seems to be because goal as a complement of source is not compatible with result. It seems that a new relation is unavoidable, but what to call it, and how to make it sufficiently distinct from SOURCE? The answer was provided by Hatcher, who included sun worship in her category \( A \rightarrow B \), pointing out that “the sun is that toward which the worship is directed” (see Figure 46, emphasis added). Now, it is frequently the case that the verb used to express the paraphrase can serve in nominalized form as the name of the relation itself (as in the case of ‘possess’ > POSSESSION). The solution to the problem of how to name the new relation is thus given: ‘direct’ > DIRECTION, understood as a relation which relates a starting point or origin and an endpoint or goal (111):

(111) DIRECTION
(a) \( \text{DIR: “(an) H whose goal is (an) M” sun worship } A \leftarrow B \)
(b) \( \text{DIR-R: “(an) H that is the goal of (an) M” sales target } A \rightarrow B \)

Adding such a relation to Bourque’s scheme can be justified on two grounds (over and above the desire to accommodate sun worship): firstly, it is very general and thus in a different league from Jackendoff’s PROTECT (FROM); and secondly,
we can assume that the ability to conceptualize direction is an important part of the human cognitive endowment. Further research may show that DIRECTION is rarely encountered in binominals, but it may turn out to be more important when action-roots are admitted (as in sun worship). Thus, whether or not the relation is frequent enough in word-formation to justify its inclusion remains an empirical question. And in any case, my plans for the Hatcher-Bourque classification are much broader than the domain of binominals,¹ and that in itself is reason to include the relation, now that its existence has been recognized.

**Reversibility**

The second set of changes to Bourque’s scheme is motivated by a desire to remove a certain arbitrariness in his choice of which form of a reversible relation to regard as basic and which to regard as reversed. Recall from Table 28 that Bourque has *career girl* as an example of the Basic Possession relation and *family estate* as an example of the Reversed Possession relation. No reason is given for regarding “an H that possesses M” as being more basic than “an H that M possesses”; the decision appears to be arbitrary. Similarly, *table leg* exemplifies the Basic Part relation and *wheelchair* the Reversed Part relation; again the choice of “an H that is part of M” as basic and “an H that M is part of” appears to be arbitrary.

Recall now, as stated in §6.2.3, that I intend to map Bourque’s classification to Hatcher’s to create a two-tier classification. If I do so on the basis of the original Bourque classification, *career girl* (basic) will map to Hatcher’s ₩ (A is contained in B) and *family estate* (reversed) will map to ₩ (B is contained in A). On the other hand, *table leg* (basic) will map to ₩ and *wheelchair* (reversed) will map to ₩. This will be confusing; it would be more consistent for every low-level relation that maps to one of Hatcher’s two “static” relations (which I recast as CONTAINMENT, cf. page 215) to be either “basic” or “reversed”. For this reason, I propose to invert Bourque’s POSSESSION relation such that the direction regarded as basic is the one in which the head element (B) is “somehow, to some extent, contained, comprehended in” the modifying element (A); *career girl* and its like will then map to ₩ instead of ₩ and the overall model will be consistent. The same problem applies to Bourque’s COMPOSITION, where *sugar cube* (“an H made of M”), which is regarded as basic, maps to ₩ while the reverse, *sheet metal* (“H that M is made of”), maps to ₩. This relation, too, will be inverted.

¹ This is the topic of Chapter 9, *A model of associative relations*. 
Similar considerations apply to those of Bourque’s relations that map to Hatcher’s two “dynamic” relations (recast above as CAUSATION). In this case *honey bee* (Basic PRODUCTION) maps to A←B (the bee is the “source” of the honey), but *cane sugar* (Basic SOURCE) maps to Hatcher’s A→B (the cane is the source of the sugar); for consistency it is preferable that the Basic Source relation also maps to A←B, and this is achieved by inverting Bourque’s model. The same is done to the USE relation: *hand brake* (“an H that M uses”) will be regarded as basic and *steamboat* (“an H that uses M”) as reversed. As a result all of the “dynamic” Basic relations map to A←B and all the “dynamic” Reversed relations map to A→B.

It is important to understand that the labels Basic and Reversed are still arbitrary, in the sense that they do not indicate any kind of psychological reality, but they are now used in a consistent manner. Having said that, when the data is analysed in §6.4.2, it will transpire that Hatcher’s Ⓑ (our HinM) is far more common than her Ⓐ (MinH), and that her A←B (HtoM) is far more common than her A→B (MtoH); in others words “basic” and “reversed” will turn out to correspond to “frequent” and “infrequent”.

**Codes**

Bourque did not publish codes for his 25 relations, although he presumably uses them in his database. I therefore devised my own and I document them in this dissertation so that they can be reused by other researchers and used as labels in the discussion to follow. The initial set of codes used abbreviations of the name of the basic relation, with the number 2 appended for reversed relations (e.g. PROD and PROD2, cf. footnote 1 on page 221); examples can be found in the previous section, both in the Bourquifier screenshots and in the text. After revising Bourque’s system (as described in the present section), and in particular after inverting Basic and Reversed for Bourque’s POSSESSION, COMPOSITION, SOURCE and USE, I revised the codes, appending -R instead of the number 2 for reversed relations (thus PROD2 became PROD-R). These new codes are used in the remainder of this dissertation, in the database, and in the R source code used for statistical analysis. Note that screenshots will henceforth be from a version of the Bourquifier that has been updated accordingly.

**Templates**

In addition to the templates introduced along with the new relations CONTAINMENT and DIRECTION, I propose certain other changes to templates. One is to the template for COORDINATION and is motivated by the desire to bring it in line with the other
templates. Thus instead of Bourque’s “a C is an H and an M” I opt for “(an) H that is also (an) M”.

Changes to the templates for SOURCE are prompted by the Caijia word for queen (112).

(112) Caijia

\[ k^h\eta \, ^{55}j\,a^{21} \ [\text{sky.old\_woman}] \ \text{QUEEN} \]

In trying to figure out the motivation behind this word, I recalled how the emperor of China is addressed by both The Unknown Prince (Calàf) and Turandot as *Figlio del cielo* ‘Son of the sky’ in Puccini’s opera *Turandot*. I assume (but do not know for sure) that royalty were in those days believed to have descended from heaven. If so, the sky must be the point of origin of Caijia’s queen, which means that SRC-R is the most appropriate relation. But the paraphrase leaves somewhat to be desired: “an old woman (made) from sky”. This is so jarring that I was tempted to create a new relation, ORIGIN, with the template “(an) H that originates in (an) M”, and of course, there is nothing in principle to prevent me from doing so: as the analyst, I get to decide my own level of granularity, as pointed out earlier. But this would cause confusion, since ORIGIN and SOURCE mean pretty much the same thing.

A better solution is to reformulate the template. If we avoid the verb ‘to make’ in doing so, we also solve the problem that the same verb is used in the templates of two other relations, and that can be confusing. My proposal is therefore to use the following templates, which also work for Bourque’s examples:

- **SRC**: “(an) H that is a source of (an) M” (e.g. *sugar cane*)
- **SRC-R**: “(an) H whose source is (an) M” (e.g. *cane sugar*)

The paraphrase for \[ k^h\eta \, ^{55}j\,a^{21} \] thus becomes “(an) old woman whose source is (a) sky”.

The other two relations whose templates use the verb ‘to make’ are COMPOSITION, which uses ‘made of’, and PRODUCTION, which uses ‘makes’. To avoid confusion I propose reformulating these using the verbs from which the name of the relation is derived:

- **COMPOSITION** (basic and reversed): ‘made of’ \(\rightarrow\) ‘composed of’
- **PRODUCTION** (basic and reversed): ‘makes’ \(\rightarrow\) ‘produces’

The final changes are to the templates for Bourque’s HYPERNYMY relation, which I rechristened TAXONOMY above. The original templates are shown in (113).
(113) HYPERNYMY
   (a) HYP: “an H of kind M” oak tree
   (b) HYP2: “an H that M is a kind of” bear cub

These paraphrases confused me no end when considering Bourque’s example bear cub (“??a bear cub is a cub that bear is a kind of”), and I fared no better with the corresponding paraphrases provided by Jackendoff (2010) for his reversible KIND (114).

(114) KIND (X, Y)
   (a) \([Y^n_2: [\text{KIND (}X_1, \alpha)]], \text{‘an } N_2 \text{ of kind } N_1’\)
      puppy dog, ferryboat, pine tree, gemstone, limestone, girl child
   (b) \([Y^n_2, [\text{KIND (}X, X_1)]]], \text{‘an } N_2 \text{ that is a kind of } N_1’\)
      seal pup, bear cub

Notice that (113a) and (114a) are essentially identical (\(N_2 = H,\) and \(N_1 = M\)), but that (113b) is the opposite of (114b). The points here go to Jackendoff: a (seal) pup is a kind of seal and a (bear) cub is a kind of bear, but a bear cub is not a cub that bear is a kind of, as Bourque would have it. The reader’s head may be spinning (I know mine was, and I think it was due to the formulation of all three paraphrases). I propose therefore to amend Bourque’s templates as shown in (115).

(115) TAXONOMY
   (a) TAX: “(an) M is a kind of H” oak tree
   (b) TAX-R: “(an) H is a kind of M” bear cub

<table>
<thead>
<tr>
<th>Bourquifier3</th>
<th>Binomial (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAX</td>
<td>oak tree</td>
<td>oak</td>
<td>oak tree</td>
<td>TAX-R</td>
</tr>
<tr>
<td>TAX</td>
<td>oak tree</td>
<td>oak tree</td>
<td>oak tree</td>
<td>TAX</td>
</tr>
<tr>
<td>TAX</td>
<td>oak tree</td>
<td>oak tree</td>
<td>oak tree</td>
<td>TAX</td>
</tr>
</tbody>
</table>

Figure 50: The Bourquifier – OAK TREE
The typology and semantics of binominal lexemes

The new formulations produce the results shown in Figure 50 and Figure 51. The populated templates make it clear that oak tree belongs under TAX (“(an) oak is a kind of tree”) and not under TAX-R (“(*“(a) tree is a kind of oak”), and that bear cub belongs under TAX-R (“(a) cub is a kind of bear”) and not under TAX (“(*“(a) bear is a kind of cub”).

<table>
<thead>
<tr>
<th>Bourquifier3</th>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>bear cub</td>
<td>bear cube</td>
<td>bear</td>
<td>cub</td>
<td>TAX-R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation Roles</th>
<th>Example</th>
<th>B2</th>
<th>H2</th>
<th>Reversed template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAXONOMY</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COORDINATION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSSESSION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOCATION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMPORALITY</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPIC</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIRECTION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOURCE</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUSATION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USAGE</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUNCTION</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PURPOSE</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation</th>
<th>Roles</th>
<th>Example</th>
<th>B2</th>
<th>H2</th>
<th>Reversed template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAX</td>
<td></td>
<td>oak tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAX-R</td>
<td></td>
<td>bear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIM</td>
<td></td>
<td>boy king</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIM-R</td>
<td></td>
<td>kidney bean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONT</td>
<td></td>
<td>orange seed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONT-R</td>
<td></td>
<td>seed orange</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td></td>
<td>family estate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS-R</td>
<td></td>
<td>career girl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MER</td>
<td></td>
<td>car motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MER-R</td>
<td></td>
<td>motor car</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC</td>
<td></td>
<td>house music</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC-R</td>
<td></td>
<td>music hall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMP</td>
<td></td>
<td>summer job</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMP-R</td>
<td></td>
<td>golf session</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMP</td>
<td></td>
<td>cube sugar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMP-R</td>
<td></td>
<td>sugar cube</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOP</td>
<td></td>
<td>history book</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOP-R</td>
<td></td>
<td>non-reversible</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIR</td>
<td></td>
<td>sun worship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIR-R</td>
<td></td>
<td>sales target</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOR</td>
<td></td>
<td>super care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOR-R</td>
<td></td>
<td>care sugar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS</td>
<td></td>
<td>urine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS-R</td>
<td></td>
<td>vomit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROD</td>
<td></td>
<td>long fuel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROD-R</td>
<td></td>
<td>oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USE</td>
<td></td>
<td>lamp oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USE-R</td>
<td></td>
<td>oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUC</td>
<td></td>
<td>buffer store</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUC-R</td>
<td></td>
<td>oil temp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUR</td>
<td></td>
<td>animal doctor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUR-R</td>
<td></td>
<td>(non-reversible)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 51: The Bourquifier – BEAR CUB**

While on the subject of bear cub, it is worth noting that there is a subtle difference between the relation it exhibits – Jackendoff’s KIND (X,Y), our TAX – and the one exhibited by Jackendoff’s reversed type – KIND (X,Y), exemplified with puppy dog. For while there can be no doubt that a puppy is a kind of dog, it is not the case that a cub is a kind of bear in the same sense: dictionary.com (accessed 2018-05-29) defines puppy, quite simply, as “a young dog”, whereas cub is defined as “the young of certain animals, as the bear, lion, or tiger”. So every puppy is a dog, but not every cub is a bear: it may be a lion, a tiger, or even a shark. We are reminded of Hatcher’s two subtypes of apposition: species-genus (pumice stone) and cross-classification (fuel oil), cf. page 208. Recall that in the former “an object belonging to class B (stone) may be assigned to a subdivision of that class (pumice), whereas in the latter “an object seen as belonging primarily to class B (oil) may also be assigned to a different, but complementary, class A (fuel).” While puppy is clearly a subclass of dog (and only dog), cub is a subclass of multiple classes, including bear: a cub need not be a bear. So is bear cub an instance of the species-genus type or of the cross-classification type? It could go either way; it depends on whether one admits the possibility of a class (CUB) belonging to multiple superclasses.

1 At least at this level of (folk) taxonomy.
Some systems of logic allow this, others don’t. For the purists, the alternative (and I find it a good one, so perhaps I am a purist) is to classify bear cub under Hatcher’s cross-classification, which is essentially the same as Jackendoff’s BE and Bourque’s COORDINATION: “a bear cub is a cub and a bear” (cf. Figure 51). This illustrates, once again, the slipperiness of our enterprise. The good news is that whichever solution we go for, bear cub will still end up under the Similarity relation when we map from Bourque to Hatcher; this, I suggest, demonstrates the power and utility of the two-level classification.

Examples

Most of my proposals for changes to examples are motivated by pedagogical (and in a couple of cases, aesthetic) considerations. Only one of Bourque’s 25 examples can actually be said to be erroneous, and that is the use of sunburn to exemplify his Basic CAUSE with the paraphrase “an H that causes M”. It is, of course, the sun (M) that causes the burn (H), not the other way round, so this example properly belongs under Reversed CAUSE, with the paraphrase “(a) burn that (a) sun causes”. A better example for Basic CAUSE is tear gas: “(a) gas that causes (a) tear”, as the Bourquifier shows (Figure 52).

![](https://example.com/figure52.png)

It is arguable that the example provided by Bourque for SIM is not incorrect, but it is certainly suboptimal. An ant lion (or antlion) is not a lion that is similar to an ant, it is a kind of insect, albeit not exactly an ant. The name appears to be a left-headed calque from Latin formicaleo, which means that the paraphrase “an ant that is similar to a lion” does in fact work. However, as a highly exceptional left-headed compound it is unsuitable in an English context for pedagogical reasons (it works...
The typology and semantics of binominal lexemes

fine as Fr. *fourmi-lion*, which may be how Bourque got to choose it as his example. I have replaced it with *kidney bean*, which Hatcher uses to exemplify her third type of cross-classification compound; I do so in preference to *butterfly table* (see page 209) for the simple reason that it takes up less space. Space-saving is also the consideration behind the choice of *history book* instead of *history conference* for TOPIC, and *sunburn* instead of *motion sickness* for CAUS-R; real estate is at a premium not only in the Bourquifier but also in Table 31, below!

Bourque’s examples for SOURCE, *cane sugar* (SRC) and *sugar cane* (SRC-R), have the pleasing property of consisting of the same two elements in reverse order. They are what David-Antoine Williams has dubbed “boathouse words”; these were the topic of a Language Log blog posting¹ by Marc Liberman, inspired in turn by an *xkcd* cartoon (Figure 53). Like the cartoonist, Randall Munroe, I rather like such words and think we should apply the scheme more consistently; in fact, I would like to go one step further and use boathouse words to exemplify every kind of relation in Bourque’s system, and not just the SOURCE relation.

![Diagram of boathouse words]

*Figure 53: Boathouse words*
Source: https://xkcd.com/2043/

Unfortunately *boathouse* and *houseboat* themselves are not suitable, because they exemplify the same (reversed) form of the CONTAINMENT (or “holds”) relation

---

¹ https://thelifeofwords.uwaterloo.ca/boathouse-words/; http://languagelog.ldc.upenn.edu/nll/?p=39951
For CONTAINMENT I therefore adopt Hatcher’s *orange seed* (CONT) and *seed orange* (CONT-R). For COMPOSITION Bourque already has *sugar cube*, which is complemented nicely by *cube sugar*. For PRODUCTION I go with *song bird* and *bird song* instead of *honey bee* and *beeswax*. *Oil lamp* and *lamp oil* will replace *hand brake* and *steamboat* as examples of USG and USG-R; and *car motor* and *motor car* can do the job of *table leg* (MER) and *wheel chair* (MER-R). Finding a suitable pair for LOCATION is more difficult: the closest I have come is *house music* and *music hall*, and so far I have failed to come up with anything at all for POSSESSION, TEMPORALITY and CAUSATION.

Summary

The changes to Bourque’s scheme, summarized in Table 30, are mostly cosmetic, but since they involve extending the classification from 25 to 29 relations, I will henceforth refer to the revised system as Bourque2.

---

1 In point of fact a houseboat is not really a boat that “holds” (or contains) a house: it is a boat that serves as a house, so the relation is actually FUNCTION, but we grant Randall poetic license.

2 Rolf Theil (p.c.) has suggested Norwegian *skadeflom* [damage flood] ‘flood that causes damage’ (CAUS) and *flomskade* ‘flood damage’ (CAUS-R), but the former doesn’t work in English. There’s a pint going for anyone who can come up with suitable English boathouse pairs for any of the four relations LOCATION, POSSESSION, TEMPORALITY and CAUSATION.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Added CONTAINMENT, exemplified by orange seed and seed orange</td>
</tr>
<tr>
<td>• Added DIRECTION, exemplified by sun worship and sales target</td>
</tr>
<tr>
<td>• Renamed HYPERNYMY → TAXONOMY</td>
</tr>
<tr>
<td>• Renamed PART → MERONOMY</td>
</tr>
<tr>
<td>• Renamed PURPOSE AND PROPER FUNCTION → PURPOSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Inverted Basic and Reversed in POSSESSION, COMPOSITION, SOURCE and USAGE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Added 3-4 letter mnemonic codes (with ‘R’ appended for reversed relations)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>• TAXONOMY (née HYPERNYMY), basic and reversed: reformulated</td>
</tr>
<tr>
<td>• COORDINATION: changed to “(an) H that is also (an) M” (e.g. boy king)</td>
</tr>
<tr>
<td>• COMPOSITION (basic and reversed): replaced ‘made of’ with ‘composed of’</td>
</tr>
<tr>
<td>• SOURCE (basic and reversed): replaced ‘(made) from’ with ‘source of’</td>
</tr>
<tr>
<td>• PRODUCTION (basic and reversed): replaced ‘makes’ with ‘produces’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>• ant lion → kidney bean (SIM)</td>
</tr>
<tr>
<td>• history conference → history book (TOP)</td>
</tr>
<tr>
<td>• table leg → car motor (MER) wheelchair → motor car (MER-R)</td>
</tr>
<tr>
<td>• window seat → house music (LOC); bedroom → music hall (LOC-R)</td>
</tr>
<tr>
<td>• sheet metal → cube sugar (COMP); cf. sugar cube (COMP-R)</td>
</tr>
<tr>
<td>• sunburn → tear gas (CAUS); motion sickness → sunburn (CAUS-R)</td>
</tr>
<tr>
<td>• honey bee → song bird (PROD); beeswax → bird song (PROD-R)</td>
</tr>
<tr>
<td>• steamboat → lamp oil (USG), hand brake → oil lamp (USG-R)</td>
</tr>
</tbody>
</table>

Table 30: Summary of changes to Bourque’s system

6.3.3 The Hatcher-Bourque classification

Once Hatcher’s four-way system covering non-appositional compounds had been extended to a five-way system that also covers appositional compounds, mapping the revised set of 27 relations in Bourque2 was entirely unproblematic, confirming, once again, the robustness of Hatcher’s system. It should be noted that the relations of CONTAINMENT and DIRECTION, added in Bourque2 (page 224), was not used for coding the binominals data. As we will see when the various semantic relations have been arranged in a hierarchy (cf. Figure 80 on page 331), CONTAINMENT turns out to be the supertype for those relations that map to one of Hatcher’s two “static” relations (i.e. POSSESSION, MERONOMY, LOCATION, TEMPORALITY, CONTAINMENT, COMPOSITION and TOPIC). In the database these subtypes (and not the supertype) are used for coding the data. The relation DIRECTION, as noted above, was added on the evidence of a word that was not in my database.
The results of the mapping are shown in Table 31. The information contained in this table differs in detail from that in Bourque’s original 25-way system (cf. Table 28 on page 197) but the two systems are fully commensurate, apart from the addition of the new reversible relations CONTAINMENT and DIRECTION. However, a number of new columns have been added:

- The **B2** column contains database-friendly codes for the Bourquifier relations. These will be used extensively in the discussion in §6.4.1.
- The **Hatcher2** column shows how the 29 Bourquifier relations map to the five-way high-level system (Hatcher2) and the three-way system this gives rise to when Hatcher’s original four relations are viewed as pairs (Similarity-Containment-Causation).
- The **H2** column provides database-friendly codes for the Hatcher2 relations. These too will be used in the discussion in §6.4.2. Note how the words linking \( H \) and \( M \) – is, in and to – represent Similarity, Containment and Causation, respectively.

The relations have also been reordered in order to permit grouping by the three top-level relations Similarity, Containment and Causation and I have underlined the head constituent in the Template and Example columns, for readability. As noted earlier, the Bourquifier has been upgraded to support the new classification (Figure 54).

<table>
<thead>
<tr>
<th>Bourquifier3</th>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Similarity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>Roles</td>
<td>B2</td>
<td>H2</td>
<td>Example</td>
</tr>
<tr>
<td>TAXONOMY</td>
<td>TAX</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>sap tree</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COORD</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>bay leaf</td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>CONT</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>orange seed</td>
</tr>
<tr>
<td>MERONOMY</td>
<td>MER</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>kidney bean</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>house music</td>
</tr>
<tr>
<td>TEMPORALITY</td>
<td>TEMP</td>
<td>(symmetrical)</td>
<td>&amp; (non-reversible)</td>
<td>summer job</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COM</td>
<td>(non-reversible)</td>
<td>&amp;</td>
<td>cube sugar</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>(non-reversible)</td>
<td></td>
<td>history book</td>
</tr>
<tr>
<td><strong>Containment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| CAUSATION | CAUS | (symmetrical) | | tar 
| SOURCE | SRC | (symmetrical) | | sugar cane |
| PRODUCTION | PROD | (symmetrical) | | sugar cane |
| USAGE | USG | | | sugar cane |
| FUNCTION | FUNC | | | sugar cane |
| PURPOSE | PURP | | | sugar cane |
| **Causation** | | | | |
| CAUSATION | CAUS | (symmetrical) | | tar 
| SOURCE | SRC | (symmetrical) | | sugar cane |
| PRODUCTION | PROD | (symmetrical) | | sugar cane |
| USAGE | USG | | | sugar cane |
| FUNCTION | FUNC | | | sugar cane |
| PURPOSE | PURP | | | sugar cane |

*Figure 54: The Bourquifier v 3.0*
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Bouque2</th>
<th>B2</th>
<th>Hatcher2</th>
<th>H2</th>
<th>Template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAXONOMY</td>
<td>TAX</td>
<td>M↓H</td>
<td>Mish</td>
<td>an M is a kind of H</td>
<td>oak tree</td>
</tr>
<tr>
<td></td>
<td>TAX-R</td>
<td>M↑H</td>
<td>Mish</td>
<td>an H is a kind of M</td>
<td>bear cub</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COOR</td>
<td>M=H</td>
<td>Mish</td>
<td>an H that is also an M</td>
<td>boy king</td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>M≈H</td>
<td>Mish</td>
<td>an H that is similar to M</td>
<td>kidney bean</td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>CONT</td>
<td>M⊂H</td>
<td>MinH</td>
<td>an H that is contained in an M</td>
<td>seed orange</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POSS</td>
<td>M⊃H</td>
<td>MinM</td>
<td>an H that is possessed by an M</td>
<td>family estate</td>
</tr>
<tr>
<td>MERONOMY</td>
<td>MER</td>
<td>M⊃H</td>
<td>MinM</td>
<td>an H that is part of M</td>
<td>car motor</td>
</tr>
<tr>
<td></td>
<td>MER-R</td>
<td>M∼H</td>
<td>MinH</td>
<td>an H that an M is part of</td>
<td>motor car</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>M⊃H</td>
<td>MinM</td>
<td>an H located at/near/in an M</td>
<td>house music</td>
</tr>
<tr>
<td>TEMPORALITY</td>
<td>TEMP</td>
<td>M⊃H</td>
<td>MinM</td>
<td>an H that occurs at/during an M</td>
<td>summer job</td>
</tr>
<tr>
<td></td>
<td>TEMP-R</td>
<td>M∼H</td>
<td>MinH</td>
<td>an H at/during which M occurs</td>
<td>golf season</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>M⊃H</td>
<td>MinM</td>
<td>an H that an M is composed of</td>
<td>cube sugar</td>
</tr>
<tr>
<td></td>
<td>COMP-R</td>
<td>M∼H</td>
<td>MinH</td>
<td>an H composed of an M</td>
<td>sugar cube</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>M∼H</td>
<td>MinH</td>
<td>an H that is about an M</td>
<td>history book</td>
</tr>
<tr>
<td>DIRECTION</td>
<td>DIR</td>
<td>M→H</td>
<td>HtoM</td>
<td>an H whose goal is an M</td>
<td>sun worship</td>
</tr>
<tr>
<td></td>
<td>DIR-R</td>
<td>M←H</td>
<td>HtoM</td>
<td>an H that the goal of an M</td>
<td>sales target</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>M←H</td>
<td>HtoM</td>
<td>an H that is a source of an M</td>
<td>sugar cane</td>
</tr>
<tr>
<td></td>
<td>SRC-R</td>
<td>M→H</td>
<td>HtoM</td>
<td>an H whose source is an M</td>
<td>cane sugar</td>
</tr>
<tr>
<td>CAUSATION</td>
<td>CAUS</td>
<td>M→H</td>
<td>MtoM</td>
<td>an H that causes an M</td>
<td>tear gas</td>
</tr>
<tr>
<td></td>
<td>CAUS-R</td>
<td>M←H</td>
<td>MtoH</td>
<td>an H that an M causes</td>
<td>sunburn</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>M→H</td>
<td>MtoM</td>
<td>an H that produces an M</td>
<td>song bird</td>
</tr>
<tr>
<td></td>
<td>PROD-R</td>
<td>M←H</td>
<td>MtoH</td>
<td>an H that an M produces</td>
<td>birdsong</td>
</tr>
<tr>
<td>USAGE</td>
<td>USG</td>
<td>M→H</td>
<td>MtoH</td>
<td>an H that an M uses</td>
<td>lamp oil</td>
</tr>
<tr>
<td></td>
<td>USG-R</td>
<td>M←H</td>
<td>MtoH</td>
<td>an H that uses an M</td>
<td>oil lamp</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>M←H</td>
<td>HtoM</td>
<td>an H that serves as an M</td>
<td>buffer state</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>M←H</td>
<td>HtoM</td>
<td>an H that is intended for an M</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

Table 31: Hatcher-Bourque classification
6.4 Data analytics

In §5.6 the binominal data were analysed in terms of the morphosyntactic strategies employed by the languages of the sample, that is, the typology of nine binominal types. In the following sections I analyse the distribution of both low-level and high-level semantic relations, in terms of their overall frequency and their distribution across languages, meanings, morphosyntactic strategies, semantic types and semantic fields. It will become apparent that some relations are far more frequent than others, a fact that has the potential to provide insights into the way in which humans conceptualize the world. This is a topic which has hardly been addressed in the typological literature at all; to my knowledge, the only researcher to even approach the question from a cross-linguistic perspective is Bauer (2001; see §2.1.1), who has the following to say regarding tatpurusa (i.e. determinative) compounds:

In a detailed survey of just three languages, Bauer (1978: 147) points out that underlying semantic relationships of location appear to be the most common relationships in those languages. The same is true with the sample discussed here. Compounds in which the head is the location of the entity denoted in the modifier (e.g. English furniture store) or where the head denotes an entity located at the modifier (e.g. English bone cancer) are the types most frequently illustrated or commented on for the languages in my sample across all areas. The next most frequent type to be illustrated is the type where the head is made from the material in the modifier (e.g. English sandcastle). Other meanings are illustrated or commented on far more sporadically. While this does not show that other meanings are not also in common use, it does suggest that compounds may be used prototypically to indicate location or source (especially if ‘made from’, ‘made by’, ‘belonging to’ and ‘coming from’ are all interpreted as sources).

Under the Bourque scheme, Bauer’s three examples (furniture store, bone cancer and sandcastle) would be classified as LOC-R, LOC and COMP-R, respectively (“a store that (a) furniture is located at/near/in”, “a cancer located at/near/in a bone” and “a castle composed of sand”). The present data provide an opportunity to test Bauer’s conjecture. This is done in §6.4.1, which is concerned with the low-level relations of the Bourque scheme (represented in the database by the variable stype). Then, in §6.4.2, I investigate the high-level relations of the Hatcher scheme (represented by the variable htype).¹

¹ All the R scripts used in this dissertation are available from the Tromsø Repository of Language and Linguistics, https://dataverse.no/dataverse/trolling.
6.4.1 Low-level semantic relations

**Overall distribution across binominals**

The overall frequency of low-level semantic relations in the database is shown in Figure 55 and is summarized in the following scale:

\[(116) \quad \text{MER} \gg \text{PURP} > \text{COOR} > \text{LOC} > \text{COMP-R}, \text{POSS} > \text{USG-R} > \text{TEMP} > \ldots\]

We note that MER (i.e. part-whole), a relation that Bauer does not consider in his discussion, is far and away the most frequent relation in absolute terms, at least in the present data. This is largely due to its extreme frequency in binominals that denote body parts, as we will see below. For this reason, Figure 55 also shows the values when body parts are excluded: apart from the greatly reduced frequency of MER and a slightly reduced frequency for LOC, the differences are minimal, so while it may be the case that method employed for selecting the set of meanings used in this study overstates the overall prevalence of MER, it is clearly among the most important relations and even with body parts excluded it is just as frequent as LOC.

![Figure 55: Overall frequency of low-level semantic relations](image)

Bauer’s suggestion that the second most frequent type is where the head is made from the material in the modifier (e.g. *sandcastle*) – this equates to the Reversed COMPOSITION relation, COMP-R – is also not supported out by the data, which put it at joint fifth in terms of overall frequency. Instead, the next most frequent relation is PURPOSE (PURP), also not mentioned by Bauer. As we will see below, this relation is especially prevalent in binominals that belong to the domain *Modern*
World and/or denote entities that fall into the semantic type Advanced technology (or concept).

The third most frequent relation is COORDINATION. This might seem surprising, given that the set of meanings was expressly designed to exclude binominals whose constituents are in a coordinate relation (see §4.1). However, this applied only to what Wälchli (2005) terms ‘co-compounds’: “word-like units consisting of two or more parts which express natural coordination”, such as Hmong Daw zaub.mov [vegetable.rice] FOOD and Vietnamese bô mẹ [father mother] PARENTS. The COORDINATION relation inherited from Bourque is broader than this: it is used for binominals that refer to “combinations that, from a semantic perspective, seem to involve both elements equally” (p.180). In our binominals data, the latter type is often encountered with items that denote animates of a certain age (Hawaiian kao keiki [goat child] KID), gender (Mbyá Guaraní kavaju kunha [horse woman] MARE), or both (Ket qīm.dūl [woman.child] GIRL); cases such as these account for over 90% of binominals that exhibit the COORDINATION relation.

The LOCATION relation, in its LOC form as “(an) H located at/near/in (an) M” (e.g. TEAR < ‘eye’ + ‘water’ in several languages), is only the fourth most frequent. Together with its inverse, LOC-R, “(an) H that (an) M is located at/near/in” (e.g. Hupdē y̥ɔ̃h mɔy [medicine house] HOSPITAL), it is found in 428 binominals, i.e. 12% of the data. Thus Bauer’s suggestion that this is the most common kind of relation appears to be unfounded.

**Distribution across languages**

We can also look at relations in terms of the number of languages in which each relation is attested (Figure 56).
The typology and semantics of binominal lexemes

Figure 56: Number of languages that exhibit a particular relation

The frequency scale here is:

\[(117)\quad \text{MER} > \text{LOC} > \text{COOR}, \text{POSS}, \text{PURP} > \text{COMP-R} > \text{USG-R}, \text{LOC-R}, \text{PROD} > \ldots\]

The same six relations predominate in both distributions, albeit with slightly different rankings. Note that one relation (COMP, “(an) H that (an) M is composed of”, e.g. cube sugar) is not attested at all in the database, and that a further four (CAUS, SRC-R, TAX-R and TEMP-R) are rare.

**Distribution across meanings**

The distribution across meanings (Figure 57) shows a generally similar scale, but now with the TAX relation displaying far greater prominence. USG now appears among the top six, with COMP-R and POSS relegated to joint 9th and 11th place:

\[(118)\quad \text{MER} > \text{COOR}, \text{PURP}, \text{TAX}, \text{LOC} > \text{USG}, \text{SIM}, \text{USG-R}, \text{POSS}, \text{LOC-R}, \text{COMP-R} \ldots\]
Figure 57: Number of meanings that exhibit a particular relation

This suggests that while TAX (“(an) M is a kind of H”, e.g. oak tree) is not especially common, it is rather versatile; conversely, while COMP-R and POSS are rather frequent, their scope of application is relatively limited. It is also worth noting that of the 45 binominals that exhibit the TAX relationship, 18 employ the der strategy. In many cases the gloss indicates an (apparently redundant) nominalizer or diminutive affixed to a root whose meaning is the same as that of the derived form, as in Lithuanian spen.elis [nipple.DIM] NIPPLE OR TEAT.

Overall, the data indicate that the most frequent low-level semantic relations cross-linguistically, at least as far as binominal lexemes are concerned, are the following:

<table>
<thead>
<tr>
<th>MER</th>
<th>977</th>
<th>(an) H that is part of (an) M</th>
<th>table leg</th>
<th>HinM</th>
<th>M ⊳ H</th>
</tr>
</thead>
<tbody>
<tr>
<td>PURP</td>
<td>540</td>
<td>(an) H intended for (an) M</td>
<td>animal doctor</td>
<td>HtoM</td>
<td>M← H</td>
</tr>
<tr>
<td>COOR</td>
<td>358</td>
<td>(an) H that is also (an) M</td>
<td>boy king</td>
<td>MisH</td>
<td>M ≈ H</td>
</tr>
<tr>
<td>LOC</td>
<td>321</td>
<td>an H that (an) M is located at/near/in</td>
<td>music hall</td>
<td>HinM</td>
<td>M ⊳ H</td>
</tr>
<tr>
<td>COMP-R</td>
<td>216</td>
<td>(an) H composed of (an) M</td>
<td>sugar cube</td>
<td>MínH</td>
<td>M ⊂ H</td>
</tr>
<tr>
<td>POSS</td>
<td>211</td>
<td>(an) H that (an) M possesses</td>
<td>family estate</td>
<td>HinM</td>
<td>M ⊳ H</td>
</tr>
</tbody>
</table>

Table 32: Most frequent low-level semantic relations

Distribution across morphosyntactic strategies

Figure 58 shows how many morphosyntactic strategies are used to express each kind of relation. Comparison with the overall frequency scale extracted from Figure 55 (above) shows that the most frequent relations can be expressed by any one of the nine binominal types; this suggests that – at this level of analysis – there is no
overall correlation between morphosyntactic strategy and semantic relation. As the data become more sparse, the number of strategies are reduced; thus, at the lower end of the scale, we find TAX-R, TEMP-R and SRC-R, each of which is expressed by just two strategies. However, since each of these three relations is represented in the database by just two or three exemplars, this does not constitute evidence against the lack of overall correlation.

But even if there is no overall correlation, there are some interesting tendencies to be observed in terms of which strategies are favoured for the different semantic relations. Figure 59 shows how the six most frequent relations (MER, PURP, COOR, LOC, POSS and COMP-R) – those that account for 5% or more of the data – distribute across the nine binominal types.

The first point to notice is that the fingerprints for five of the nine strategies (jxt, cmp, prp, gen and con) are remarkably similar and generally mirror the overall distribution of the six relations, i.e. MER >> PURP > COOR > LOC > COMP-R, POSS (cf. 116). These five thus appear to be more-or-less equally adept at expressing any relation. On the other hand, the fingerprints for der, adj, cls and dbl show some divergence:

- **der** frequently encodes the COOR relation and rarely encodes MER
- **adj** is less used for MER and more for PURP and COMP-R
- **cls** favours COOR and, in particular, MER, but disprefers PURP
- **dbl** also favours MER and disprefers PURP and COOR

These discrepancies will be discussed further below in the next section.
Distribution across semantic types

The frequency of different relations varies according to the semantic type of the referent. Figure 60 shows the proportional distribution of the six most common relations – MER, PURP, COOR, LOC, COMP-R and POSS – across the seven semantic types (cf. Figure 17b on page 85 and Table 18 on page 78). The results for Animal, Natural phenomenon and Location should be approached with caution, since these semantic types represent only 7, 5 and 12 of the 100 meanings, respectively, but the differences between the other four types are striking.

As noted above, in binominals that denote Body parts the MER relation accounts for 85% of the data; the only significant alternative is LOC, which is the preferred
The typology and semantics of binominal lexemes

relation for naming bodily substances, such as EARWAX and TEAR. On the other hand, MER is rarely used to denote an Advanced technology (or concept), such as BICYCLE PUMP, KEYWORD or RAILWAY; instead, the PURP relation predominates, accounting for over 80% of the data, with COMP-R the most widely used alternative (as in many words for RAILWAY, which is often conceptualised as a road composed of iron). In short, there is a strong tendency to name (secondary) body parts/fluids in terms of the (primary) body parts they are a part of/located at, and to name advanced concepts in terms of either their intended function or the material they are made of.

The semantic type Basic technology (or concept) is more mixed: as with Advanced technology (or concept), PURP and COMP-R are the most widespread relations, but the two are now equally common; however, in contradistinction to the latter type, MER and LOC are also quite frequent. These are also the most widely used relations for Natural phenomena – together with POSS, which expresses the relation between a spider and its web, or bees and their hive, and phenomena viewed as belonging to supernatural beings, such as Ket Albara kàn ‘Alba’s hunting trail’ MILKY WAY and Assamese ramdhenu ‘Lord Rama’s bow’ RAINBOW.

**Distribution across semantic fields**

![Graph showing distribution of low-level relations across semantic fields](image)

*Figure 61: How low-level relations vary across semantic fields*
A similar variation is found across semantic fields (see page 77). Figure 61 shows the frequency of the six most common semantic relations across the nine most frequent semantic fields. We note again that MER plays the dominant role in The body, but also in Agriculture and vegetation and Food and drink; and, as expected, Modern world is dominated by the PURP relation. We see also that the patterning in Animals and Kinship is remarkably similar: binominals in these fields have an overwhelming preference for either COOR or POSS. The latter is also widely used in Social and political relations. Finally, the LOC relation that Bauer assumed to be most widespread is in fact largely confined to the fields of The physical world and Clothing and grooming.

6.4.2 High-level semantic relations

We turn now from the low-level semantic relations of the Bourque scheme to the high-level relations of the (revised) Hatcher scheme. For ease of reference, Table 33 provides a summary of the five high-level relations and the 27 low-level relations that map to them (recall that CONT and CONT-R are not used in the database). The terms used in the column ‘role’ will sometimes be used in the following in order to simplify the discussion. The importance of roles will become clear when the discussion of semantic relations is broadened in §9.3.

<table>
<thead>
<tr>
<th>htype</th>
<th>stypes</th>
<th>role</th>
</tr>
</thead>
<tbody>
<tr>
<td>MisH</td>
<td>$M \cong H$</td>
<td>TAX, TAX-R, COOR, SIM</td>
</tr>
<tr>
<td>HinM</td>
<td>$M \supset H$</td>
<td>CONT, POSS, MER, LOC, TEMP, COMP, TOP</td>
</tr>
<tr>
<td>HtoM</td>
<td>$M \leftarrow H$</td>
<td>SRC, CAUS, PROD, USG, FUNC, PURP</td>
</tr>
<tr>
<td>MtoH</td>
<td>$M \rightarrow H$</td>
<td>SRC-R, CAUS-R, PROD-R, USG-R</td>
</tr>
</tbody>
</table>

Table 33: Summary of high- and low-level semantic relations

The first four plots in the previous section showed how the low-level relations distribute across the database as a whole (with and without body parts), and across languages, meanings and morphosyntactic strategies. Figure 62 provides similar information for the high-level relations. Predictably, the amount of information is considerably reduced; on the other hand, the categories are considerably more balanced and therefore more amenable to statistical analysis.
The typology and semantics of binominal lexemes

Figure 62: Binominals, languages, meanings and strategies by htype

Plot (a) shows that the high-level relation $HinM$ (Hatcher’s “B is contained in A”)\(^1\) accounts for nearly half of the data (a). This comes as no surprise, given that this relation subsumes MER; if body parts are excluded it has approximately the same frequency as $HtoM$ (Hatcher’s “A is the destination of B”), which subsumes the PURP relation, among others. With body parts included, the overall scale is

\[(119)\]  
$HinM >> HtoM > MisH, MinH > MtoH$

and with body parts excluded, the scale is

\[(120)\]  
$HinM, HtoM > MisH, MinH > MtoH$

The two most frequent low-level relations ($HinM$ and $HtoM$) account for two-thirds of the data and thus suggest a pronounced tendency for a non-basic meaning to be conceptualized in terms of either its container or its goal (i.e. destination) – both of which should be interpreted in Hatcher’s very broad sense.

Plot (b) tells us that $HinM$ is ubiquitous, occurring in every language in the sample. However, the other four low-level relations are also widespread across languages and they are probably also ubiquitous. The fact that they are not attested in every language is almost certainly due to the paucity of data for some languages: it would be highly unlikely that a language that is represented by fewer than, say, 10 data

---

\(^1\) Recall from §6.2.2 that Hatcher uses the letters A and B to denote the first and second constituents of a typical right-headed English compound, so A is the modifier (M) and B is the head (H). To avoid confusion, think of Hatcher’s letters as denoting the Attribute (modifier) and Base (head).
points – of which there are five (Gurindji, Puyuma, Selice Romani, Datooga and Tuwari) – would exhibit all five high-level relations.

The distribution of relations across meanings (Plot c) shows a scale similar to the two preceding ones,

\[(121) \quad \text{HinM} > \text{HtoM} > \text{MisH} > \text{MinH} > \text{MtoH}\]

but the values more spread out: HinM is less dominant, while MisH, the similarity-based relation added to Hatcher’s original four (see page 206ff), is higher up the scale (in the sense that it is significantly more widespread across meanings than MinH). This reflects what was referred to as the versatility of the TAX relation, shown in (118), above. More worthy of mention, though, is the fact that none of the high-level relations appears suited for conceptualizing anything like the full range of meanings. Even HinM, which is found in every language and accounts for over 45% of all binominals in the database, is used with only just over half of the 100 meanings: in other words, there are limits to the versality of conceptualizations that are based on how an entity is (in the broadest sense) “contained”.

Finally, plot (d) shows that every one of the nine morphosyntactic strategies is attested in the data as expressing each of the five high-level relations; this provides additional evidence that there is no overall, cross-linguistic correlation between morphosyntactic strategies and semantic relations. However, the differing behaviour of der, adj, cls and dbl revealed with respect to low-level relations in Figure 59 is also apparent when viewed from the perspective of the high-level relations (Figure 63).

![Figure 63: High-level semantic relations across binominal types](image-url)
We observe that \textit{cls} and \textit{der} are largely appositional (\textit{HisM}); this is due to the classificatory nature of the former and to the tendency for derivational affixes to denote a broad class of referents such as males, females, diminutives, etc.

The fact that \textit{dbl} is used overwhelmingly for \textit{HinM} seems to be due to two factors. The first is that over 40\% of the \textit{dbl} data comes from Akkadian, a language which is unrepresentative in two senses: (i) the COOR relation (\textit{HisM}) is hardly found in the data (this seems to be because gender alternations have not been coded by the contributor, e.g. \textit{šarratum} QUEEN is given as monomorphemic even though it is clearly derived from \textit{šarrum} KING, cf. Worthington 2010: 31); (ii) similarly, the PURP relation (\textit{HtoM}), which is overwhelmingly found in words denoting concepts in the domain Modern world, is (unsurprisingly) rare in this ancient language. The second factor is that the other languages in which the \textit{dbl} strategy is preferred (for example Oroqen, Seri and Takia) tend to borrow such words when they need them.

The last strategy to be considered here is \textit{adj}, which shows unusually low frequencies for \textit{MisH} and \textit{HinM} and correspondingly higher frequencies for \textit{MinH} and \textit{HtoM}. This is accounted for by the fact that the \textit{adj} strategy is always involved in competition with other strategies (usually \textit{cmp} or \textit{gen}) and tends to be specialized for contents, source and goal relations rather than similarity and container relations.

Thus, while it is possible to observe some weak tendencies at the cross-linguistic level, the overall picture is one in which there are no strong correlations between morphosyntactic strategy and semantic relations. This does not, however, mean that there are no such correlations at the level of individual languages. That issue will be explored with a more fine-grained analysis in §7.3.

Turning to semantic types, Figure 64 shows clearly that containment is central to the conceptualization of (secondary) Body parts and also important for concepts that express Location or denote Basic technologies (or concepts) or entities in the Natural world. On the other hand, it is marginal to the conceptualization of Persons and of almost no use when it comes to Animals and Advanced technologies (or concepts). With the semantic types Animal and Person (and only those) the similarity-based \textit{HisM} relation is most important, whereas conceptualizations that are goal-oriented – indicated by the \textit{HtoM} relation – are most frequent with Advanced technologies (and concepts), but also encountered with other semantic types (albeit only rarely with Body parts and Natural phenomena).
Conceptualization of an entity in terms of its contents (MinH) is considerably less common than the inverse and never the dominant form; it is found most often with semantic types that denote Basic and Advanced technologies (and concepts) and Locations, rarely with Body parts and never with Animals. As for source-based conceptualizations, they are mostly found with Persons (in particular, professions), Natural phenomena, and Advanced technologies (and concepts).
Similar patterns emerge with respect to semantic fields (Figure 65, the high-level equivalent of Figure 61). Whereas the low-level plot highlights similarities between Animals and Kinship, the new one reveals additional commonalities, in particular between The body, The physical world and Food and drink. In all of these, Hatcher’s ⊗ predominates: there is a tendency for conceptualizations where (to quote Hatcher) the target concept, B, “is somehow, to some extent, contained, comprehended in” the modifying concept, A.

In sum, and referring back to the notion of roles introduced in Table 33, we see that containment (HinM) is particularly important for The body, Food and drink and The physical world; goal-oriented conceptualization (HtoM) for the Modern world; similarity (MisH) for Kinship and Animals; contents (MinH) for both Clothing and grooming and Social and political relations; and source-orientation (MtoH) for Agriculture and vegetation.

6.5 Chapter summary

In this chapter I started out by providing a brief overview of previous studies on semantic relations in compounding. I then showed how Bauer and Tarasova’s work, together with that of Janda, provides grounds to believe that the same relations apply to binominal lexemes in general.

I described in some detail the systems developed by Bourque and Hatcher that I have harnessed in the present study. Bourque’s classification was revised and extended with two new relations, CONTAINMENT and DIRECTION, for a total of 29 relations, 12 reversible and five unidirectional. Hatcher’s classification was also revised by the addition of a fifth high-level relation, SIMILARITY, in order to cover appositional and coordinate compounds. The two revised classifications were then unified to create the two-tiered Hatcher-Bourque classification, and a tool called the Bourquifier was developed to assist in the slippery task of classifying individual binominals. Both the Hatcher-Bourque classification and the Bourquifier app are offered to the research community in order to promote collaboration in the field of semantic relations.

The analysis of the data provides insights into the ways in which humans tend to conceptualize the world. It suggests, contra Bauer, that MERONOMY and PURPOSE are far more widespread, and thus more important, than the LOCATION relation. Of the two types of MERONOMY – Basic (MER) and Reversed (MER-R) – the former is far more frequent than the latter, which indicates that the conceptualization of a complex meaning is much more likely to involve modification by the whole (or,
more generally, the container) than modification by the parts (or, more generally, the contents). The Basic MERONYM relation (MER) occurs most frequently with body parts and in the semantic field of agriculture and vegetation. It can express about one third of the 100 meanings used in this survey; it is found in all 106 languages of the sample; and it can be expressed using any one of the nine nominal modification strategies.

Bauer’s suggestion that the next most frequent type is where the head is made from the material in the modifier is also not supported out by the data: both PURPOSE and COORDINATION are much more common than COMPOSITION. PURPOSE is most often encountered in the semantic field Modern world to denote advanced technological concepts; it only occurs in 89 of the 106 languages, no doubt because some of the languages in the sample do not have words for concepts of that kind; significantly, the only morphosyntactic strategy that does not occur with this relation is the classifier strategy, cls, but this also the most sparsely populated of all strategies. COORDINATION is used primarily to denote animates of a certain age, gender or both; it is therefore not surprising that it occurs mostly in the domains of kinship, animals, agriculture and vegetation. Cases such as these account for over 90% of binominals that exhibit this relation, and once again, every morphosyntactic strategy is attested in the data.

The Basic LOCATION relation is the fourth most frequent type overall and occurs three times as often as the Reversed relation; in other words, it is more usual to conceptualize an object in terms of where it is located than what is located at, near or in it. It is found in almost all of the languages of the sample (97 out of 106) and can be expressed by any of the nine strategies. It is most often encountered in the fields of the natural world and basic technologies and concepts.

The other frequent relations are Basic POSSESSION and Reversed COMPOSITION, but the Reversed form of POSSESSION is uncommon, and the Basic form of COMPOSITION does not occur in the data at all. The range of meanings theses relations can express is limited: only 12% in each case. Nevertheless, they can be expressed by any morphosyntactic strategy.

Apart from COMP, every one of the 25 relations was found in the data, but some were very rare, in particular CAUS, SRC-R, TEMP-R, TAX-R. While these are clearly peripheral in noun-noun compounds and their functional equivalents, they may be more common in other types of compounds, for example those in which the head or the modifier is an action-morph rather than a thing-morph.
The data for the low-level relations suggests that there is no overall correlation between morphosyntactic strategy and semantic relation: many relations can be expressed by every strategy, most are expressed by almost every strategy, and those that are expressed by just a few strategies are those where the data is sparse. This impression is confirmed by the analysis of high-level relations: every one of the five relations of Hatcher2 are attested with every one of the nine strategies, so we can state quite categorically that there is no such overall correlation. It is thus not the case that some strategies are used to express some relations, while other strategies are used for other relations.

However, while this applies cross-linguistically, it does not mean that there are no such correlations within individual languages. In fact, the opposite is the case: As I showed in Pepper (2010), the Cameroonian language Nizaa (ISO 639: sgi) uses left-headed and right-headed compounds for two distinct sets of relations. Zúñiga (2014) reports something similar for Mapudungun (ISO 639: arn), as does Atoyebi (2010) for Oko (ISO 639: oks). Bourque himself (p. 253) compares N N and N à N binominals in French and shows that the two constructions have very different profiles (for example, PURPOSE and USE account for 47.5% of all French N à N binominals in his database, but only 12.7% of his N N binominals). This would be a fruitful topic for further research. I return to Nizaa and Mapudungun repeatedly in the next two chapters, which discuss typological and conceptual generalizations, respectively.
Chapter 7. Typological generalizations

7 Typological generalizations

The analyses presented in the preceding two chapters resulted in a number of cross-linguistic generalizations concerning morphosyntactic strategies and semantic relations. In this chapter I investigate three typological issues in more detail. §7.1 concerns word order typology; §7.2 looks at the relationship between binominals and possessive constructions (or more precisely, nominal modifier constructions); and in §7.3 I explore possible interactions between morphosyntactic strategies and semantic relations.

7.1 Word order typology

Word order – or more properly, constituent order – has been a staple of modern typological research ever since the foundational work of Greenberg. Clearly, it is also important in the study of binominals, given that the latter consist of a head and a modifier and can thus exhibit either head-initial or head-final order. In the following I discuss constituent order in general terms (§7.1.1) and then in terms of consistency across binominal types (§7.1.2). Finally I look in more detail at some languages that do not exhibit any basic order (§7.1.3).

7.1.1 Constituent order

The order of constituents in compounds was first studied by Bauer (2001), who found “a slight preference” for head-final structures across compounds in general. Bauer does not present the actual numbers in his paper; however, his Table 51.2 (reproduced as Table 3 on page 25), in which head-modifier order in compounds is compared to the order of noun and adjective, allows us to glean that at least 20 of the 36 languages in his study were judged to be right-headed and at least 11 left-headed. For the remaining five languages there was “insufficient data”, but it is not clear whether the lack of data applies to compounds or adjectives. In my replication of Bauer’s study (see page 25), in which I only considered noun-noun compounds, I found a 20-10 split; three languages exhibited mixed order (Cantonese, Kanuri and Tzutujil), and a further three (Mara, Hixkaryana and West Greenlandic) did not have appear to have compounds.
Quite different results are reported from the Morbo/Comp project (cf. §2.1.3 and §2.1.5), which like Bauer’s study covered the whole gamut of compounding, not just noun-noun compounds. The two principal sources give somewhat different sets of numbers (Table 34), but they agree in reporting much higher frequencies (roughly 6:1 and 10:1) for right-headed compounds than the 2:1 ratio found by Bauer and myself.

<table>
<thead>
<tr>
<th>Source</th>
<th>RH</th>
<th>LH</th>
<th>none</th>
<th>both</th>
<th>RH:LH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guevara &amp; Scalise (2009)</td>
<td>55.89</td>
<td>9.72</td>
<td>22.05</td>
<td>6.26</td>
<td><strong>5.75</strong></td>
</tr>
<tr>
<td>Scalise &amp; Fábregas (2010)</td>
<td>66.7</td>
<td>6.8</td>
<td>16.3</td>
<td>5.9</td>
<td><strong>9.81</strong></td>
</tr>
</tbody>
</table>

Table 34: Constituent order pace Morbo/Comp

Turning to the binominal data of the present study, 2,426 of the 3,738 binominals are right-headed and 1,312 are left-headed. These data are, of course, not strictly comparable with those of Bauer and Morbo/Comp since, on the one hand, they only include compounds consisting of two nouns and, on the other hand, they also include functional equivalents of the latter. That being said, it is striking that the ratio of right-headed to left-headed binominals across the database as a whole (1.9:1) is very close to the 2:1 figure obtained by Bauer and myself.

It might seem appropriate to compare this ratio to that between the order of object and verb. In the WALS data (Dryer 2013c) the ratio of OV to VO is almost exactly 1:1; the object preceding the verb in 713 languages (OV) and following it in 705 languages (VO). There is thus a discrepancy between the almost 2:1 ratio obtained for binominals and Dryer’s 1:1 ratio. However this should not be accorded too much weight, for two reasons. The first is that Dryer does not claim that his sample of 1,519 languages is balanced, and nor is my own of 106. In my sample there are 56 VO languages and 40 OV languages (in addition to seven languages with no dominant order and three for which no information is available), giving a ratio of 7:5 in favour of VO. The higher frequency of the latter in my sample as compared to Dryer’s is mostly due to the number of Indo-European languages, of which there are 24: 17 VO and 5 OV (plus one each for no dominant order and no available information). Secondly, the data points on which the 2:1 ratio is based represent binominals, not languages, so we are in any case comparing apples and oranges (in addition to which the distribution of binominals across the languages of the sample is very uneven (see the discussion of Binominal frequency on page 124). In order to compare like with like we would need to have a single data point that expresses the basic constituent order in binominals for each language (with a value of either
left, right, mixed or NA). However, as we noted in §5.6.2, most languages – in fact, all but nine of the 106 in our sample – employ more than one binominal strategy (cf. Figure 41). For example, the two principal strategies in Greek are gen and cmp; the former is consistently left-headed (e.g. *istos araxni.s* [web spider.GEN] SPIDER WEB) and the latter is equally consistently right-headed (e.g. *sidıır.o.δromos* [iron.LE.road] RAILWAY). As a result, it is impossible to determine a basic order for many languages and consequently the exercise of comparing constituent order in binominals with that of object and verb has little point.

![Figure 66: Constituent order by binominal type](image)

More interesting facts are revealed by breaking the numbers for constituent order down by morphosyntactic strategy, as shown in Figure 66. We start by observing that 81% of cmp binominals are right-headed, while 19% are left-headed, a ratio of more than 4:1. Again, these figures cannot strictly be comparable to those of Bauer and Morbo/Comp, for two reasons: first of all, they do not include compounds of types other than N+N (e.g., N+A, synthetic, verbal, coordinate, etc.); secondly, ‘compounds’ consisting of two words are excluded (e.g. *handbag* is included but *hand brake* is not). The second issue can be remedied by combining the two compounding strategies, cmp and jxt. The latter, as the plot shows, are almost equally divided in terms of head position. The difference between cmp and jxt is quite striking in this regard, and justifies analytical decision to separate them, even though there is no clear cross-linguistic definition of wordhood (see §5.1.2); it would seem that two nouns have a greater tendency to fuse when the head is on the right.
Combining cmp and jxt reduces the ratio of right- to left-headed compounds to slightly over 2:1 (69% vs. 31%). The new numbers tend to support Bauer as against Morbo/Comp, and suggest that the latter’s inflated view of the prevalence of right-headedness is probably due to areal and genetic bias (recall from page 34 that 21 of the 25 languages in Morbo/Comp’s sample are spoken in Europe, 17 are Indo-European and all but one, Mandarin, belong to the same large linguistic area, Eurasia). This is confirmed in the breakdown by geographical area (Figure 67) which shows right-headed compounding to be more frequent in Eurasia than in other parts of the world. Lest it be thought that the preponderance of right-headed compounds in Eurasia is due to the large number of Indo-European languages in the sample (cf. Table 20, page 91), it turns out that excluding IE languages results in an increase of the percentage of right-headed compounds in Eurasia, from 90% to 99%.

![Figure 67: Constituent order in compounds by area](image)

The other numbers in Figure 66 should be viewed with caution because of the relatively low proportion of types other than cmp and jxt in the database. In particular, the preponderance of right-headed cls binominals is based on evidence from just three Amazonian languages (Harakmbut, Murui Huitoto and Trinitario); these account for 27 of the 37 instances of cls, while the 10 (left-headed) instances of cls are from Āiwoo, Bandial and Swahili; clearly the sample is too unbalanced and the data too sparse to make any useful generalizations. The values for der demonstrate the overwhelming preponderance of suffixation as opposed to prefixation in the binominal lexicon. Prefixes that denote things are few and far between: ten of the 23 binominals containing preposed thing-affixes are found in Malagasy, and most of these consist of the agentive prefix mpaN-, as in mpan. jòno [AGT.bait] FISHERMAN.
The preference for right-headedness in \textit{gen} and \textit{adj} constructions, both of which involve dependent-marking, approaches that of \textit{cmp}; on the other hand there is a slight preference for left-headedness in \textit{con} and \textit{dbl}, both of which involve head-marking (the latter in addition to dependent-marking, of course).

The figures for \textit{prp} are as striking as those for \textit{der}: 93\% are left-headed and only 7\% right-headed. As is well-known, prepositions correlate very highly with head-initial word order and postpositions with head-final word order, so the figures reflect an imbalance between prepositional and postpositional constructions in the data. That imbalance is partly due to the lack of balance in the sample; the three Romance languages, French, Italian and Romanian, together yield almost 30\% of all \textit{prp} binominals and these are, of course, all left-headed (as in \textit{chemin de fer}).

However, that is not the whole story. A closer inspection of the data shows that the only language with more than one or two right-headed \textit{prp} binominals is Hindi; an example is given in (75a) on page 167. On the other hand, Nepali, a very close relative of Hindi, has no binominals of this type at all; instead it has 6 instances of \textit{gen} (cf. 75b). The explanation is diachronic: the postposition found in the common ancestor of the two languages has stayed as such in Hindi but developed through grammaticalization into a case marker in Nepali. This leads to an explanation for the paucity of right-headed \textit{prp} binominals as follows:

Adpositions usually occupy the position between the head and the modifier, so the two basic patterns are left-headed \textit{Head [PREP Mod]} and right-headed \textit{[Mod POSTP] Head}. Now, it is much more common for a particle that follows its co-constituent (as \textit{POSTP} follows \textit{Mod}) to fuse with the latter than for one that precedes it (as \textit{PREP} precedes \textit{Mod}); this is the well-known “suffixing preference” (Sapir 1921: 67; Bybee, Pagliuca & Perkins 1990). A postposition therefore develops more easily into a postposed case affix than a preposition does into a preposed case affix. As a result there are many more left-headed binominals of type \textit{prp}\(^1\) than there are left-headed.

An argument could be made for combining \textit{prp} and \textit{gen} into a single type (which could be labelled \textit{flg} on the basis of Haspelmath’s flag/index terminology, or \textit{rel} on the basis of Croft’s relational/indexical dichotomy, cf. §5.1.2), but this would obscure the oppositions found in languages like Polish (§5.6.2). It would also give Irish a binominal fingerprint very similar to that of the Romance languages (§5.6.3).

\(^1\) Recall from §5.3.2 that the \textit{prp} type, its label notwithstanding, represents adpositional strategies in general, not just the prepositional strategy.
Figure 69: Distribution of right-headed and left-headed binominals
7.1.2 Consistency

The degree to which languages are consistent with respect to constituent order can be gauged from Figure 69. Here each language is represented by one or two vertical bars that indicate the percentage of left- and right-headed binominals respectively. At the left-hand edge of the plot are languages that are consistently left-headed, and at the right-hand edge those that are consistently right-headed. Languages at the two extremes are represented by a single bar. Languages in the middle, where the colours overlap, exhibit both right- and left-headed binominals; I will use the term ‘inconsistent’ for these.¹

Generally speaking, in languages that have both right- and left-headed binominals there is more than one kind of binominal construction, and each such construction is consistently either left-headed or right-headed. For example, the most common types of binominal in Modern Greek are cmp and gen; the former are consistently right-headed (cmpR), as in siðir.o.ðromos [iron.LE.road] RAILWAY, and the latter are consistently left-headed (genL), as in istos araxni.s [web spider.GEN] SPIDER WEB. However, there are a few rare cases of languages in which the same strategy results in both head-initial and head-final order. This is the case with both Vietnamese Compounds (122) and Polish Adjectival constructions (123). In the case of Vietnamese, this is the result of borrowing: native compounds are left-headed, whereas loans from Chinese are right-headed (Nguyễn 1997: 72, 77). The situation in Polish has traditionally been explained by the difference between qualifying and classifying adjectives (Cetnarowska 2014).

(122) Vietnamese

a. cmpL  Head   Mod  xe liía [vehicle fire] TRAIN  
b. cmpR  Mod   Head   hoá xa [fire vehicle] TRAIN (from Chinese)

(123) Polish

a. adjL  Head Mod.ADJZ droga mlecz.na [road milk.ADJZ] MILKY WAY  
b. adjR  Mod.ADJZ Head kamien.ny most [stone.ADJZ bridge] STONE BRIDGE

Some situations in which a language is very nearly but not quite consistent are a result of the analytical decision I took during the annotation of the data to use the ISA test to determine the head of a binominal (§4.2). As a consequence, when

¹ Note that since the purpose of the diagram is to give an impression of the overall trend, the individual language names are not of crucial importance; they are included as a convenience for readers who can zoom using the electronic version of this work.
MIDDAY is denoted through a combination of MIDDLE and DAY, it is the former that must be regarded as the head (cf. the discussion on page 112). In many cases, the result of such an analysis conforms to the patterns otherwise found in the language. For example in Irish, where the dominant pattern is genL, it is meán lae [middle day:GEN]; in Indonesian, where almost all binominals are of type cmpL, the word for MIDDAY is tengah hari [middle day]; and in Yakut, kün orto.to [day middle.3SG] conforms to the majority conR pattern. But in the Germanic and Finnic languages, which are robustly right-headed, it is the first constituent that carries the meaning MIDDLE, as in Dutch mid.dag, and Estonian kesk.päev. In order to maintain analytical consistency, such words forms be classed as left-headed. These anomalies are paralleled in the database by three words with the meaning WEDNESDAY: Ger. mitt.woch, OHG mitta.wehha and Fin. keski.viikko, all glossed as [middle.week]; outside the database we find more such words that denote the midpoint of a time period (e.g. Norwegian midsommer, English midweek, etc.) and occasionally the middle of something else (e.g. English midship). The explanation for these exceptions (if they are exceptions and not analytical errors) is likely to be found in diachrony, but that is beyond the scope of the present work. In the present context, the point is that, had it not been for such “anomalies”, many more languages would have shown complete consistency with respect to constituent order and the right-hand tail of the middle section of Figure 69 would be much less pronounced.

Inconsistent ordering is found in 66 of the 106 languages in the sample, but it is somewhat peripheral in many languages, as witness the size of the tails (especially the right-hand tail) of the middle section of the plot. For languages where it is not peripheral, it can be of interest to explore the possible causes of inconsistency in more detail, in case this leads to the discovery of phenomena similar to that in Nizaa, described in §1.1.2. This requires us to investigate individual languages. While this might be counter to traditional Greenbergian typology, since we are no longer looking for generalizations of constructions across languages, the insights we gain could feed back into more typologically oriented research projects.

In order to investigate what is going on when inconsistency is not peripheral, we can set a threshold for the number of binominals of each type (left-headed and right-headed) that we require in order for a language to qualify for closer inspection. A threshold of seven results in the exclusion of languages in the two tails of Figure 69 and produces a manageable list of 15 languages in which inconsistent ordering is a significant feature. These are listed in Table 35, together with the binominal types that they exhibit (here using the 18-way classification that includes the head position parameter), and with types grouped by head position.
For some languages the threshold of seven is only achieved for right-headed items because of the contribution of der binominals. Now, we know from Figure 37 (page 170) that the der type is relatively common; from Figure 38 we see that it is fairly widespread; and Figure 43 shows that it is attested in well over half of the languages in the sample. Moreover, Figure 66 tells us that 95% of such binominals are right-headed. That the derR type should often be the cause of inconsistent ordering should therefore come as no surprise. It is therefore particularly instructive to see which languages are inconsistent if we ignore der, and that is the significance of the Boolean value in the final column. In the top six languages, inconsistency is found even if der is ignored, whereas in the bottom nine the inconsistency is due to the presence of right-headed der constructions along with one or more left-headed, non-derivational types. The latter would seem to be somewhat less interesting, so I will have no more to say about them here.

Of potentially greater interest are the six cases in which derivation does not play a role. Of these, two (Vietnamese and Mapudungun) show an opposition between jxtL and jxtR; in Greek the opposition is mainly between genL and cmpR, as we saw earlier; in Barain it is between prpL and conR. In Kalamang it is between jxtL, cmpR and conR. In Trinitario the opposition is less clear-cut: most right-headed binominals are of type cls, but left-headed binominals are a mix of cmp, con and jxt. In the next section I will briefly discuss the four cases in which the

<table>
<thead>
<tr>
<th>language</th>
<th>left-headed</th>
<th>right-headed</th>
<th>ignore derR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barain</td>
<td>dbL, jxtL, prpL</td>
<td>conR, prpR</td>
<td>TRUE</td>
</tr>
<tr>
<td>Greek</td>
<td>genL, jxtL</td>
<td>adjR, cmpR, derR</td>
<td>TRUE</td>
</tr>
<tr>
<td>Kalamang</td>
<td>adjL, genL, jxtL, prpL</td>
<td>cmpR, conR, jxtR</td>
<td>TRUE</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>jxtL</td>
<td>cmpR, derR, genR, jxtR</td>
<td>TRUE</td>
</tr>
<tr>
<td>Trinitario</td>
<td>cmpL, conL, jxtL</td>
<td>clsR, cmpR, derR</td>
<td>TRUE</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>jxtL</td>
<td>jxtR</td>
<td>TRUE</td>
</tr>
<tr>
<td>French</td>
<td>adjL, jxtL, prpL</td>
<td>cmpR, derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Gawwada</td>
<td>genL</td>
<td>derR, jxtR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Irish</td>
<td>genL</td>
<td>cmpR, derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Italian</td>
<td>adjL, cmpL, jxtL, prpL</td>
<td>cmpR, derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Kanuri</td>
<td>adjL, cmpL, genL</td>
<td>derR, jxtR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Polish</td>
<td>adjL, cmpL, genL, prpL</td>
<td>adjR, cmpR, derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Romanian</td>
<td>adjL, dbL, jxtL, prpL</td>
<td>derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Welsh</td>
<td>adjL, cmpL, jxtL, prpL</td>
<td>cmpR, derR</td>
<td>FALSE</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>adjL, cmpL, conL, dbL, jxtL</td>
<td>adjR, cmpR, derR, jxtR</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

Table 35: Languages with significant degree of mixed order
opposition involves just two types, that is, Vietnamese, Mapudungun, Greek and Barain. The case of Kalamang will be investigated in more depth in §8.1.

7.1.3 Mixed languages

Two of the four languages – Vietnamese and Mapudungun – have both left-headed and right-headed compounds: both exhibit two subtypes of the jxt strategy, jxtL and jxtR. They are thus cases of inconsistency in compounding, a phenomenon that has not been investigated cross-linguistically but appears to be rare. In any given language, compounds tend to be either left-headed or right-headed, but there are exceptions. Examples cited in the literature include Mandarin (Packard 2000; Ceccagno & Scalise 2006; cf. also Table 21 in Scalise & Fábregas 2010), Breton (Ternes 2016), Oko (okoo1245; Atoyebi 2010), Welsh (Awbery 2004) and Western Farsi (Tehranisa 1987), as well as the languages under consideration here (and, of course, Nizaa, to which I will return in §7.3 and §8.1).

The explanations for such inconsistency vary. In some languages different orders are exhibited by different compound types, as is the case with Mandarin, which has left-headed verbal compounds and right-headed nominal compounds. This kind of explanation obviously does not hold in the domain of binominals, since only noun-noun compounds qualify as binominals. A second explanation is diachronic word order change; this is the case with Breton, Welsh and Western Farsi. A third reason can be language contact – more precisely, borrowing – and this, as was noted above, is the case with Vietnamese, in which native compounds are left-headed, and loans from Chinese are right-headed (cf. example 122 on page 261).

As for Mapudungun, we will see in §7.1.3 that the system here is similar to that in Nizaa, and that the position of the head correlates with the type of semantic relation being expressed, which gives us a fourth kind of explanation. In Nizaa, attributive compounds (to use Scalise and Bisetto’s term) are left-headed, whereas subordinate compounds are right-headed. In Mapudungun the situation is the reverse, as shown in the contingency table below (Table 36). Unfortunately, there is too little data from Mapudungun in the current database to provide statistically significant evidence, but the tendency for jxtL to express one kind of relation (HinM) and jxtR (or cmpR) to express others (MinH and HtoM) is clear.2

---

1 Here I ignore exceptions that prove the rule, like Eng. attorney-general (cf. page 175).
2 The exception that proves the rule in this table is malle ñawe [father’s_brother daughter] NIECE, a right-headed construction (jxtR) that expresses an HinM relation (POSS).
Greek and Barain constitute a different case, since the head ordering opposition is between different morphosyntactic strategies. In Greek, compounds are right-headed while genitival binominals are left-headed (Table 37); the latter, as Ralli (2013: 246) points out, resemble Greek Noun phrases, from which they no doubt develop historically. However, to the best of my knowledge, no-one has investigated why some binominal concepts are expressed by the cmp strategy and others by the gen strategy and the present data is too sparse to offer any hints.

In Barain the competition is between left-headed prp strategy and the right-headed con strategy (Table 38).
There are no significant correlations to be observed, as witness the high $p$-values. However, if we restrict the analysis to the **conR** and **prpL** strategies and investigate low-level instead of high-level relations, an interesting pattern does emerge (Table 39).

Table 39: Strategies and low-level relations in Barain

<table>
<thead>
<tr>
<th>Strategy</th>
<th>COOR</th>
<th>POSS</th>
<th>MER</th>
<th>LOC</th>
<th>TEMP</th>
<th>POSS-R</th>
<th>PROD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>conR</strong></td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>prpL</strong></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

| Chi-squared $p = 0.0307320697274852$ |
| X-squared $= 13.9035714285714$, df $= 6$ |
| Fisher’s exact $p = 0.00949525237381309$ |

Of the seven relations exemplified in the (admittedly rather sparse) data, only two are found in the **con** strategy (COOR and MER) and all the remainder are found only with the **prp** strategy (POSS, LOC, TEMP, POSS-R and PROD). Examples of each are given in (124a-g).

(124) a. **con** COOR *duma non.ji* [sheep child.POSS:3SG:M] LAMB
    b. **con** MER *sinja guma.geti* [nose hole.POSS:3SG:F] NOSTRIL
    c. **prp** POSS *gera ge ŋ gargar* [home REL:3SG:M ASS spider] SPIDER WEB
    d. **prp** TEMP *tii de ŋ bonte* [eat:INF REL:SG:F ASS morning] BREAKFAST
    e. **prp** POSS-R *mee de mer.geti* [women REL:SG:F husband.POSS:3SG:F] MARRIED WOMAN
    f. **prp** LOC *looli ta lutta* [dirt PURP ear] EARWAX
    g. **prp** PROD *mooso ta peye* [cow PURP milk] DAIRY COW
    h. **prp** MER *assi ge ŋ kee* [bone REL:3SG:M ASS head] SKULL

Only one data point disturbs the consistency of this pattern, and that is (124h), in which head + bone has been analysed as a meronomic relation. However, there is an alternative interpretation for this conceptual combination, which is of “a bone that is similar to a head” (SIM). Exactly this case was noted by Pepper (*2016: 301–302*) in Nizaa, where two words for SKULL were attested in the data source (125).

(125) a. *kôw bûn ‘skull’ < kôw ‘bone’ + bûn ‘head’
    b. *bûn kôw ‘skull’ < bûn ‘head’ + kôw ‘bone’

As I wrote then:

During the first step of the analysis there was uncertainty as to whether the relation between ‘bone’ and ‘head’ should be interpreted as PART-WHOLE or RESEMBLANCE. During Step 2, these compounds (and the remainder of those yet to be resolved) were used to test the hypothesis that had emerged during Step 1, and it turned out that they
could indeed be accommodated in the system that had emerged. If the two conceptualizations of ‘skull’ (as a bone that resembles a head, or as a bone that is part of the head) are assigned to compounds [125a and b] as LH and RH, respectively, then the sub-regularity is upheld.

If (124h) is given a resemblance interpretation (SIM), a consistent pattern emerges:

\[
\text{Table 40: Strategies and low-level relations in Barain (rev.)}
\]

This pattern can be summarized as follows: In Barain the con strategy always and only encodes either a MER relation or a COOR relation; all other relations are encoded by the prp strategy. This can be taken one step further by examining the actual data points for the COOR relation. There are four altogether (126) and they all permit of a possessive reading (as well as one of coordination). With such a reading, the generalization becomes: In Barain the con strategy always and only encodes a MER or POSS relation; all other relations are encoded by the prp strategy.

(126)  

a. peesi non.ji [horse child.POSS:3SG:M] FOAL OR COLT  
b. nokuno non.ji [goat child.POSS:3SG:M] KID  
c. duma non.ji [sheep child.POSS:3SG:M] LAMB  
d. mon mee.ji [chief woman.POSS:3SG:M] QUEEN

Whether this generalization, based on a very limited set of data, holds up across the language as a whole is an empirical question. Lovestrand (2012) discusses the con strategy in the context of adnominal possession, but not as a source of lexical items, and semantic relations are not mentioned. Denominal derivation (der) and noun compounds (cmp) are covered, but as so often, other means of enriching the nominal lexicon seem to fall between the cracks – a consequence, I believe, of the traditional division of grammar into morphology and syntax. It should be mentioned that eight of the 10 instances of the prp strategy use the morphosyntax of relative clauses (cf. 124c-e, h). It could therefore be argued that these are not lexical items at all, but simply descriptive phrases. However, this seems unlikely given the kind of meanings that they denote (BREAKFAST, DINNER, LUNCH, MARRIED WOMAN, SKULL, SPIDER WEB, SPINE and SUPPER), most of which tend cross-linguistically to be lexicalized.
7.2 Binominals and possessives

The mention of adnominal possession in the previous paragraph leads to a topic that has been alluded to throughout this dissertation, but so far not confronted head-on: the relationship between binominals and possessives. Evidence of a significant relationship has been gradually accumulating. The first hint for me personally came during my work on Nizaa (Pepper 2010b), when I discovered that the semantic relations found in right-headed compounds in Nizaa overlap considerably with those typical of possessive constructions (PART-WHOLE, KINSHIP, POSSESSION, etc.). I discussed this finding in terms of Langacker’s (1991) reference point model, and noted that right-headed compounds somehow “correspond to” possessive constructions, but I failed to make the connection in terms of grammaticalization.

For the present project I knew from the start that I wanted to include constructions like French chemin de fer alongside typical Germanic compounds like Eisenbahn, and that the preposition de is used in archetypical French Possessive constructions like la plume de ma tante.1 As I started compiling the data, I came across more and more examples, like those in (126) above, where the glosses indicate some kind of possessive function. In all, the database contains some 400 instances of binominals whose gloss includes one of the abbreviations 3SG, POSS, AL or INAL (and this by no means exhausts the list of possessive morphemes). This again indicates some kind of connection between the two kinds of construction.

Another hint is to be found in the discussion of the correlation between constituent order in compounds and constituent order in possessive constructions. In §2.1.1, we saw that Bauer’s (2001) data show a tendency toward a biconditional universal of the type Poss-N ≡ Mod-N; in my own replication of Bauer’s study, which was restricted to noun-noun compounds, I found an exceptionless pattern: either the head is on the left in both compounds and possessive constructions, or it is on the right in both (cf. Table 6b on page 27). Even more significant are the similarities between the morphosyntactic strategies found in adnominal possessive constructions and those of binominal lexemes. Six of the seven strategies found by Koptjevskaja-Tamm in her study of possessive noun phrases in Europe (§5.1.1) are identical to six of the nine binominal types,2 and the absence of her “linking pronouns” (the missing binominal type that I labelled prn in §5.4.1) can be accounted for.

1 Life magazine once called this “the most idiotically useless phrase in a beginner’s French textbook”.
2 The seventh PNP strategy, “possessive compounding”, does not figure in Koptjevskaja-Tamm’s typology (Figure 32) because in her sample it was “mainly restricted to Northern Swedish”.

---

1 Life magazine once called this “the most idiotically useless phrase in a beginner’s French textbook”.
2 The seventh PNP strategy, “possessive compounding”, does not figure in Koptjevskaja-Tamm’s typology (Figure 32) because in her sample it was “mainly restricted to Northern Swedish”.
All of this evidence suggests the hypothesis that the binominal lexeme construction often grammaticalizes from the adnominal possession construction. One prediction from this is that there should be a correlation between the strategy used for the binominal lexeme construction and the adnominal possession construction; in other words, that the binominal lexeme construction recruited its strategy from the adnominal possession construction. In this section I investigate that hypothesis. For reasons of time and space the investigation will not be definitive, but it will lead to a number of interesting observations and to some provisional conclusions.

As Rijkhoff (2001: 528) points out, “possessive constructions have been the subject of several typological studies (Chappell & McGregor 1996; Heine 1997; Manzelli 1990; Plank 1991; Seiler 1983; Ultan 1978)”. Since Rijkhoff wrote, this list has grown to include Baron, Herslund & Sørensen (2001), Koptjevskaja-Tamm (2002; 2003; 2004), McGregor (2009), Stassen (2009), Koch (2012), Aikhenvald & Dixon (2013), Carlier & Verstraete (2013) and Johanson, Mazzitelli & Nevskaia (2019), to cite some of the more important contributions. However, the majority of these studies have been concerned with predicative rather than adnominal (or attributive) possession. The latter seemed to Heine “to present a relatively simple structure: it consists essentially of two noun phrases linked to one another in a specific way” (1997: 143). Moreover, even those studies that deal with adnominal possession tend to focus on semantics, for example alienability, or Heine’s set of eight event schemas “that account for the majority of possessive constructions in the languages of the world” (Table 41).

<table>
<thead>
<tr>
<th>Formula</th>
<th>Label of event schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>X takes Y</td>
<td>Action</td>
</tr>
<tr>
<td>Y is located at X</td>
<td>Location</td>
</tr>
<tr>
<td>X is with Y</td>
<td>Companion</td>
</tr>
<tr>
<td>X's Y exists</td>
<td>Genitive</td>
</tr>
<tr>
<td>Y exists for/to X</td>
<td>Goal</td>
</tr>
<tr>
<td>Y exists from X</td>
<td>Source</td>
</tr>
<tr>
<td>As for X, Y exists</td>
<td>Topic</td>
</tr>
<tr>
<td>Y is X's (property)</td>
<td>Equation</td>
</tr>
</tbody>
</table>

*Table 41: Heine’s set of eight event schemes*
The most substantive work on the morphosyntactic typology of adnominal possession is to be found in Koptjevskaja-Tamm’s series of papers on possessive noun phrases (PNPs) from the early 2000s. These are limited to the languages of Europe, but they will provide our starting point here, as they did in Chapter 6.

7.2.1 Anchoring and typifying possessive constructions

Koptjevskaja-Tamm’s 2002 paper, entitled *Adnominal possession in the European languages*, provides an overview of the structure of PNPs, introduces the distinction between anchoring and non-anchoring relations and discusses animacy/referentiality splits, alienability and nominalizations; the 2003 paper (*Possessive noun phrases in the languages of Europe*) discusses synthetic and analytic constructions in some depth, along with “cross-categorial” uses of construction markers; and the 2004 paper (*Maria’s ring of gold: Adnominal possession and non-anchoring relations in European languages*) focuses on anchoring vs. non-anchoring relations. For the remainder of this section, all references will be to these three papers unless otherwise stated.

In the present context we are primarily concerned with the distinction between anchoring and non-anchoring relations – and with the relation between the latter and binominals. Koptjevskaja-Tamm does not provide explicit definitions, but characterizes anchoring and non-anchoring relations on the basis of two sets of examples from Lithuanian (2004: 155-156):

As is well known, so called “adnominal possessive constructions” are never dedicated to expressing possession stricto sensu (LEGAL OWNERSHIP or DISPOSAL, but are used for various relations by which the head’s referent is identified via the possessor’s referent (e.g. BODY-PART vs. person-relations, or KIN relations).

(1) Petr-o nāmas / pīrštas / brolis
    Peter-GEN house:NOM finger:NOM brother:NOM
    ‘Peter’s house / finger / brother’

Typical possessors, such as the noun in the genitive in ex. (1) from Lithuanian, act thus as anchors or reference point entities for identification of the head, and the whole construction can be said to denote anchoring relations. In ex. (2), however, genitive nouns serve as MATERIAL (2a) or PURPOSE (2b) modifiers of the head; in other words, the same pattern is used for expressing non-anchoring relations, in which the nominal dependent is to classify, describe or qualify the class of entities denoted by the head.

(2) a. auks-o žedas
    gold-GEN ring:NOM
    ‘a golden ring’

    b. kav-os puodelis
    coffee-GEN cup:NOM
    ‘a coffee cup’
What anchoring and non-anchoring constructions have in common is that both types of adnominal dependent characterize entities via their relations to other entities. Non-anchoring (or *typifying*, to use a less apophatic term) adnominals differ, however, in the words of Croft (forthc.), in that

1. the object modifier is only type identifiable;
2. the modifier-head combination refers to a subclass of a broader class and often functions as a classificatory label for it, suggesting that the modifier and the head together correspond to one concept, but
3. the head cannot be identified via its relation to the modifier.

From this it is clear that binominals, as defined in the present study, are essentially typifying nominal modifier constructions, albeit ones in which a diachronic process of lexicalization has proceeded to the point where the binominal is coming to be a unitary lexeme; binominals are basically lexicalized typifying constructions that represent the penultimate stage in a continuum of constructions from anchoring object modification to reference (127):

(127) anchoring construction → typifying construction → binominal lexeme → simple noun construction

*modification* ............................................................ *reference*

The difference between binominals and the general case of typifying constructions is minimal and based on the degree of lexicalization, which is hard to measure. It is therefore more instructive to compare the morphosyntactic strategies between binominals and anchoring constructions. Koptjevskaja-Tamm does more or less this in her 2004 paper, but for typifying (“non-anchoring”) constructions in general and for European languages only. Table 42 reproduces the table from her 2002 and 2004 papers in which Koptjevskaja-Tamm summarizes the results of her investigation.¹ Five cases are distinguished, one of them with two subtypes, as follows:

1. Identical structures
2. Similar structures
3. Differing complexity on (i) the modifier or (ii) the head
4. Phrase vs. one word
5. NP vs. adjectival modifier

¹ I have modified Koptjevskaja-Tamm’s numbering slightly in order to separate “identical structures” from “similar structures” and used the term typifying instead of non-anchoring. I have also simplified some glosses (without losing any critical information) for reasons of clarity, and corrected a minor error: 3.2 concerns different degrees of H’s (not D’s) morphological complexity.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Representative languages</th>
<th>Structural patterns and examples</th>
<th>Non-anchoring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Lithuanian</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identical structures</td>
<td>D in the genitive case</td>
<td>D in the genitive case</td>
</tr>
<tr>
<td></td>
<td><em>mokytojas</em> name</td>
<td><em>duonos</em> name</td>
</tr>
<tr>
<td></td>
<td>‘the teacher’s name’</td>
<td>‘a bread knife’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.1. Italian</strong></td>
<td>Articles in D</td>
<td>No articles in D</td>
</tr>
<tr>
<td>Similar structures</td>
<td><em>la casa di un professore</em></td>
<td><em>la casa di pietra</em></td>
</tr>
<tr>
<td>(D = de-phrases)</td>
<td>‘the house of a teacher’</td>
<td>‘the house of stone’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.2. Scottish Gaelic</strong></td>
<td>Articles in D;</td>
<td>No articles in D</td>
</tr>
<tr>
<td>Similar structures</td>
<td>H incompatible with articles</td>
<td>H compatible with articles</td>
</tr>
<tr>
<td>(D in the genitive case)</td>
<td><em>an/O cù na caileige</em></td>
<td><em>an/O cù fiodha</em></td>
</tr>
<tr>
<td></td>
<td>‘the girl’s dog’</td>
<td><em>DEF/O dog DEF.GEN girl:GEN</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘the/a wooden dog’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3.1. Albanian</strong></td>
<td>D = Agreeing preps + genitive</td>
<td>D in the ablative case</td>
</tr>
<tr>
<td>Different degrees of</td>
<td><em>buk-a e grur-it</em></td>
<td><em>bukë grur-i</em></td>
</tr>
<tr>
<td>D’s morphological</td>
<td><em>bread:DEF ATTR wheat:DEF.GEN</em></td>
<td><em>bread:DEF GEN</em></td>
</tr>
<tr>
<td>complexity</td>
<td>‘the wheat bread’</td>
<td>‘wheat bread’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3.1. Rumanian</strong></td>
<td>D in the genitive case</td>
<td>D = <em>de-phrase</em></td>
</tr>
<tr>
<td>Different degrees of</td>
<td><em>fiul regelui</em></td>
<td><em>fiul de rege</em></td>
</tr>
<tr>
<td>D’s morphological</td>
<td><em>son:DEF king:DEF.GEN</em></td>
<td><em>son:DEF of king</em></td>
</tr>
<tr>
<td>complexity</td>
<td>‘the son of the king’</td>
<td>‘the royal son’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3.1. Turkish</strong></td>
<td>Double-marking</td>
<td>Head-marking</td>
</tr>
<tr>
<td>Different degrees of</td>
<td><em>kadın-in kitab-i</em></td>
<td><em>kadın kitab-i</em></td>
</tr>
<tr>
<td>D’s morphological</td>
<td><em>woman-GEN book-POSS</em></td>
<td><em>woman book-POSS</em></td>
</tr>
<tr>
<td>complexity</td>
<td>‘the woman’s book’</td>
<td>‘a women’s book’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3.2. Érzya-Mordvin</strong></td>
<td>Double-marking</td>
<td>Dependent-marking</td>
</tr>
<tr>
<td>Different degrees of</td>
<td><em>sazor-ont’ éjkakš-ozo</em></td>
<td><em>éjkakš-ö kni-i-t</em></td>
</tr>
<tr>
<td>H’s morphological</td>
<td><em>elder.sister-DEF.GEN child-POSS</em></td>
<td><em>child-GEN book-PL NOM</em></td>
</tr>
<tr>
<td>complexity</td>
<td>‘the elder sister’s child’</td>
<td>‘books for children’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4. Swedish:</strong></td>
<td>Dependent-marking</td>
<td>Compounding</td>
</tr>
<tr>
<td>Presence v. absence</td>
<td><em>student-en-s rum</em></td>
<td><em>student-rumm-et</em></td>
</tr>
<tr>
<td>of morphological markers;</td>
<td><em>student-DEF GEN room</em></td>
<td><em>student-room-DEF</em></td>
</tr>
<tr>
<td>phrase v. one word</td>
<td>‘the student’s room’</td>
<td>‘the room for students’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5. Russian:</strong></td>
<td>D in the genitive case</td>
<td>D = derived adjectives</td>
</tr>
<tr>
<td>Depends on NPs vs.</td>
<td><em>plat’e źenš’iny</em></td>
<td><em>źen-sk-oe plat’e</em></td>
</tr>
<tr>
<td>adjectives</td>
<td><em>dress woman:GEN</em></td>
<td><em>woman-ADD-NEUT</em></td>
</tr>
<tr>
<td></td>
<td>‘a/the woman’s dress’</td>
<td><em>dress</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 42: Anchoring/non-anchoring comparison table (Koptjevskaja-Tamm 2004; D = dependent)*

272 Submitted version 2020-05-26
These five cases can be briefly characterized as follows:

**Identical structures.** The same morphosyntactic strategy is used for both anchoring and typifying constructions. Exemplified by Lithuanian, Georgian, Daghestanian, Russian and Finnish, which use a genitive modifier and inflect nouns for case but lack articles and a grammaticalized definiteness-indefiniteness opposition.

**Similar structures.** The same morphosyntactic strategy is used for both anchoring and typifying constructions but articles (markers of definiteness) on the modifier are permitted with the former but not the latter. Exemplified by Italian (prepositional strategy) and Scottish Gaelic (genitival strategy).

**Differing morphological complexity.** Typifying constructions are morphologically less complex and/or looser than anchoring constructions. Exemplified by Albanian, Rumanian, Turkish and Kirmandji (dependent-marking) and Mordvin and Armenian (head-marking).

**Loss of nominal autonomy: compounding.** The relational or indexical marker found in anchoring constructions is lost in typifying constructions, leading to a compound or juxtaposition strategy for the latter. Exemplified by the Uralic and Germanic languages.

**Loss of nominal autonomy: relational adjectives.** Typifying constructions use a derived, adjectival form of the modifying noun instead of case markers (i.e. case affixes or adpositions). Exemplified by Slavic and – to a lesser extent – Romance languages.

### 7.2.2 Comparing non-binary typologies

Before we can compare binominal constructions with typifying constructions in a quantitative manner, two issues need to be addressed. The first concerns the number of types in the respective typologies. In most studies of the Greenbergian type, the constructions under comparison offer a binary choice: for example, the possessor can either follow (N-Poss) or precede (Poss-N) the possessum, and the modifier of a compound can either follow (N-Mod) or precede (Mod-N) the head. In such a scenario most languages can be assigned a single value for each construction, e.g. GenN and OV. Ignoring the few mixed languages, the comparison can then be represented as a tetrachoric (2x2) table, such as that in Table 43, reproduced from page 27. From this can be deduced either a bidirectional universal (as in this case) or an implicational universal.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th></th>
<th>N-Mod</th>
<th>Mod-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Poss</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Poss-N</td>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 43: Noun-possessor tetrachoric tables (Pepper 2015)

In the case of both anchoring nominal modifier constructions and binominals there is no binary choice: Koptjevskaja-Tamm’s PNP typology contains seven types, and our binominal typology contains nine. Representing this as a 7x9 table would clearly not be very helpful. Moreover, many languages employ multiple strategies in order to represent attributive possession, just as they do to express binominals (see the discussion of intralingual competition in §5.6.2 on page 173 ff). This complicates the comparison even more. A different approach is thus required, one that Koptjevskaja-Tamm has already pioneered. Notice that her comparison table and its five primary categories do not focus on the values assigned to each language on the basis of her PNP typology, but on a characterization of the relation between each language’s primary anchoring and typifying strategy: for each language those are essentially described as “identical”, “similar” and “differing”, with the latter amenable to subcategorization such that it encompasses the two cases of “loss of autonomy” as well as differing morphological complexity. In my analysis I will follow Koptjevskaja-Tamm’s lead, but with a small adjustment to make the set of categories more amenable to statistical analysis.

The five categories listed above may be used as simple nominal variables. However, the adjectives used to describe the first three categories suggest a potential for representation as ordinal variables: identical → similar → different. In the hope that this might provide new insights I have chosen to replace Koptjevskaja-Tamm’s three “orderable” categories with a more fine-grained system of five categories that express the degree of similarity between anchoring and binominal constructions. Not surprisingly, since this is about different grades of a property (similarity), a naming system based on adjectives is less useful. I will therefore adopt one based on adverbs that qualify the adjective ‘identical’ (128).

---

1 Nominal variables, as Levshina (2015: 16–17) explains are two or more categories that are mutually exclusive. They represent the least precise and informative level of measurement.

2 Levshina uses the five-point Likert scale (‘strongly disagree’ – ‘disagree’ – ‘neither agree nor disagree’ – ‘agree’ – ‘strongly agree’) as an example of an ordinal variable and points out that “the categories thus differ in order, but we do not know yet by how much. We cannot say, for example, that the difference between ‘disagree’ and ‘neither agree nor disagree’ is the same as the difference between ‘strongly disagree’ and ‘disagree’”.

---
always \rightarrow mostly \rightarrow sometimes \rightarrow rarely \rightarrow never

Clearly, these categories need to be defined precisely for the task at hand, but before doing so there is second issue that needs to be addressed: that of mixed languages. As we saw in §5.6.2, most languages have more than one binominal strategy available to them; some have as many as six (and some, like Polish, as many as nine if the order of constituents is taken into account); most have at least four; and only seven of the 106 languages in our sample are entirely monogamous (see Figure 41 on page 173). The question thus arises which strategy to select for the comparison with anchoring constructions. Fortunately, almost every language in the sample shows a preference for one type of binominal type or another, and 70 of the 106 can be said to have a dominant type according to Dryer’s criteria for dominance (§5.6.1). My comparison will therefore be based on what I will term the ‘primary binominal strategy’, defined as the type that occurs most frequently; languages that have no clear preference (Äiwoo, Galibi Carib and Selice Romani, see Table 26 on page 172) will be deemed to have no such strategy. In addition, I will employ the term ‘secondary binominal strategy’ for any non-primary strategy that is ‘common’ (defined as occurring in at least 10% of the data for any given language, cf. §5.6.3). Having defined these terms, the five grades of ‘identicality’ in (128) can be defined as shown in Table 44.

<table>
<thead>
<tr>
<th>grade</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>always</td>
<td>the primary binominal strategy is identical to the primary anchoring strategy and there are no secondary binominal strategies</td>
</tr>
<tr>
<td>mostly</td>
<td>the primary binominal strategy is identical to the primary anchoring strategy but there also are secondary binominal strategies</td>
</tr>
<tr>
<td>sometimes</td>
<td>a secondary binominal strategy is identical to the primary anchoring strategy, or the primary binominal strategy is identical to a secondary anchoring strategy</td>
</tr>
<tr>
<td>rarely</td>
<td>a secondary binominal strategy is identical to one of the secondary anchoring strategies</td>
</tr>
<tr>
<td>never</td>
<td>binominal strategies and anchoring strategies are quite different</td>
</tr>
</tbody>
</table>

Table 44: Pepper scale for non-binary typological comparison

The definitions themselves are, of course, particular to the actual constructions that we are investigating, but the Pepper scale\(^1\) itself has universal validity and could

---

\(^1\) I venture to name the scale in line with the convention established by psychologist Rensis Likert. The Pepper scale should not be confused with the Scoville scale!
provide an additional tool for typologists, alongside tetrachoric tables and semantic maps, for use when comparing non-binary typologies.

7.2.3 Data analytics

Before proceeding with the quantitative analysis of the data, it should be noted that the Pepper scale loses some of the qualitative detail in Koptjevskaja-Tamm’s model (for example, differences in complexity and morphological tightness), but it is perfectly possible to add that back in, by subdividing the basic categories (I do so below with “never” in order to capture grammaticalization). Table 45 shows the results of applying the scale when comparing anchoring and binominal strategies (labelled A and B). The 106 languages of the sample are grouped according to the grade they exhibit on the scale, in the decreasing order of similarity shown in (128), along with short comment that explains the situation that gives rise to the grading:

*always:* Hausa, Hebrew, Akkadian, etc. are graded “always” because they use the same primary strategy for both anchoring possessive and binominals and there are no other strategies for the latter.

*mostly:* Lower Sorbian, Russian, Anindilyakwa, etc. are graded “mostly” because, although they use the same primary strategy for both anchoring possessives and binominal, there are additional strategies for the latter.

*sometimes:* Here there are two possible scenarios: the primary anchoring strategy is a secondary binominal strategy, or the primary binominal strategy is a secondary anchoring strategy; for example, Galibi Carib’s con primary anchoring strategy is its secondary binominal strategy, whereas in Maltese the situation is the reverse.

*rarely:* In Croatian, an instance of the “rarely” category, adj is found among the secondary strategies of each language, but the primary strategies (gen and der, respectively) differ.

*never:* The “never” category has been split, to highlight cases where the principal binominal strategy is a grammaticalized form of the principal anchoring strategy, following one of nine pathways that involve fusion or loss of a single marker:

(129) jxt gen gen con con prn prn dbl dbl cmp cmp jxt cmp cmp jxt cmp con gen

Each pathway in (129) represents a single step: either fusion, as in the case of jxt → cmp, or loss of a single morpheme (as in gen → cmp). Other pathways are conceivable (e.g. prp → gen), but these are not attested in the data.
### ALWAYS

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausa</td>
<td>con</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Hebrew</td>
<td>con</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Akkadian</td>
<td>dbl</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Bezhta</td>
<td>gen</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Kambaata</td>
<td>gen</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Ceq Wong</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Datooga</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Hmong Daw</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Indonesians</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Seych. Creole</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Srenge</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Walmat</td>
<td>jxt</td>
<td>primary both A and B</td>
</tr>
<tr>
<td>Malayalam</td>
<td>gen</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Selice Romani</td>
<td>gen</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Somali</td>
<td>gen</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Ticuna</td>
<td>gen</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Kam</td>
<td>jxt</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Murui Huitoto</td>
<td>jxt</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Saramaccian</td>
<td>jxt</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Hindi</td>
<td>prp</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Maltese</td>
<td>prp</td>
<td>(primary A is secondary B)</td>
</tr>
<tr>
<td>Seri</td>
<td>dbl</td>
<td>(secondary A is primary B)</td>
</tr>
<tr>
<td>Czech</td>
<td>adj</td>
<td>(secondary A is primary B)</td>
</tr>
<tr>
<td>Yaqvi</td>
<td>jxt</td>
<td>(secondary A is primary B)</td>
</tr>
<tr>
<td>Thai</td>
<td>cmp</td>
<td>(secondary A is primary B)</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>jxt</td>
<td>(secondary A is primary B)</td>
</tr>
<tr>
<td>Welsh</td>
<td>jxt</td>
<td>(secondary A is primary B)</td>
</tr>
</tbody>
</table>

### MOSTLY

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Sorbian</td>
<td>adj</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Russian</td>
<td>adj</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>con</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Iraqw</td>
<td>con</td>
<td>both prim.; + jxt.der in B</td>
</tr>
<tr>
<td>Kechki</td>
<td>con</td>
<td>both prim.; + jxt.der in B</td>
</tr>
<tr>
<td>Kupsabiny</td>
<td>con</td>
<td>both prim.; + prp in B</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>con</td>
<td>both prim.; + jxt cmp adj dbl in B</td>
</tr>
<tr>
<td>Wolof</td>
<td>con</td>
<td>both prim.; + prp in B</td>
</tr>
<tr>
<td>Yakut</td>
<td>con</td>
<td>both prim.; + jxt.der in B</td>
</tr>
<tr>
<td>Amharic</td>
<td>gen</td>
<td>both prim.; + jxt in B</td>
</tr>
<tr>
<td>Archi</td>
<td>gen</td>
<td>both prim.; + jxt adj in B</td>
</tr>
<tr>
<td>Assamese</td>
<td>gen</td>
<td>both prim.; + jxt cmp in B</td>
</tr>
<tr>
<td>Estonian</td>
<td>gen</td>
<td>both prim.; + cmp in B</td>
</tr>
<tr>
<td>Greek</td>
<td>gen</td>
<td>both prim.; + cmp, der in B</td>
</tr>
<tr>
<td>Irish</td>
<td>gen</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Kanuri</td>
<td>gen</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Latvian</td>
<td>gen</td>
<td>both prim.; + cmp, der in B</td>
</tr>
<tr>
<td>Nepali</td>
<td>gen</td>
<td>both prim.; + cmp in B</td>
</tr>
<tr>
<td>Sidamo</td>
<td>gen</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Wawa</td>
<td>gen</td>
<td>both prim.; + jxt in B</td>
</tr>
<tr>
<td>Bandial</td>
<td>jxt</td>
<td>both prim.; + cls, prp in B</td>
</tr>
<tr>
<td>Cabécar</td>
<td>jxt</td>
<td>both prim.; + con in B</td>
</tr>
<tr>
<td>Hupdè</td>
<td>jxt</td>
<td>both prim.; + cmp in B</td>
</tr>
<tr>
<td>Wichi</td>
<td>jxt</td>
<td>both prim.; + cmp, der in B</td>
</tr>
<tr>
<td>French</td>
<td>prp</td>
<td>both prim.; + der in B</td>
</tr>
<tr>
<td>Italian</td>
<td>prp</td>
<td>both prim.; + der, adj in B</td>
</tr>
<tr>
<td>Swahili</td>
<td>prp</td>
<td>both prim.; + cmp in B</td>
</tr>
<tr>
<td>Tagalog</td>
<td>prp</td>
<td>both prim.; + jxt, der in B</td>
</tr>
<tr>
<td>Tarifit</td>
<td>prp</td>
<td>both prim.; + gen in B</td>
</tr>
</tbody>
</table>

### RARELY

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croatian</td>
<td>adj</td>
<td>(secondary A is secondary B)</td>
</tr>
<tr>
<td>Polish</td>
<td>adj</td>
<td>(secondary A is secondary B)</td>
</tr>
<tr>
<td>Slovak</td>
<td>adj</td>
<td>(secondary A is secondary B)</td>
</tr>
<tr>
<td>Bambara</td>
<td>gen</td>
<td>(secondary A is secondary B)</td>
</tr>
<tr>
<td>German</td>
<td>gen</td>
<td>(secondary A is secondary B)</td>
</tr>
</tbody>
</table>

### GRAMMATICALIZED

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mbya Guarani</td>
<td>con</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Turkish</td>
<td>dbl</td>
<td>grammaticalized to B con</td>
</tr>
<tr>
<td>Basque</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>English</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Korean</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Norwegian</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Old H German</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Tuwari</td>
<td>gen</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Inb. Quechua</td>
<td>gen</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Kildin Sami</td>
<td>gen</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Manange</td>
<td>gen</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Warta Thunai</td>
<td>gen</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Wik-Mungkan</td>
<td>gen</td>
<td>grammaticalized to B jxt</td>
</tr>
<tr>
<td>Baa</td>
<td>jxt</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Chakali</td>
<td>jxt</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Cuijia</td>
<td>prp</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Quer. Otomi</td>
<td>prp</td>
<td>grammaticalized to B cmp</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>prp</td>
<td>grammaticalized to B jxt</td>
</tr>
</tbody>
</table>

### NEVER

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aiwoow</td>
<td>con</td>
<td>diff. from B NA (+ jxt, cls, prp)</td>
</tr>
<tr>
<td>Gawaiwada</td>
<td>con</td>
<td>diff. from B gen (+ jxt, der)</td>
</tr>
<tr>
<td>Zin. Tzotzil</td>
<td>con</td>
<td>diff. from B gen (+ jxt)</td>
</tr>
<tr>
<td>Barain</td>
<td>dbl</td>
<td>diff. from B prp (+ con)</td>
</tr>
<tr>
<td>Romanian</td>
<td>dbl</td>
<td>diff. from B prp (+ der, adj)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>dbl</td>
<td>diff. from B cmp (+ der, adj)</td>
</tr>
<tr>
<td>Dutch</td>
<td>prp</td>
<td>diff. from B cmp (+ gen)</td>
</tr>
<tr>
<td>Mam. Senoufo</td>
<td>prp</td>
<td>diff. from B cmp</td>
</tr>
<tr>
<td>Mandarin</td>
<td>prp</td>
<td>diff. from B cmp (+ jxt)</td>
</tr>
<tr>
<td>Ho-Chunk</td>
<td>dbl</td>
<td>diff. from B jxt (+ cmp, der)</td>
</tr>
<tr>
<td>Western Mari</td>
<td>dbl</td>
<td>diff. from B jxt (+ cmp)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>prp</td>
<td>diff. from B jxt (+ gen)</td>
</tr>
</tbody>
</table>

### (NO DATA)

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gurindji</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 45: Comparing anchoring and binominal strategies
Finally, the “never” category contains languages in which no anchoring strategy is the same as any of the binominal strategies. In Table 45 this category is divided into two parts by a dotted line. Above the dotted line there is no discernible pattern. Below it, the primary binominal strategy is either cmp or jxt, i.e. compounding.

In all four categories except “never”, binominal constructions can be said to recruit one of the anchoring strategies, sometimes across the board (“always”), sometimes to a lesser degree (“mostly”, “sometimes”, “rarely”), and sometimes in a more “grammaticalized” form. When the anchoring strategy is not recruited, more often than not, it is compounding that fills the void.

Figure 69 plots the six categories used in Table 45. We see that all but 12 of the 105 languages for which data was available (almost 90%) recruit an anchoring strategy for use in the formation of binominals. Of the twelve languages in the sample that do not, the majority (58%) use a compounding strategy.

In conclusion, I posit the following universals on the basis of the present data:

(130) With overwhelmingly greater than chance frequency, languages recruit at least one of their binominal strategies from anchoring nominal modifier constructions.

---

1 Äiwoo is one of the three languages for which there is no primary binominal strategy (cf. Table 26 on page 170). The other two “mixed” languages, Galibi Carib and Selice Romani, fall into the “sometimes” category, since their primary anchoring strategies are secondary binominal strategies.

2 Collection of the data regarding anchoring strategies was somewhat cursory, due to time constraints, and there may be errors. However, it seems unlikely that these will have influenced the overall picture in a major way. The typology of adnominal possession deserves a research project of its own.
If a language does not recruit at least one of its binominal strategies from an anchoring nominal modifier constructions, there is a strong tendency for it to use compounding for this purpose.

7.3 Strategies and semantics

In Chapter 6 we saw that there is no overall correlation, as we might have expected, between morphosyntactic strategies and semantic relations. In this section I delve more deeply into the question. I start by explaining why we should not expect such a correlation after all and then suggest alternative ways in which the question might be approached.

The category of ‘compound’ was abandoned in Chapter 1 on the grounds that it is a formal, language-specific descriptive category and therefore unsuitable for cross-linguistic comparison. The comparative concept binominal lexeme, on the other hand, is based on function, was easy to apply across languages and resulted in a nine-way typology of morphosyntactic strategies. But while we can identify these nine strategies across languages, we cannot assume that they always carry the same functional load; the case of Nizaa and Mapudungun is instructive in this respect.

Recall from §1.1.2 that Nizaa exhibits two types of nominal compound, which we denote as jxtL and jxtR, and furthermore that these comprise two disjunct sets of semantic relations. Left-headed compounds involve relations such as LOCATION, PURPOSE, ACTIVITY and APPEARANCE that can be characterized as attributive, whereas right-headed compounds involve relations such as PART, KIN, POSSESSION and CONTAINER, that can be similarly characterized as subordinative. In terms of the high-level taxonomy of semantic relations, this equates to the following:

\[
\begin{align*}
jxtL & : \text{attributive} \Rightarrow MinH, MisH, HtoM \\
jxtR & : \text{subordinative} \Rightarrow HinM, MtoH
\end{align*}
\]

Now, Mapudungun is similar to Nizaa in that it exhibits both left- and right-headed nominal compounds, as reported by Harmelink (1996) and summarized by Zúñiga (2014):

complex expressions consisting of at least two nouns are either head-final […] or head-initial […]. While in the former group the first element characterizes the second in an unspecified way, the latter group is more restricted in semantic terms, viz. it consists of part-whole relationships or of those in which “the former noun is an element of the latter.”
However, from the description given it is clear that the functions of \textit{jxtR} (head-final) and \textit{jxtL} (head-initial) are the exact opposite of those in Nizaa: the former encode attributive relations, whereas the latter encode subordinative relations.

Two observations can be made on the basis of these facts: First of all, constituent order cannot be disregarded when examining patterns in individual languages: left- and right-headed variants should not be lumped together under one morphosyntactic category. Secondly, care must be taken when doing cross-linguistic comparisons: if one tries to generalize across Nizaa and Mapudungun, the interesting patterns will cancel each other out, and those found when observing each language alone will be lost.

There are two ways to address this problem. One is to look for correlations between morphosyntactic strategies and semantic relations within individual languages rather than across languages, i.e. intralinguistically. Such an approach is admittedly not typological, but it can be justified on the grounds that it has the potential to reveal insights that may form the basis of later typological investigations. The other way to address the problem is to leverage the known facts of morphosyntactic structure differently, in a way that is cross-linguistically valid. In the following two sections I explore these two approaches.

7.3.1 Intralinguistic patterns

This investigation into possible associations between morphosyntactic strategies and semantic relations was inspired by Koch’s (2001) use of a two-dimensional grid to investigate lexical motivation (see also Koch & Marzo 2007a). The grid, reproduced in Table 46, operationalizes an insight that led Koch to refine Ullmann’s motivational typology, which I referred to briefly on page 123, namely:

Ullmann’s distinction between ‘morphological’ and ‘semantic’ motivation turns out to correspond not to a clear-cut opposition between disjunct motivational devices, but to \textit{two cross-classified dimensions} of the motivation problem in general: vertical axis = formal ‘morphological’ dimension and horizontal axis = cognitive ‘semantic’ dimension (Koch 2001: 1158; emphasis added).
Table 46: Koch’s motivational grid

<table>
<thead>
<tr>
<th>formal relations</th>
<th>conceptual identity</th>
<th>contiguity</th>
<th>metaphorical similarity</th>
<th>cotaxonomic similarity</th>
<th>taxonomic superordination</th>
<th>taxonomic subordination</th>
<th>conceptual contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>formal identity/polysemy</td>
<td>00</td>
<td>01</td>
<td>02</td>
<td>03</td>
<td>04</td>
<td>05</td>
<td>06</td>
</tr>
<tr>
<td>tone alternation</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>reduplication</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>number alternation</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>gender alternation</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>voice alternation</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>stem alternation</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>word-class alternation</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
</tr>
<tr>
<td>suffixation</td>
<td>80</td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>86</td>
</tr>
<tr>
<td>prefixation</td>
<td>90</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>compounding</td>
<td>100</td>
<td>101</td>
<td>102</td>
<td>103</td>
<td>104</td>
<td>105</td>
<td>106</td>
</tr>
<tr>
<td>serial verb</td>
<td>110</td>
<td>111</td>
<td>112</td>
<td>113</td>
<td>114</td>
<td>115</td>
<td>116</td>
</tr>
<tr>
<td>lexical syntagm</td>
<td>120</td>
<td>121</td>
<td>122</td>
<td>123</td>
<td>124</td>
<td>125</td>
<td>126</td>
</tr>
<tr>
<td>idiom</td>
<td>130</td>
<td>131</td>
<td>132</td>
<td>133</td>
<td>134</td>
<td>135</td>
<td>136</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The vertical dimension of this grid consists of an open set of formal devices that includes affixation, composition and lexical syntagm, all of which are employed in binominal word-formation, as well as formal identity, reduplication, alternation, idiom and more. The horizontal dimension constitutes “a closed set of universal cognitive relations”, viz: conceptual identity, contiguity, metaphorical similarity, co-taxonomic similarity, taxonomic superordination, taxonomic subordination and co-taxonomic contrast, which Koch arrived at by differentiating and combining the Aristotelian relations of contiguity, similarity and contrast. Aristotle’s “three principles of remembering” were mentioned in Chapter 6 and I will have occasion to return to them again in Chapter 9. “The numbers 00, 01, 02, etc., 10, 11, 12, etc. etc.”, writes Koch, “are purely arbitrary and only serve as means for identifying the different squares.”
Koch’s goal was to study lexical motivation in its full generality; the present study, on the other hand, is concerned only with binominals. Consequently, the categories Koch uses for his formal (vertical) and semantic (horizontal) axes are too coarse-grained for our purposes. I will therefore use the set of 9x2 binominal types (i.e. those that take account of constituent order) in place of his list of formal devices, and the set of five high-level semantic relations in place of his set of universal cognitive relations.

To illustrate this approach, consider the contingency table generated from the Polish data (Table 47). We encountered such tables earlier (§7.1.3) and can now appreciate that they are essentially “motivational grids” in which each row represents a morphosyntactic strategy (Koch’s formal relation) and each column a semantic relation (Koch’s cognitive relation). The numbers inside the grid show how many binominals are found in the data for each combination of strategy (ftype2) and high-level semantic relation (htype).

<table>
<thead>
<tr>
<th></th>
<th>Htype MisH HinM MinH HtoM MtoH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ftype2</strong></td>
<td></td>
</tr>
<tr>
<td>adjL</td>
<td>0  3  3  8  2</td>
</tr>
<tr>
<td>adjR</td>
<td>0  0  3  0  0</td>
</tr>
<tr>
<td>cmpL</td>
<td>0  1  0  0  0</td>
</tr>
<tr>
<td>cmpR</td>
<td>0  1  0  0  0</td>
</tr>
<tr>
<td>derR</td>
<td>14 2  2  3  3</td>
</tr>
<tr>
<td>genL</td>
<td>0  2  0  0  1</td>
</tr>
<tr>
<td>prpL</td>
<td>0  1  0  3  0</td>
</tr>
<tr>
<td><strong>chi-squared p = 0.000125218006240017</strong></td>
<td></td>
</tr>
<tr>
<td><strong>x-squared = 57.9106150793651 df = 24</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Fisher's exact p = 0.000499750124937531</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 47: Strategies and high-level relations in Polish

Polish is represented in the database by 54 binominals (NN), distributed across seven strategies (adjL, adjR, etc.), that express all five semantic relations (MisH, HinM, etc.). This gives us a 7x5 “motivational grid” (Table 47). Looking at the first row of the grid we observe that, of the 16 Polish binominals of type adjL, three express the HinM relation, three the MinH relation, eight the HtoM relation and two the MtoH relation: From the second row we see that all three binominals of type adjR express in MinH relation, and so forth. Having constructed this grid from the data, we can now apply statistical methods to test whether the formal type (ftype2) and the semantic relation (htype) are independent of one another. The null hypothesis \( H_0 \) is that there is no association between the two; in other words, that the kind of semantic relation has no influence on the choice of construction.
A chi-squared test of independence conducted using R gives a highly significant \( p \)-value with 24 degrees of freedom and an \( \chi^2 \) value of 57.9, which would mean that the null hypothesis must be rejected. However, R issues a warning that the chi-squared approximation “may be incorrect”. This is because the chi-squared test has certain limitations, as explained by Agresti:

The \( X^2 \) and \( G^2 \) chi-squared tests also have limitations in the types of data sets for which they are applicable. For instance, they require large samples. The sampling distributions of \( X^2 \) and \( G^2 \) get closer to chi-squared as the sample size \( n \) increases, relative to the number of cells \( IJ \). The convergence is quicker for \( X^2 \) than \( G^2 \). The chi-squared approximation is often poor for \( G^2 \) when some expected frequencies are less than about 5. When \( I \) or \( J \) is large, it can be decent for \( X^2 \) when some expected frequencies are as small as 1. To play safe, you can instead use a small-sample procedure whenever at least one expected frequency is less than 5 (Agresti 2005: 40).

With a sample size \( n \) of 54, an \( IJ \) value of 35 (i.e. 7×5) and many cells where the frequency is zero, the Polish data is clearly too small and too sparse for the chi-squared test to be used reliably. We turn therefore to a small-sample procedure, as recommended by Agresti: Fisher’s Exact Test. Because of the size of the Polish contingency table (7×5), this is performed using a Monte Carlo simulation, which requires less computational power. The resulting \( p \)-value of 0.0005, is again highly significant and confirms that there is, indeed, structure to be found in the Polish data: the kind of semantic relation being expressed does influence the choice of construction to a significant extent. To understand how, we can tentatively examine the residuals (that is, the differences between expected and observed values), which are shown in Figure 70. The larger the residual (in either positive or negative direction), the stronger the effect of the association.

<table>
<thead>
<tr>
<th>htype</th>
<th>ftype2</th>
<th>MISH</th>
<th>MINM</th>
<th>Mihn</th>
<th>HTom</th>
<th>MtoH</th>
</tr>
</thead>
<tbody>
<tr>
<td>adjL</td>
<td>-2.0754981</td>
<td>-0.0438529</td>
<td>0.3432032</td>
<td>1.7789984</td>
<td>0.1132277</td>
<td></td>
</tr>
<tr>
<td>adjR</td>
<td>-0.8987170</td>
<td>-0.7595545</td>
<td>3.7365142</td>
<td>-0.8987170</td>
<td>-0.5883484</td>
<td></td>
</tr>
<tr>
<td>cmlL</td>
<td>-0.5188745</td>
<td>1.8418218</td>
<td>-0.3922323</td>
<td>-0.5188745</td>
<td>-0.3396831</td>
<td></td>
</tr>
<tr>
<td>cmpR</td>
<td>-0.5188745</td>
<td>1.8418218</td>
<td>-0.3922323</td>
<td>-0.5188745</td>
<td>-0.3396831</td>
<td></td>
</tr>
<tr>
<td>derR</td>
<td>2.9656149</td>
<td>-1.2173953</td>
<td>-0.8807048</td>
<td>-1.3617619</td>
<td>0.1386750</td>
<td></td>
</tr>
<tr>
<td>genL</td>
<td>-0.8987170</td>
<td>1.8735678</td>
<td>-0.6793662</td>
<td>-0.8987170</td>
<td>1.1113248</td>
<td></td>
</tr>
<tr>
<td>prplL</td>
<td>-1.0377490</td>
<td>0.2631174</td>
<td>-0.7844645</td>
<td>1.8531233</td>
<td>-0.6793662</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 70: Residuals for cells in the Polish contingency table*

The largest residual is that between \( \text{adjR} \) and \( \text{MinH} \), with a value of 3.74, which indicates a strong tendency for right-headed adjectival constructions to be used to express relations in which the modifier “somehow, to some extent” denotes the contents of the head. Compare this with the residual between \( \text{adjL} \) and \( \text{HtoM} \) (1.78). The contrast appears to confirm the traditional Polish distinction between
The typology and semantics of binominal lexemes

(right-headed) classificatory adjectives (133a) and (left-headed) qualifying relational adjectives (133b).

(133) a. złot.y pierścionek [gold.ADJZ ring] GOLD RING \( \text{COMP-R} = \text{MinH} \) adjR
    b. pompa wod.na [pump water.ADJZ] WATER PUMP \( \text{PURP} = \text{HtoM} \) adjL

The next largest residual is that between derR and MisH, with a value of 2.99, which indicates a strong tendency for suffixation to be used to express coordinate and taxonomic (i.e. appositional) relations. Another residual of note is that between genL and HinM (1.94), which shows that genitives tend to express relations in which the modifier denotes some kind of container (or possessor, whole, place, etc.), in other words: MER, LOC, etc.

Since residuals are produced by the chi-squared test, they cannot be taken at face value; they can, however, form the basis of new hypotheses that could be tested using larger samples. To that end, Table 48 shows the results of applying the same tests to every language in the database. It lists all 22 languages for which the chi-squared test returned a \( p \)-value of less than 0.01, and gives the sample and table sizes, the ratio between these two \((n/IJ)\), the chi-squared \( p \)-value, the number of degrees of freedom \((df)\), the value of the \( \chi^2 \) statistic and the Fisher statistic. The three languages that failed Fisher’s Exact test despite passing the chi-squared test are marked by an asterisk.

These results indicate that there may be a significant association between the kind of semantic relation expressed by a binominal and the morphosyntactic strategy in at least 19 of the 106 languages in the sample. A one-by-one investigation of these languages is beyond the scope of the present work, but could lead to insights into the various ways in which languages divide the cake of semantic relations which may then feed back into broader typological research.
Table 48: $\chi^2$-squared and Fisher tests for 22 languages

<table>
<thead>
<tr>
<th>Language</th>
<th>NN</th>
<th>ftypes</th>
<th>htypes</th>
<th>n/IJ</th>
<th>p ($\chi^2$)</th>
<th>df</th>
<th>$X^2$</th>
<th>p (Fisher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anindilyakwa</td>
<td>15</td>
<td>5</td>
<td>5</td>
<td>0.60</td>
<td>0.000</td>
<td>16</td>
<td>45.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Chakali</td>
<td>24</td>
<td>3</td>
<td>4</td>
<td>2.00</td>
<td>0.000</td>
<td>6</td>
<td>25.60</td>
<td>0.000</td>
</tr>
<tr>
<td>*Dutch</td>
<td>54</td>
<td>5</td>
<td>5</td>
<td>2.16</td>
<td>0.000</td>
<td>16</td>
<td>45.40</td>
<td>0.014</td>
</tr>
<tr>
<td>German</td>
<td>75</td>
<td>6</td>
<td>5</td>
<td>2.50</td>
<td>0.000</td>
<td>20</td>
<td>50.59</td>
<td>0.000</td>
</tr>
<tr>
<td>Hebrew</td>
<td>44</td>
<td>5</td>
<td>5</td>
<td>1.76</td>
<td>0.000</td>
<td>16</td>
<td>49.52</td>
<td>0.001</td>
</tr>
<tr>
<td>*Iraqw</td>
<td>21</td>
<td>3</td>
<td>4</td>
<td>1.75</td>
<td>0.009</td>
<td>6</td>
<td>17.10</td>
<td>0.012</td>
</tr>
<tr>
<td>Kalamang</td>
<td>56</td>
<td>7</td>
<td>5</td>
<td>1.60</td>
<td>0.001</td>
<td>24</td>
<td>52.63</td>
<td>0.000</td>
</tr>
<tr>
<td>Kanuri</td>
<td>39</td>
<td>5</td>
<td>5</td>
<td>1.56</td>
<td>0.000</td>
<td>16</td>
<td>42.85</td>
<td>0.000</td>
</tr>
<tr>
<td>Ket</td>
<td>29</td>
<td>4</td>
<td>5</td>
<td>1.45</td>
<td>0.003</td>
<td>12</td>
<td>29.90</td>
<td>0.002</td>
</tr>
<tr>
<td>Malagasy</td>
<td>58</td>
<td>5</td>
<td>5</td>
<td>2.32</td>
<td>0.001</td>
<td>16</td>
<td>41.22</td>
<td>0.001</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td>105</td>
<td>3</td>
<td>5</td>
<td>7.00</td>
<td>0.000</td>
<td>8</td>
<td>39.66</td>
<td>0.000</td>
</tr>
<tr>
<td>Polish</td>
<td>52</td>
<td>7</td>
<td>5</td>
<td>1.49</td>
<td>0.000</td>
<td>24</td>
<td>57.91</td>
<td>0.000</td>
</tr>
<tr>
<td>Russian</td>
<td>43</td>
<td>6</td>
<td>5</td>
<td>1.43</td>
<td>0.000</td>
<td>20</td>
<td>54.93</td>
<td>0.000</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>31</td>
<td>4</td>
<td>5</td>
<td>1.55</td>
<td>0.000</td>
<td>12</td>
<td>36.18</td>
<td>0.003</td>
</tr>
<tr>
<td>Seri</td>
<td>16</td>
<td>5</td>
<td>3</td>
<td>1.07</td>
<td>0.000</td>
<td>8</td>
<td>29.60</td>
<td>0.000</td>
</tr>
<tr>
<td>Sidamo</td>
<td>29</td>
<td>4</td>
<td>4</td>
<td>1.81</td>
<td>0.003</td>
<td>9</td>
<td>24.59</td>
<td>0.000</td>
</tr>
<tr>
<td>Turkish</td>
<td>64</td>
<td>4</td>
<td>5</td>
<td>3.20</td>
<td>0.000</td>
<td>12</td>
<td>38.17</td>
<td>0.000</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>47</td>
<td>9</td>
<td>5</td>
<td>1.04</td>
<td>0.000</td>
<td>32</td>
<td>67.39</td>
<td>0.005</td>
</tr>
<tr>
<td>Wik-Mungkan</td>
<td>24</td>
<td>3</td>
<td>4</td>
<td>2.00</td>
<td>0.001</td>
<td>6</td>
<td>23.43</td>
<td>0.003</td>
</tr>
<tr>
<td>Wolof</td>
<td>21</td>
<td>2</td>
<td>5</td>
<td>2.10</td>
<td>0.000</td>
<td>4</td>
<td>21.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Yakut</td>
<td>41</td>
<td>5</td>
<td>5</td>
<td>1.64</td>
<td>0.000</td>
<td>16</td>
<td>47.01</td>
<td>0.000</td>
</tr>
<tr>
<td>*Yaqui</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>1.06</td>
<td>0.001</td>
<td>9</td>
<td>28.90</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Table 49: Number of markers per binominal type

<table>
<thead>
<tr>
<th>Markers</th>
<th>cmp</th>
<th>der</th>
<th>cls</th>
<th>adj</th>
<th>gen</th>
<th>prp</th>
<th>con</th>
<th>dbl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

7.3.2 Cross-linguistic patterns

Having investigated how intralinguistic patterns can be investigated using the grid approach, we turn now to cross-linguistic comparison. We noted earlier that it is not possible to simply compare strategies across languages because patterns found in one language may cancel out those found in another. However, cross-linguistic comparisons based on a formal criterion can be made by abstracting away from binominal types and considering only the amount of grammatical marking used. Recall that each structural type is located on one of three “tiers” depending on the number of additional morphs over and above the two obligatory thing-morphs (cf. Figure 34 on page 143). Each binominal can therefore be assigned one of three values (0, 1 or 2) for a variable called markers, as shown in Table 49.
Every binominal belonging to level 0, where there is no additional marking, (i.e. whose type is \textit{jxt}, \textit{cmp}, \textit{der} and \textit{cls}) is given the value 0; those on level 1 (\textit{adj}, \textit{gen}, \textit{prp} and \textit{con}) get the value 1; and those of type \textit{dbl} (the only type on level 2) are assigned the value 2. The examples shown in Table 50 all have the meaning NOSTRIL.

<table>
<thead>
<tr>
<th>language</th>
<th>word [gloss]</th>
<th>construction</th>
<th>ftype</th>
<th>markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murui Huitoto</td>
<td>defo [nose:CLF(cavity)]</td>
<td>Base.CLF</td>
<td>cls</td>
<td>0</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>wechoz yu [hole nose]</td>
<td>Head Mod</td>
<td>jxt</td>
<td>0</td>
</tr>
<tr>
<td>Central Yupik</td>
<td>pacig.uaq [gills.UAQ]</td>
<td>Base.UAQ</td>
<td>der</td>
<td>0</td>
</tr>
<tr>
<td>Basque</td>
<td>sudur.tzulo [nose:hole]</td>
<td>Mod.Head</td>
<td>cmp</td>
<td>0</td>
</tr>
<tr>
<td>Wolof</td>
<td>pax.u bakkan [hole.PER nose]</td>
<td>Head.PER Mod</td>
<td>con</td>
<td>1</td>
</tr>
<tr>
<td>Czech</td>
<td>nos.nl dirka [nose:ADJZ hole]</td>
<td>Mod.ADJZ Head</td>
<td>adj</td>
<td>1</td>
</tr>
<tr>
<td>Sidamo</td>
<td>sano.te giddo [nose:GEN inside]</td>
<td>Mod.GEN Head</td>
<td>gen</td>
<td>1</td>
</tr>
<tr>
<td>Tagalog</td>
<td>butas ng ilong [hole:LK nose]</td>
<td>Head LK Mod</td>
<td>prp</td>
<td>1</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>surax.e bin.i [hole:EZ nose:ADJZ]</td>
<td>Head EZ Mod.ADJZ</td>
<td>dbl</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 50: Coding of binominals for number of markers

Having given every binominal a value for markers, weighted averages can now be calculated for each language (Table 51). Languages like Akkadian, which only have \textit{dbl} binominals, receive the maximum value, 2; those like Caijia, which only has \textit{cmp}, receive the minimum value, 0. All other languages receive a value somewhere in between.

<table>
<thead>
<tr>
<th>area</th>
<th>family</th>
<th>genus</th>
<th>markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Äiwoo</td>
<td>O</td>
<td>Oceanic</td>
<td>0.3751429</td>
</tr>
<tr>
<td>Akkadian</td>
<td>A</td>
<td>Semitic</td>
<td>2.0000000</td>
</tr>
<tr>
<td>Amharic</td>
<td>A</td>
<td>Semitic</td>
<td>0.7872340</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>G</td>
<td>Gunwinyguan</td>
<td>0.6250000</td>
</tr>
<tr>
<td>Archi</td>
<td>E</td>
<td>Lezgic</td>
<td>0.7857143</td>
</tr>
<tr>
<td>Assamese</td>
<td>E</td>
<td>Indo-European</td>
<td>0.5000000</td>
</tr>
<tr>
<td>Baa</td>
<td>A</td>
<td>Gur</td>
<td>0.4255319</td>
</tr>
<tr>
<td>Bambara</td>
<td>A</td>
<td>Mande</td>
<td>0.2820513</td>
</tr>
<tr>
<td>Bandial</td>
<td>A</td>
<td>Atlantic-Congo North-Central Atlantic</td>
<td>0.2500000</td>
</tr>
<tr>
<td>Barain</td>
<td>A</td>
<td>Chadic</td>
<td>1.0476190</td>
</tr>
<tr>
<td>Basque</td>
<td>E</td>
<td>Basque</td>
<td>0.1090909</td>
</tr>
<tr>
<td>Bezhta</td>
<td>E</td>
<td>Nakh-Daghestanian Avar-Andic-Tsezic</td>
<td>0.8285714</td>
</tr>
<tr>
<td>Cabécar</td>
<td>S</td>
<td>Chibchan</td>
<td>0.2121212</td>
</tr>
<tr>
<td>Caijia</td>
<td>O</td>
<td>Macro-Bai</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

Table 51: Coding of languages for weighted average of markers (excerpt)

Weighted averages can also be calculated for each meaning:

<table>
<thead>
<tr>
<th>semField</th>
<th>semType</th>
<th>markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ankle</td>
<td>The body body part</td>
<td>0.29787234</td>
</tr>
<tr>
<td>arctic lights</td>
<td>The physical world</td>
<td>natural 0.8055556</td>
</tr>
<tr>
<td>backpack</td>
<td>Modern world basic</td>
<td>0.50000000</td>
</tr>
<tr>
<td>bee</td>
<td>Animals animal</td>
<td>0.12000000</td>
</tr>
<tr>
<td>beehive</td>
<td>Animals natural</td>
<td>0.46428571</td>
</tr>
<tr>
<td>beeswax</td>
<td>Animals natural</td>
<td>0.42000000</td>
</tr>
<tr>
<td>bicycle</td>
<td>Modern world advanced</td>
<td>0.07692308</td>
</tr>
<tr>
<td>bicycle pump</td>
<td>Modern world advanced</td>
<td>0.73809524</td>
</tr>
</tbody>
</table>

Table 52: Coding of meanings for weighted average of markers (excerpt)

Now it is possible to look for structure in groupings of languages and meanings.
Number of markers by language

In the case of languages, the only relevant grouping is geographical area; the sample of languages is spread too thinly and unevenly across the 42 families and 72 genera for any useful analysis on the basis of genetic affiliation. There are six geographical areas (we exclude Pidgins and creoles since the grouping is non-areal and only contains two languages), and we start by investigating the distribution of the variable markers across languages:

```r
> summary(markers)
     Min. 1st Qu.  Median    Mean 3rd Qu. Max.
0.00000 0.04625 0.38033 0.40454 0.67500 2.00000
> sd(markers)
[1] 0.3746556
```

Values range from 0 to 2 with a median of 0.39, and the first 25% of the data fall between 0 and 0.046, so they are clearly not normally distributed. This is confirmed by the plots in Figure 71, which show nothing like a bell-shaped curve, and by the Shapiro-Wilks test, where a p-value of 4.063e-07 indicates that the null hypothesis of normally distributed data must be rejected, so we cannot use parametric methods on this data.

![Histogram](image1.png) ![Density plot](image2.png)

*Figure 71: Non-normal distribution of marking across languages*

We therefore use the non-parametric Kruskal-Wallis test to find out whether there are any significant areal differences in the degree of marking across areas:

```r
> kruskal.test(markers ~ area, data=l) # p < 0.05 # H0 cannot be rejected

Kruskal-Wallis rank sum test

data:  markers by area
Kruskal-Wallis chi-squared = 26.904, df = 5, p-value = 5.955e-05
```

The p-value provides only weak evidence for rejecting the null hypothesis that the medians of all groups are equal, and this is confirmed by the pair-wise Wilcoxon rank sums test, which finds no significant differences between any of the groups:
The typology and semantics of binominal lexemes

> pairwise.wilcox.test(jitter(markers), area2, p.adjust.method = "BH")

Pairwise comparisons using wilcoxon rank sum test
data:  jitter(markers) and area2

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Eurasia</th>
<th>Oceania/SE Asia</th>
<th>NG/Australia</th>
<th>North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurasia</td>
<td>0.017</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oceania/SE Asia</td>
<td>0.012</td>
<td>0.016</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NG/Australia</td>
<td>0.012</td>
<td>0.011</td>
<td>0.764</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North America</td>
<td>0.147</td>
<td>0.438</td>
<td>0.461</td>
<td>0.364</td>
<td>-</td>
</tr>
<tr>
<td>South America</td>
<td>0.011</td>
<td>0.011</td>
<td>0.764</td>
<td>0.471</td>
<td>0.364</td>
</tr>
</tbody>
</table>

P value adjustment method: BH

In conclusion, the present data do not provide evidence of significant differences in the number of markers across the five main geographical areas.

**Number of markers by meanings**

The situation with respect to the distribution of markers across meanings is quite different. Values range from 0 to 0.81 with a median of 0.36, which is very near the centre of the range, and the first 25% of the data fall between 0 and 0.181. Also, the Shapiro-Wilks test returns a p-value of 0.1 indicates that the null hypothesis of normally distributed data cannot be rejected.

> summary(markers)

```
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    
0.0000  0.1814  0.3549  0.3500  0.4837  0.8056
```

> sd(markers)

```
[1] 0.188504
```

> # test for normality using Shapiro-Wilks
> shapiro.test(markers)

```
Shapiro-Wilk normality test

data:  markers
W = 0.97866, p-value = 0.1044
```

The first two plots in Figure 72 show a very slightly bimodal distribution and the QQ-plot has a whiff of an S shape, but the Shapiro-Wilks test returns a p-value of 0.09, which means, again, that the null hypothesis cannot be rejected. Moreover, the box plot of markers by semantic type indicates considerable separation between the groups. Parametric tests are thus permissible and we proceed to fit a linear model and carry out an analysis of variance:

> lm1 <- lm(markers ~ semType, data=m)
> anova(lm1)

```
Analysis of Variance Table

Response: markers

   Df Sum Sq Mean Sq F value    Pr(>F)
semType  6  1.0484 0.174734  6.5806 8.069e-06 ***
Residuals 93  2.4694 0.026553

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
The ANOVA shows highly significant differences in the value of markers across the various semantic types ($p < 0.001$). A summary of the linear model shows where those differences mainly lie: the groups Body part, Natural, Basic and Advanced all show significantly more marking than Person (the intercept).

```r
> summary(lm1)
Call:
lm(formula = markers ~ semType, data = m)

Residuals:
   Min      1Q  Median      3Q     Max
-0.37573 -0.11825 -0.00612  0.11615  0.41366

Coefficients:          Estimate Std. Error t value Pr(>|t|)
(Intercept)             0.17338    0.03841   4.514  1.86e-05 ***
semTypeanimal           0.07963    0.07258   1.097   0.27545
semTypelocation        0.08693    0.08238   1.055   0.29401
semTypebody part       0.21075    0.05294   3.981  5.77e-06 ***
semTypenatural         0.29218    0.06073   4.811  5.77e-06 ***
semTypebasic           0.19613    0.05294   3.705  1.50e-05 ***
semTypeadvanced        0.27927    0.05432   5.141  1.50e-06 ***
---
Signif. codes:  < *** 0.001 *** 0.01 ** 0.05 * 0.1  ' ' 1

Residual standard error: 0.163 on 93 degrees of freedom
Multiple R-squared: 0.298, Adjusted R-squared: 0.2527
F-statistic: 6.581 on 6 and 93 DF, p-value: 8.069e-06
```

Figure 72: Normal distribution of marking across meanings
An alternative to using the function `anova()` on a linear model is to use `aov()`:

```r
> a1 <- aov(markers ~ semType)
> summary(a1)

    Df Sum Sq Mean Sq F value   Pr(>F)
semType      6  1.048  0.17473   6.581 8.07e-06 ***
Residuals   93  2.469  0.02655
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

This confirms that the effects are significant ($p < 0.001$), so we can perform post-hoc pairwise comparisons (using the conservative Bonferroni correction) to find out which pairs of groups are significantly different from each other:

```r
> pairwise.t.test(markers, semType, p.adj = "bonf")

Pairwise comparisons using t tests with pooled SD

data:  markers and semType

  person animal location body part natural basic
animal  1.00000
location 1.00000 1.00000
body part 0.00286 1.00000 1.00000
natural 0.00012 0.15352 0.42082 1.00000
basic   0.00754 1.00000 1.00000 1.00000 1.00000
advanced 3.2e-05 0.15019 0.45575 1.00000 1.00000 1.00000

P value adjustment method: bonferroni
```

These comparisons confirm the previous results: the semantic types Body part, Natural phenomenon, Basic technology (or concept) and Advanced technology (or concept) all give significant results when compared to Person, which suggest the following basic scale for degree of marking:

(134)  Person < Concrete (Body part, Natural, Basic, Advanced)

Further confirmation is provided using Tukey’s Honest Significant Difference test. But since there is some uncertainty as to whether the data are normally distributed, we double-check the result using the Kruskal-Wallis test. This confirms that there are significant differences between the groups ($p < 0.001$):

```r
> kruskal.test(markers ~ semType, data=m)

Kruskal-Wallis rank sum test

data:  markers by semType
Kruskal-Wallis chi-squared = 30.454, df = 6, p-value = 3.221e-05
```

Finally, we perform further multiple pairwise comparisons between groups (with correction for multiple testing and jittering to avoid warnings due to ties) using the Wilcoxon rank sum test:
In addition to confirming previous results, this test reveals differences at lower levels of significance ($p < 0.01$) between Animal on the one hand and Body part, Natural phenomenon and Advanced technology (or concept) on the other. There are also significant differences between Location on the one hand and Body part and Natural phenomenon on the other. This gives the following approximate scale:

(135) person, animal, location < basic < body part, natural, advanced

In conclusion, we have established that the number of markers a language uses does not differ significantly across geographical areas, but the number of markers used to express meanings does differ significantly across certain semantic types. This can be summed up as follows:

(136) Words denoting body parts, natural phenomena and advanced technologies (and concepts) tend to use significantly more marking than words denoting people, animals and locations. Words denoting basic technologies (and concepts) fall in between the two extremes.

7.4 Chapter summary

In this chapter I investigated three typological issues in some detail: constituent order, the relationship between binominals and possessive constructions, and the interplay between morphosyntactic strategies and semantics.

The data suggest that right-headed binominals are approximately twice as frequent as left-headed, a ratio which is in line with that found by Bauer and at odds with that arrived at in the Morbo/Comp project. In terms of morphosyntactic strategies, a number of interesting observations can be made. The der type is overwhelmingly (95%) right-headed, and right-headedness is also favoured by cmp (81%), adj (79%), cls (73%) and gen (70%); conversely, left-headedness is favoured by prp (93%), con (63%) and dbl (60%), though the numbers for cls and dbl, in particular, should be taken with a grain of salt due to the small sample sizes for these types.
Perhaps the most striking finding at this level is that the jxt strategy, with a 51:49 split, exhibits no particular preference. When languages exhibit “inconsistent” (or mixed) order, this is often due to the effect of including the predominantly right-headed derivational strategy (der), but six languages are mixed even without der. For one of these, Vietnamese, the explanation is clear (loanwords from Chinese); for the others, some evidence of a correlation between morphosyntactic strategy and semantic relation was found, but these require further investigation with more data.

To test the hypothesis that binominal lexeme constructions recruit their strategy from adnominal possession constructions I developed a methodology for comparing non-binary typologies. This involved defining a scale for expressing the degree of similarity between two constructions, the Pepper scale, which consists of five grades: always, mostly, sometimes, rarely and never. I then provided a set of definitions for these five grades suitable for the comparison of binominal and anchoring possessive constructions that allowed me to operationalize the comparison in a purely objective manner. I also showed how the scale could be adapted for more fine-grained analysis by splitting cases of grammaticalization out from the never category. As a result, I was able to define two universals:

(130) With overwhelmingly greater than chance frequency, languages recruit at least one of their binominal strategies from anchoring nominal modifier constructions.

(131) If a language does not recruit at least one of its binominal strategies from an anchoring nominal modifier constructions, there is a strong tendency for it to use compounding for this purpose.

I believe the Pepper scale has the potential for more general application in the field of language typology, whenever there is a need to compare two constructions that cannot be reduced to pairs of binary values suitable for expression using tetrachoric tables.

After explaining why we should not expect to find overall, crosslinguistic correlations between morphosyntactic strategies and semantics, I approached the question from two directions, one intralinguistic, the other cross-linguistic, using statistical methods. Drawing on Koch’s idea of a motivational grid I showed that, even with the limited set of data available for each language, it is possible to find significant structure. I illustrated this with the Polish data and then produced a table listing 22 languages that appear to be deserving of further investigation in this respect.
For the investigation of cross-linguistic patterns I abstracted away from the set of nine binominal types and focused on the number of markers. Doing so allowed me to assign a value of 0, 1 or 2 to each binominal and then calculate weighted values for each language and for each meaning. It was then possible to test groupings of languages and meanings for statistically significant differences in terms of degree of marking. For languages I focused on areal distribution (since the data is too sparse for comparisons based on genetic affiliation) and found nothing significant. For meanings, on the other hand, there were clear differences in terms of the amount of marking found in binominals used to denote meanings of different semantic types, which I summed up as follows:

(136) Words denoting body parts, natural phenomena and advanced technologies (and concepts) tend to use significantly more marking than words denoting people, animals and locations. Words denoting basic technologies (and concepts) fall in between the two extremes.

In sum, the data revealed a number of interesting patterns that are worthy of further investigation, but many of them require more data. Some could be studied using fewer meanings but would require coverage of more languages, others necessitate more data for each language (in other words, more meanings), but not such a large sample. 100 meanings from 100 languages was thus in some sense the ideal size for an exploratory, hypothesis-generating study, even though it was sometimes too small to produce immediate results.
8 Conceptual generalizations

Despite the exploratory, hypothesis-generating nature of the present study, there has been from the outset the germ of a hypothesis waiting to be tested. This is the idea, first formulated in Pepper (2010), that there are two fundamentally distinct ways of providing mental access to a complex concept that involves two other, less-complex concepts. In this section I revisit the original claim and then examine it in the light of evidence available in the binominals data (§8.1). Following this I approach the hypothesis from the perspective of salience: in §8.2 I look into the relative salience of head and modifier, and in §8.3 the salience of elaboration sites. Finally, in §8.4 I permit myself an excursion into a more tangential topic (albeit one related to head-modifier salience), where I suggest the possibility of a “species-attribute typology”.

8.1 The two-paths hypothesis

The basis for the two-paths hypothesis (originally called the “dual strategy”) was the presence of two types of compound in Nizaa, left-headed (jxtL) and right-headed (jxtR). I discovered that these exhibited two disjunct sets of semantic relations (Table 53). The relations that I found amongst left-headed compounds included LOCATION, PURPOSE, ACTIVITY, APPEARANCE, etc., while those found amongst right-headed compounds included PART, KIND, LOCATED and POSSESSION, etc.

For the Nizaa study I reinvented the wheel by developing my own set of semantic relations. These do not correspond directly to the ones used in the present study, but their names are fairly transparent if one bears in mind that they generally denote the role played by one of participants in the relation. For example, LOCATION is a relation in which the modifier denotes a location (137a), while LOCATED is a relation in which the modifier denotes something which has a location (137b). The non-transparent term IDEM was used for any kind of appositional relation (137c).

(137) Nizaa (Pepper 2010)
    a. LOCATION  siyw nim [bird water] ‘duck’
    b. LOCATED    yir nim [eye water] ‘tear’
    c. IDEM        siyw ɲʊdɔ̀ [bird cock] ‘cock’
What was striking about the results of my analysis was that there was no overlap between the two sets of relations: none of those found in left-headed compounds ever occur in right-headed compounds, and vice versa. The results might seem too good to be true, but there can be little doubt about the overall trend.

This clearly called for an explanation, which I attempted to provide from within the framework of Cognitive Grammar. I observed that cmpR relations (i.e. those found in right-headed compounds) were those typical of possessive constructions, which Langacker analyses in terms of a cognitive facility called the reference point ability; in such compounds, mental access to the target concept (e.g. ‘tear’) is provided via a related concept (e.g. ‘eye’). In the case of cmpL relations, I argued, a different cognitive facility appears to be at work, namely the ability to categorise (and sub-categorise); in this case, mental access to the target concept (e.g. ‘duck’) is provided via a more general concept (e.g. ‘bird’), and its extension reduced through reference to a characteristic feature (e.g. ‘water’).

---

1 I address this question in Pepper (2016: 301). It would be interesting to reanalyse the Nizaa data with the aid of the Bourquifier and (preferably) multiple raters.
In my 2016 paper I restate this analysis in the context of the three-way classification of compounds developed by Scalise and Bisetto. As discussed earlier (§2.1.4), this classification identifies three basic types, subordinate, attributive and coordinate, but fails to define them and, as a result, the scheme gets applied inconsistently. (I discuss the case of windmill, which is classified as subordinate by the authors and as attributive by Lieber.) This raises the following question:

What does “subordination” actually mean in the context of the relation between two nominals? In the case of a verb and a nominal its meaning is fairly clear: if the nominal can be regarded as an argument of the verb, then it is subordinate to it, but in what sense is mill subordinate (or not, as the case may be) to wind in windmill? (p. 290).

My proposal is that the notion of subordination should be tied to that of reference points. Doing so allows us to rephrase the question as follows:

Is it more likely that WINDMILL is conceptualized primarily as a MILL of a particular kind, or as an entity closely associated with the (more salient) concept of WIND? The former would point towards an attributive interpretation and support the classification proposed by Lieber (2010); the latter would indicate a subordinative interpretation and support that of Bisetto and Scalise (2005) (p. 306).

In my paper I leave the decision to the reader, on the grounds that “such judgment calls inevitably involve an element of subjectivity”, and that there is reason to believe that the two types of compound sometimes shade into each other. Here I can reveal that in the present study the relation between WIND and MILL is regarded as one of USG-R (“a windmill is a mill that uses wind”), which, as the Bourquifier shows, maps to Hatcher’s MtoH, is subordinative (Figure 73).

<table>
<thead>
<tr>
<th>Bourquifier3</th>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>windmill</td>
<td>wind</td>
<td>mill</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 73: The Bourquifier – WINDMILL**

2020-05-26 Submitted version 297
The evidence from Nizaa for the two-paths hypothesis resides solely in the fact that the two strategies surface directly in the grammar (in the form of two types, \texttt{jxtL} and \texttt{jxtR}). Importantly, it does not hinge on the order of constituents; in Nizaa the two constructions do in fact iconically diagram the mental route that the conceptualizer needs to follow in order to identify the target concept (to adapt the words of Taylor 1996: 18). Thus, when a reference point is involved, it precedes the target (as ‘eye’ precedes ‘water’ in 137b), whereas when (sub-)categorisation is employed, the broader concept precedes the modifier (as ‘bird’ precedes ‘water’ in 137a). While this is noteworthy, such iconicity is not crucial to the argument, as we shall see below.

It is also important to note that the claim relates to the process of conceptualization and does not require that the two paths always surface in the grammar. In other words, I hypothesise that the two strategies are used by speakers of every language, including those that do not provide explicit syntactic or morphological evidence of the fact. However, it was already clear to me that Nizaa is not alone in providing such evidence, and I envisaged that a cross-linguistic investigation into binominals could bring more to light.

I expected evidence for the two-paths hypothesis to take the form of a significant one-way correlation between morphosyntactic strategy (which would have to be at the level of my variable \texttt{ftype2} in order to capture differences between left- and right-headed forms of the same higher level \texttt{ftype}) and semantic relation (either at the lower-level of Bourque2 or at the higher level of Hatcher2). Moreover, it would have to be possible to plausibly relate the kinds of semantic relation involved in the correlation to the notions of reference point and subcategorization.

Unfortunately for the hypothesis, no such statistically significant correlations were found across languages. It is true that the analysis in §6.4 revealed some minor differences between binominal strategies in terms of the kinds of semantic relation that they express. However, these differences have not been pursued, for the reason given §7.1.3: that the morphosyntactic strategies of a comparative concept are not functionally commensurable across languages.

Compared to this, the analysis in §7.3.1 using the chi-squared and Fisher’s Exact tests on language-specific motivational grids is more theoretically sound, and the results indicate that there may be significant associations between the relations expressed by binominals and the choice of morphosyntactic strategy in as many as 22 of the 106 languages. However, this analysis came up against the problems of
sparse data and small sample size. This is well illustrated by the fact that the one
language that on the basis of previous knowledge was expected to provide evidence
in support of the hypothesis, Mapudungun, did not even make it onto the list of 22
languages in Table 48 (page 285). This is surely because of the sparsity of the data,
as shown in the motivational grid on page 265 (Table 36).

In sum, the present data provide no significant evidence in support of the two-paths
hypothesis. But this, in itself, does not disprove the hypothesis. It could just be that
more data is required. A more likely alternative is that the semantic relation is just
one of several factors that play a role in influencing the user’s choice of construction.
The most obvious competing motivation is alienability, which is known to be an
important factor in determining the surface form of possessive constructions in
many languages (e.g. Chappell & McGregor 1996; Aikhenvald & Dixon 2013;
Haspelmath 2017b). The kinds of relation found in prototypical possessive con-
structions (ownership, kinship, part-whole, etc.) are exactly those that involve the reference point ability. Giving these relations different surface realizations depending on the alienability of the possessum will interfere with the simple two-way
distinction predicted by the two-paths hypothesis. If there are other, uncorrelated
factors at play, the interference will be even stronger.

In order to investigate what the various factors might be, and how they interact, we
can examine in more depth a language where there are several competing construc-
tions. Such a language is Kalamang, mentioned previously in §7.1.3 as a language that exhibits no basic order.

### Table 54: Motivational grid for Kalamang with row sums

<table>
<thead>
<tr>
<th>htype</th>
<th>Mish</th>
<th>HinM</th>
<th>MinH</th>
<th>HtoM</th>
<th>MtoH</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftype2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjL</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>cmpR</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>conR</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>genL</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>jxtL</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>jxtR</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>prpL</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

chi-squared p = 0.000646754003969361
x-squared = 52.6318355973528 df = 24
Fisher’s exact p = 0.00199900049975012

Table 54 shows the motivational grid for Kalamang, here extended with row sums.

---

1 For my analysis of Nizaa I had some 200 compounds to work with. In the present study the average number of binominals per language is a mere 35.3 (see Table 23 on page 116).
jxtL and conR account for slightly over 50% of the data and the third most frequent type is jxtL. Since the latter has almost the same profile (or fingerprint) as cmpL, we are justified in combining the two, ignoring the different degrees of fusion that are indicated by the presence of a word space in the former and its absence in the latter. This combined type, which I will label cmbL, comprises all left-headed “compounds” (in the broad sense) and has 18 instances. In order to avoid confusion, I will also relabel jxtL, as cmbL, even though there are no instances of cmpL to combine it with. In the resulting system, 47 of the 56 binominals in the Kalamang data set are distributed almost equally across three structural types: cmbL, cmbR and conR (Table 55). The question we wish to answer, is what kind of system motivates this distribution.

<table>
<thead>
<tr>
<th>htype</th>
<th>MisH</th>
<th>HinM</th>
<th>MinH</th>
<th>HtoM</th>
<th>MtoH</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftype2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cmbL</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>cmbR</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>conR</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 55: Simplified motivational grid for Kalamang

The principal difference between cmbL and cmbR is that the former favours the MisH relation and the latter the HinM relation, that is, attributive and subordinative relations, respectively. More striking, however, are the profiles of the two right-headed constructions, cmbR and conR, which tend to encode the same high-level relation, HinM. I will therefore focus on the data points that underpin this particular statistic. Table 56 lists all 32 binominals that instantiate these two types, grouped by type, together with the corresponding meaning and semantic type, the gloss for head and modifier, and both types of semantic relation: in short, every property for which these binominals have been annotated in the database.

Now, if any of these properties were responsible for the choice of morphosyntactic strategy, there should be no overlap between the values of that property across the two types, just as there was no overlap between the semantic relations exhibited by the cmpL and cmpR compounds in Nizaa. In fact, it turns out that there are overlaps in every one of these properties. We already know from Table 54 that the high-level relation HinM is found in both types; Table 56 shows that the low-level relation MER is also common to both, so it is clearly not a matter of the semantic relation. Likewise, the semantic types Body part and Basic are common to both, so the property semantic type can also not be the criterion by which concepts are encoded using different morphosyntactic strategies.
Nor does alienability appear to play a role. If it did, we would not expect the same concept to appear as the head in both sets of binominals, but it does: *kang* ‘bone’
can is the head of \textit{(conR) aknar kang.un} COLLARBONE, as well as various \textit{(cmbR)} words denoting the body parts RIB, SHOULDERBLADE, SKULL and SPINE. Bones that are part of the human anatomy are surely equally inalienable, irrespective of the particular body part to which they belong?

As to the role and type of the modifier, both ‘eye’, ‘hand’ and ‘fire’ appear as the first (left-hand) constituent of both types. There is even a minimal pair (138) of words that both use the MER relation to denote a Body part, where the modifier is ‘eye’ and the heads are both body-related (‘hair’ and ‘skin’), but that exhibit two different structural types.

\begin{enumerate}
\item \textit{cmbR} \textit{kanggir nenen} [eye hair] EYELASH
\item \textit{conR} \textit{kanggir pul.un} [eye skin.3POSS] EYELID
\end{enumerate}

To sum up: None of the recorded properties of these binominals can account for the choice of structural type used to express them. Of course, it is unlikely that the choice is random, but understanding the actual motivation has to remain a topic for further research. Although our attempt to explain the motivation for the two main structural types in Kalamang ended in failure, the method applied here may prove fruitful with other languages, However, that, too, is a topic for future research.

8.2 Head-framing vs. modifier-framing

Another way to think about the two-paths hypothesis is to pose the question \textit{why} one path of mental access would be used in preference to the other. One plausible answer relates to salience. There are two aspects to this:

\begin{itemize}
\item If one of the two concepts to be combined is significantly more salient than the other, then it might constitute a better starting point for the path (of mental access) than the other.
\item Alternatively, if one of the elaboration sites (or “slots”) of the two concepts is significantly more salient than the other, then the concept it belongs to might constitute the better starting point.
\end{itemize}

I discuss the first of these here and the second in the next section. Both discussions require an understanding of the concept of salience and a way to operationalise it.

Salience is defined as “the quality of being particularly noticeable or important; prominence”.\footnote{https://en.oxforddictionaries.com/definition/salience} Koch (2001: 1151) notes how a major issue in prototype theory is “the discovery of salience effects in the vertical dimension of (folk) taxonomies:
the ‘basic level’ of categorization, e.g. BIRD, is cognitively more salient than the ‘superordinate level’, e.g. ANIMAL, and the ‘subordinate level’, e.g. ROBIN.” For the sake of simplicity, I will assume that salience correlates to a large degree with frequency: that things that are more salient are attended to more frequently. With this as our starting point, consider Table 57, which lists every gloss for the head and modifier respectively, for each of the three meanings EYELID, TRAIN and MARE.

Amongst binominals denoting EYELID there are 31 different head constituents and just four different modifier constituents. The contrast between the multiplicity of head concepts and the paucity of modifier concepts is striking, and this would still be the case if synonyms like animal skin ~ skin ~ skin/hide were to be consolidated into a single concept: the count of 31 might be reduced to 10. The contrast in what could be called “conceptual variation” is even more striking given that the modifiers face, lid and surface only occur once each: of the 70 binominals denoting EYELID, 67 use a word meaning EYE as the modifier. I call this pattern, where there are few modifiers and many heads, ‘modifier-framed’ (MF) and denote it symbolically as (1,10); the latter will shortly be used as coordinates in a scatter plot.

<table>
<thead>
<tr>
<th>EYELID (4,31) → (1,10)</th>
<th>TRAIN (9,5) → (10,1)</th>
<th>MARE (6,6) → (2,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>modifiers</strong></td>
<td><strong>heads</strong></td>
<td><strong>modifiers</strong></td>
</tr>
<tr>
<td>eye, face, lid, surface</td>
<td>animal skin, back, body</td>
<td>electricity, fire, ground, land, outside, rail, railway, smoke, steam</td>
</tr>
<tr>
<td></td>
<td>hair, brim, cover, cup-like shape, DIM, eyelid, fabric, flap, hair, half, hand, hut, layer, leaf, lid, little cap, meat, rim, roof, root, sack/bag, sheath, shell, skin, skin/hide, small piece, superior part, top, up</td>
<td>boat, car, cart, tree, vehicle</td>
</tr>
<tr>
<td>4 → 1</td>
<td>31 → 10</td>
<td>5 → 1</td>
</tr>
</tbody>
</table>

Table 57: Conceptual variation in heads and modifiers

The pattern exhibited by the 19 binominals that denote TRAIN is quite different. Here there are five different heads and nine different modifiers. Three of the five heads (boat, car and cart) are hyponyms of the fourth (vehicle), and the word glossed as ‘tree’ in Baa kí.sà [tree.outside] is also used more generally to denote...
vehicle (cf. *kí.kpà* [tree.water] ‘canoe’ and *kí.inyi* [tree.sky] ‘airplane’, Mirjam Möller, p.c.). So there is essentially just one kind of head. Modifiers, by contrast, vary substantially, so the pattern here (with many modifiers and few heads) is ‘head-framed’ (HF), which I will denote symbolically as (10,1).

The meaning MARE exemplifies a third type. At a superficial level there are six different heads and six different modifiers, but both heads and modifiers can be reduced to just two basic concepts, HORSE and FEMALE, giving a ‘neutrally-framed’ (NF) pattern (few heads, few modifiers), which can be denoted as (2,2). Note that, in this analysis, both HORSE and FEMALE can occur as either the head or the modifier. The implications of this are discussed in §8.4.

The very different patterns of conceptual variation exemplified by EYELID, TRAIN and MARE can be illustrated schematically as shown in Figure 74, in which the horizontal axis represents the diversity of modifier concepts and the vertical axis the diversity of head concepts. A meaning like EYELID (MF) belongs in the top left corner of the quadrant; TRAIN (HF) belongs in the bottom right; and MARE (NF) near the bottom left. As implied above, MARE constitutes a special case that will be discussed in depth later; it is the contrast between EYELID and TRAIN that concerns us here, since it relates to the two-paths hypothesis.

![Figure 74: Conceptual variability of head and modifier](image)

My proposal is that the same two paths of mental access (via a related concept and via a superordinate concept, or hypernym) are inherent in the conceptualizations of EYELID and TRAIN. EYELID has a strong tendency to be conceptualized as an object *related to* the EYE, whereas TRAIN tends to be conceptualized as a VEHICLE of a particular *(sub)type*. This influences the degree of conceptual variation in the head and modifier, as follows:
• Given the choice of EYE as the reference point (modifier) for EYELID, there are relatively few constraints on the choice of head; its task is simply to distinguish EYELID from other eye-related concepts, such as EYEBALL, EYEBROW or EYELASH. As a consequence, there is room for considerable variety in the head; skin, cover, hut, shell, cup-like shape: all of these and more will do the job. The MF pattern thus corresponds to the reference point path.

• With TRAIN, the situation is the reverse: once the hypernym VEHICLE, has been selected as the starting point for the conceptualization, the only constraint on the modifier is that it should serve to distinguish TRAIN from other types of vehicle. This can be done in numerous ways and, consequently, there is room for considerable variety in the choice of modifier: electricity, fire, railway, iron, steam, etc. The HF pattern thus corresponds to the hypernym path.

EYE and VEHICLE are chosen as starting points for the two conceptualizations, I suggest, because they are salient, basic-level concepts, and this is reflected in the small amount of conceptual variation amongst the modifiers of EYELID binominals and the heads of TRAIN binominals.

This hypothesis can be explored by comparing the degree of conceptual variation in the head and modifier concepts of all 100 meanings. The ideal way to proceed would be to first consolidate synonyms and near-synonyms into synonym sets. This could be done through introspective analysis (as I did with EYELID, TRAIN and MARE, above). A more principled approach would be to use a thesaurus, such as WordNet, a large lexical database of English, in which synonyms are grouped into “synsets”, which themselves are connected via hyponymy relations (Fellbaum 2010). Both of these approaches are beyond the scope of the present study and must be left for future work. Instead I take a less nuanced approach, in which each word is counted, irrespective of its status as a synonym. For the three meanings in Table 57 this results in the following counts: EYELID (4,31), TRAIN (9,5) and MARE (6,7). The advantage of this approach is that the counts can be performed directly on the data in the database, without additional qualitative analysis. The results of doing so with all 100 meanings is shown in Figure 75 and Figure 76.
Figure 75 shows a scatterplot of all 100 meanings using the first four letters of the name as labels. The vertical axis gives the number of different glosses used for the head and the horizontal axis the number of different glosses used for the modifier. Meanings are grouped by semantic type, which is indicated by different symbols and different colours. Thus, for example, the Body part EYELID is indicated by a blue × and the label eye1, and positioned at coordinates x=4, y=31. We observe a strong tendency for body parts to cluster with EYELID in the top left (MF) corner of the plot, with EARLOBE and PALM (OF THE HAND) exhibiting even more variability in their heads than EYELID. The tendency is for body part binominals to be named in terms of a more salient body part. The main exception to this pattern is the bony body parts; here the more salient concept is BONE, the variety of heads is much less, and such meanings are found in the lower portion of the plot.

The Advanced concept TRAIN is found at coordinates (9,5), which turns out to be higher up the modifier scale compared to other Advanced concepts, but it is by no means the concept with the highest number of modifiers; that honour goes to three Person concepts: SORCERER OR WITCH, NIECE and MOTHER-IN-LAW (OF A MAN). The feature that characterizes Advanced concepts is not so much the multiplicity of modifiers but the paucity of heads. This is no doubt because most new-fangled objects are conceptualized as specialized subtypes of a more familiar (and more general) concept.
The different ways in which semantic types distribute across the head/modifier variability space can be observed more clearly in the heatmap shown in Figure 76. Here the seven semantic types are separated out and dots are replaced by hexagons of varying densities. As we just observed, Advanced concepts tend strongly to have rather few heads and this is indicated by the dense hexagon in the far left panel. In the far right panel we see that most of the Person category cluster around the (10,10) region; these are mostly professions. Body parts, as previously noted, tend to occupy the MF (top left) area with few modifiers and many heads, while most animals, including MARE, are to be found in the lower-left, NF (neutrally-framed) area.

![Figure 76: Heatmap of heads and modifiers by semantic type](image)

Meanings might also be positioned on a one-dimensional scale based on the ratio of heads to modifiers. On such a scale, TRAIN would have the value (5/9=) 0.56 (i.e. towards zero); MARE would have the value (6/6=) 1.0 (exactly one); and EYELID would have the value (31/4=) 7.75 (considerably more than one). The lowest value would be for PADDLE WHEEL (1/10=) 0.1 and the highest EARLOBE (35/3=) 11.67. However, the ratio-based method would obscure the difference between MARE and a concept like RAINBOW, which shows almost the same ratio (1.07) but far greater multiplicity of heads and modifiers (15 and 14, respectively).

This investigation demonstrates marked differences across meanings in terms of the number of heads and modifiers employed in conceptualization: some meanings are routinely conceptualized in terms of one and the same modifier concept (e.g. EYE in EYELID), others in terms of the same head concept (e.g. VEHICLE in TRAIN). These correspond precisely to the two hypothesized paths of mental access – via a reference point and via a superordinate concept – and provide evidence in support of the two paths hypothesis. The case of MARE and its ilk, where we find the same two concepts for head and modifier, will be examined in §8.4.
8.3 E-sites and the relational/sortal cline

In the previous section I focused simply on the salience of heads and modifiers, but salience has other aspects as well. At a doctoral defence in 2016, Ron Langacker wondered whether the salience of “elaboration sites” in noun-noun compounds could be measured empirically. He gave mill as an example (139).

(139) product (pepper mill, salt mill, flour mill, paper mill)
      power source (windmill, water mill)
      part, use (saw mill)

Langacker (1991) defines elaboration site as follows: “In a construction, those facets of one component structure that the other component serves to elaborate”. In the later wording of Langacker (2008: 198), an e-site is a “schematic substructure” of one component of a construction, “which the other component serves to elaborate, i.e. characterize in finer-grained detail”. E-site thus equates to ‘slot’ in the slot-filler theories described on page 185. Of Langacker’s examples, the most pertinent in the context of binominals is jar lid, where “lid evokes a schematic container specified in finer detail by jar”. In our examples, mill evokes both a PRODUCT, a POWER SOURCE, and (perhaps to a lesser degree) some COMPONENT that is of central importance to its function. Each of these constitutes an e-site; in (139) they are elaborated by pepper, wind, saw, etc.

Implicit in the notion of e-site, but not discussed by Langacker, is the existence of a semantic relation between the component to which the e-site belongs (e.g. mill) and the component that elaborates it (e.g. pepper). In the case of pepper mill, the Bourquifier tells us that this is PROD: a pepper mill is a mill that produces (or processes) pepper; in the case of windmill it is USG-R; and in the case of saw mill it is MER-R.¹ Given the kind of data collected and annotated for the present project, it is therefore possible to identify e-sites empirically, by querying for the kinds of semantic relation that a head constituent participates in. The raw results of such a query are shown in Table 58, which lists the ten head concepts that participate in the greatest diversity of relations. Thus, house participates in seven kinds of relation; head, place, string, thread, tree and wheel in five, etc. (Note that I have not attempted to consolidate synonyms in the manner discussed in the previous section, otherwise string and thread might have been combined, and a synset that includes skin, hide, pelt, and perhaps even bark, cf. 9 on page 57, might have made it into the Top Ten.)

¹ The relations LOC-R and USG-R are acceptable alternatives here. MER-R and LOC-R both map to MinH, whereas USG-R maps to MtoH. As observed earlier, there is often overlap between different relations.
Table 58: Salience of elaboration sites (by relation)

Interpreting Table 58 is not entirely straightforward since the codes are mnemonics for the kind of relation and not the role played by the head in that relation. For example, while POSS clearly indicates the relation of POSSESSION, the code itself does not tell us whether the head constituent is the possessor or the possessum. In order to figure that out, we need to refer back to Table 31 on page 238, where the template for POSS tells us that the relation involves “(an) H that (an) M possesses”. If the head is house, as it is in the first row of the table, then the e-site indicated by the POSS relation is the one that would be elaborated by a modifier indicating the possessor. Conversely the template for POSS-R (“(an) H that possesses (an) M”), which we find with the head father in the penultimate row, tells us that the e-site would be elaborated by a modifier indicating the possessum.

Similarly, the template for LOC-R (“(an) H that (an) M is located at/near/in”) tells us that the e-site concerns the kinds of thing that might be located at, near or in the house, where the house itself is the location; it does not concern the location of the house (which would be indicated by the inverse relation LOC).

In turns out that it is the role played by the modifier (the elaborating component) that best characterizes the e-site. This suggests that it would be useful to supplement the Hatcher-Bourque classification scheme with standardized labels for the roles played by the head and modifier respectively. I will return to this matter in §9.3. If we relabel Table 58 using the modifier role instead of the relation (Table 59), the nature of each e-site becomes clear.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>head</th>
<th>elaboration sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSE (7)</td>
<td>Possessor, Located, Purpose, Material, Location, Part, Whole</td>
</tr>
<tr>
<td>HEAD (5)</td>
<td>Whole, Possessor, Located, Result, Material</td>
</tr>
<tr>
<td>PLACE (5)</td>
<td>Located, Purpose, Cause, Location, Result</td>
</tr>
<tr>
<td>STRING (5)</td>
<td>Possessum, Possessor, Part, Purpose, Whole</td>
</tr>
<tr>
<td>THREAD (5)</td>
<td>Possessor, Purpose, User, Whole, Possessum</td>
</tr>
<tr>
<td>TREE (5)</td>
<td>Part, Whole, Located, Location, User</td>
</tr>
<tr>
<td>WHEEL (5)</td>
<td>Part, Purpose, Whole, Location, Used</td>
</tr>
<tr>
<td>BONE (4)</td>
<td>Whole, Location, Located, Possessum</td>
</tr>
<tr>
<td>FATHER (4)</td>
<td>Possessor, Possessum, Product, Whole</td>
</tr>
<tr>
<td>HOLE (4)</td>
<td>Whole, Located, Purpose, Possessor</td>
</tr>
</tbody>
</table>

*Table 59: Salience of elaboration sites (by role)*

We see at once that HOUSE has seven e-sites, for its possessor, the things located in it, its purpose, material and location, its parts, and the whole of which it is a part. These seven sites are elaborated by the components shown in Table 60, where the numbers in the first column show the number of different components per e-site, e.g. Purpose (13), and in the second column the number of tokens of each component, e.g. excrement (3). (Recall that these results are based on the raw, unnormalized data, so both ‘animal’ and ‘animals’ occur under Purpose, ‘bee’ and ‘bees’ under Possessor, etc.) Given some familiarity with the meaning list it is easy to infer just from this table that the meanings TOILET and STABLE OR STALL utilize the Purpose e-site of HOUSE; HOSPITAL and COOKHOUSE utilize the Location e-site, SPIDER WEB and BEEHIVE the Possessor e-site, etc.

<table>
<thead>
<tr>
<th>e-site</th>
<th>elaborating component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose (13)</td>
<td>excrement (3), urine (3), bovine (2), horse (2), livestock (2), animal (1), animal/cattle (1), animals (1), donkey (1), hand (1), horses (1), pig (1), poo (1)</td>
</tr>
<tr>
<td>Located (12)</td>
<td>doctor (5), medicine (3), food (2), hospital (2), bath (1), disease (1), fire (1), illnesses (1), kitchen (1), oven (1), pain (1), stove (1)</td>
</tr>
<tr>
<td>Possessor (3)</td>
<td>spider (11), bee (9), bees (1)</td>
</tr>
<tr>
<td>Material (2)</td>
<td>leaf (1), straw (1)</td>
</tr>
<tr>
<td>Location (1)</td>
<td>side (1)</td>
</tr>
<tr>
<td>Part (1)</td>
<td>incense (1)</td>
</tr>
<tr>
<td>Whole (1)</td>
<td>nose (1)</td>
</tr>
</tbody>
</table>

*Table 60: Components that elaborate the concept HOUSE*
With this approach, it is possible to measure the salience of e-sites empirically: as the frequency with which it is used in concept combination. Based on the present data, we see that the most salient e-sites of the concept HOUSE are Purpose (i.e., what the house is used for) and Located (what it contains). The other five e-sites (Possessor, Material, Location, Part and Whole) are also inherent in the concept, but are much less frequently used to denote a more specialized meaning of HOUSE, and can hence be said to be less salient. Saliency can be measured either by the type frequency of the elaborating components (cf. the first column of Table 60), which for HOUSE would give a value of 3 for Possessor (2 after normalization); or by the sum of the token frequencies of the elaborating components (cf. the second column), which would give a value of 21 (11+9+1) for Possessor. For Purpose the corresponding values would be 13 (reduced to perhaps 3 after normalization and synonym consolidation) and 20; for Location they would be 12 (9) and 20; etc.

Admittedly, the present data are limited by the fact that they are based on just 100 meanings. Out of these, an element glossed as [house] occurs as the head of 11 meanings which are expressed by elaborating the seven e-sites listed above: BEEHIVE, BEESWAX, COOKHOUSE, GLOVE, HOSPITAL, NEIGHBOUR, NOSTRIL, SPIDERWEB, STABLE OR STALL, THATCH and TOILET. A different sample of 100 meanings would lead to a different list of meanings in which [house] occurs as the head of the binominal, and this may lead to somewhat different results in terms of e-site salience. However, since the original selection of 100 meanings was made entirely independently of any considerations regarding the kinds of semantic relation they embody or the elaboration sites that might be involved, there is no reason to believe that the results attained above are not reliable.

To conclude this section, I would like to reflect on the fact that in traditional grammar nouns are divided into two types: sortal and relational:

Nouns have two basic interpretations. Taken in isolation they can be considered either sortal nouns or relational nouns. Sortal nouns classify objects, whereas relational nouns describe objects as standing in a certain relation to others. These are two fundamentally different ways to characterize objects, and one cannot be reduced to another (Löbner 1985: 292).

Löbner illustrates the difference with the minimal pair woman and wife, the former sortal, the latter relational, since “a “wife” is always the wife of someone”. Kinship terms like brother, uncle and daughter are typically used to illustrate the notion of relational noun, but Barker (2011) lists other pairs of nouns that “contrast minimally with respect to the sortal-relational distinction”:

2020-05-26 Submitted version 311
sortal relational
a. a day (*of someone) a birthday of someone
b. a person (*of someone) a child of someone
c. an animal (*of someone) a pet of someone

This prompts the question, what is the status of house: is it sortal or relational? If pet is relational because a pet can be owned, surely house is also relational, and not just because it, too, can be owned, but also because it has a purpose (its TELIC, in Pustejovsky’s 1995 terms) and contents (its CONSTITUTIVE). Pustejovsky himself (p. 96) talks of “implicitly relational nouns”, such as door, book, newspaper and window. It would seem that the distinction between sortal and relational nouns is not binary at all, but rather a cline in which concepts exhibit differing degrees of relationality. If that is the case, then the method developed here can provide a measure of the degree to which individual concepts are relational. This measure could be as simple as a count of the number of e-sites (cf. the head column of Table 59), which would allow us to claim that HOUSE is the most relational concept in our data, followed by HEAD, PLACE, STRING, etc.

In conclusion, the kind of data collected for the present study lends itself rather well to the task of identifying elaboration sites and measuring their salience, and also to measuring the degree of relationality of a concept along a cline from purely sortal to highly relational.

8.4 Species-framing vs. attribute-framing

The discussion of the relative salience of head and modifier in §8.2 concerned the contrast between the modifier-framed pattern exemplified by EYELID and the head-framed pattern exemplified by MARE. In this section I consider the neutral pattern exemplified by MARE. What is striking about this example is not so much that we find relatively little conceptual variation in either the head or the modifier: this is a property, as can be seen in Figure 76, of many other meanings, in particular Advanced concepts. The outstanding feature of MARE is that the same two concepts, HORSE and FEMALE, can occur as either the head or the modifier, as illustrated in (140), where the head constituent is underlined for reasons of clarity. In (a) and (b) the head is HORSE and in (c) and (d) it is FEMALE.

(140) a. Amharic ye.seti feres [GEN.female horse] GEN.Mod Head
     b. Hawaiian lio wahine [horse female] Head Mod
     c. Tagalog babai.ng kabayo [female.LK horse] Head.LK Mod
     d. Wichi yel’ataj atsinya [horse.female] Mod Head
This raises the question of how one determines the head in such cases. The ISA test tells us that a mare is both a horse and a female, so should not the examples in (140) be analysed as coordinate compounds, which are often said to have two heads (or, alternatively, none)? Doing so with (a) and (c) would be imprudent, since the morphology indicates the head and modifier quite explicitly. The genitive prefix in Amharic always attaches to the modifier, and the Tagalog linker always attaches to the head, so we have clear evidence that the head of (a) is HORSE and the head of (c) is FEMALE. In the case of (b) and (d) there is internal evidence: compounds are consistently left-headed in Hawaiian and right-headed in Wichi.

These observations suggest that MARE is primarily regarded as a kind of HORSE in some languages, and as a kind of FEMALE in others: in some languages the concept is ‘framed’ in terms of the species, and in others it is ‘framed’ in terms of the gender. Now, MARE is not the only meaning used in the present study that has this property of what we might call ‘interchangeable’ heads and modifiers. It is found more generally in meanings of the type FEMALE X and also in meanings of the type MALE X. Moreover, a similar situation pertains with words denoting young animals, such as KID, where we find a combination of species and a different attribute, this time denoting age rather than gender (141). In (a) and (b) the head is GOAT, and in (c) and (d) it is CHILD or some equivalent. In (c) the evidence for this analysis is explicit, while in (a), (b) and (d) it is internal.

(141) a. Imbabura Quechua wawa chita [child goat] Mod Head  
    b. Hawaiian kao keiki [goat child] Head Mod  
    c. Tagalog bata.ng kambing [child.LK goat] Head.LK Mod  
    d. Wichi kaila lhos [goat.breeding/son] Mod Head

Observe that the three languages that appear in both (140) and (141) appear to be consistent in terms of framing: Hawaiian is ‘species-framed’ (the heads are HORSE and GOAT); whereas both Tagalog and Wichi are ‘attribute-framed’ (the heads are FEMALE and CHILD or its equivalent). This raises two questions:

1. Are languages always consistent in this respect?
2. What is the distribution of the two types, species-framed and attribute-framed?

The binominal data allow us to address these two questions on the basis of eight different meanings (142).
Words that denote one or more of these meanings using the COOR relation\(^1\) are found in 84 of the languages in the sample. Of these, 11 (13\%) are consistently species-framed, 57 (68\%) are consistently attribute-framed, and 16 (19\%) are mixed. Their geographical distribution is shown in Figure 77.

All 16 of the languages that mix species- and attribute-framing are consistent in terms of how they frame meanings of the type “young X”, with the exception of Mandarin Chinese, Trinitario and Western Farsi. In Mandarin, the meaning LAMB can be expressed as either 羔羔 gāoyáng or 羊羔 yánghāo (where 羊 is glossed as sheep/goat and 羔 as lamb/kid). Apart from this exception, Mandarin is consistently species-framed.

In Trinitario, three different constructions are employed (143). The internal evidence for which constituent of (b) and (c) to regard as the head is extremely meagre, but whatever the case, these two are not consistent so further investigation is required.

\[
\begin{align*}
(142) & \quad \text{a. } \text{FEMALE } \text{X:} \text{ MARE, QUEEN, NIECE} \\
& \quad \text{b. } \text{MALE } \text{X:} \text{ COCK/ROOSTER, WIDOWER} \\
& \quad \text{c. } \text{YOUNG } \text{X:} \text{ FOAL OR COLT, KID, LAMB}
\end{align*}
\]

\[
\begin{align*}
(143) & \quad \text{a. } \text{kwoy.gira [horse.DIM]} \text{ FOAL OR COLT} \\
& \quad \text{b. } \text{chiwa.gira } \text{'moyo [goat.DIM child]} \text{ KID} \\
& \quad \text{c. } \text{'moyo } \text{'vesa [child sheep]} \text{ LAMB}
\end{align*}
\]

\(^1\) Other relations that are often used to express such meanings are TAX for young animals, POSS for NIECE, and TOP for WIDOWER. Recall also Caijia’s old woman from the sky (SRC-R), cf. page 228.
And in Western Farsi, there are two constructions (144), cmp and jxt, and one of the constituents is not glossed, so here too it is not possible to adjudicate the matter.

(144) a. boz.qāle [goat.???] KID
    b. korre asb [young horse] FOAL OR COLT

There is less consistency amongst mixed languages in how they frame meanings of the type “female X”: MARE and QUEEN are framed differently in Barain and Ket, MARE and NIECE are framed differently in Oroqen and Tagalog, and NIECE and QUEEN are framed differently in Turkish. In Saramaccan, QUEEN can be expressed as either kônu.mujêê [king.woman] or mujêê.kônu [woman.king]; assuming both of these are right-headed, the former would be classed as attribute-framed and the latter as species-framed. It may be significant that all of these exceptions involve a human referent. In five of the remaining six ‘inconsistent’ languages (Kalamang, Kam, Malagasy, Malayalam and Wawa) the opposition is between age and gender, i.e. young X (which is attribute-framed) and (fe)male X (species-framed). The same would be true of Kildin Sami were it not for the case of liessk-kâll’es’ [widow-old_man] WIDOWER, which is attribute-framed; here again we perhaps see the system breaking down when applied to human referents.

Overall, the striking imbalance between species-framing and attribute-framing and the areal tendencies hinted at in Figure 77 are worthy of further investigation, but that is beyond the scope of the present study.

8.5 Chapter summary

In this chapter I pursued issues related to the two-paths hypothesis that I originally developed to account for data from Nizaa. I started by restating the hypothesis and suggested that the lack of direct support for it in the present data may be due to sparse data or competing motivations. A search for the latter based on properties recorded in the database for Kalamang, a language in which three morphosyntactic strategies are very frequent, proved fruitless, so the explanation must be sought elsewhere. On the other hand, the investigation into the salience of heads and modifiers can be said to have provided important supporting evidence for the hypothesis, perhaps even opened up a new field of research in semantic typology.

I also developed a method for quantifying the salience of elaboration sites and for measuring the degree of relationality of a concept. And finally, I looked into the ways in which languages combine concepts of species and their attributes, showing a strong preference for attribute-framing as compared to species-framing.
In Chapter 6 (pages 216–218) I allowed myself a rant about how researchers working on the semantic relations of compounding tend to reinvent the wheel by devising their own classification systems instead of reusing (and refining) existing systems, and I pleaded guilty to the same offence in my 2010 study of compounding in Nizaa. My point was that science cannot progress unless we build on each other’s work. To atone for my own transgression and to set a good example, I harnessed (and refined) two existing classifications in the present study, those of Hatcher and Bourque, and integrated them into a single two-tier system, the Hatcher-Bourque classification (see Table 31 on page 238). The system is shown in hierarchical form in Figure 78, now with the addition of a top-level tier representing Aristotle’s three principles of remembering, reference to which was made several times in Chapters 6 and 7. (The third principle, contrast, is included in the diagram as a placeholder, in the expectation that there may be a use for it later.)

![Figure 78: Semantic relations as a hierarchy](image)

There is an important difference between the two representations of the Hatcher-Bourque classification that should be pointed out. In both the hierarchical and the tabular representation, SIMILARITY and CONTIGUITY are distinguished, and the latter is subdivided into CONTAINMENT and CAUSATION. Beyond that the subdivision of contiguity differs. In Table 31 it is by relation (CONTAINMENT, POSSESSION, etc.);
in Figure 78 it is according to the two pairs of relations in the Hatcher classification (H⊆M vs. M⊆H and H→M vs. M→H), which amounts to subdividing by Basic vs. Reversed relation. There are advantages and disadvantages to each of these representations. Subdividing by relation means that Basic and Reversed forms are kept together (thus, for example, CONT and CONT-R are grouped together under CONTAINMENT in Table 31). However, the commonality between CONT, POSS and MER (etc.) on the one hand, and CONT-R, POSS-R and MER-R (etc.) on the other is lost. Subdividing by Hatcher’s pairs highlights this commonality, as the diagram shows, but the connection between Basic and Reversed forms is lost and we have to indicate them in the diagram by dashed lines.

In the process of developing the Hatcher-Bourque classification and applying it to the binominal data set, it became clear that many of the semantic relations found in compounds (and binominals more generally) are present in other domains as well. This would indicate that these relations instantiate a domain general cognitive ability to perceive relationships, an ability that is related to the associative nature of human thought. Moreover it suggests the potential for a unified system of relations that may have much wider application, both within linguistics and beyond. The purpose of this chapter is to follow up on those insights. The end result will be a proposal for a unified model of associative relations that can applied to concept combination in word-formation, metonymy, and lexical semantics, and extended to use in the fields of knowledge representation. Jackendoff’s suggestion that “we are dealing with a common stock of rather primitive semantic relations that can be expressed through various (morpho)syntactic frames, compounding among them”, cited on page 188, will turn out to have underestimated the potential.

In § 9.1 I build on Janda’s insight (§ 6.1.3) regarding the role of metonymy in word-formation. Drawing on the work of Peirsman & Geeraerts (2006) I show that many (perhaps most) metonymic relations are parallel to the semantic relations found in binominals and can be accommodated by the Hatcher-Bourque classification. In §9.2 I turn my attention to the kinds of cognitive relation discussed by Peter Koch and Andreas Blank in the domain of lexical semantics and show how these, too, can be accommodated by the Hatcher-Bourque classification with only minimal extensions. Finally, in §9.3, I return to Topic Maps, which I first mentioned in Chapter 1 in the context of my journey towards the topic (sic) of the present study.

---

1 This is in the nature of a hierarchical representation, which is why I avoided it §5.2.
This will provide further insights, not least into the importance of roles, and enable me to articulate a general model of associative relations.

9.1 Metonymic relations (Peirsman & Geeraerts 2006)

Metonymy is defined by Kövecses (2002: 145) as “a cognitive process in which one conceptual entity, the vehicle, provides mental access to another conceptual entity, the target, within the same domain, or idealized cognitive model (ICM)” (emphasis added). It involves two “entities”, the vehicle and the target, which are connected in such a way as to enable mental access from the one to the other.

To underpin her claim that metonymy is at work in word-formation, Janda (2011) uses a classification system modelled after that found in Peirsman & Geeraerts (2006) and I follow her in taking the latter as my starting point for this discussion. Before doing so I should note that Peirsman & Geeraerts’ purpose was to present metonymy as a prototypical category. They argue that

metonymies, starting from spatial part-whole contiguity as the core of the category, can be plotted against three dimensions: strength of contact (going from part-whole containment over physical contact to adjacency without contact), boundedness (involving an extension of the part-whole relationship towards unbounded wholes and parts), and domain (with shifts from the spatial to the temporal, the spatiotemporal and the categorial domain) (p. 269).

I will present a different perspective, one that does not contradict theirs, and that may turn out to be complementary. In the first instance, it is P&G’s inventory of ‘metonymical patterns’ that interests me. It is shown here in Table 61, along with examples, but without the references supplied by P&G. The list was compiled from five studies of metonymy spanning the period 1880 to 1981 but was not intended to be exhaustive. The goal was “merely to define an empirical basis” for the exercise that they pursue in their paper. I will use it for the same purpose here.

The unstructured inventory consists of 23 patterns of the type SPATIAL PART & WHOLE, LOCATION & LOCATED, PRODUCER & PRODUCT, etc. The ampersand in the name indicates that no direction is specified for the meaning shift. When the direction of the meaning shift is relevant, P&G use the more conventional form VEHICLE FOR TARGET. Thus, for example, the pattern PRODUCER & PRODUCT covers two actual metonymies: PRODUCER FOR PRODUCT and PRODUCT FOR PRODUCER.
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Metonymical pattern</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SPATIAL PART &amp; WHOLE</td>
<td><em>Tony Blair is the Prime Minister of England</em></td>
</tr>
<tr>
<td>2. TEMPORAL PART &amp; WHOLE</td>
<td>Dutch and German <em>morgen</em> “morning” for “tomorrow”</td>
</tr>
<tr>
<td>3. LOCATION &amp; LOCATED</td>
<td>German <em>Das ganze Haus wurde aus dem Schlaf geschreckt</em> “the whole house was startled from sleep”</td>
</tr>
<tr>
<td>4. ANTECEDENT &amp; CONSEQUENT</td>
<td>Greek <em>phobos</em> “flight” for “fear”</td>
</tr>
<tr>
<td>5. SUBEVENT &amp; COMPLEX EVENT</td>
<td><em>Mother is cooking potatoes</em></td>
</tr>
<tr>
<td>6. CHARACTERISTIC &amp; ENTITY</td>
<td>French <em>beauté</em> “beauty”</td>
</tr>
<tr>
<td>7. PRODUCER &amp; PRODUCT</td>
<td><em>I’m reading Shakespeare</em></td>
</tr>
<tr>
<td>8. CONTROLLER &amp; CONTROLLED</td>
<td><em>Schwarzkopf defeated Iraq</em></td>
</tr>
<tr>
<td>9. CONTAINER &amp; CONTAINED</td>
<td>French <em>aimer la bouteille</em> “love the bottle”</td>
</tr>
<tr>
<td>10. MATERIAL &amp; OBJECT</td>
<td>French <em>carton</em> “cardboard” for “cardboard box”</td>
</tr>
<tr>
<td>11. CAUSE &amp; EFFECT</td>
<td><em>unlock the prisons</em> for “let the prisoners free”</td>
</tr>
<tr>
<td>12. LOCATION &amp; PRODUCT</td>
<td><em>china</em></td>
</tr>
<tr>
<td>13. POSSESSOR &amp; POSSESSED</td>
<td><em>the long straw starts</em> for “the person with the long straw”</td>
</tr>
<tr>
<td>14. ACTION &amp; PARTICIPANT</td>
<td><em>to author a book</em></td>
</tr>
<tr>
<td>15. PARTICIPANT &amp; PARTICIPANT</td>
<td><em>the pen is mightier than the sword</em> for “the writer is mightier than the soldier”</td>
</tr>
<tr>
<td>16. PIECE OF CLOTHING &amp; PERSON</td>
<td>French <em>une vieille perruque</em> “an old wig” for “an old person”</td>
</tr>
<tr>
<td>17. PIECE OF CLOTHING &amp; BODY PART</td>
<td>German <em>Sohle</em> “sole” for “sole”</td>
</tr>
<tr>
<td>18. SINGLE ENTITY &amp; COLLECTION</td>
<td>German <em>Imme</em> “swarm of bees” for “bee”</td>
</tr>
<tr>
<td>19. TIME &amp; ENTITY</td>
<td>French <em>un mardi-gras</em> “a Shrove Tuesday” for “a disguised man”</td>
</tr>
<tr>
<td>20. OBJECT &amp; ENTITY</td>
<td>French <em>un quart</em> “a quarter” for “a tin of sardines in oil”</td>
</tr>
<tr>
<td>21. CENTRAL FACTOR &amp; INSTITUTION</td>
<td><em>the press</em></td>
</tr>
<tr>
<td>22. POTENTIAL &amp; ACTUAL</td>
<td><em>Can you see him?</em></td>
</tr>
<tr>
<td>23. HYPONYM &amp; HYPERONYM</td>
<td>*the pill for “the contraceptive pill”</td>
</tr>
</tbody>
</table>

Table 61: Peirsman & Geeraerts’ metonymical patterns

The are many parallels with Bourque’s reversible relations here. Table 62 shows the Bourquifier templates for the basic and reversed forms of PRODUCTION, along with the examples, *song bird* and *bird song*. In both relations, *song* is the product and *bird* is the producer, irrespective of the direction of the relation. However, the roles played by the head and modifier can be either of these, depending on the direction of the relation. Thus the general relation PRODUCTION corresponds precisely to the metonymic pattern PRODUCER & PRODUCT, and its basic and reversed forms (PROD and PROD-R) correspond to the directed metonymies PRODUCT FOR PRODUCER and PRODUCER FOR PRODUCT, respectively.
I use the term ‘corresponds to’ advisedly since I do not claim that metonymy is at work in binominals, as Janda does in her paper. But this is mostly a terminological issue, as Janda points out. My claim is that the same kinds of relation are found not only in metonymy and derivation, but also in binominal word-formation. One could extend the term metonymic relation (or semantic relation) to cover all three domains, but it seems preferable to me to employ a new term, ‘associative relation’, which is underutilized in linguistics, rather than extend the use of terms which have long-established traditions and communities of research interest.

The correspondence between the metonymic pattern PRODUCER & PRODUCT (#7) and the semantic relation PRODUCTION is transparent because the same verb, to produce, is used for both. This applies to another eight of P&G’s patterns, coded as (T) in Table 63, viz. MERONOMY (#1), LOCATION (#3), CONTAINMENT (#9), COMPOSITION (#10), CAUSATION (#11), POSSESSION (#13), TEMPORALITY (#19), TAXONOMY (#23): none of these require further explication.

Lest this evidence be deemed insufficient to support my claim that the same kinds of associative relation underpin metonymy and word-formation, I will discuss each of the remaining 14 patterns, showing that there are correspondences of one kind of another with relations in Bourque’s system for all but one of them. The reason many of these are less transparent than the nine listed above is because the metonymic pattern either specializes (S) or generalizes (G) the corresponding relation. Specialization accounts for nine patterns, four of which specialize the MERONOMY relation. In #2 the part and whole are temporal, while those of #5 are events. The whole in #18 is special in that it consists of multiple parts of the same type, and in #21, exemplified by the press, the whole is an institution (here, that of newspapers), denoted by a salient part (the printing press). It is quite possible to define more fine-grained semantic relations to provide exact equivalents for these four, as required by the analysis (for example, the relation COLLECTION, with the roles member and group would cater more specifically to #18). If that level of detail is not required, the more general relation of MERONOMY covers all four of these patterns.

<table>
<thead>
<tr>
<th>PRODUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td>Basic</td>
</tr>
<tr>
<td>Reversed</td>
</tr>
</tbody>
</table>

Table 62: Bourquifier template for PRODUCTION

2020-05-26 Submitted version
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Metonymic pattern</th>
<th>Code Relation</th>
<th>Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SPATIAL PART &amp; WHOLE</td>
<td>(T) MERONY</td>
<td>part, whole</td>
</tr>
<tr>
<td>2. TEMPORAL PART &amp; WHOLE</td>
<td>(S) MERONY</td>
<td>part, whole (temporal)</td>
</tr>
<tr>
<td>3. LOCATION &amp; LOCATED</td>
<td>(T) LOCATION</td>
<td>location, located</td>
</tr>
<tr>
<td>4. ANTECEDENT &amp; CONSEQUENT</td>
<td>(G) CAUSATION</td>
<td>cause, effect</td>
</tr>
<tr>
<td>5. SUBEVENT &amp; COMPLEX EVENT</td>
<td>(S) MERONY</td>
<td>part, whole (events)</td>
</tr>
<tr>
<td>6. CHARACTERISTIC &amp; ENTITY</td>
<td>(A) SIMILARITY</td>
<td>likeness</td>
</tr>
<tr>
<td>7. PRODUCER &amp; PRODUCT</td>
<td>(T) PRODUCTION</td>
<td>producer, product</td>
</tr>
<tr>
<td>8. CONTROLLER &amp; CONTROLLED</td>
<td>(G) POSSESSION</td>
<td>possessor, possessed</td>
</tr>
<tr>
<td>9. CONTAINER &amp; CONTAINED</td>
<td>(T) CONTAINMENT</td>
<td>container, contained</td>
</tr>
<tr>
<td>10. MATERIAL &amp; OBJECT</td>
<td>(T) COMPOSITION</td>
<td>material, object</td>
</tr>
<tr>
<td>11. CAUSE &amp; EFFECT</td>
<td>(T) CAUSATION</td>
<td>cause, effect</td>
</tr>
<tr>
<td>12. LOCATION &amp; PRODUCT</td>
<td>(G) SOURCE</td>
<td>source, result</td>
</tr>
<tr>
<td>13. POSSESSOR &amp; POSSESSED</td>
<td>(T) POSSESSION</td>
<td>possessor, possessed</td>
</tr>
<tr>
<td>14. ACTION &amp; PARTICIPANT</td>
<td>(S) SOURCE</td>
<td>agent, action</td>
</tr>
<tr>
<td>15. PARTICIPANT &amp; PARTICIPANT</td>
<td>(G) USAGE</td>
<td>user, used</td>
</tr>
<tr>
<td>16. PIECE OF CLOTHING &amp; PERSON</td>
<td>(S) USAGE</td>
<td>wearer, clothing</td>
</tr>
<tr>
<td>17. PIECE OF CLOTHING &amp; BODY PART</td>
<td>(S) LOCATION</td>
<td>clothing, body part</td>
</tr>
<tr>
<td>18. SINGLE ENTITY &amp; COLLECTION</td>
<td>(S) MERONY</td>
<td>individual, collection</td>
</tr>
<tr>
<td>19. TIME &amp; ENTITY</td>
<td>(T) TEMPORALITY</td>
<td>time, entity</td>
</tr>
<tr>
<td>20. OBJECT &amp; QUANTITY</td>
<td>(S) CONTAINMENT?</td>
<td>object, quantity</td>
</tr>
<tr>
<td>21. CENTRAL FACTOR &amp; INSTITUTION</td>
<td>(S) MERONY</td>
<td>central part, whole</td>
</tr>
<tr>
<td>22. POTENTIAL &amp; ACTUAL</td>
<td>(X) NA</td>
<td>–</td>
</tr>
<tr>
<td>23. HYPONYM &amp; HYPERONY</td>
<td>(T) TAXONOMY</td>
<td>supertype, subtype</td>
</tr>
</tbody>
</table>

Table 63: Metonymic patterns and semantic relations

The remaining five cases in which the metonymical pattern is a specialization of an existing relation are:

- #12 LOCATION & PRODUCT is exemplified by china (in the sense of household tableware or other objects made of porcelain, which originally came from China). This, like the old woman from the sky (cf. page 230), is clearly a specialization of the SOURCE relation (not LOCATION, as might at first be thought), in which the more specific product is used in place of the more general result.\(^1\)

- #14 ACTION & PARTICIPANT: In P&G’s example, to author a book, the vehicle is an agent (the author) and the target is an action (the act of writing). We recall from page 202 that Hatcher assigned such cases (exemplified by sunshine) to her A→B, on the grounds that the subject is the source of its own activity, and

---

\(^1\) This is actually a double metonymy, since china is also the name of the material used, so the actual SOURCE relation (LOCATION & PRODUCT) occurs in the meaning shift from country to material. The second meaning shift, from material to tableware, is an example of the COMPOSITION relation (P&G’s MATERIAL & OBJECT).
that in doing so “we are merely adding Agent to Agency” (p. 365). On that basis, Bourque’s SOURCE is the most appropriate correspondence, with roles specialized, if necessary, from source and result to agent and action.

- #16 PIECE OF CLOTHING & PERSON: The relation between a piece of clothing and the person who wears it is USAGE. Again, the roles can be specialized from user to wearer and from used to clothing.

- #17 PIECE OF CLOTHING & BODY PART: This is a specialization of LOCATION with the roles clothing and body part for located and location, respectively.

- #20 OBJECT & QUANTITY: P&G’s example here is Fr. un quart ‘a quarter’ for “a tin of sardines in oil”, cf. Eng. a pint (of beer). Since binominal quantifier constructions were explicitly excluded from the present study (see page 13) and they were also not pertinent to Bourque’s, there is no precise counterpart among the existing relations. However, this is no more than a very slightly specialized version of CONTAINMENT, with container specialized as quantity.

Three of the metonymical patterns generalize one or more of Bourque’s relations:

- #4 ANTECEDENT & CONSEQUENT is a more general relation, encompassing both SOURCE and CAUSATION, that we might call SEQUENCE in a less fine-grained classification (with the roles antecedent and consequent). The example cited by P&G (Greek phobos “flight” for “fear”) is surely an instance of CAUSATION (with the roles cause and effect).

- #8 CONTROLLER & CONTROLLED. This pattern is slightly more tricky. It could be regarded as a generalization of POSSESSION, a significant aspect of which is control. Alternatively, and given the example of Schwarzkopf (standing for the coalition forces) defeating Iraq in Operation Desert Shield, the metonymy might be expressed slightly more generally as MEMBER & GROUP, for which there is no precise equivalent in Bourque’s system. However, extending the latter with MEMBERSHIP (member, group), as a subtype of MERONYM (part, whole), is perfectly possible if one needs that degree of granularity.

- #15 PARTICIPANT & PARTICIPANT: This pattern is general to the point of being almost innocuous. P&G’s gloss their example (the pen is mightier than the sword) as “the writer is mightier than the soldier”. I suggest that this could be more precisely characterized as INSTRUMENT FOR USER, which corresponds to Bourque’s USG-R.

The two remaining patterns are exceptional, each in its own way.
• #6 CHARACTERISTIC & ENTITY embodies a relation that would not be found in a classification geared towards binominals since a characteristic is a property, not a thing.¹ A broader system of associative relations would have to cater for such cases, but that is a topic for further research. Having said that, the very name of the pattern recalls Jespersen’s Characterizing Feature, which Hatcher had no hesitation in subsuming under her A ⊆ B, our MinH, so any additional relation that might be required would fit in neatly beside the others that map to MinH (i.e. POSS, CONT, MER, etc.).

• #22 POTENTIAL & ACTUAL, the only remaining pattern, is exemplified by *Can you see him?* This one is special because it involves predicational rather than referential metonymy (Panther & Thornburg 1999). Incorporating it into the Hatcher-Bourque classification would therefore involve extending the system far beyond its present scope, so I choose to ignore it in this context.

To conclude this section: I have shown that all but one highly exceptional item from P&G’s inventory of metonymical patterns have (or could have) equivalents – many of them completely transparent – in the Hatcher-Bourque classification. This constitutes strong evidence in support of my claim of a commonality between metonymy and the kinds of semantic relation found in binominals. The fact that some correspondences involve generalization or specialization simply underlines the view expressed earlier, that the granularity of the classification is dependent on the needs of the application: any two relations can always be lumped together, and any single relation can always be split into subtypes. The next section, on cognitive relations, offers further evidence for this view.

### 9.2 Cognitive relations (Koch and Blank)

The terms ‘cognitive relations’, ‘associative relations’ and ‘semantic associations’ are variously used in the fields of lexical semantics and lexical typology to denote both synchronic and diachronic relations between lexical concepts (e.g. Blank 1999, 2003; Koch 1999, 2001; Koch & Marzo 2007; Urban 2012). In the following I will employ the term cognitive relations, since it was introduced earlier in the context of motivational grids in §7.3.1.

According to Koch and Marzo, all such basic cognitive relations ultimately derive from the three associative relations of ‘contiguity’, ‘similarity’, and ‘contrast’,

¹ In Croft’s terms, the relation involves property modification, not object modification; in Štekauer’s terms it is appropriate to word-formations of Onomasiological Type 4 rather than Type 3.
that have been well established since Aristotle and have been corroborated by Husserl’s phenomenology (cf. Holenstein 1972), by gestalt psychology (cf. Wertheimer 1922/23 and Raible 1981), and by free association tests (cf. Raible 1981), that have been introduced into linguistics by Roudet (1921: 686–692) and Jakobson (1956), and that have been applied with success to problems of lexical semantic change (cf., once more Roudet, and Ullmann 1962: 211–227…) (p. 269).

Blank (2003) identifies ten subtypes of these basic relations, summarized in Table 64. Two of these, formal identity and syntagmatic contiguity, are based on form rather than meaning, and are disregarded by Koch in the works cited above, as is antiphrastic contrast, which Koch claims is very rare (2001: 1159). Thus, for the semantic axis of his motivational grid, Koch includes seven of the ten relations (those shown in bold).

<table>
<thead>
<tr>
<th>similarity</th>
<th>contrast</th>
<th>contiguity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• metaphorical similarity (foot of mountain)</td>
<td>• co-taxonomic contrast (bad &gt; good)</td>
<td>• conceptual contiguity (“spatial, temporal, or logical connection between the concepts”)</td>
</tr>
<tr>
<td>• co-taxonomic similarity (Sp. tigre &gt; jaguar in South America)</td>
<td>• (antiphrastic contrast) (“more indirectly opposed concepts”, e.g. Fr. pensionnaire ‘guest in a boarding house’ &gt; ‘convict’)</td>
<td>• syntagmatic contiguity (“the relation between the parts of complex lexical units”, e.g. motor car ~ car)</td>
</tr>
<tr>
<td>• taxonomic subordination (hound ‘dog in general’ &gt; ‘hunting dog’)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• taxonomic superordination (inverse of preceding, no example)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• conceptual identity (tautology “or certain classes of word-formation”)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• formal identity (hotel &gt; motel)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 64: Blank’s ten cognitive relations**

Blank is of the opinion that, of the three Aristotelian principles of remembering, “similarity shows by far the greatest diversity” (2003: 46),¹ and this is reflected in his classification. Observe that there are five subtypes of similarity as against four in the Hatcher-Bourque system (cf. Figure 78). Metaphorical similarity corresponds to SIMILARITY; taxonomic subordination and superordination are the equivalent of TAXONOMY, and co-taxonomic similarity equates to COORDINATION. Conceptual

¹ This despite his awareness of engynomy as “a system of concepts that exhibit a subtype of contiguity, such as part/whole, cause/consequence, producer/product, activity/place etc.”
identity does not have any direct correspondence. In §6.2.3 I characterized identity as a limiting case of similarity (in the context of Arnaud’s proposed addition of an ANALOG relation to Hatcher’s system). There was no need to add it to the Hatcher-Bourque classification since it was not needed in order to account for the binominal data. Now, however, we have reason to add it as subtype of the high-level category Similarity, alongside SIMILARITY (PROPER), COORDINATION and TAXONOMY, in order to achieve compatibility with Blank and Koch. This has the added advantage of providing a very natural home for reduplication as a word-formation process. Blank and Koch’s relation of (co-taxonomic) contrast has also not played any role in the present study for the simple reason that no examples turned up in either Bourque’s data or mine. However, as I noted on page 216, it does occur, for example in Chinese, and so it should be incorporated into our extended model.

The upshot of this discussion is that all that is required in order to accommodate Blank and Koch’s similarity- and contrast-based relations in our classification is the inclusion of IDENTITY and CONTRAST as separate relations. No such extension is required for the one remaining cognitive relation, conceptual contiguity, which Blank characterizes as a “spatial, temporal, or logical connection” between concepts, since it basically lumps together all the remaining relations under a single label.

I have now shown how the same, multilevel system of associative relations covers word-formation, metonymy, polysemy and semantic shift. It seems possible that the system may also accommodate semantic roles (cf. Kittilä & Zúñiga 2014) and semantic word-substitution speech errors (slips of the tongue, cf. Arnaud 1999), but these are topics for further research. At this point I would like to move beyond linguistics to the domain of knowledge representation, where I believe that further insights can be gleaned from Topic Maps.

9.3 Topic Maps: roles and granularity

The insights that Topic Maps can offer in the context of associative relations are of two kinds. The first relate to the underlying model: the nature of the relationship, as it were, between relations, roles and participants. The second has to do with the further understanding of granularity as it relates to the classification I have been developing.

Topic Maps, we recall from §1.1.1, is an ISO standard for knowledge organisation that incorporates significant insights from the fields of knowledge representation and artificial intelligence. In contrast to RDF (the representational model used in the CLLD, the project that includes both WOLD and WALS, and other linked data
initiatives), Topic Maps is not based on formal logic. It developed organically from a formalization of the models implicit in traditional finding aids such as back-of-book indexes, thesauri and glossaries and, in that sense, it can be called a “usage-based” model. The primary constituents of that model, which I described earlier, are topics, associations and occurrences (the TAO of Topic Maps). Our concern here is with associations.

The structure of an association was shown in Figure 2 (reproduced here as Figure 79 for convenience). The figure depicts an association (A) from the domain of opera, which asserts a relation of a particular type (labelled composed-by) between two topics (T) that represent the opera Tosca and the person Puccini, and play the roles (R) of work and composer, respectively. In short, it encodes the assertion that Tosca was composed by Puccini.

![Figure 79: The anatomy of an association](after Pepper 2010a)

The association between Puccini and Tosca can be viewed in terms of the cognitive, metonymic and semantic relations that we have been discussing as follows:

- In Koch and Blank’s model (cf. Table 64 on page 325) it is an example of conceptual contiguity (or more precisely, the engynomic relation ‘producer/product’, cf. Blank 2003: 42).
- The statement *I am listening to Puccini* when I am in fact listening to Tosca is a case of Peirsman and Geeraerts’ PRODUCER FOR PRODUCT metonymy.
- The binominal lexeme *puccini tosca* would be somewhat unusual since both Puccini and Tosca denote individuals rather than types, but the corresponding anchoring possessive construction *Puccini’s Tosca* is perfectly plausible (it would distinguish Puccini’s opera from Renoir’s film of the same name); the semantic relation it employs is Hatcher-Bourque’s PRODUCTION, which has the roles producer and product.
The only difference between the relation expressed by the association and these three statements is the level of granularity: as indicated in the diagram, composed-by specializes a more high-level relation created-by, which itself specializes the relation produced-by, i.e. PRODUCTION. Corresponding to these three levels of relation are three pairs of roles: work/composer, creation/creator and product/producer, the names of which could be reflected in three metonymies: WORK & COMPOSER, CREATION & CREATOR and PRODUCT & PRODUCER, all at different levels of granularity. Moreover, it can be seen from Figure 78 that the PRODUCTION relation can be further generalized to CAUSATION, to CONTIGUITY, and thence to the mother of all relations, the equivalent of the unspecified ‘see’ relation in book indexes that I discussed in Pepper (2002): the one, very abstract and vague relation meaning “there is a connection between” that was entertained by Bauer in 1979 (cf. page 187).

This example draws our attention to the naming of semantic relations. Following Bourque, I have mostly used nominalizations of a verb commonly used to express the relation (CONTAINMENT, POSSESSION, COMPOSITION, PRODUCTION, etc.). Such names are neutral with respect to the direction of the relation. They were used as the basis for the codes used to denote relations viewed from one particular direction (CONT, POSS, COMP and PROD, etc.). Then, in order to cater for the other direction, I adopted codes like CONT-R, POSS-R, COMP-R and PROD-R. This approach had the advantage of brevity, which is good for databases and spreadsheets and statistical software, but it will not do for humans, because we cannot tell from the label whether POSS-R stands for “(an) H that possesses (an) M” or “(an) H that (an) M possesses”. An alternative labelling convention was used (up to a point) by Warren (1978), who, we recall from page 186 used names like Source-Result, Whole-Part, Part-Whole, Size-Whole, Goal-OBJ, Place-OBJ, Time-OBJ, Origin-OBJ and Activity-Actor (but also Copula, Resemblance and Purpose). These, as we can now recognize, are based on the roles played by the participants in the relation and they correspond exactly to the naming conventions used in the study of metonymy (PART FOR WHOLE, WHOLE FOR PART, etc.). In the extended model of associative relations that I will present shortly, I adopt the same convention for ‘reversible’ relations. Adding explicit roles to the overall model will also cater for the need that we experienced in §8.3 to use the role played by the modifier in order to understand the nature of specific elaboration sites.

The preceding discussion centred on a single association type, composed-by, and how it corresponds to the semantic relations, metonymies and cognitive relations discussed in the earlier part of this chapter. That association type comes from the
**Italian Opera Topic Map** (Pepper 2009), a topic map that I developed incrementally over a period of ten years as I strove first to understand, and then to explain and further develop the Topic Maps model. In fact, the whole *Italian Opera Topic Map* grew out of the initial assertion that *Tosca* was composed by Puccini. Following that, I added the other Puccini operas, and then other composers and their operas. At that point I had to add further association types, in order to capture who wrote the librettis, the literary works these were based on, where the composers, librettists and writers were born and died, etc. In the most recent version of the topic map, which has not been touched for the last ten years, there are 22 different association types. They are listed in Table 65, along with their corresponding roles and an example of each. It is instructive to examine some of them more closely. Most of them are binary (i.e. involve two participants), one (killed-by) is ternary, and one (unfinished) is unary. In some respects these correspond to transitive, ditransitive and intransitive predicates.

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Roles</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>appears-in</td>
<td>character, work</td>
<td>Scarpia appears in <em>Tosca</em></td>
</tr>
<tr>
<td>based-on</td>
<td>source, result</td>
<td><em>Tosca</em> (the opera) is based on <em>La Tosca</em> (the play)</td>
</tr>
<tr>
<td>born-in</td>
<td>person, place</td>
<td>Puccini was born in Lucca</td>
</tr>
<tr>
<td>completed-by</td>
<td>work, composer</td>
<td><em>Turandot</em> was completed by Alfano</td>
</tr>
<tr>
<td>composed-by</td>
<td>work, composer</td>
<td><em>Tosca</em> was composed by Puccini</td>
</tr>
<tr>
<td>died-in</td>
<td>person, place</td>
<td>Puccini died in Brussels</td>
</tr>
<tr>
<td>exponent-of</td>
<td>person, style</td>
<td>Puccini is an exponent of verismo</td>
</tr>
<tr>
<td>has-voice</td>
<td>character, voice type</td>
<td><em>Floria Tosca</em> has the voice type soprano</td>
</tr>
<tr>
<td>killed-by</td>
<td>victim, perpetrator,</td>
<td><em>Tosca</em> killed Scarpia by stabbing</td>
</tr>
<tr>
<td></td>
<td>cause of death</td>
<td></td>
</tr>
<tr>
<td>libretto-by</td>
<td>opera, librettist</td>
<td><em>Tosca’s</em> libretto was written by Illica and Giacosa</td>
</tr>
<tr>
<td>located-in</td>
<td>containee, container</td>
<td>Lucca is located in Italy</td>
</tr>
<tr>
<td>part-of</td>
<td>part, whole</td>
<td><em>Vissi d’arte</em> (the aria) is part of <em>Tosca</em></td>
</tr>
<tr>
<td>premiere</td>
<td>work, place</td>
<td><em>Tosca</em> was first performed at Teatro Costanzi</td>
</tr>
<tr>
<td>published-by</td>
<td>work, publisher</td>
<td><em>Tosca</em> was published by Ricordi</td>
</tr>
<tr>
<td>pupil-of</td>
<td>pupil, teacher</td>
<td>Puccini was a pupil of Ponchielli</td>
</tr>
<tr>
<td>revision-of</td>
<td>source, result</td>
<td><em>Aroldo</em> is a revision of <em>Stiffelio</em></td>
</tr>
<tr>
<td>subtype-of</td>
<td>subtype, supertype</td>
<td>city is a subtype of place</td>
</tr>
<tr>
<td>sung-by</td>
<td>person, aria</td>
<td><em>Vissi d’arte</em> is sung by <em>Floria Tosca</em></td>
</tr>
<tr>
<td>takes-place-during</td>
<td>opera, event</td>
<td><em>Tosca</em> takes place during the Napoleonic Wars</td>
</tr>
<tr>
<td>takes-place-in</td>
<td>opera, place</td>
<td><em>Tosca</em> takes place in Rome</td>
</tr>
<tr>
<td>unfinished</td>
<td>work</td>
<td><em>Turandot</em> was unfinished</td>
</tr>
<tr>
<td>written-by</td>
<td>writer, work</td>
<td><em>La Tosca</em> was written by <em>Sardou</em></td>
</tr>
</tbody>
</table>

*Table 65: Association types in the Italian Opera Topic Map*
We observe that some association types are isomorphic with Bourque’s relations (located-in = LOCATION, part-of = MERONYM); subtype-of is equivalent to TAXONOMY; libretto-by and written-by, like composed-by, are specializations of PRODUCTION, and based-on and revision-of are more fine-grained versions of SOURCE, as their roles make clear. We see that the level of granularity tends to be higher in a topic map than in binominals, but this does not detract from the overall compatibility of the two models, and the ability of one model to inform the other. For my own part, I believe that the very explicit and carefully articulated model of Topic Maps has helped me to clarify my understanding of the relations, roles and labels in semantic, metonymic and conceptual relations, and enabled me to unify what, on the surface, seem to be very different linguistic processes.

Topic Maps has thus informed linguistics in some respects, but linguistics can also inform Topic Maps. During the development of the standard, there were repeated discussions about the extent to which we should not only standardize the basic model and interchange syntax, but also some more detailed aspects of the model. In general the ISO working group held off from standardizing too many details. We did make two exceptions, however: we standardized the association types class-instance and superclass-subclass (Pepper & Moore 2001 §2.2.4). The first expresses the relationship between a class (e.g. opera) and one of its instances (e.g. tosca), and the second expresses the relationship between a class (e.g. place) and one of its subclasses (e.g. city). The rationale for standardizing these two types of association results was to enable greater interoperability between Topic Maps systems “out of the box”. Having done so, the question we then faced was whether to standardize more than just these two types of relationship, in order to achieve even greater interoperability. We refrained, on the grounds that we did not have sufficient experience with how Topic Maps would be used, and did not know which kinds of relationship would be most useful.

Thanks to the present study we now know much more about the relative frequency of various semantic relations. We know that MERONYM, LOCATION, CAUSATION and the like are extremely frequent in binominal word-formation and metonymy, and since language is essentially a form of knowledge representation, we can take for granted that these relations should be among those prioritized if any further standardization were to take place.

---

1 The superclass-subclass relation is, of course, the equivalent of the TAXONOMY relation. The class-instance relation is not relevant to the present domain because the common nouns that we are concerned with in the domain of binominals denote classes rather than individual referents.
9.4 Putting it all together: the PHAB model

On the basis of all preceding discussion I propose the integrated model shown in Figure 80, which I call the PHAB model.1

![Figure 80: The PHAB model of associative relations]

At the top of the hierarchy is Bauer’s (1979) one relation to rule them all (see page 187), which conveys no information other than that there is some kind of connection. At the next level we find the three Aristotelian relations, Similarity, Contiguity and Contrast (page 209). Similarity is divided into four subtypes, one of which, IDENTITY, is additional to the three found in Bourque’s classification (cf. page 215). Given that TAXONOMY is a reversible relation (which is why it alone has two distinct roles), these four cover Blank’s five subtypes of Similarity listed in Table 64 on page 325; together with Contiguity and Contrast, they constitute the seven cognitive relations employed by Koch in his motivational grids (page 281).

Contiguity is subdivided into CONTAINMENT and DIRECTION, which corresponds to Hatcher’s static vs. dynamic duality in non-appositional compounds (page 201) as well as Bourque’s Basic and Reversed types (page 195); DIRECTION is used in preference to CAUSATION (cf. Figure 78, page 317) since on reflection it seems to be more general, and thus a better way to express the basic commonality between SOURCE/RESULT, CAUSATION and PRODUCTION. The latter are tentatively placed in hierarchical relation to one another since some kind of scale of intentionality or agency appears to distinguish them. (There is, of course, a certain sense in which Puccini was both the “cause” and the “source” of Tosca.) Likewise, I suggest that

---

1 The Pepper-Hatcher-Arnaud-Bourque model, pronounced ‘fab’.
COMPOSITION is a subtype of MERONOMY, in which the whole is made up of uniform parts, and that TEMPORALITY (a term I have co-opted to denote temporal location) is a subtype of LOCATION. The dotted boxes around USAGE, FUNCTION and PURPOSE indicate uncertainty as to how they relate to one another. This is yet another topic for future research.

Note, finally, that TOPIC, FUNCTION and PURPOSE each have two distinct roles, even though they appear on the basis of the available data to be non-reversible. Quite why these three, and only these three, have that property is a conundrum that awaits further investigation.

Caveats
The PHAB model as currently presented appears to cover all the semantic relations found in determinative noun-noun compounds, binominals and typifying nominal modifier constructions in general, and quite possibly also all anchoring nominal modifier constructions. In addition, it appears to cover most forms of metonymy (at least as used in word-formation), albeit sometimes at a more general level than one finds in metonymy research. It also covers all the cognitive relations identified by Koch and Blank in the domains of polysemy and semantic shift.

However, it is not yet a complete model of associative relations. For one thing, more work is required to test, and perhaps extend, the subclassification of Similarity and Contrast. Wälchli’s (2005) study of co-compounds may be a good starting point for Similarity, but I know of no equivalent study relating to Contrast.

Furthermore, the PHAB model has only been stress tested on word-formations of Onomasiological Type 3; it remains to be seen how other Onomasiological Types fit in. The characteristic feature of OT1 (e.g. truck driver) and OT2 (e.g. driver) is, as seen in Figure 3 on page 11, that the determined (actional) element is present. That element essentially describes the relationship involved in the conceptualization (for example the DRIVING relationship between the agent (-er) and the truck), at a much finer level of detail than the relations found in Hatcher-Bourque. But to what extent are they subtypes of the latter? Based on the discussion above, we can already confirm that opera composer, a parallel formation to truck driver, fits easily within the model (musical composition as a subtype of the PRODUCTION→CREATION sub-hierarchy), but does that apply to all such synthetic compounds?

Then there is the question of relations involving action nouns, which also fall outside the scope of prototypical binominals. Do the same relations apply? Hatcher was able to cater for Jespersen’s sunshine and sun worship as A→B and A←B (page 202); in
the revised Bourque scheme they would be classified as SRC-R (“(a) shine whose source is (a) sun”) and DIR (“(a) worship whose goal is (a) sun”); and further grounds for optimism are that Levi’s (1978) classification (which Bourque had recourse to when developing his system) was designed to cover all kinds of ‘complex nominal’, including nominalizations (cf. §2.3.1).

Property modification is yet another domain that remains to be investigated, and that would extend the enquiry to Onomasiological Type 4 (e.g. blackbird). Croft (forthc.) presents a ranking of property concept semantic classes according to the likelihood of their employing a distinct morphosyntactic strategy or of recruiting an object modification or action modification strategy (* = most prototypical property modifiers, + = next most prototypical property modifiers):

(145) MORE LIKELY TO RECRUIT OBJECT MODIFICATION CONSTRUCTION
1. Material/Substance: wood(en), gold(en), etc.
2. Gender: male, female
3. +Color: white, black, red, green, blue, yellow, brown
4. +Shape: round, flat, etc.
5. *Age: old, young, ripe
6. *Value: good, bad
7. *Dimension: big, little, tall, short, wide, narrow, etc.
8. Physical Properties: smooth, sharp, soft, hard, heavy, light, etc.
9. Human Propensity: happy, jealous, kind, clever, etc.

MORE LIKELY TO RECRUIT ACTION MODIFICATION CONSTRUCTION

The binominals data includes many examples from the top end of this scale. The question to be addressed is to what extent properties at the lower end of the scale can be accommodated by the PHAB model.

There is, in other words, plenty of work still to be done, but I suggest that the PHAB model provides a good starting point and I hope other researchers will take it up.

9.5 Chapter summary

In this chapter I ventured far beyond the confines of binominal lexemes in order to synthesize the results of investigating those unstated (or underspecified) semantic relations that first intrigued me about noun-noun compounds.

I looked first at metonymy, taking my cue from Janda’s 2011 paper and showed that the Hatcher-Bourque classification developed in Chapter 6 covers all of the referential metonymic relations discussed by Peirsman and Geeraerts, albeit it at
different levels of granularity in some cases. I then turned to the kind of cognitive relations found in polysemy and semantic shift, drawing on the work of Blank and Koch. In addition, I emphasized the importance of having an explicit model of relations that makes a clear distinction between relations, roles and participants.

I further demonstrated how it is possible for lumpers and splitters to coexist, since any two (or more) relations can be subsumed under some more general relation. For example, at a low level of granularity, COMPOSED BY and WRITTEN BY can be amalgamated into CREATED BY, which can have further subtypes: DIRECTED BY, PAINTED BY, etc. At a higher level of granularity, even such disparate relations as MERONOMY and CAUSATION can be combined under the umbrella Contiguity (which is precisely what Blank does, cf. footnote 1 on page 325). Conversely, any relation can be subdivided into more specific relations, as I proposed doing with the SOURCE relation in order to accommodate the metonymic pattern LOCATION & PRODUCT. It is not a matter of different sets of relations, but one of selecting the appropriate degree of granularity for any particular application.

The resulting synthesis, the PHAB model, is a general model of associative relations that I believe has applicability across several fields of linguistic enquiry and also beyond.
10 Conclusion

Typology’s remit is simple in principle, though beset with huge practical difficulties: it is (i) to chart linguistic diversity and (ii) to seek out order or even unity in diversity and to make sense of it. (Plank 2016)

Whether or not I have succeeded in fulfilling the remit of typology as Plank sees it is for others to judge. As far as binominal lexemes are concerned, I have at least made a start. I have been able to show that ‘binominal lexeme’ figures in certain cross-linguistic generalizations, so I can claim to have satisfied Edith Moravcsik’s criterion for a cross-linguistically applicable comparative concept:

It seems to me that the only consideration to help us choose from among the various alternative definitions listed by Matthew is typological implications. If a given concept can be shown to figure in crosslinguistic generalizations either by serving as an implicans or by serving as an implicatum, it is it a useful one. If it is not a term in typological implications, it is not useful (LingTyp mailing list, 19 Oct 2018).

But then again, this was only ever an exploratory study, and my own feeling is that it has thrown up enough interesting data and observations to justify the selection of binominal lexeme as a comparative concept. Now, by way of conclusion, I present a brief summary, outline what I believe has been the contribution to science of this work, and present some of my ideas for future research.

10.1 Summary

In Chapter 1 I explained the genesis of the present study and arrived at a definition of binominal lexeme. My choice of comparative concept was then further justified in Chapter 2 in the light of previous studies of compounding, word-formation and morphological complexity – and not least, a number of studies that prefigured the concept of binominals without actually recognizing it as a category. In Chapter 3, I discussed my methodology in some detail (not least in order to promote reuse of the data and replication of the study): I covered the development of the list of 100 meanings, the sample of 106 languages and the data collection. Then, in Chapter 4 I described the initial annotation of the data and considered some theoretical and analytical issues concerning identification of thing-roots and -affixes, determining the head, and defining constructions based on glosses provided by contributor. This chapter also contained an initial analysis of the data.
In Chapter 5 I developed a typological classification based on morphosyntactic strategies, identifying nine types of binominal, each of which can occur as a head-initial or head-final construction. These were presented in a two-dimensional grid in which the primary parameters were number of markers and locus of marking. I also discussed gradient phenomena and apparent gaps in the system.

I then developed a second classification in Chapter 6, devoted to semantic relations, building on the work of Anne Hatcher and Yves Bourque. I refined and extended their high- and low-level systems and then integrated the two into a single model, the Hatcher-Bourque classification, which I then applied to my binominal data.

In Chapter 7 I carried out investigations into three major typological topics, looking first at constituent order, then the relationship between binominals and possessives (or, more precisely, anchoring nominal modifier constructions), and finally various aspects of the relationship between morphosyntactic strategies and semantics.

The focus of Chapter 8 was the two-paths hypothesis: the idea that there are two fundamentally distinct ways of providing mental access to a complex concept that involves two other, less-complex concepts. Here it was the notion of head-framing versus modifier-framing in concept combination that provided most support for the hypothesis. I developed a method for quantifying the salience of elaboration sites and measuring the degree of relationality of an object concept, and I also put forward the idea that languages might be more or less consistent in their application of species-framing versus attribute-framing (when concept combination does not clearly follow either the reference point path or the hypernym path).

In Chapter 9 I expanded the focus quite radically from binominals and the semantic relations they embody to other domains of linguistics (specifically metonymy and lexical semantics) and to a domain that might seem far beyond linguistics, namely knowledge representation. The purpose was to demonstrate the feasibility of a very general, reusable model of associative relations with much wider application.

10.2 Contribution to science

It is for others to judge the contribution to science of a work like this, but since it is customary in dissertations for the researcher to express an opinion on the matter, I shall set out my own thoughts. My contributions are in the two main areas named in the title – typology and semantics – and in the development of some new research methods.
10.2.1 Typology

The most important contribution to typology, as I see it, is in identifying binominal lexeme as a comparative concept and developing the nine-way typological classification.

**Binominal lexemes.** Most previous work in this field has focused on subsets of binominals: either compounds (e.g. Bauer 2001; Scalise & Bisetto 2009; Guevara & Scalise 2009; Scalise & Fábregas 2010), or phrasal lexemes (e.g. Masini & Benigni 2012), or denominal derivation (multiple studies). These are traditionally seen as belonging to different compartments of grammar – syntax or morphology – or, in the case of compounding, as being somehow on the cusp between the two (Jackendoff 2010). The present work is the first systematic investigation of all three types of construction as a single category.

Over recent decades, more and more linguists have come to question the traditional division of language into grammar and lexicon, and of grammar into syntax and morphology (see, in particular, various constructionist approaches, e.g. Langacker 1987; 1991; Goldberg 1995; 2006; Croft 2001; Masini 2009). The boundary line between roots and affixes has been shown to be a fuzzy one (e.g. Tuggy 1992), and so has that between derivation and compounding (e.g. Bauer 2005; Booij 2005; Štekauer 2005). The present study is part of this trend. The very object of study constitutes strong evidence that the compartmental view of grammar is artificial and that the constructionist approach accords better with the facts.

**Binominal typology.** The binominals database provides evidence to support a nine-way classification of morphosyntactic strategies (jxt, cmp, cls, der; gen, prp, adj; con; dbl), the names of which are purely mnemonic and not to be taken literally (thus, for example, prp includes postpositions as well as prepositions). These nine types are arranged on a two-dimensional grid, rather than hierarchically; the vertical axis represents the number of markers (0, 1 or 2) and the horizontal axis (on level 2 only) the degree of fusion. A second layer of the grid is used when it is desirable to take constituent order into account (jxtL, jxtR, etc.).
Of the many findings, large and small, which I have attempted to summarize at the end of each chapter, I believe the following stand out:

**Strategy recruitment.** With overwhelmingly greater than chance frequency, languages recruit at least one of their binominal strategies from anchoring nominal modifier constructions; when they don’t, there is an overwhelming tendency to use compounding (§7.2). The figure below shows how the languages distribute on the Pepper scale (with cases of grammaticalization split out of the category *never*).

![Pepper Scale Diagram]

**Non-universality of compounding.** Noun-noun compounding, the union of the jxt and cmp strategies, is not an absolute universal, but it is by far the most widespread form of binominal word-formation, accounting for roughly half of all binominals. If a language has just one binominal strategy, that strategy is almost always compounding (§5.6.3). If noun-noun compounding can serve as a proxy for compounding in general, then compounding in general is not universal.
Prevalence of right-headedness. The ratio of right- to left-headed constructions is approximately 2:1, for both binominals in general and noun-noun compounds, the latter defined as the union of jxt and cmp strategies. The so-called “righthand head rule” thus fails to account for fully one third of the data. The ratios for jxt and cmp are 1:1 and 4:1, respectively, demonstrating a strong tendency for the constituents of head-final compounds to fuse (§7.1). The predilection for right-headedness is strongest in Eurasia (90%), NG/Australia (87%) and South America (84%); in Oceania/SE Asia it is weakest (35%); and in Africa and North America the figures there is no strong predilection either way (60% and 53% right-headed, respectively).

10.2.2 Semantics
The most important contributions in the field of semantics are the Hatcher-Bourque classification (and the accompanying tool, the Bourquifier), the PHAB model of associative relations, and various scales of preference for semantic relations.
The typology and semantics of binominal lexemes

**Hatcher-Bourque.** A well-documented and reusable two-level classification of semantic relations, based on revisions of Hatcher’s and Bourque’s classifications. It comprises 29 relations (of which 12 reversible) at the lower level of granularity and five at the higher level (§6.3.3, cf. Table 31).

The Bourquifier. An Excel-based tool for assisting classification by automatically populating the templates of the Hatcher-Bourque classification.

<table>
<thead>
<tr>
<th>Bourquifier3</th>
<th>Binominal (B)</th>
<th>Modifier (M)</th>
<th>Head (H)</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relation</strong></td>
<td><strong>Roles</strong> B2 H2 <strong>TWO-LEVEL CLASSIFICATION</strong></td>
<td><strong>TWO-LEVEL CLASSIFICATION</strong></td>
<td><strong>TWO-LEVEL CLASSIFICATION</strong></td>
<td><strong>TWO-LEVEL CLASSIFICATION</strong></td>
</tr>
<tr>
<td><strong>TAXONOMY</strong></td>
<td>TAX</td>
<td>(a) wind is a kind of mill</td>
<td>oak tree</td>
<td>TAX-R</td>
</tr>
<tr>
<td><strong>COORDINATION</strong></td>
<td>COORD</td>
<td>(a) mill that is also (a) wind</td>
<td>boy king</td>
<td>COORD-R</td>
</tr>
<tr>
<td><strong>SIMILARITY</strong></td>
<td>SIM</td>
<td>(a) mill that is similar to (a) wind</td>
<td>kidney bean</td>
<td>SIM-R</td>
</tr>
<tr>
<td><strong>CONTAINMENT</strong></td>
<td>CONT</td>
<td>(a) mill that is contained in (a) wind</td>
<td>orange seed</td>
<td>CONT-R</td>
</tr>
<tr>
<td><strong>MERONOMY</strong></td>
<td>MER</td>
<td>(a) mill that is part of (a) wind</td>
<td>car motor</td>
<td>MER-R</td>
</tr>
<tr>
<td><strong>LOCATION</strong></td>
<td>LOC</td>
<td>(a) mill located at/near/in (a) wind</td>
<td>house music</td>
<td>LOC-R</td>
</tr>
<tr>
<td><strong>SIMILARITY</strong></td>
<td>SIM</td>
<td>(a) mill that occurs at/during (a) wind</td>
<td>summer job</td>
<td>SIM-R</td>
</tr>
<tr>
<td><strong>COMPANION</strong></td>
<td>COMP</td>
<td>(a) mill that (a) wind is composed of</td>
<td>tube sugar</td>
<td>COMP-R</td>
</tr>
<tr>
<td><strong>TOPIC</strong></td>
<td>TOP</td>
<td>(a) mill that is about (a) wind</td>
<td>history book</td>
<td>TOP-R</td>
</tr>
<tr>
<td><strong>DIRECTION</strong></td>
<td>DIR</td>
<td>(a) mill whose goal is (a) wind</td>
<td>sun worship</td>
<td>DIR-R</td>
</tr>
<tr>
<td><strong>SOURCE</strong></td>
<td>SRC</td>
<td>(a) mill whose source is (a) wind</td>
<td>cane sugar</td>
<td>SRC-R</td>
</tr>
<tr>
<td><strong>PRODUCTION</strong></td>
<td>PROD</td>
<td>(a) mill that produces (a) wind</td>
<td>long bird</td>
<td>PROD-R</td>
</tr>
<tr>
<td><strong>USAGE</strong></td>
<td>USG</td>
<td>(a) mill that uses (a) wind</td>
<td>lamp oil</td>
<td>USG-R</td>
</tr>
<tr>
<td><strong>PURPOSE</strong></td>
<td>PURP</td>
<td>(a) mill intended for (a) wind</td>
<td>animal doctor</td>
<td>PURP-R</td>
</tr>
</tbody>
</table>

The PHAB model. A synthesis of semantic, metonymical and conceptual relations found in the literature into a holistic system of associative relations. It is not yet a complete model, and more collaborative work is required in order to test, refine and extend it in various ways. The model may have application in the study of semantic roles, case systems and semantic word-substitution speech errors, and also beyond linguistics, in the domain of knowledge representation.
Frequency of semantic relations. The analysis of semantic relations can provide insights into the ways in which humans tend to conceptualize the world. The present data disconfirm Bauer’s suggestion that the most frequent semantic relations are LOCATION and COMPOSITION. In fact, by far the most widespread semantic relation is MERONOMY, followed by PURPOSE. Of the two types of MERONOMY, Basic (MER) and Reversed (MER-R), the former is considerably more frequent, which indicates that the conceptualization of a complex meaning is much more likely to involve modification by the whole than modification by the parts.

<table>
<thead>
<tr>
<th>MER</th>
<th>977</th>
<th>(an) H that is part of (an) M</th>
<th>table leg</th>
<th>HinM</th>
<th>$M \supset H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PURP</td>
<td>540</td>
<td>(an) H intended for (an) M</td>
<td>animal doctor</td>
<td>HtoM</td>
<td>$M \leftarrow H$</td>
</tr>
<tr>
<td>COOR</td>
<td>358</td>
<td>(an) H that is also (an) M</td>
<td>boy king</td>
<td>MisH</td>
<td>$M \approx H$</td>
</tr>
<tr>
<td>LOC</td>
<td>321</td>
<td>an H that (an) M is located at/near/in</td>
<td>music hall</td>
<td>HtoM</td>
<td>$M \supset H$</td>
</tr>
<tr>
<td>COMP-R</td>
<td>216</td>
<td>(an) H composed of (an) M</td>
<td>sugar cube</td>
<td>MinH</td>
<td>$M \subset H$</td>
</tr>
<tr>
<td>POSS</td>
<td>211</td>
<td>(an) H that (an) M possesses</td>
<td>family estate</td>
<td>HinM</td>
<td>$M \supset H$</td>
</tr>
</tbody>
</table>

Morphosyntactic strategies and semantic relations. There is no general, cross-linguistic correlation between morphosyntactic strategies and semantic relations. It is thus not the case that some strategies are used to express some relations, while other strategies are used for other relations. Every one of the high-level relations is expressed by each of the nine strategies, and the same seems to apply to low-level relations, although sparsity of data for some strategies and a low frequency for some low-level relations (in particular TEMP-R, SRC-R and TAX-R) means that not every combination of binominal type and relation is attested.
Head- vs. modifier-framing. Binominal meanings tend to be conceptualized in one of three ways: as concepts related to a more salient concept that serves as a reference point (modifier-framed); as concepts that are subtypes of a more salient, superordinate concept (head-framed); or as combinations of concepts that are roughly equally salient (neutrally-framed) (§8.2).

Degree of relationality of a concept. If the distinction between sortal nouns and relationals nouns is a cline, the degree of relationality of a concept can be measured by the number of elaboration sites (or “slots”) that it employs when combining with other concepts. This equates to the number of different low-level relations the concept participates in. E-sites are best expressed as the role played by the modifier in the combined expression, as defined by the Hatcher-Bourque classification. In the raw, unnormalized data used in the present study, the most relational concept is HOUSE, followed by HEAD, PLACE, STRING, THREAD, TREE and WHEEL (§8.3).
Elaboration site salience. The salience of the elaboration site of a concept such as HOUSE can be measured by the either type frequency (i.e. the number of different elaborating components) or their token frequency. The elaborating components should ideally be normalized before being counted. With the data shown here, this would result in values of 2 (type frequency) and 21 (token frequency) for the e-site labelled Possessor.

e-site                     elaborating component
Purpose (13)               excrement (3), urine (3), bovine (2), horse (2), livestock (2), animal (1), animal/cattle (1), animals (1), donkey (1), hand (1), horses (1), pig (1), poo (1)
Located (12)               doctor (5), medicine (3), food (2), hospital (2), bath (1), disease (1), fire (1), illnesses (1), kitchen (1), oven (1), pain (1), stove (1)
Possessor (3)              spider (11), bee (9), bees (1)
Material (2)               leaf (1), straw (1)
Location (1)               side (1)
Part (1)                   incense (1)
Whole (1)                  nose (1)

10.2.3 Methods

I hope that some of the methodologies I developed will be of use to other researchers. These include:

Constructing a meaning list. The list of 100 meanings that formed the basis for data collection was constructed in a semi-principled manner (§3.1). With hindsight it is possible to list four main desiderata for such a list:
The typology and semantics of binominal lexemes

1. maximal structural diversity
2. maximal yield of binominals
3. cross-categorial balance
4. cross-linguistic representation

The iterative method by which the list was constructed favoured the first two, and it is fair to say that the latter two did not receive sufficient attention.

**Typological fingerprints.** With the method of fingerprinting it possible to perform groupwise comparisons of languages (for example by geographical area or genetic affiliation) even when many of them exhibit multiple values for an attribute such as morphosyntactic strategy (§5.6.4).

![Typological fingerprints diagram]

**The Pepper scale.** This provides a means of comparing two non-binary typologies, and is particularly useful for investigating the degree to which languages recruit strategies from one construction (such as the anchoring possessive construction) for use in another (such as the binominal construction). The scale consists of five grades of similarity; these are expressed as adverbs qualifying the property ‘identical’ and must be given clear definitions for the purpose at hand, preferably in a form that can be operationalized in a non-subjective manner (§7.2.2).

always → mostly → sometimes → rarely → never
Motivational grids. The motivational grid approach pioneered by Koch is very well suited to investigating correlations between morphosyntactic strategies and semantic relations in individual languages (§7.3.1). The present data was mostly too sparse to produce interesting results, but with sufficient data points, coded for binominal type and semantic relation, significant patterns can be expected in many languages. The method involves generating simple contingency tables, testing for significance, and examining residuals.

<table>
<thead>
<tr>
<th>Language</th>
<th>Polish (52x NN, 7x5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftype2</td>
<td>htype M</td>
</tr>
<tr>
<td>adjL</td>
<td>0</td>
</tr>
<tr>
<td>adjR</td>
<td>0</td>
</tr>
<tr>
<td>cmpL</td>
<td>0</td>
</tr>
<tr>
<td>cmpR</td>
<td>0</td>
</tr>
<tr>
<td>derR</td>
<td>14</td>
</tr>
<tr>
<td>genL</td>
<td>0</td>
</tr>
<tr>
<td>prpl</td>
<td>0</td>
</tr>
</tbody>
</table>

R scripts. Almost all the tables and figures in this dissertation (including the one above) were generated directly using R. The scripts are publicly available and I encourage other researchers to adapt and use them for their own purposes.

Binominals database. All the data used for this project is freely available in the form of an Excel workbook at https://dataverse.no/dataverse/trolling. The database table structures are described in Appendix F. Please reuse it, please inform me of any errors (bearing in mind that I am not an expert on the 106 languages of the sample), and please contact me if you want to contribute more data to this project, especially from languages of NG/Australia, North America and South America.
10.3 Further research

This study was the first systematic investigation into the typology and semantics of binominal lexemes and, as such, it has been mostly exploratory. Nevertheless it has produced a number of new insights and interesting results, as the preceding section shows. But, as befits an exploratory study, it has thrown up many more new questions than it has answered and provided plenty of scope for further research.

The data set consists of over 10,000 data points: lexical items from 106 languages that express 100 meanings. 3,738 of these are binominals, all of which have been annotated for a number of morphosyntactic and semantic properties: morphological structure, gloss, language-specific construction, binominal type, head, modifier, constituent order and semantic relation; and, of course, language and meaning, each of which bring with them additional properties, including areal and genetic affiliation, semantic type and semantic field. That is a lot of data and I was able to do quite a lot with it. Nevertheless, some investigations I wished to make required more. Furthermore, the language sample was not balanced; this had the advantage that I was able to investigate some microvariation (e.g. the binominal fingerprints of Indo-European languages, §5.6.4), but it also meant that I was unable to make robust typological claims regarding certain matters. The first task for me would therefore be to extend the sample for better genetic and areal balance. This means expanding the representation of NG/Australia, North America and South America; adding data from a further 11 languages would allow me to create an areally balanced sample of 72 languages, 12 per major geographical area. So if anyone is in a position to supply me with data from a language from one of these parts of the world, in particular, one from a genus that is not already represented in the database,
I would love to hear from you. I will do the annotation and number crunching once you have filled in the questionnaire in Appendix G.

Other topics for further research call for even more collaboration. Foremost among these is testing and refining the PHAB model and its central component, the Hatcher-Bourque classification. The latter is ideal for investigating the semantic relations of noun-noun compounds and other binominals in individual languages. So, instead of reinventing the wheel, I urge future researchers to come together and contribute to the development of a system that makes it possible to compare our results across languages. With the help of the Bourquifier, it is very easy to apply Hatcher-Bourque consistently and it gives you two levels of granularity for the price of one. If you find a bug or a shortcoming, let me know and together we can fix it.

The PHAB model is even more of a work-in-progress. It needs to be tested against more appositional compounds, more onomasiological types and more metonymical patterns. The subtypes of SIMILARITY and CONTRAST require particular attention. It would also be interesting to see how semantic roles, case systems and semantic word-substitution speech errors can be accommodated, and there are a number of other questions that need to be addressed, including:

- Is it possible to define subhierarchies of CONTAINMENT and CAUSATION based on logically consistent criteria?
- What is the logical relationship between USAGE, FUNCTION and PURPOSE?
- Why are TOPIC, FUNCTION and PURPOSE non-reversible?
- Which semantic types (as discussed on page 81) can play the different roles in the various relations?

If such questions could be answered, we could be confident that we have a model that truly provides interesting insights into the associative nature of human thought.

More work remains to be done also on the typological classification of binominals. The cls type, in particular, needs to be fleshed out with many more examples from many more languages, in order to understand the real nature of this category. A good challenge would also be to apply Croft’s classification of morphosyntactic strategies instead of mine and see if they throw up any more interesting insights. Then there is the matter of the “missing types”, labelled (prn) and (nml) in the diagram at the beginning of §10.2.1. The former may well exist in some language out there; if so, please help me find it! The latter, I claim, is a logical impossibility, but I am willing to be proven wrong.
The question of competition between different binominal constructions in one and the same language has been raised at several points and is a suitable topic for studies that focus on a single language. Many factors can play a role: alienability, borrowing, diachronic word order change, the analogical pull of items already in the lexicon, and more … but also the semantic relation inherent in the conceptualization. The latter can be investigated using the motivational grid approach, but that will require samples of data larger than the average 35 binominals per language I had available. For anyone wanting to take on such a task, the Hatcher-Bourque classification, the Bourquifier and the binominal typology will be of great assistance.

In terms of cross-linguistic studies, both the species- vs. attribute-framing and the head- vs. modifier-framing dichotomies are worth exploring further, but they too require more data. For the latter there is also the need to develop a methodology (perhaps based on WordNet) for normalizing the raw data, so as to avoid counting, say, bee and bees as two separate concepts.

These are just some ideas for further research relating to the typology and semantics of binominal lexemes.

10.4 Envoie

In writing this work I have had the honour to stand on the shoulders of giants. My greatest sources of inspiration have been Bill Croft, Ron Langacker and Martin Haspelmath. Without Masja Koptjevskaja-Tamm’s detailed analysis of possessive noun phrases and Yves Bourque’s tremendous job of synthesizing many previous attempts to classify semantic relations, my task would have been much harder. Without Anna Granville Hatcher’s distillation of Jespersen, the innovative work of Laura Janda on metonymy in word-formation, and Pierre Arnaud’s brilliant idea of mapping a low-level classification to a higher one, this work would not have been possible in its present form. To all these linguists, living and deceased, I express my profound gratitude. And in parting, I would like to express my particular thanks, once again, to all those who contributed data for this project. Their names are listed in the Acknowledgements, in Appendix B (Sources), and in the Name index.
References


The typology and semantics of binominal lexemes


References


The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


Egmond, Marie-Elaine van. forthc. New types of binominal lexeme in Anindilyakwa (Australia). In Francesca Masini, Simone Mattiola & Steve Pepper (eds.), *When noun meets noun: Binominal lexemes in cross-linguistic perspective*. Berlin: de Gruyter.


The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


References


Lesage, Jakob. in prep. A grammar and lexicon of Kam (àŋwɔ́m), a Niger-Congo language of central eastern Nigeria. Paris: INALCO.


The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


Næss, Åshild. forthc. Seven ways for nouns to meet nouns in Äiwoo. In Francesca Masini, Simone Mattiola & Steve Pepper (eds.), When noun meets noun: Binominal lexemes in cross-linguistic perspective. Berlin: de Gruyter.


Pepper, Steve. 2007. As We REALLY May Think: Memex, Topic Maps, and subject-centric computing. Presented at the AToMS 2007, Kyoto.
References


References


The typology and semantics of binominal lexemes


https://doi.org/10.1418/30932.
Stone, Robert. 1976. K’ekchi kwa:nk, the genitive, and oblique cases. In Sandra
Stundžia, Bonifacas. 2016. Lithuanian. In Peter O. Müller, Ingeborg Ohnheiser, Susan
Olsen & Franz Rainer (eds.), Word-formation: An international handbook of the
Routledge.
World Loanword Database. Leipzig: Max Planck Institute for Evolutionary
Szuupert, Andrzej. 2012. Zur internen Semantik der substantivischen Komposita im
Dänischen. Wydawnictwo Naukowe UAM.
Press.
Szymanek, Bogdan. 2015. A Panorama of Polish Word-Formation. 2nd Edition. Lublin:
Wydawnictwo KUL.
Takada, Hareo. 2008. Le mot composé: étude contrastive de certains types de mots
composés japonais et français. Niigata University.
Tarasova, Elizaveta. 2013. Some new insights into the semantics of English N+N
compounds. Victoria University of Wellington PhD dissertation.
formation. Uppsala: Uppsala University.
University.
Oxford: Oxford University Press.
Tehraniya, Seyed Hassan. 1987. A syntactic and semantic analysis of compound nouns in
Teng, Stacy Fang-Ching. 2007. A reference grammar of Puyuma, an Austronesian
language of Taiwan. Canberra: Australian National University.
Franz Rainer (eds.), Word-formation: An international handbook of the languages of
Theil Endresen, Rolf. nd. Field notes on Nizaa: Word lists and accompanying notes.
The typology and semantics of binominal lexemes


The typology and semantics of binominal lexemes


Appendix A.

Appendices

A. Languages 383
B. Sources 387
C. Meanings 393
D. Strategies and constructions 395
E. Binominal data set 448
F. Database structures 485
G. Questionnaire 487
H. Additional tables 489
## A. Languages

This appendix lists all the languages in the sample, ordered by ISO code\(^1\) in order to facilitate looking up language names by ISO code. The genetic classification is based on Glottolog 2.7 and the geographic classification on Dryer (1992). For a list of sources, see Appendix B.

<table>
<thead>
<tr>
<th>Code</th>
<th>Language</th>
<th>Family</th>
<th>Genus</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AKK</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>2</td>
<td>AMH</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>3</td>
<td>AMR</td>
<td>Harakmbut</td>
<td>Harakmbut</td>
<td>South America</td>
</tr>
<tr>
<td>4</td>
<td>AOI</td>
<td>Gunwinyguan</td>
<td>Gunwinyguan</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>5</td>
<td>AQC</td>
<td>Archi</td>
<td>Lezgic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>6</td>
<td>ARN</td>
<td>Mapudungun</td>
<td>Araucanian</td>
<td>South America</td>
</tr>
<tr>
<td>7</td>
<td>ASM</td>
<td>Assamese</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>8</td>
<td>BAM</td>
<td>Atlantic-Congo</td>
<td>Mande</td>
<td>Africa</td>
</tr>
<tr>
<td>9</td>
<td>BQJ</td>
<td>Atlantic-Congo</td>
<td>North-Central Atlantic</td>
<td>Africa</td>
</tr>
<tr>
<td>10</td>
<td>BVA</td>
<td>Barain</td>
<td>Chadic</td>
<td>Africa</td>
</tr>
<tr>
<td>11</td>
<td>CAI</td>
<td>Caijia</td>
<td>Macro-Bai</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>12</td>
<td>CAR</td>
<td>Galibi Carib</td>
<td>Guianan</td>
<td>South America</td>
</tr>
<tr>
<td>13</td>
<td>CES</td>
<td>Czech</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>14</td>
<td>CJP</td>
<td>Cabécar</td>
<td>Chibchan</td>
<td>South America</td>
</tr>
<tr>
<td>15</td>
<td>CLI</td>
<td>Chakali</td>
<td>Gur</td>
<td>Africa</td>
</tr>
<tr>
<td>16</td>
<td>CMN</td>
<td>Mandarin Chinese</td>
<td>Sinitic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>17</td>
<td>CRS</td>
<td>Seychelles Creole</td>
<td>French-based</td>
<td>Pidgins/Croles</td>
</tr>
<tr>
<td>18</td>
<td>CWG</td>
<td>Ceq Wong</td>
<td>Aslian</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>19</td>
<td>CYM</td>
<td>Welsh</td>
<td>Celtic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>20</td>
<td>DEU</td>
<td>German</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>21</td>
<td>DSB</td>
<td>Lower Sorbian</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>22</td>
<td>ELL</td>
<td>Greek</td>
<td>Greek</td>
<td>Eurasia</td>
</tr>
<tr>
<td>23</td>
<td>ENG</td>
<td>English</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>24</td>
<td>EST</td>
<td>Estonian</td>
<td>Finnic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>25</td>
<td>EUS</td>
<td>Central Yupik</td>
<td>Yupik</td>
<td>North America</td>
</tr>
<tr>
<td>26</td>
<td>EUS</td>
<td>Basque</td>
<td>Basque</td>
<td>Eurasia</td>
</tr>
<tr>
<td>27</td>
<td>FIN</td>
<td>Finnish</td>
<td>Finnic</td>
<td>Eurasia</td>
</tr>
</tbody>
</table>

\(^1\) Caijia does not have an ISO 639 code so I have co-opted CAI in this work, since it is currently unused in ISO 639. The Glottocode, caij1234, should be used in database applications.
# The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Code</th>
<th>Language</th>
<th>Family</th>
<th>Genus</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>FRA</td>
<td>French</td>
<td>Romance</td>
<td>Eurasia</td>
</tr>
<tr>
<td>29</td>
<td>GLE</td>
<td>Irish</td>
<td>Celtic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>30</td>
<td>GNT</td>
<td>Warta Thuntai</td>
<td>Morehead-Wasur</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>31</td>
<td>GOH</td>
<td>Old High German</td>
<td>Indo-European</td>
<td>Germanic</td>
</tr>
<tr>
<td>32</td>
<td>GUE</td>
<td>Gurindji</td>
<td>Desert Nyungic</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>33</td>
<td>GUN</td>
<td>Mbyá Guaraní</td>
<td>Tupian</td>
<td>North America</td>
</tr>
<tr>
<td>34</td>
<td>GWD</td>
<td>Gawwada</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>35</td>
<td>HAU</td>
<td>Hausa</td>
<td>Chadic</td>
<td>Africa</td>
</tr>
<tr>
<td>36</td>
<td>HAW</td>
<td>Hawaiian</td>
<td>Oceanic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>37</td>
<td>HEB</td>
<td>Hebrew</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>38</td>
<td>HIN</td>
<td>Hindi</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>39</td>
<td>HRV</td>
<td>Croatian</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>40</td>
<td>HUN</td>
<td>Hungarian</td>
<td>Hungarian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>41</td>
<td>HUU</td>
<td>Murui Huitoto</td>
<td>Huitotoan</td>
<td>South America</td>
</tr>
<tr>
<td>42</td>
<td>IND</td>
<td>Indonesian</td>
<td>Malayo-Sumbawan</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>43</td>
<td>IRK</td>
<td>Iraqw</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>44</td>
<td>ITA</td>
<td>Italian</td>
<td>Romance</td>
<td>Eurasia</td>
</tr>
<tr>
<td>45</td>
<td>JPN</td>
<td>Japanese</td>
<td>Japanese</td>
<td>Eurasia</td>
</tr>
<tr>
<td>46</td>
<td>JUP</td>
<td>Hupdë</td>
<td>Nadahup</td>
<td>South America</td>
</tr>
<tr>
<td>47</td>
<td>KAP</td>
<td>Bezhta</td>
<td>Avar-Andic-Tsezic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>48</td>
<td>KDX</td>
<td>Kam</td>
<td>Adamawa-Ubangi</td>
<td>Africa</td>
</tr>
<tr>
<td>49</td>
<td>KEK</td>
<td>Kekchí</td>
<td>Core Mayan</td>
<td>North America</td>
</tr>
<tr>
<td>50</td>
<td>KET</td>
<td>Ket</td>
<td>Yeniseian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>51</td>
<td>KGV</td>
<td>Kalamang</td>
<td>West Bomberai</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>52</td>
<td>KNC</td>
<td>Kanuri</td>
<td>Saharan</td>
<td>Africa</td>
</tr>
<tr>
<td>53</td>
<td>KOR</td>
<td>Korean</td>
<td>Koreanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>54</td>
<td>KPZ</td>
<td>Kupsabiny</td>
<td>Nilotic</td>
<td>Africa</td>
</tr>
<tr>
<td>55</td>
<td>KTB</td>
<td>Kambaata</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>56</td>
<td>KWB</td>
<td>Baa</td>
<td>Gur</td>
<td>Africa</td>
</tr>
<tr>
<td>57</td>
<td>LAV</td>
<td>Latvian</td>
<td>Indo-European</td>
<td>Baltic</td>
</tr>
<tr>
<td>58</td>
<td>LET</td>
<td>Lithuanian</td>
<td>Indo-European</td>
<td>Baltic</td>
</tr>
<tr>
<td>59</td>
<td>LSR</td>
<td>Srenge</td>
<td>Nuclear Torricelli</td>
<td>Nuclear Torricelli</td>
</tr>
<tr>
<td>60</td>
<td>MAL</td>
<td>Malayalam</td>
<td>Dravidian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>61</td>
<td>MLT</td>
<td>Maltese</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>62</td>
<td>MRJ</td>
<td>Western Mari</td>
<td>Uralic</td>
<td>Mari</td>
</tr>
<tr>
<td>63</td>
<td>MWW</td>
<td>Hmong Daw</td>
<td>Hmong-Mien</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>64</td>
<td>MYK</td>
<td>Mamara Senoufo</td>
<td>Atlantic-Congo</td>
<td>Kwa</td>
</tr>
<tr>
<td>65</td>
<td>MZH</td>
<td>Wichita</td>
<td>Matacoan</td>
<td>South America</td>
</tr>
<tr>
<td>66</td>
<td>NAV</td>
<td>Navajo</td>
<td>Athabaskan-Eyak-Tlingit</td>
<td>Athabaskan-Eyak-Tlingit</td>
</tr>
<tr>
<td>67</td>
<td>NFL</td>
<td>Aiwooo</td>
<td>Oceanic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>Code</td>
<td>Language</td>
<td>Family</td>
<td>Genus</td>
<td>Area</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>NLD</td>
<td>Dutch</td>
<td>Indo-European</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>NMM</td>
<td>Manange</td>
<td>Sino-Tibetan</td>
<td>Bodic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>NOR</td>
<td>Norwegian</td>
<td>Indo-European</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>NPI</td>
<td>Nepali</td>
<td>Indo-European</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>ORH</td>
<td>Oroqen</td>
<td>Altaic</td>
<td>Tungusic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>OTQ</td>
<td>Querétaro Otomi</td>
<td>Otomanguean</td>
<td>Otopamean</td>
<td>North America</td>
</tr>
<tr>
<td>PES</td>
<td>Western Farsi</td>
<td>Indo-European</td>
<td>Iranian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>PLT</td>
<td>Malagasy</td>
<td>Austronesian</td>
<td>Greater Barito</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>POL</td>
<td>Polish</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>PYU</td>
<td>Puyuma</td>
<td>Austronesian</td>
<td>Formosan</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>QVI</td>
<td>Imbabura Quechua</td>
<td>Quechuan</td>
<td>Quechua II</td>
<td>South America</td>
</tr>
<tr>
<td>RIF</td>
<td>Taritit</td>
<td>Afro-Asiatic</td>
<td>Berber</td>
<td>Africa</td>
</tr>
<tr>
<td>RMC</td>
<td>Selice Romani</td>
<td>Indo-European</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>RON</td>
<td>Romanian</td>
<td>Indo-European</td>
<td>Romance</td>
<td>Eurasia</td>
</tr>
<tr>
<td>RUS</td>
<td>Russian</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>SAH</td>
<td>Yakut</td>
<td>Altaic</td>
<td>Turkic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>SEI</td>
<td>Seri</td>
<td>Seri</td>
<td>Seri</td>
<td>North America</td>
</tr>
<tr>
<td>SID</td>
<td>Sidamo</td>
<td>Afro-Asiatic</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>SID</td>
<td>Kildin Sami</td>
<td>Uralic</td>
<td>Saami</td>
<td>Eurasia</td>
</tr>
<tr>
<td>SLK</td>
<td>Slovak</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>SOM</td>
<td>Somali</td>
<td>Afro-Asiatic</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>SRM</td>
<td>Saramaccan</td>
<td>Pidgins &amp; Creoles</td>
<td>English-based</td>
<td>Pidgins/Creoles</td>
</tr>
<tr>
<td>SWH</td>
<td>Swahili</td>
<td>Atlantic-Congo</td>
<td>Bantoid</td>
<td>Africa</td>
</tr>
<tr>
<td>TBC</td>
<td>Takia</td>
<td>Austronesian</td>
<td>Oceanic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>TCA</td>
<td>Ticuna</td>
<td>Ticuna-Yuri</td>
<td>Ticuna-Yuri</td>
<td>South America</td>
</tr>
<tr>
<td>TCC</td>
<td>Datooga</td>
<td>Nilo-Saharan</td>
<td>Nilotic</td>
<td>Africa</td>
</tr>
<tr>
<td>TGL</td>
<td>Tagalog</td>
<td>Austronesian</td>
<td>Greater Central Philippine</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>THA</td>
<td>Thai</td>
<td>Tai-Kadai</td>
<td>Kam-Tai</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>TRN</td>
<td>Trinitario</td>
<td>Arawakan</td>
<td>Arawakan</td>
<td>South America</td>
</tr>
<tr>
<td>TUR</td>
<td>Turkish</td>
<td>Altaic</td>
<td>Turkic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>TWW</td>
<td>Tuwari</td>
<td>Walio</td>
<td>Walio</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>TOZ</td>
<td>Zinacantán Tzotzil</td>
<td>Mayan</td>
<td>Core Mayan</td>
<td>North America</td>
</tr>
<tr>
<td>VAN</td>
<td>Walman</td>
<td>Nuclear Torricelli</td>
<td>Nuclear Torricelli</td>
<td>Eurasia</td>
</tr>
<tr>
<td>VIE</td>
<td>Vietnamese</td>
<td>Austro-Asiatic</td>
<td>Vietic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>WIM</td>
<td>Wik-Mungkan</td>
<td>Pama-Nyungan</td>
<td>Paman</td>
<td>NG/Australia</td>
</tr>
<tr>
<td>WIN</td>
<td>Ho-Chunk</td>
<td>Siouan</td>
<td>Siouan</td>
<td>North America</td>
</tr>
<tr>
<td>WOL</td>
<td>Wolof</td>
<td>Atlantic-Congo</td>
<td>North-Central Atlantic</td>
<td>Africa</td>
</tr>
<tr>
<td>WWW</td>
<td>Wawa</td>
<td>Atlantic-Congo</td>
<td>Bantoid</td>
<td>Africa</td>
</tr>
<tr>
<td>YAQ</td>
<td>Yaqui</td>
<td>Uto-Aztecan</td>
<td>Southern Uto-Aztecan</td>
<td>North America</td>
</tr>
</tbody>
</table>
B. Sources

This appendix lists the sources used for each language, ordered by language name (in order to facilitate looking up ISO codes from language names). Contributors who have provided data specifically for the present project, either for the full set of meanings, or for meanings supplementary to WOLD, are listed by name. Other sources are:

- **W**: WOLD vocabulary
- **D**: dictionary or wordlist
- **G**: grammar, grammar sketch or other grammatical information

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Äiwoo</td>
<td>NFL</td>
<td>Åshild Næss</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Wurm 1981; Næss 2006; 2018; forthc.)</td>
</tr>
<tr>
<td>Akkadian</td>
<td>AKK</td>
<td>Rodrigo Hernaiz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Ungnad 1992; Worthington 2010)</td>
</tr>
<tr>
<td>Amharic</td>
<td>AMH</td>
<td>Mengistu Amberber</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Leslau 1995)</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>AOI</td>
<td>Marie-Elaine Van Egmond</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (van Egmond 2012; forthc.)</td>
</tr>
<tr>
<td>Archi</td>
<td>AQC</td>
<td>W: (Chumakina 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Kibrik 1994)</td>
</tr>
<tr>
<td>Assamese</td>
<td>ASM</td>
<td>Bipasha Patgiri</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Goswami &amp; Tamuli 2003)</td>
</tr>
<tr>
<td>Baa</td>
<td>KWB</td>
<td>Mirjam Möller Nwadigo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Möller Nwadigo 2018)</td>
</tr>
<tr>
<td>Bambara</td>
<td>BAM</td>
<td>Sekou Coulibaly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Dumestre 2003)</td>
</tr>
<tr>
<td>Bandial</td>
<td>BQJ</td>
<td>Rachel Watson</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Watson 2015)</td>
</tr>
<tr>
<td>Barain</td>
<td>BVA</td>
<td>Joey Lovestrand</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Lovestrand 2012)</td>
</tr>
<tr>
<td>Basque</td>
<td>EUS</td>
<td>Koldo Garai</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Saltarelli 1988; Coyos 2004; Rijk 2008; Artiagoitia, Hualde &amp; Urbina 2016)</td>
</tr>
<tr>
<td>Bezhta</td>
<td>KAP</td>
<td>Bernard Comrie and Madzhid Khalilov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Comrie &amp; Khalilov 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Kibrik &amp; Testelets 2004)</td>
</tr>
<tr>
<td>Cabécar</td>
<td>CJP</td>
<td>Guillermo González</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Quesada 2007; Verhoeven 2012)</td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caijia1</td>
<td>CAI</td>
<td>Shanshan Lü</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Central Yupik</td>
<td>ESU</td>
<td>D: (Jacobson 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Jacobson 1995; Miyaoka 2012)</td>
</tr>
<tr>
<td>Ceq Wong</td>
<td>CWG</td>
<td>Nicole Kruspe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Kruspe 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Kruspe, Burenhult &amp; Wnuk 2015)</td>
</tr>
<tr>
<td>Chakali</td>
<td>CLI</td>
<td>Jonathan Brindle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Brindle 2017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Brindle 2017)</td>
</tr>
<tr>
<td>Croatian</td>
<td>HRV</td>
<td>Višnja Cicin-Sain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Grubišić 2007)</td>
</tr>
<tr>
<td>Czech</td>
<td>CES</td>
<td>Pavlína Pešková</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Janda &amp; Townsend 2002; Naughton 2005; Štichauer 2009)</td>
</tr>
<tr>
<td>Datooga</td>
<td>TCC</td>
<td>Alice Mitchell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Rottland 1983)</td>
</tr>
<tr>
<td>Dutch</td>
<td>NLD</td>
<td>W: (van der Sijs 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Donaldson 1981; Booij 2002; Don 2009)</td>
</tr>
<tr>
<td>English</td>
<td>ENG</td>
<td>W: (Grant 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Dixon 2005)</td>
</tr>
<tr>
<td>Estonian</td>
<td>EST</td>
<td>Reili Argus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Harms 1962; Tauli 1973; 1983)</td>
</tr>
<tr>
<td>Finnish</td>
<td>FIN</td>
<td>Matti Miestamo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Sulkala &amp; Karjalainen 1992; Karlsson 2013; Mahieu 2013)</td>
</tr>
<tr>
<td>French</td>
<td>FRA</td>
<td>Yves Bourque</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Corréard et al. 2009; Bourque 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Lang &amp; Perez 2004; Booth 2010; Bourque 2014)</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>CAR</td>
<td>W: (Renault-Lescure 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Courtz 2008)</td>
</tr>
<tr>
<td>Gawwada</td>
<td>GWD</td>
<td>Mauro Tosco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Tosco 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Tulu 2003; Tosco 2007)</td>
</tr>
<tr>
<td>German</td>
<td>DEU</td>
<td>Anne Krause</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Whittle et al. 2011)</td>
</tr>
<tr>
<td>Greek</td>
<td>ELL</td>
<td>Katerina Fragkopoulou</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Holton et al. 2012; Ralli 2013)</td>
</tr>
<tr>
<td>Gurindji</td>
<td>GUE</td>
<td>W: (McConvell 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: –</td>
</tr>
<tr>
<td>Harakmbut</td>
<td>AMR</td>
<td>An Van linden</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Van linden 2019; forthc.)</td>
</tr>
<tr>
<td>Hausa</td>
<td>HAU</td>
<td>W: (Awagana &amp; Wolff 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Newman 1990; Newman 2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Newman 2000; Jagger 2001; Ahmad 1994; Abraham 1968)</td>
</tr>
</tbody>
</table>

1 Caijia (glottocode caij1234) does not have an ISO 639 code. CAI is currently unused in ISO 639 and has therefore been co-opted in the present work.
### Appendix B. Sources

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
</table>
| Hawaiian       | HAW  | W: (Jones 2009)  
                |       | D: (Pukui & Elbert 1971)  
                |       | G: (Elbert & Pukui 1979)  |
| Hebrew         | HEB  | Julian Lysvik  
                |       | G: (Levi 1976; Glinert 1989; Borer 2009)  |
| Hindi          | HIN  | Claus Peter Zoller  
                |       | D: (McGregor 1993; Verma & Sahai 2003)  
                |       | G: (Kachru 2006)  |
| Hmong Daw      | MWW  | Martha Ratliff  
                |       | W: (Ratliff 2009)  
                |       | D: (Heimbach 1979)  
                |       | G: (Mottin 1978)  |
| Ho-Chunk       | WIN  | Iren Hartmann  
                |       | D: (Marino 1968)  
                |       | G: (Lipkind 1945; Helmbrecht 2003)  |
| Hungarian      | HUN  | Zsofia Schön  
                |       | G: (Tompa 1985; Rounds 2001)  |
| Hupdë          | JUP  | Patience Epps  
                |       | W: (Epps 2009)  
                |       | G: (Epps 2008)  |
| Imbabura Quechua | QVI   | W: (Rendón 2009)  
| Indonesian     | IND  | W: (Tadmor 2009)  
                |       | D: (Echols & Shadily 1975)  
                |       | G: (Sneddon 1996)  |
| Iraqw          | IRK  | Maarten Mous  
                |       | W: (Mous 2009)  
                |       | D: (Mous, Qorro & Kiessling 2002)  
                |       | G: (Nordbustad 1988; Mous 1993)  |
| Irish          | GLE  | Cormac MacAindir  
                |       | G: (Ahlqvist 1985; Doyle 2001; Craig 2014)  |
| Italian        | ITA  | Francesca Masini  
                |       | D: (Bareggi 2010)  
                |       | G: (Maiden 2007; Masini & Scalise 2012; Proudfoot & Cardo 2013; Masini 2016)  |
| Japanese       | JPN  | W: (Schmidt 2009)  
| Kalamang       | KGV  | Eline Visser  
                |       | G: (Visser 2016)  |
| Kam            | KDX  | Jakob Lesage  
                |       | G: (Lesage in prep)  |
| Kambaata       | KTB  | Yvonne Treis and Deginet Wotango  
                |       | G: (Treis 2008)  |
| Kanuri         | KNC  | W: (Löhr & Wolff 2009)  
                |       | G: (Lukas 1937; Hutchison 1981; Cyffer 1998)  |
| Kekchí         | KEK  | W: (Wichmann & Hull 2009)  
                |       | G: (Stone 1976; Tzoc 2003)  |
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ket</td>
<td>KET</td>
<td>Edward Vajda and Andrey Nefedov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Vajda &amp; Nefedov 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Maksunova 2003; Vajda 2004; Georg 2007; Smith 2014)</td>
</tr>
<tr>
<td>Kildin Sami</td>
<td>SJD</td>
<td>Michael Rießler</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Rießler 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Rießler 2007; Rießler 2009; to appear)</td>
</tr>
<tr>
<td>Korean</td>
<td>KOR</td>
<td>Petter Mæhlum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Sohn 1994; Song 2005; Yeon &amp; Brown 2011)</td>
</tr>
<tr>
<td>Kupsabiny</td>
<td>KPZ</td>
<td>Kazuhiro Kawachi</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Montgomery 1966; O’Brien &amp; Cuypers 1975)</td>
</tr>
<tr>
<td>Latvian</td>
<td>LAV</td>
<td>Agnē Navickaitė-Klišauskienē</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Turkina 1982; Belzēja 1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Mathiassen 1997; Pauļiņš 2012; Navickaitė-Klišauskienē 2016)</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>LIT</td>
<td>Agnē Navickaitė-Klišauskienē</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Mathiassen 1996; Ambrazas 1997; Stundžia 2016)</td>
</tr>
<tr>
<td>Lower Sorbian</td>
<td>DSB</td>
<td>Hauke Bartels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Bartels 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Stone 1993; Schuster-Šewc 1996; Pohontsch 2016)</td>
</tr>
<tr>
<td>Malagasy</td>
<td>PLT</td>
<td>W: (Adelaar 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Rasoloson &amp; Rubino 2005)</td>
</tr>
<tr>
<td>Malayalam</td>
<td>MAL</td>
<td>Árathi Raghunathan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Ravindran 1973; Asher &amp; Kumari 1997)</td>
</tr>
<tr>
<td>Maltese</td>
<td>MLT</td>
<td>Benjamin Saade</td>
</tr>
<tr>
<td>Mamara Senoufo</td>
<td>MYK</td>
<td>Sekou Coulibaly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Manange</td>
<td>NMM</td>
<td>W: (Hildebrandt 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Hildebrandt 2004)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>CMN</td>
<td>Shuqiong Wu</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td>W: (Wiebusch 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Yip &amp; Rimmington 2004; Ceccagno &amp; Basciano 2009; Luo 2013)</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>ARN</td>
<td>Antonio Chihuaicura, Belén Villena, Fernando Zúñiga</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Golluscio, Fraguas &amp; Mellico 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Smeets 2008; Baker &amp; Fasola 2009; Zúñiga 2000; 2014)</td>
</tr>
<tr>
<td>Mbyá Guaraní</td>
<td>GUN</td>
<td>Bob Dooley</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Martins 2003)</td>
</tr>
<tr>
<td>Murui Huitoto</td>
<td>HUU</td>
<td>Kasia Wojtylak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Wojtylak 2015; Wojtylak 2017)</td>
</tr>
<tr>
<td>Navajo</td>
<td>NAV</td>
<td>Steve Pepper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Young &amp; Morgan 1980)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Reichard 1951; Young &amp; Morgan 1980; Goossen 1995)</td>
</tr>
<tr>
<td>Nepali</td>
<td>NPI</td>
<td>Krishna Prasad Parajuli</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Jayaraj 1990)</td>
</tr>
<tr>
<td>Norwegian</td>
<td>NOR</td>
<td>Steve Pepper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Faarlund, Lie &amp; Vannebo 1997; Golden, Mac Donald &amp; Ryen 2008; Askedal 2016)</td>
</tr>
</tbody>
</table>
## Appendix B. Sources

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old High German</td>
<td>GOH</td>
<td>W: (Schuhmann 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Oroqen</td>
<td>ORH</td>
<td>W: (Fengxiang Li &amp; Whaley 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Nedjalkov 1997)</td>
</tr>
<tr>
<td>Polish</td>
<td>POL</td>
<td>Pavel Urbanik</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Linde-Usiekniwicz &amp; Smith 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Swan 2002; Szymanek 2009; 2015; Cetnarowska 2014; Nagórko 2016)</td>
</tr>
<tr>
<td>Puyuma</td>
<td>PYU</td>
<td>Stacy Fang Ching Teng</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Cauquelin 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Teng 2007)</td>
</tr>
<tr>
<td>Querétaro Otomi</td>
<td>OTQ</td>
<td>W: (Bakker &amp; Hekking 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Palancar 2006; 2009)</td>
</tr>
<tr>
<td>Romanian</td>
<td>RON</td>
<td>W: (Schulte 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Mallinson 1986; Grossmann 2012; Giurigea &amp; Dobrovie-Sorin 2013;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nicolae 2013)</td>
</tr>
<tr>
<td>Russian</td>
<td>RUS</td>
<td>Svetlana Sokolova</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Timberlake 2004; Wade 2011; Petegham &amp; Paykin 2013)</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>SRM</td>
<td>Jeff Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Good 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Veenstra 2006; McWhorter &amp; Good 2012)</td>
</tr>
<tr>
<td>Selice Romani</td>
<td>RMC</td>
<td>W: (Elšik 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Holzinger 1995; Koptjevskaia-Tamm 2000)</td>
</tr>
<tr>
<td>Seri</td>
<td>SEI</td>
<td>Steve Marlett</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Marlett 1981)</td>
</tr>
<tr>
<td>Seychelles Creole</td>
<td>CRS</td>
<td>W: (Michaelis 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Alsdorf-Bollée 1977; Corne 1977; Arends et al. 2006)</td>
</tr>
<tr>
<td>Sidamo</td>
<td>SID</td>
<td>Kjell-Magne Yri</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Kawachi 2007)</td>
</tr>
<tr>
<td>Slovak</td>
<td>SLK</td>
<td>Pavol Štekauer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Mistrík 1983; Naughton 1997)</td>
</tr>
<tr>
<td>Somali</td>
<td>SOM</td>
<td>Nina Hagen Kaldhol</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Saeed 1993; 1999)</td>
</tr>
<tr>
<td>Srenge</td>
<td>LSR</td>
<td>Lea Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Swahili</td>
<td>SWH</td>
<td>W: (Schadeberg 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Polomé 1967; Van de Velde 2013)</td>
</tr>
<tr>
<td>Tagalog</td>
<td>TGL</td>
<td>Mark Donohue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (English 1977; 1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Schachter &amp; Otanes 1972)</td>
</tr>
<tr>
<td>Takia</td>
<td>TBC</td>
<td>Malcolm Ross</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Ross 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Ross 2002)</td>
</tr>
<tr>
<td>Tarifit</td>
<td>RIF</td>
<td>Maarten Kossmann and Khalid Mourigh</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Kossmann 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Theil Endresen 1990; Kossmann 2000; Putten 2013)</td>
</tr>
</tbody>
</table>
### The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thai</td>
<td>THA</td>
<td>Titima Suthiwan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W: (Suthiwan 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Singnoi 2000; Smyth 2002; Iwasaki &amp; Ingkaphirom 2005)</td>
</tr>
<tr>
<td>Ticuna</td>
<td>TCA</td>
<td>Denis Bertet</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Anderson 1962)</td>
</tr>
<tr>
<td>Trinitario</td>
<td>TRN</td>
<td>Françoise Rose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Rose 2015)</td>
</tr>
<tr>
<td>Turkish</td>
<td>TUR</td>
<td>Ashl Gürer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Akdikmen 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Kornfilt 1997; Göksel &amp; Kerslake 2005; Göksel &amp; Haznedar 2007; Göksel 2009)</td>
</tr>
<tr>
<td>Tuwari</td>
<td>TWW</td>
<td>Sylvain Loiseau</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>VIE</td>
<td>W: (Kruspe 2009)</td>
</tr>
<tr>
<td>Walman</td>
<td>VSN</td>
<td>Lea Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: –</td>
</tr>
<tr>
<td>Warta Thuntai</td>
<td>GNT</td>
<td>Kyla Quinn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Quinn 2014)</td>
</tr>
<tr>
<td>Wawa</td>
<td>WWW</td>
<td>Marieke Martin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Martin 2012)</td>
</tr>
<tr>
<td>Welsh</td>
<td>CYM</td>
<td>Gwen Awbery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Hawke, Fychan &amp; Roberts 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Jones &amp; Thomas 1977; Thorne 1993; King 2003; Awbery 2004; Awbery 2014)</td>
</tr>
<tr>
<td>Western Mari</td>
<td>MRT</td>
<td>Polina Pleshak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Riese et al. 2010; Riese 2016; Pleshak 2018)</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>PES</td>
<td>Cormac MacAindir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Tehranisa 1987; Mahootian 1997; Mace 2003; Foroodi-Nejad &amp; Paradis 2009)</td>
</tr>
<tr>
<td>Wichi</td>
<td>MZH</td>
<td>W: (Vidal &amp; Nercesian 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Nercesian 2014)</td>
</tr>
<tr>
<td>Wik-Mungkan</td>
<td>WIM</td>
<td>David Osgarby</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Milham 1974; Kilham et al. 1986)</td>
</tr>
<tr>
<td>Wolof</td>
<td>WOL</td>
<td>Sokhna Bao</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Fal 1990; Diouf 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Ngom 2003)</td>
</tr>
<tr>
<td>Yakut</td>
<td>SAH</td>
<td>Innokentij Novgorodov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D: (Nedjalkov 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Stachowski &amp; Menz 1998)</td>
</tr>
<tr>
<td>Yaqui</td>
<td>YAQ</td>
<td>W: (Fernández 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G: (Lindenfeld 1973; Fernández &amp; Gonzalez 2008)</td>
</tr>
<tr>
<td>Zinacantán</td>
<td>TZO</td>
<td>W: (Brown 2009)</td>
</tr>
<tr>
<td>Tzotzil</td>
<td>TZO</td>
<td>G: (Cowan 1969; Haviland 2001)</td>
</tr>
</tbody>
</table>
C. Meanings

This appendix lists all 100 meanings in alphabetical order, together with the size of the sub-vocabulary, the number of analysable forms and the number of binominals (NN). These data are plotted in order of sub-vocabulary size in Figure 27 on page 126. Information regarding how the meanings are categorized by semantic field and semantic type can be found in Table 18 on page 78. Additional statistics are given in §4.4.2.

<table>
<thead>
<tr>
<th>meaning</th>
<th>size</th>
<th>anal.</th>
<th>NN</th>
<th>meaning</th>
<th>size</th>
<th>anal.</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANKLE</td>
<td>111</td>
<td>58</td>
<td>47</td>
<td>FISHERMAN</td>
<td>101</td>
<td>86</td>
<td>28</td>
</tr>
<tr>
<td>ARCTIC LIGHTS</td>
<td>54</td>
<td>46</td>
<td>36</td>
<td>FISHING LINE</td>
<td>94</td>
<td>61</td>
<td>34</td>
</tr>
<tr>
<td>BACKPACK</td>
<td>62</td>
<td>37</td>
<td>30</td>
<td>FLAME</td>
<td>102</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>BEE</td>
<td>126</td>
<td>34</td>
<td>25</td>
<td>FLEA MARKET</td>
<td>42</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>BEEHIVE</td>
<td>105</td>
<td>68</td>
<td>56</td>
<td>FOAL OR COLT</td>
<td>89</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>BEESWAX</td>
<td>82</td>
<td>55</td>
<td>50</td>
<td>FOOTPRINT</td>
<td>103</td>
<td>63</td>
<td>47</td>
</tr>
<tr>
<td>BICYCLE PUMP</td>
<td>57</td>
<td>51</td>
<td>42</td>
<td>GIRL</td>
<td>115</td>
<td>58</td>
<td>38</td>
</tr>
<tr>
<td>BICYCLE</td>
<td>99</td>
<td>33</td>
<td>13</td>
<td>GLOVE</td>
<td>84</td>
<td>49</td>
<td>38</td>
</tr>
<tr>
<td>BLACKSMITH</td>
<td>92</td>
<td>54</td>
<td>26</td>
<td>GOLD RING</td>
<td>66</td>
<td>64</td>
<td>56</td>
</tr>
<tr>
<td>BOY</td>
<td>116</td>
<td>52</td>
<td>34</td>
<td>HAND BRAKE</td>
<td>52</td>
<td>46</td>
<td>40</td>
</tr>
<tr>
<td>BRACELET</td>
<td>103</td>
<td>50</td>
<td>30</td>
<td>HANDBAG</td>
<td>79</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>BREAKFAST</td>
<td>96</td>
<td>62</td>
<td>29</td>
<td>HANDKERCHIEF OR RAG</td>
<td>114</td>
<td>50</td>
<td>31</td>
</tr>
<tr>
<td>CARPENTER</td>
<td>90</td>
<td>48</td>
<td>25</td>
<td>HERDSMAN</td>
<td>91</td>
<td>58</td>
<td>14</td>
</tr>
<tr>
<td>CHIEFTAIN</td>
<td>119</td>
<td>53</td>
<td>27</td>
<td>HOSPITAL</td>
<td>99</td>
<td>55</td>
<td>22</td>
</tr>
<tr>
<td>COCK/ROOSTER</td>
<td>103</td>
<td>40</td>
<td>33</td>
<td>HOST</td>
<td>82</td>
<td>49</td>
<td>27</td>
</tr>
<tr>
<td>COLLARBONE</td>
<td>84</td>
<td>57</td>
<td>46</td>
<td>KEYWORD</td>
<td>46</td>
<td>42</td>
<td>31</td>
</tr>
<tr>
<td>COOKHOUSE</td>
<td>83</td>
<td>53</td>
<td>19</td>
<td>KID</td>
<td>92</td>
<td>62</td>
<td>46</td>
</tr>
<tr>
<td>DAIRY COW</td>
<td>60</td>
<td>48</td>
<td>33</td>
<td>LAMB</td>
<td>99</td>
<td>47</td>
<td>34</td>
</tr>
<tr>
<td>DINNER</td>
<td>83</td>
<td>55</td>
<td>40</td>
<td>LICENSE PLATE</td>
<td>70</td>
<td>50</td>
<td>36</td>
</tr>
<tr>
<td>DOORPOST</td>
<td>85</td>
<td>52</td>
<td>44</td>
<td>LUNCH</td>
<td>93</td>
<td>51</td>
<td>36</td>
</tr>
<tr>
<td>EARLOBE</td>
<td>83</td>
<td>69</td>
<td>60</td>
<td>MAGIC</td>
<td>101</td>
<td>34</td>
<td>11</td>
</tr>
<tr>
<td>EARRING</td>
<td>104</td>
<td>66</td>
<td>41</td>
<td>MAIL BOX</td>
<td>53</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>EARWAX</td>
<td>95</td>
<td>73</td>
<td>68</td>
<td>MARE</td>
<td>86</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>EYEBROW</td>
<td>110</td>
<td>57</td>
<td>44</td>
<td>MARRIED WOMAN</td>
<td>86</td>
<td>64</td>
<td>15</td>
</tr>
<tr>
<td>EYELASH</td>
<td>100</td>
<td>66</td>
<td>54</td>
<td>MIDDAY</td>
<td>108</td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>EYELID</td>
<td>99</td>
<td>76</td>
<td>70</td>
<td>MILKY WAY</td>
<td>65</td>
<td>56</td>
<td>48</td>
</tr>
<tr>
<td>FARMER</td>
<td>99</td>
<td>63</td>
<td>32</td>
<td>MOTHER-IN-LAW</td>
<td>104</td>
<td>56</td>
<td>30</td>
</tr>
<tr>
<td>FIREPLACE</td>
<td>97</td>
<td>48</td>
<td>35</td>
<td>NATIVE COUNTRY</td>
<td>88</td>
<td>70</td>
<td>37</td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>meaning</th>
<th>size</th>
<th>anal.</th>
<th>NN</th>
<th>meaning</th>
<th>size</th>
<th>anal.</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEIGHBOUR</td>
<td>95</td>
<td>48</td>
<td>14</td>
<td>SUGAR CANE</td>
<td>91</td>
<td>46</td>
<td>35</td>
</tr>
<tr>
<td>NIECE</td>
<td>105</td>
<td>60</td>
<td>49</td>
<td>SUNDAY</td>
<td>68</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>NIPPLE OR TEAT</td>
<td>108</td>
<td>57</td>
<td>53</td>
<td>SUPPER</td>
<td>86</td>
<td>56</td>
<td>40</td>
</tr>
<tr>
<td>NOSTRIL</td>
<td>104</td>
<td>80</td>
<td>76</td>
<td>TEAR</td>
<td>101</td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>PADDLE WHEEL</td>
<td>35</td>
<td>31</td>
<td>26</td>
<td>THATCH</td>
<td>73</td>
<td>37</td>
<td>31</td>
</tr>
<tr>
<td>PALM OF HAND</td>
<td>85</td>
<td>64</td>
<td>60</td>
<td>THUMB</td>
<td>107</td>
<td>72</td>
<td>33</td>
</tr>
<tr>
<td>POSTCARD</td>
<td>65</td>
<td>38</td>
<td>30</td>
<td>TOE</td>
<td>110</td>
<td>75</td>
<td>64</td>
</tr>
<tr>
<td>POTTER</td>
<td>84</td>
<td>65</td>
<td>35</td>
<td>TOILET PAPER</td>
<td>57</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>QUEEN</td>
<td>81</td>
<td>47</td>
<td>41</td>
<td>TOILET</td>
<td>114</td>
<td>47</td>
<td>24</td>
</tr>
<tr>
<td>RAILWAY</td>
<td>67</td>
<td>57</td>
<td>57</td>
<td>TOOL</td>
<td>98</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>RAINBOW</td>
<td>107</td>
<td>56</td>
<td>39</td>
<td>TOOLBOX</td>
<td>54</td>
<td>49</td>
<td>44</td>
</tr>
<tr>
<td>RIB</td>
<td>109</td>
<td>41</td>
<td>27</td>
<td>TOOTHBRUSH</td>
<td>69</td>
<td>59</td>
<td>38</td>
</tr>
<tr>
<td>SHOEMAKER</td>
<td>79</td>
<td>57</td>
<td>21</td>
<td>TRAIN</td>
<td>85</td>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>SHORE</td>
<td>122</td>
<td>54</td>
<td>41</td>
<td>TREE TRUNK</td>
<td>103</td>
<td>60</td>
<td>52</td>
</tr>
<tr>
<td>SHOULDERBLADE</td>
<td>102</td>
<td>58</td>
<td>46</td>
<td>VEIN OR ARTERY</td>
<td>113</td>
<td>42</td>
<td>29</td>
</tr>
<tr>
<td>SKULL</td>
<td>107</td>
<td>63</td>
<td>52</td>
<td>VINE</td>
<td>85</td>
<td>34</td>
<td>26</td>
</tr>
<tr>
<td>SORCERER OR WITCH</td>
<td>121</td>
<td>67</td>
<td>38</td>
<td>WATER PUMP</td>
<td>61</td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td>SPECTACLES/GLASSES</td>
<td>94</td>
<td>47</td>
<td>23</td>
<td>WEDNESDAY</td>
<td>65</td>
<td>44</td>
<td>21</td>
</tr>
<tr>
<td>SPIDER WEB</td>
<td>95</td>
<td>73</td>
<td>66</td>
<td>WIDOWER</td>
<td>94</td>
<td>55</td>
<td>18</td>
</tr>
<tr>
<td>SPINE</td>
<td>107</td>
<td>73</td>
<td>58</td>
<td>WINDMILL</td>
<td>47</td>
<td>43</td>
<td>36</td>
</tr>
<tr>
<td>STABLE OR STALL</td>
<td>93</td>
<td>29</td>
<td>26</td>
<td>WRIST</td>
<td>105</td>
<td>66</td>
<td>57</td>
</tr>
<tr>
<td>STONE BRIDGE</td>
<td>59</td>
<td>56</td>
<td>51</td>
<td>YOLK</td>
<td>88</td>
<td>63</td>
<td>43</td>
</tr>
</tbody>
</table>
D. Strategies and constructions

This appendix provides a summary of every (cross-linguistic) strategy and every (language-specific) construction employed by each language in the sample for both anchoring (i.e. possessive) and non-anchoring (i.e. binominal) relations. Headings provide a summary of the binominal data in the following format:

**Tarifit (RIF): 96W / 18B / 4C ► prp:14 gen:3 jxt:1**

The heading consists of the language name (here, Tarifit), followed by the ISO code and three numbers that indicate the number of words (W) in the data set for the language in question (in the case of Tarifit, 96), the total number of binominals (18B), and the number of different binominal constructions (4C). Following the arrow ► is a list of the strategies employed and the number of instances of each (prp:14 gen:3 cmp:1).¹

There are two sections for each language. The first provides a brief description of attributive possessive construction(s), together with examples; the focus here is on anchoring possessives where the possessor is a full nominal. The second section comments on binominal preferences and lists every binominal construction found in the data (in order of frequency), along with the binominal type, an example and the number of occurrences. Finally there is a characterization of the relationship between possessives and binominals in terms of the Pepper scale (cf. Table 44).

**Africa**

(Afro-Asiatic) Berber

**Tarifit (RIF): 96W / 18B / 4C ► prp:14 gen:3 jxt:1**

**Possessives:** prp strategy using ‘la préposition génitiv’ with the modifier in the annexed state (Kossmann 2000: 107-108).

*axxam n wəryaz* [house of man:STC] ‘the man’s house’

*bbwà as n mḥənd* [father of Mhend:STC] ‘Mhend’s father’

¹ Note that a strategy may be employed by more than one Construction, as is the case with Tarifit, where both Head PREP Mod:STC and Head.3SG PREP Mod:STC are considered to exemplify the prp strategy.
Binominals: Strong preference for prp. The only difference between the construction used for NIECE and the dominant construction is that the former has additional marking on the head.

13  prp  Head PREP Mod:STC: abrid n mašina [road of train:STC] RAILWAY
3   gen  Head Mod:STC: tahr’us’t umožżun [ring? ear:STC] EARLOBE
1   jxt  Head Mod: frīnu manu [brake hand] HAND BRAKE
1   prp  Head.3SG PREP Mod:STC: dyoqgi.s n uma [daughter.3SG of brother:STC] NIECE

► Recruitment of anchoring strategy: mostly

(Afro-Asiatic) Chadic

Barain (BVA): 57W / 21B / 7C ► prp:11 con:7 dbl:2 jxt:1

Possessives: dbl strategy. “The possessor agreement suffix attaches to the possessum, and indexes the person, gender (if singular), and number of the possessor of the noun to which the marker is suffixed. It can express ownership or immediate possession. It can also express intangible attributes or a relationship to a person or thing… When the possessor is explicitly stated, it follows the possessum with the oblique preposition iŋ preceding it” (Lovestrand 2012: 75-79).

gerá,ji iŋ úmar [village.POSS:3M ASS Oumar] ‘the village of Oumar’

Binominals: prp and con are both common; dbl is marginal. Two prepositions occur: (i)ŋ (associative) and ta (purposive). Of the seven binominals that involve a relativizer as well as a preposition, four could be descriptive phrases rather than naming units, but the other three (including the example below) are probably not.

7   prp  Head REL ASS Mod: gera ge ŋ gargar [home REL:3SG:M ASS spider] SPIDER WEB
7   con  Mod Head.POSS: sinja guma.geti [nose hole.POSS:3SG:F] NOSTRIL
2   prp  Head PURP Mod: looli ta lutta [dirt PURP ear] EARWAX
2   dbl  Head.POSS ASS Mod: bug.eti ŋ mooro [mouth.POSS:3SG:F ASS river] SHORE
1   jxt  Head Mod: peesi mee [horse woman] MARE
1   prp  Mod ASS Head.POSS: golmo ŋ bal.ti [house ASS back.POSS:3SG:F] TOILET
1   prp  Head REL Mod.POSS: mee de mer.geti [women REL:SG:F husband.POSS:3SG:F] MARRIED WOMAN

► Recruitment of anchoring strategy: never

Hausa (HAU): 100W / 43B / 4C ► con:40 der:2 jxt:1

Possessives: con strategy. “The linker has two main variants, a free particle na(a) / ta(a), which has two grammatically conditioned allomorphs differing in vowel length, and a bound clitic -n / -r̃ (which is written connected to its host)” (Newman 2000: 300). Based on the examples given, adnominal possession usually involves the free particle.
gidā na Sulè [house LK Sule] ‘Sule’s house’
bàkām na mahàrībi [bow LK hunter] ‘the hunter’s bow’
kibiyàt ta mahàrībi [arrow LK hunter] ‘the hunter’s arrow’

**Binominals:** Overwhelming preference for con using the bound clitic linker.

40 con Head.LK Mod: kàfá.r háncìi [orifice.LK nose] NOSTRIL
1 der Base.ABST: māi.táa [sorcerer.ABST] MAGIC
1 der Base.F: sàráu.niyáa [king.F] QUEEN
1 jxt Mod Head: ráanáa tsákàa [sun mid] MIDDAY

► Recruitment of anchoring strategy: **always**

**(Afro-Asiatic) Cushitic**

**Gawwada (GWD): 74W / 17B / 4C ► gen:8 der:6 jxt:3**

**Possessives:** con strategy. “The possessor / the possessed noun relationship in Gawwada is shown by vowel length, whereby the final vowel of the head noun is affected” (Tulu 2003: 25), cf. (a-b). However, Tosco (2007: 526) cites a use of the locative case as genitive, cf. (c-d).

(a) laalle ongaaye.e [clothe Ongaye.GEN] ‘Ongaye’s clothe’ (sic)
(b) emte kalato.o [sheep kalato.GEN] ‘Kalato’s sheep’
(c) pāko gárm.ito [mouth lion.LOC:M] ‘the lion’s mouth’
(d) pāso hullulitt atte [field guinea-fowl.LOC:F] ‘the guinea-fowl’s field’

**Binominals:** Preference for gen. The examples from WOLD are glossed with ASSOC (M,F,P), while those in Tosco (2007) are glossed with LOC and described as both locative and genitive. However, the forms are identical, viz. -ito (M), -atte (F) and -ete (PL). Tulu (2003) uses the term ‘Source and Purposive genitive cases’ (and also regards the initial vowel as epenthetical).

8 gen Head Mod.ASS: hawdatto sikk.ete [craftsman pots.ASS] POTTER
4 der Base.M: šad.unko [breast.SG:M] NIPPLE OR TEAT
3 jxt Mod Head: paso pako [lake mouth] SHORE

► Recruitment of anchoring strategy: **never**

**Iraqw (IRK): 69W / 22B / 5C ► con:15 der:4 jxt:3**

**Possessives:** con strategy. “The possessed noun occurs in the construct form followed by the possessor noun in the basic form” (Nordbustad 1988: 105). **Note:** Sources use different orthographies.

huurusmoó waawutmo [cook:STC chief] ‘the chief’s cook’
ságw hee [head:STC person] ‘a person’s head’
maraá wahaangw [houses:STC people] ‘people’s houses’
The typology and semantics of binominal lexemes

**Binominals:** Strong preference for **con**.

| **con** | Head.CON Mod: *foxár duunga* [hole:of nose] NOSTRIL |
| **jxt** | Head Mod: *afa tlawi* [mouth lake] SHORE |
| **der** | Base.AGT: *tlee'.usmo* [clay.AGT] POTTER |
| **der** | Base.F: *wawi’ita.o’o* [king.SG:F] QUEEN |
| **der** | Base.M: *koonk.amo* [hen.SG:M] COCK/ROOSTER |

► Recruitment of anchoring strategy: **mostly**

**Kambaata (KTB):** 100W / 34B / 4C ► **gen:**31 **der:**2 **cmp:**1

**Possessives:** **gen** strategy. “The genitive is the case of nominal modifiers. There is an array of different semantic relations between modifying and modified noun, the most prototypical being the possessor-possessed relationship or the whole-part-relationship. Kambaata makes little use of compounding for the creation of nouns. Therefore, virtually all compound words of the meta-languages Amharic and English are expressed by genitive constructions in Kambaata” (Treis 2008: 118-119).


*mesel.ée oddishsh.áta* [girl.F:GEN clothes.F:ACC] ‘clothes of the / a girl’


**Binominals:** Overwhelming preference for **gen**.

| **gen** | Mod.GEN Head: *baabur.i maar.a* [train.M:GEN line.M:ACC] RAILWAY |
| **der** | Base.AGT: *hoga’.aan.ch.u* [plough.AGT.SGLT.M:ACC] FARMER |
| **gen** | Mod.GEN.POSS Head: *hiz.ee.i beet.uta* [sibling.M:GEN.1SGLT:POSS child.F:ACC] NIECE |
| **cmp** | Mod.Head: *magan.zeeb.uta* [God.??:F:ACC] RAINBOW |

► Recruitment of anchoring strategy: **always**

**Sidamo (SID):** 98W / 29B / 6C ► **gen:**20 **der:**7 **jxt:**1 **adj:**1

**Possessives:** **gen** strategy. “The genitive case is marked with both a suffix and a supraftix” (Kawachi 2007: 352:356-57).

*beett.ú farašš.i* [child.GEN:M horse.NOM:MOD:M] ‘the boy’s horse’

**Binominals:** Overwhelming preference for **gen**.

| **gen** | Mod.GEN Head: *baabur.u doogo* [train.GEN road] RAILWAY |
| **der** | Base.F: *moot.itte* [king.DER_F] QUEEN |
| **der** | Base.SUF: *boos.allo* [fireplace.DER] COOKHOUSE |
| **der** | Base.M: *lukk.oʧʧo* [hen.SGLT_M] COCK/ROOSTER |
| **jxt** | Head Mod: *taalo barra* [middle day] MIDDAY |
| **adj** | Mod.ADJ Head: *k’ar.aame k’ale* [sharp_blade.ADJZ wheel] PADDLE WHEEL |

► Recruitment of anchoring strategy: **mostly**
Somali (SOM): 95W / 19B / 8C ➤ dbl:7 gen:5 cmp:4 der:3

Possessives: **con** and **gen** strategies. “Noun phrases expressing possession occur in two types: the first… involves the use of a special possessive determiner on the possessed noun. The order of nominals is possessor-possessed” (a). In the second type (b) the order is possessed-possessor and “the possessed nominal occurs in the absolutive case while the possessor occurs in the genitive” (Saeed 1999: 175).

(b) *gūriga Calī* [house:ABS Ali:GEN] ‘Ali’s house’

**Binominals:** **gen,** **dbl** and **cmp** strategies all common. The **dbl** strategy involves determiners and is unrelated to the **gen** strategy.

5 **gen** Head Mod:GEN: *guri shinni* [house bee:GEN] BEEHIVE
4 **dbl** Head.DEF Mod.DEF: *deegaan.ta roob.ka* [arch.DEF rain.DEF] RAINBOW
3 **cmp** Head.Mod: *xuub.caaro* [sloughed_snakeskin.spider] SPIDER WEB
2 **der** Base.F: *boqor.ad* [king.F] QUEEN
2 **dbl** Head.DEF Mod:GEN: *ilays qudbi.ga waqooyi* [light pole.DEF north:GEN] ARCTIC LIGHTS
1 **der** Base.OWN: *dheri.yoo.ley* [pot.PL.OWN] POTTER
1 **dbl** Mod.DEF Head.POSS: *geed.ka gun.tiisa* [tree.DEF base.POSS] TREE TRUNK
1 **cmp** Mod.Head: *beer.a.le.y* [farm.PL.owner.PL] FARMER

► Recruitment of anchoring strategy: sometimes

(Afro-Asiatic) Semitic

Akkadian (AKK): 61W / 19B / 2C ➤ dbl:19

Possessives: **dbl** strategy. “The genitive always follows its governing noun, which is in the construct state … Instead of a construct chain one may also use a formation with *ša*. This is necessary when a genitive is separated from its governing noun by an attributive, e.g. *zērum dārium ša šarrūtim* (OB) ‘eternal seed of kingship’” (Ungnad 1993: 112-13).

*būt awīl-im* [house:STC citizen.GEN] ‘the house of the citizen’

**Binominals:** **dbl** only.

18 **dbl** Head:STC Mod:GEN: *piliš app.im* [hole:STC nose.GEN] NOSTRIL
1 **dbl** Head:STC Mod.OBL: *bīt habūb.āti* [house:STC bees.OBL] BEEHIVE

► Recruitment of anchoring strategy: **always**

Amharic (AMH): 100W / 47B / 4C ➤ gen:37 jxt:9 cmp:1

Possessives: **gen** strategy. “Possession is expressed by the element *ỳā* ‘of’ followed by the possessor. The *ỳā*+possessor structure functions as a qualifier and as such it precedes the possessed” (Leslau 1995: 191-192). **Note:** Sources use different transliterations.

*yā.tāmari māšaf* [GEN.student book] ‘a student’s book’
The typology and semantics of binominal lexemes

Binominals: Strong preference for *gen*.

37 *gen* GEN.Mod Head: ye.baburi ḥādīdi [GEN.train way] RAILWAY
8 *jxt* Mod Head: shērerīṭi diri [spider web] SPIDER WEB
1 *jxt* Head Mod: ikule k’eni [half day] MIDDAY
1 *cmp* Mod.Head: hakim.beti [doctor.house] HOSPITAL

► Recruitment of anchoring strategy: *mostly*

Hebrew (HEB): 100W / 44B / 6C ► con:35 der:4 dbl:3 jxt:1 prp:1

Possessives: *con*, *prp* and *dbl* strategies. “Hebrew has a range of ‘genitive’ constructions, i.e. constructions that are often called ‘possessives’ in a very loose sense (though only some of them are strictly possessive). There are three main types... ‘construct phrases’ are a particular juxtaposition of nouns (a)... ‘שֶׁל shel phrases’ use the preposition ‘שֶׁל shel ‘of’ to link two nouns (b)... The third type of genitive, ‘construct + שֶׁל shel phrases (‘double genitives’), is a particular blend of the two constructions (c)” (Glinert 1989: 24).

(a) *bigd.ey ha.tinok* [clothes.STC DET.baby] ‘the baby’s clothes’
(b) *ha.bgd.im shel ha.tinok* [DET.clothes.PL PREP DET.baby] ‘the baby’s clothes’
(c) *bgd.av shel ha.tinok* [clothes.3M:POSS:PL PREP DET.baby] ‘the baby’s clothes’

Binominals: Strong preference for *con* using the construct state.

35 *con* Head.STC Mod: mesil.at barzel [track.STC iron] RAILWAY
4 *der* Base.SUF: malk.a [king.F] QUEEN
2 *dbl* Head.STC DEF.Mod: sevil he.xalav [path:STC DEF.milk] MILKY WAY
1 *jxt* Head Mod: tik gav [bag back] BACKPACK
1 *prp* Head PREP:DEF.Mod: kešet be.’anan [arc in:DEF.sky] RAINBOW
1 *dbl* Head.STC Mod.ADJZ: or.ot arkt.iyim [light.STC arctic.ADJZ] ARCTIC LIGHTS

► Recruitment of anchoring strategy: *always*

Maltese (MLT): 96W / 32B / 8C ► prp:25 cmp:2 dbl:2 der:1 gen:1 con:1

Possessives: *con* and *prp* strategies. The “two structurally different possessive NPs in Maltese – Construct State-NPs (a) and constructions with ‘analytical genitives’ (b) ... are primarily distinguished as referring to *alienable vs. inalienable possession” (Koptjevskaja-Tamm 1996: 245).

(a) *bin is-sultān* [son DEF-king] ‘the king’s son’
(b) *is-siġġu ta’ Pietru* [DEF-chair of Peter] ‘Peter’s chair’

Binominals: Strong preference for *prp*. Fabri (2007) uses the term ‘lexical construct’ for Head DEF.Mod and calls Head PREP(:DEF)-Mod constructions ‘periphrastic compounds’. The latter are identical to Italian constructions like *dito del piede* [finger of:DEF foot] TOE.

20 *prp* Head PREP:DEF-Mod: qolla tan-nahal [hive of:DEF-bees] BEEHIVE
3 *prp* Head PREP Mod: saba’ ta’ sieq [digit of foot] TOE
Recruitment of anchoring strategy: sometimes

(Atlantic-Congo) Adamawa-Ubangi

Kam (KDX): 69W / 27B / 3C ► cmp:21 jxt:6

Possessives: cmp strategy. “I have not found any formal differences between attributive (or adnominal) possession constructions and N+N compounds. The construction works with regular juxtaposition: there are no linking morphemes and regular tone rules and phonological rules apply (H tone spread and deletion/centralization of prosodically weak initial â). These rules apply in all parts on the language, and also for example in N+V constructions and preposition+N constructions. The order is always possessee + possessor (which aligns with the general head + modifier order in noun phrases in the language)” (Jakob Lesage, p.c.)

ângwôg (i)wàn [house chief] ‘house of (the) chief’
àlôg mè mì [maternal_relative mother 1SG:NONS] ‘My mother’s relatives’
zûrûné wô [claw dog] ‘claw of dog’
àkùb (i)tábé [bone fish] ‘bone of fish’

Binominals: Strong preference for cmp; jxt also common.

13 cmp Head.Mod: à.fì.ghàn [NOM.child/boy.female/woman/wife] GIRL
8 cmp Head.LE.Mod: à.gù.m.i.mbìrèg [NOM.hole.LE.nose] NOSTRIL
6 jxt Head Mod: ñwè màrì [mouth river] SHORE

Recruitment of anchoring strategy: sometimes

(Atlantic-Congo) Bantoid

Swahili (SWH): 104W / 52B / 5C ► prp:39 cmp:6 cls:4 jxt:3

Possessives: prp strategy. “The connective particle {a} preceded by the pronominal concord constitutes essentially two types of complexes with a following nominal complement. [In the first type] the pronominal concord agrees with the preceding noun in compliance with the usual class-concord rules… The first type either shows the possessor of the thing under reference or describes one of its characteristic features…” (Polomé 1967: 132).

jembe la mkulima [hoe CON peasant] ‘the peasant’s hoe’
viatu vya bwana Fulani [shoes CON Mr Fulani] ‘Mr. Fulani’s shoes’
The typology and semantics of binominal lexemes

Binominals: Overwhelming preference for prp.
39 prp  Head CON Mod: mwanzi wa pua [bamboo CON nose] NOSTRIL
6 cmp  Head.Mod: kinu.upepo [mill.wind] WINDMILL
4 cls  CL.Base: fumo [CL9:spear] CHIEFTAIN
2 jxt  Head Mod: dalili mguu [sign foot] FOOTPRINT
1 jxt  Mod Head: jua kati [sun middle] MIDDAY
► Recruitment of anchoring strategy: mostly

Wawa (www): 27W / 13B / 4C ► gen:9 jxt:4

Possessives: Two strategies, both gen. “There are two possibilities to mark possession. One possibility is to order the constituents of the NP as possessor-possessed and mark the possessor with the -ɔ̀ suffix (a). The second possibility is to change the word order to possessed-possessor and mark the possessor with the suffix -wòì (b)” (Martin 2012: 211-212).

(a) mūn.ɔ̀ gùr [child.POSS foot] ‘the child’s foot’
(b) gùr mūn.ɔ̀ì [foot child.ASS] ‘the child’s foot’

Binominals: Strong preference for gen; jxt also common.
8 gen  Head Mod.POSS.ASS: jūg nāb.ɔ̀ì [house food.POSS.ASS] COOKHOUSE
3 jxt  Head Mod: mūn nugwɛ̀ɛ̀ [child woman] GIRL
1 gen  Head Mod.LOC.ASS: nābɔ̀ tɔ̀n.ì [fufu ear.LOC.ASS] EARWAX
1 jxt  Mod Head: sāŋgā tɛ̀ŋgɔ̀ [sun middle] MIDDAY
► Recruitment of anchoring strategy: mostly

(Atlantic-Congo) Gur

Baa (KWB): 90W / 47B / 5C ► cmp:26 gen:20 jxt:1

Possessives: jxt strategy. No alienability distinction (Mirjam Möller Nwadigo p.c.).

bà ̀̀ Hàdà [father Ada] ‘Ada’s father’
ìkùlù kpà̀n [chair king] ‘king’s chair’

Binominals: Strong preference for cmp; gen also common.
26 cmp  Head.Mod: krà.kì.sà [road.tree.outside] RAILWAY
14 gen  Head Mod.INAL: pà.jì.vì [mouth.milk.INAL] NIPPLE OR TEAT
5 gen  Head LK.Mod: bìsìn rà.nà.t [horse LK.female] MARE
1 jxt  Head Mod: gè̀ jàŋgà [day Yanga] WEDNESDAY
1 gen  Head Mod.AL: là.bà.lìt [land.father.AL] NATIVE COUNTRY
► Recruitment of anchoring strategy: grammaticalized
Appendix D. Strategies and constructions

Chakali (CLI): 51W / 25B / 5C ▶ cmp:20 der:5
súgló pābīī [Suglo hoe._blade] ‘Suglo’s hoe blade’
báál nū̀ [man head] ‘a man’s head’
Binominals: Strong preference for cmp.

Mamara Senoufo (MYK): 81W / 28B / 2C ▶ cmp:26 der:2
Possessives: prp and jxt strategies for both alienable and inalienable possession. “In some Senoufo languages (see Carlson 1994:477 for Supyire and Dombrowsky-Hahn 1999: 228–234 for the Minyanka dialect spoken in Yorosso) the genitive construction with the possessive marker always has a contrastive focus function. The same meaning exists in the Minyanka dialect spoken in Penesso (the dialect on which I’m interested), but in many cases, there is not any semantic distinction between the genitive construction with possessive marker and the one without any marker (by juxtaposition)” (Sekou Coulibaly, p.c.).
Sékò mò pì.kè.šèyi [Sekou POSS house.INDF:CLk.DEF:CLk] ‘Sekou’s house’
Sékò pì.kè.šèyi [Sekou house.INDF:CLk.DEF:CLk] ‘Sekou’s house’
lōfà.nù mó firi.mi [donkey.DEF:CLw POSS urine.DEF:CLk] ‘donkey’s urine’
lōfànù firiìm [donkey.DEF:CLw urine.DEF:CLk] ‘donkey’s urine’
Binominals: Overwhelming preference for cmp.

Bambara (BAM): 81W / 39B / 6C ▶ cmp:25 gen:10 jxt:3 der:1
Possessives: prp and dbl strategies for alienable and inalienable possession, respectively (Sekou Coulibaly, p.c.).
Músò ká sò [woman:DEF POSS money:DEF] ‘the woman’s house’
Músò sèn` [woman:DEF foot:DEF] ‘the woman’s foot’
Binominals: Strong preference for cmp; gen also frequent.
The typology and semantics of binominal lexemes

(Atlantic-Congo) North-Central Atlantic

**Bandial (BQJ):** 72W / 16B / 3C ► jxt:9 prp:4 cls:3

Possessives: jxt and prp strategies. “The relation between two nouns that stand in a possessor-possessee relation can be encoded in two ways. The first is direct juxtaposition, with the possessee preceding the possessor (a)… A possessive relation can also be marked with the connector AGR-a, which appears between the possessed and possessor, and where AGR corresponds to agreement with the possessee (b)” Watson (2015: 150-151).

(a) ka.at e.be [CL:ka. foot CL:e. cow] ‘cow’s foot’
(b) ka.ñen k.a pa.i [CL:ka. hand AGR:k. CON father.2S.POSS] ‘your father’s hand’

Binominals: Preference for jxt; prp also found.

| 9 | jxt | Head Mod: ka. liba e.mit [CL:knife CL:god/sky] RAINBOW |
| 4 | prp | Head CON Mod: (sinaŋ) sa këbujom [rice AGR:CON morning] BREAKFAST |
| 3 | cls | CLF.Base: ji.jamen [CLF:goat] KID |

► Recruitment of anchoring strategy: mostly

**Wolof (WOL):** 64W / 23B / 2C ► con:18 prp:5

Possessives: con strategy. “Wolof suffixes the morpheme ‘-u’ (singular) and ‘-i’ (plural) to the first noun [when] ending with a consonant to indicate the possessive relationship between two nouns (a)… If the noun does not end with a consonant, the singular genitive morpheme ‘-u’ is deleted, and the plural genitive morpheme ‘-i’ is replaced by ‘-y’” (Ngom 2003: 6). In the absence of a class of adjectives, “relative clauses are used to describe, modify or qualify a substantive or NP (noun phrase)” (Ngom 2003: 50). The relativizer -u is prefixed by a class marker that agrees with the noun class of the head (b).

(a) kër.u Mbaye [house.PER Mbaye] ‘Mbaye’s house’
(b) fas wu rëy [horse REL big] ‘big horse’

Binominals: Strong preference for con; prp also found.

| 18 | con | Head.PER Mod: pax.u bakkan [hole.PER nose] NOSTRIL |
| 5  | prp | Head REL Mod: gone gu jigéen [child REL girl] GIRL |

► Recruitment of anchoring strategy: mostly
Appendix D. Strategies and constructions

(Nilo-Saharan) Nilotic

**Datooga (TCC): 29W / 1B / 1C ► jxt:1**

**Possessives:** jxt strategy. “Pronominal possession involves suffixes but nominal possession involves juxtaposition... There are a couple of associative constructions with an extra word in between the possessee and possessor and sometimes it’s unclear what the semantic difference is between the two constructions” (Alice Mitchell, p.c.).

diibiga huda [children daughter] ‘daughter’s children’
diibiga bea huda [children ASSOC daughter] ‘daughter’s children’

**Binominals:** Based on the data received, binominals appear to be rare in Datooga; of the 29 items in the list, only one is a binominal. To the extent that they occur, however, the preferred strategy seems to be jxt.¹

1 jxt  Head.SGLT.SG Mod.SG: ñu.wéan.dá bádày.da [thread.SGLT.SG back.SG] SPFNE

► Recruitment of anchoring strategy: *always*

**Kupsabiny (KPZ): 91W / 43B / 4C ► con:35 prp:8**

**Possessives:** con strategy. “There are only three prepositions which are not incorporated in verb forms: *am, af,* and *pp*. Of these, *ap* and *pp* are alternants of a possessive morpheme, *ap* occurring only between the nouns and *pp* occurring elsewhere. Further, *ap* has the morphophonemic alternant *a*- before /m/ and /p/” (Montgomery 1966: 46). Note: The sources use different orthographies.

korket *áp sámintét* [woman POSS beggar] ‘the beggar’s woman’
rupet *á maynà* [famine POSS Mayna] ‘the Mayna famine’
lekwet *áp Cerop* [child POSS Cerop] ‘Cerop’s child’ (O’Brien & Cuypers 1975: 45)

**Binominals:** Overwhelming preference for con.

35 con  Head.POSS Mod: areet.aap karitaaap.maata [road.POSS train] RAILWAY
5 prp  Head PREP Mod: tiinkeet nyépo yoomaat [machine for air] WINDMILL
2 prp  Head REL Mod: lolleet nyéé céekáséené [bag REL:SG back] BACKPACK
1 prp  Mod REL Head: makeyooanteet wuloo mayayi [egg where yellow] YOLK

► Recruitment of anchoring strategy: *mostly*

¹ Other examples from the original (201 meaning) data set include uthùu.dá qée.da [head.SGLT house.SGLT] ROOF, hú.dá íiyà [daughter.SGLT mother] SIBLING, qèang’.dá éen.da [eye.SGLT river.SGLT] SPRING OR WELL, all jxt.
The typology and semantics of binominal lexemes

(Nilo-Saharan) Saharan

**Kanuri (KNC): 83W / 39B / 7C ► gen:23 der:12 jxt:2 cmp:1 adj:1**

**Possessives:** gen strategy. “The normal order of a NP with a modifying genitive postpositional phrase is head noun, followed by genitive postpositional phrase (a)… When genitive PPs are stacked within a given NP then the final modifier phrase is marked by only one occurrence of the genitive postposition (b)... Both of the genitive postpositions in such a stacked NP will be indicated however in the event that a determiner element intervenes between the two (c)” (Hutchison 1981: 197-198).

(a) ƒə̂r Álì.bè [horse Ali.GEN] ‘Ali’s horse’
(b) fótò bänkì kúrà Nìjérìyà.bè [foto bank central Nigeria.GEN] ‘the photo of the central Bank of Nigeria’
(c) fótò bänkì kúrà Nìjérìyà.bè Èkkóbè dòbè [foto bank central Nigeria.GEN Lagos.GEN DEM] ‘the photo of that central Bank of Nigeria of Lagos’

**Binominals:** Overwhelming preference for gen.

23 gen Head Mod.GEN: súwúlí kónznà.bè [opening nose.GEN] NOSTRIL
6 der Base.AGT: njè.má [pot.OWN] POTTER
5 der Base.LOC: kànñù.rám [fire.LOC] COOKHOUSE
2 jxt Mod Head: fə̂r kùrwúdì [horse female] MARE
1 der Base.NMLZ: kə̀mə̀gə̀n.mí [honey.NMLZ] BEESWAX
1 adj Head Mod.ADJZ: kámú nyìyá.à [woman marriage.ADJZ] MARRIED WOMAN
1 cmp Head.Mod: kúlì.kə̀mágə̀n [insect.honey] BEE

► Recruitment of anchoring strategy: *mostly*

Eurasia

(Altaic) Tungusic

**Oroqen (ORH): 60W / 24B / 7C ► der:8 dbl:7 jxt:6 gen:3**

**Possessives:** dbl and con strategies (?). No grammar of Oroqen was available for this project, but for the closely related Evenki (EVN) Nedjalkov (1997: 158-159) distinguishes two major variants of the con strategy. Inalienable possession “is expressed by the nominative case on the possessor nominal plus the marker of possession on the possessum (a)... Alienable possession is marked by the suffix -ng(i) added to the possessum. The noun with the alienable possession marker obligatorily must also have the possession marker of either personal or reflexive possession (b).”

(a) ollomimni.Ø d’av.in [fisherman.NOM boat.3SG:POSS] ‘the boat of a/the fisherman’
(b) bejumimni ulle.ngi.n [hunter meat.AL.3SG:POSS] ‘the hunter’s meat’ (i.e. which he got during hunting)
However, “the possessive relation may [also] be expressed by means of the suffix -ngi affixed to the possessor, but this is rare. In this case the possessum has the marker of personal possession (c). This type of possession with the ‘old genitive’ was preferable some fifty years ago, as the folklore texts show, but nowadays people use such possessive phrases mostly without the suffix -ngi”. The old genitive construction exemplifies the dbl strategy and is identical to the preferred binominal strategy in Oroqen.

(c) atyrkan.ngi gerbi.n [old_woman.POSS name.3SG:POSS] ‘the name of the old woman’

**Binominals:** Preference for **dbl; jxt** also common.

7 **dbl** Mod.GEN Head.POSS: әңәктәү.нә dә:.n [nose.GEN inside.3SG:POSS] NOSTRIL
5 **der** Base.DIM: una:dʒι.kan [young_woman.DIM] GIRL
5 **jxt** Mod Head: ғәә.әә нәәр [female horse] MARE
3 **gen** Mod.GEN Head: tәә.мәә бәәгә [morning.GEN food] BREAKFAST
2 **der** Base.AGT: tәәгәn.әә [farm_land.AGT] FARMER
1 **der** Base.REC: аәәмәә.әәкә [stool.REC] TOILET
1 **jxt** Head Mod: вәәү дәәлән [day middle] MIDDAY

► Recruitment of anchoring strategy: sometimes

(Altaic) Turkic

**Turkish (TURE): 100W / 66B / 6C ► con:47 jxt:8 der:8 cmp:3**

**Possessives: dbl** strategy. “The possessive noun phrase places the possessor in the genitive case, and the possessed element as the head of the construction. Suffixed to that head is the possessive agreement suffix, agreeing with the possessor in person and number” (Kornfilt 1997: 185).

Hasan.ин kitab.ин [Hasan.GEN book.3SG] ‘Hasan’s book’

**Binominals:** Overwhelming preference for **con.** The element suffixed to the head element is formally identical to the third person singular possessive marker and is sometimes called a linking element (Göksel & Haznedar 2007) or a compound marker (van Schaaik 2002).

47 **con** Mod Head.3SG: demәә.рәәў.у [iron.road.3SG] RAILWAY
8 **jxt** Mod Head: тәәә.көәәрү [stone bridge] STONE BRIDGE
6 **der** Base.AGT: çөөләк.әә [pot.AGT] POTTER
3 **cmp** Mod.Head: kәәвәән.әәвәәде [brother-in-law.mother] MOTHER-IN-LAW (OF A MAN)
1 **der** Base.F: kәәләңә [king.F] QUEEN
1 **der** Base.SUF: гәә.ләәк [eye.SUF] SPECTACLES/GLASSES

► Recruitment of anchoring strategy: grammaticalized
The typology and semantics of binominal lexemes

Yakut (SAH): 109W / 41B / 7C ▶ con:22 jxt:10 gen:1 adj:1

Possessives: con strategy. “A construction containing two nouns, of which the second one bears a possessive suffix, can function as a possessive nominal construction (a)” (Stachowski & Menz 1998: 428). Note: A remnant of the dbl strategy involving the old genitive suffix -(t)In is found in possessive chains (b).

(a) učūtal jiete [teacher house:3SG] ‘the house of the teacher’
(b) kini ayat.in ẓaray.in ūta [(s)he father.GEN eye.GEN water:3SG] ‘his/her father’s tears’

Binominals: Strong preference for con; jxt also common.

22 con Mod Head.3SG: murun ẓayayah.a [nose hole.3SG] NOSTRIL
10 jxt Mod Head: timir suol [iron path] RAILWAY
4 der Base.AGT: küös.cüt [pot.AGT] POTTER
2 der Base.NMLZ: ẓos.pox [room.NMLZ] COOKHOUSE
1 adj Mod.ADJZ Head: tualet.nay kumaːγi [toilet.ADJZ paper] TOILET PAPER
1 gen Mod.PROP Head: saːχar.daːχ trostnik [sugar.PROP cane] SUGAR CANE

Basque


Possessives: gen strategy. “For the translation of [English Possessives] into Basque careful analysis is needed, as Basque will use either the genitive in -(r)en or the adnominal in -ko, depending on the nature of the meaning relation in question” (Rijk 2008: 100).

sorgina.ren etxea [witch.GEN house] ‘the witch’s house’

Binominals: Strong preference for cmp; jxt also common.

27 cmp Mod.Head: tren.bide [train.way] RAILWAY
15 jxt Mod Head: itsas bazter [sea corner] SHORE
6 adj Mod.ADJZ Head: harri.zko zubi [stone.ADJZ bridge] STONE BRIDGE
4 der Base:AGT: buzți.nari [mud.AGT] POTTER
2 der Base.LOC: erla.tegi [bee.LOC] BEEHIVE
1 der Base:SUF: gos.ari [morning.SUF] BREAKFAST

Dravidian

Malayalam (MAL): 100W / 46B / 5C ▶ cmp:41 jxt:4 gen:1


meeʃay.u̠te kaalɔ [table.GEN leg] ‘the leg of the table’
kụṭṭi.yu̠te amma [child.GEN mother] ‘the child’s mother’ (ibid p.193)
Binominals: Overwhelming preference for cmp. Binominals containing a linking element are classed as cmp rather than gen on the grounds that the gemination involved is a morphophonological process and in order not to obscure the lack of a gen strategy parallel to that of possessives.

31 cmp Mod.Head: tī.vanți.pāṭa [fire.vehicle.way] RAILWAY
9  cmp Mod.LE.Header: cevi.kāja [ear.LE.wax] EARWAX
4   jxt Mod Head: svarṇa mōtirām [gold ring] GOLD RING
1 cmp Head.Mod: kaṇan.kāl [part.foot] ANKLE
1 gen Mod.GEN.Header: āṭṭ.in.kūṭṭi [goat.GEN.child] LAMB
▶ Recruitment of anchoring strategy: sometimes

(Indo-European) Baltic

Latvian (LAV): 102W / 59B / 8C ► gen:30 cmp:16 der:11 jxt:2


koka zari [tree:GEN branches] ‘the branches of the tree’
latviešu valoda [Latvian:GEN:PL language] ‘the Latvian language’

Binominals: Strong preference for gen; cmp also common.

30 gen Mod.GEN Head: zirnekļ.a tīkls [spider.GEN net] SPIDER WEB
15 cmp Mod.Head: dzelz.celš [iron.way] RAILWAY
4   der Base.AGT: pod.nieks [pot.AGT] POTTER
4   der Base.DIM: kaz.l.ēns [goat.LE.DIM] KID
2   der Base.F: karal.iene [king.F] QUEEN
2   jxt Mod Head: atsēg.kauls [key.bone] COLLARBONE
1   der Base.INS: acenes [eye:INS] SPECTACLES/GLASSES
1   cmp Head.Mod: pus.diena [half.day] MIDDAY
▶ Recruitment of anchoring strategy: mostly

Lithuanian (LIT): 102W / 47B / 10C ► der:22 gen:15 adj:7 cmp:3

Possessives: gen strategy. The functions of the genitive include “the possessive (in a broad sense encompassing both possession and belonging)” (Mathiassen 1996: 179).
sesers pinigai [sister:GEN money] ‘(the) sister’s money’
žmogaus koja [man:GEN foot] ‘the man’s foot’.

Binominals: Preference for gen; adj also common.

12 der Base.NMLZ: puodž.ius [pot.NMLZ] POTTER
The typology and semantics of binominal lexemes

10 gen  Mod.GEN Head: aus.ų vaškas [ear.GEN wax] EARWAX
7 adj  Mod.ADJZ.Head: gelež.in.kelis [iron.ADJZ.way] RAILWAY
5 gen  Mod.LE.Head: vor.a.tinklis [spider.LE.web] SPIDER WEB
3 der  Base.DIM: spen.elis [nipple.DIM] NIPPLE OR TEAT
2 der  Base.AGT: šeim.ininkas [family.AGT] HOST
2 cmp  Mod.Head: šon.kaulis [side.bone] RIB
1 der  Base.COLL: kaim.ynas [village.COLL] NEIGHBOUR
1 cmp  Head.Mod: vidur.dienis [middle.day] MIDDAY
► Recruitment of anchoring strategy: sometimes

(Indo-European) Celtic

Irish (GLE): 100W / 40B / 6C ► gen:27 der:10 cmp:3
Possessives: gen strategy. “A dependent genitive follows the head N (a)... when both Ns are definite, the article may not appear on the first” (Doyle 2001: 63).

Teach Mháiire [house Mary:GEN] ‘Mary’s house’
Hata an fhír [hat DET man:GEN] ‘the man’s hat’

Binominals: Overwhelming preference for gen.

27 gen  Head Mod.GEN: poll sróine [hole/pool nose:GEN] NOSTRIL
5 der  Base.DIM: caille.in [veil.DIM] GIRL
3 der  Base.AGT: feirm.óir [farm.AGT] FARMER
3 cmp  Mod.Head: iarn.ród [iron.road] RAILWAY
1 der  Base.NMLZ: draí.acht [magician.NMLZ] MAGIC
1 der  Base.SUF: roth.ar [wheel.SUF] BICYCLE
► Recruitment of anchoring strategy: mostly

Possessives: jxt strategy. “Basically, the genitive construction involves the ordering of a noun which indicates the thing possessed, which can be labelled possessum, and a noun which indicates the possessor. They are ordered in such a way that the possessum precedes the possessor... It is to be noted that features of definiteness involving definite determiners, such as the definitie article, relate only to the possessor... The possessum itself is never modified by a definite marker” (Jones & Thomas 1977: 192-193).1

1 “The possessum is thus always indefinite in genitive constructions. In passing, a distinction can be introduced between a genitive construction and similar-looking nominal compounds which can take up an initially-positioned determiner. An example such as het plismon (‘a policeman’s hat’) is ambiguous. We can be referring to an ordinary hat which belongs to a policeman: here we have a genitive construction, and a definite article only occurs medially to give het y plismon (‘the policeman’s hat’). Or, we can be referring to a special sort of hat which is part of a policeman’s uniform: in this sense we
car yr athro [car DET teacher] ‘the teacher’s car’
het merch [hat girl] ‘a girl’s hat’
brawd Gwil [brother Gwil] ‘Gwil’s brother’

Binominals: Strong preference for jxt; cmp and prp also occur.

40 jxt Head Mod: gwe pryf cop [web insect spider] SPIDER WEB
6 cmp Mod.Head: rheil.ffordd [rail.road] RAILWAY
5 prp Head DET Mod: bwa ’r Drindod [bow DET trinity] RAINBOW
4 der Base.AGT: crochan.ydd [cauldron.AGT] POTTER
2 adj Head Mod.ADJZ: llwybr llaeth.og [path milk.ADJZ] MILKY WAY
2 cmp Head Mod: bôn.cyff [base.trunk] TREE TRUNK
1 der Base:F: brenin.es [king.F] QUEEN
1 der Base:NMLZ: pen.aeth [chief.NMLZ] CHIEFTAIN
1 der Base:SUF: cwmwd.og [district.SUF] NEIGHBOUR
1 prp Head PREP Mod: papur lle chwech [paper for toilet] TOILET PAPER

► Recruitment of anchoring strategy: sometimes

(Indo-European) Germanic

Dutch (NLD): 105W / 54B / 8C ► cmp:41 gen:8 der:3 adj:2

Possessives: prp strategy; gen also occurs. “The English possessive ‘s’ is known to Dutch also but is not used extensively in Dutch. Generally speaking it is only commonly used after proper nouns [e.g. Anneke.s boek]. Close relatives preceded by a possessive can employ this s too [e.g. mijn moeder.s keuken]. All other nouns can better employ a van construction, however” (Donaldson 1981: 41-42).

de hoofdstad van Frankrijk [DET capital PREP France] ‘the capital of France’

Binominals: Strong preference for cmp. Some gen that exhibit a linking element derived from case markers.

40 cmp Mod.Head: spoor.weg [track.way] RAILWAY
7 gen Mod.LE:Head: noord.er.licht [north.LE.light] ARCTIC LIGHTS
2 adj Mod.ADJZ Head: sten.en brug [stone.ADJZ bridge] STONE BRIDGE
1 der Base:DIM: geit.je [goat.DIM] KID
1 der Base:F: koning.in [king.F] QUEEN
1 der Base.M: weduw.naar [widow.M] WIDOWER
1 cmp Head Mod: mid.dag [mid.day] MIDDAY
1 gen Mod Head: woen.GEN.dag [Woden.GEN.day] WEDNESDAY

► Recruitment of anchoring strategy: never

have a nominal compound and the determiner can occur initially to give yr het plismon (‘the police-man’s hat’) – though, of course, in this case the determiner cannot occur medially” (Jones & Thomas 1977: 193).
The typology and semantics of binominal lexemes

English (ENG): 105W / 44B / 5C ► cmp:26 jxt:13 der:3 adj:2

Possessives: gen and prp strategies. “When a possessor is a noun phrase [there are] two ways of marking it – by suffix ’s on the possessor (which precedes the possessed) or by the preposition of before the possessor (which follows the possessed)” (Dixon 2005: 318).

my friend’s sister [my friend.GEN sister]
the table’s leg [DET table.GEN leg]
the sister of my friend [DET sister PREP my friend]
the leg of the table [DET leg PREP DET table]

Binominals: Strong preference for cmp; jxt also common.

25 cmp Mod.Head: rail.way RAILWAY
13 jxt  Mod Head: stone bridge STONE BRIDGE
3  der  Base.AGT: pot.AGT POTTER
2  adj  Mod.ADJZ Head: milk.ADJZ way MILKY WAY
1  cmp  Head.Mod: middle.day MIDDAY

► Recruitment of anchoring strategy: grammaticalized

German (DEU): 131W / 75B / 8C ► cmp:52 gen:11 der:10 jxt:1 adj:1

Possessives: prp and gen strategies: “the genitive indicates possession or the relationship between nouns (a)” but it “is nowadays less common in spoken German, where the use of prepositions tends to be preferred (b)... The so-called Anglo-Saxon genitive with s (c) is found in both spoken and written German” (Whittle et al. 2011: 6, 48).

(a) das alte Auto des Herrn Zeiler [DET old car DET:GEN Mr. Z] ‘Mr. Z’s old car’
(b) das alte Auto von Herrn Zeiler [DET old car PREP Mr. Z] ‘Mr. Z’s old car’
(c) Herrn Zeiler.s alte.s Auto [Mr. Z:GEN old:GEN car] ‘Mr. Zeiler’s old car’

Binominals: Strong preference for cmp. Some gen that exhibit a linking element derived from case markers.

51 cmp Mod.Head: eisen.bahn [iron.way] RAILWAY
11 gen  Mod.LE.Head: nase.n.loch [nose.LE.hole] NOSTRIL
4  der  Base.DIM: mäd.chen [maid.DIM] GIRL
3  der  Base.AGT: töpf.er [pot.AGT] POTTER
2  der  Base.NMLZ: schrein.er [cupboard.NMLZ] CARPENTER
2  cmp  Head.Mod: mit.tag [middle.day] MIDDAY
1  der  Base.F: könig.in [king.F] QUEEN
1  adj  Mod.ADJZ Head: pol.ar.lichter [pole.ADJZ.lights] ARCTIC LIGHTS

► Recruitment of anchoring strategy: rarely
Norwegian (NOR): 100W / 57B / 3C ► cmp:52 gen:5

Possessives: gen strategy. The genitive is used to express possession (a), belonging more broadly (b), in names of institutions (c) and in the expression of weights and measures (d). Preposisjons are often used instead of the genitive for expressing possession (d, e; cf. a, b) (Golden, Mac Donald & Ryen 2008: 17-18).

(a)  Eva.s bøker [Eva:GEN books] ‘Eve’s books’
(b)  hus.et.s fasade [house.DEF.GEN façade] ‘the façade of the house’
(c)  Norge.s Bank [Norway:GEN bank] ‘the Bank of Norway’
(d)  to minutter.s stillhet [two minutes.GEN silence] ‘two minutes’ silence’
(e)  bøkene til Eva [book:PL:DEF PREP(to) Eva] ‘Eve’s books’
(f)  fasade.n på hus.et [façade.DEF PREP(on) house.DEF] ‘the façade of the house’

Binominals: Strong preference for cmp. Some gen that exhibit a linking element derived from case markers.

51 cmp Mod.Head: jern.bane [iron.way] RAILWAY 
5 gen Mod.LE.Head: øye.n.bryn [eye.LE.brow] EYEBROW 
1 cmp Head.Mod: mid.dag [middle.day] MIDDAY

► Recruitment of anchoring strategy: grammaticalized

Old High German (GOH): 63W / 23B / 5C ► cmp:18 der:5


(a)  stimma ruafentes in wuastinну
(b)  truhtenis ist diu erda
(c)  theiz allaz sīnes fater was

Binominals: Strong preference for cmp.

16 cmp Mod.Head: spinna.webbi [spider.fabric] SPIDER WEB 
3 der Base.AGT: hafan.âri [pot.AGT] POTTER 
2 cmp Head.Mod: mitti(l).tag [middle.day] MIDDAY 
1 der Base.DIM: zickîn [goat:DIM] KID 
1 der Base.F: kunig.in [king.SUF] QUEEN

► Recruitment of anchoring strategy: grammaticalized

(Indo-European) Greek

Greek (ELL): 100W / 30B / 7C ► gen:12 cmp:9 der:5 adj:3 jxt:1

Possessives: gen strategy. “When a noun phrase or pronoun is in the genitive it may indicate that the person, thing, etc. which it denotes is the possessor of the person, thing, etc., which is denoted by the noun on which it depends” (Holton et al. 339-340).

to spíti tis Loukía.s [DEF house DEF:GEN Lucy.GEN] ‘Lucy’s house’
ta kladiá ton déntro.n [the branches the:GEN tree.GEN:PL] ‘the branches of the trees’
The typology and semantics of binominal lexemes

Binominals: Both gen and cmp are common. Note that the latter contain a linking element -o- which Ralli (2013) argues is both phonologically and morphologically conditioned. Whereas binominals with a linking element in Germanic languages are coded as gen (in order to mark the contrast with binominals of type cmp), these Greek binominals are coded as cmp in order to mark the contrast with binominals of type gen.

- gen Head Mod: istos araxni.s [web spider] SPIDER WEB
- cmp Mod: siodir.o.dromos [iron.road] RAILWAY
- der Base: proi.ino [morning] BREAKFAST
- gen Head DEF Mod: keri tu aft.pos [wax] EARWAX
- adj Mod: uran.io tokso [sky] RAINBOW
- der Base: katsik.a [goat] KID
- jxt Head: leksi-kliki [word-key] KEYWORD

► Recruitment of anchoring strategy: mostly

(Indo-European) Indo-Aryan

Assamese (ASM): 87W / 34B / 4C ► gen:17 cmp:12 jxt:5

Possessives: gen strategy. The function of expressing the notion of possession is performed by the genitive which precedes the possessum (Goswami & Tamuli 2003: 419;433).

- ma.ra.r [mother.2FAM.GEN] ‘of your mother’
- ram.pr ei tini.khun dami kitap [Ram.GEN this three.DEF costly book] ‘these three costly books of Ram’

Binominals: Preference for gen; cmp also common.

- 17 gen Mod: nak.or phuta [nose] NOSTRIL
- 12 cmp Mod: ram.dhenu [Lord_Rama.bow] RAINBOW
- 4 jxt Mod: dak sihi [postal letter] POSTCARD
- 1 jxt Head: ga.gors [body.tree] TREE TRUNK

► Recruitment of anchoring strategy: mostly

Hindi (HIN): 95W / 30B / 4C ► cmp:13 prp:10 jxt:6 der:1

Possessives: prp strategy. “The postposition का kā … indicates possession; it agrees in the same way as an adjective in -ā with nouns” (McGregor 1986: 9).

- us strī kā beṭā [DEM woman GEN:M:SG son] ‘that woman’s son’
- us strī ke beṭe [DEM woman GEN:M:PL sons] ‘that woman’s sons’

Binominals: cmp and prp strategies both common.

- 13 cmp Mod: karna.mal [ear] EARWAX
- 10 prp Mod GEN: makṛi kā jālā [spider GEN] SPIDER WEB
- 6 jxt Mod: pavan cakkā [wind mill] WINDMILL
Appendix D. Strategies and constructions

1 der Base.AGT: lohā.r [iron.AGT] BLACKSMITH

► Recruitment of anchoring strategy: sometimes


Possessives: gen strategy. “Genitive modifiers are marked by the genitive case suffix -ko of nouns… The genitive case markers -ko, -ro and -no show their allomorphs… -kā, -rā and -nā when the genitive modifiers modify the nouns in plural number, or nouns in oblique cases… Since the genitive modifiers function as adjectives, they show inflections not just for number but also for gender. For instance, the genitive case markers -ko, -ro and -no show their allomorphs… -kī, -rī and -nī when they stand in construction with the nouns of feminine gender” (Jayaraj 1990: 112-115).

subhadrā.ko kokha [Subhadra.GEN womb] ‘Subhadra’s womb’
devirama.ṇa.ā̃ khā [Deviraman.GEN eyes] ‘Deviraman’s eyes’
ghara.ṇi purāṇī cākarnī [home.GEN old maid] ‘the old maid at home’

Binominals: gen and cmp strategies both common.

6 gen Mod.GEN Head: mākurā.ṇa jālo [spider.GEN web] SPIDER WEB
5 cmp Mod.Head: kāne.ṇuji [ear.??] EARPAX

► Recruitment of anchoring strategy: mostly

Selice Romani (RMC): 88W / 8B / 4C ► der:4 gen:4

Possessives: gen strategy. “The ending marking the ‘possessor’ is formed by adding -kr/-gr-(Sg./Pl.) to the accusative (oblique) ending. The possessive form is marked for gender, [number] and case in agreement with the corresponding noun (the ‘possessum’) like an adjective” (Holzinger 1995: 16 on Sinte Romani RMO).

ko dramaskr.es.kr.i tušni [DEM doctor.OBL.GEN.F bottle] ‘that doctor’s bottle’

Binominals: gen only.

4 gen Mod.GEN Head: kan.en.ner.e khula [ear.OBL.GEN.PL shit.PL] EARPAX
2 der Base.DIM: čha.j.óri [Gypsy_child.F.DIM] GIRL
1 der Base.ABST: ćohan.ipe [witch/sorcerer.ABST] MAGIC
1 der Base.F: kirā.čkiňa [king.F] QUEEN

► Recruitment of anchoring strategy: sometimes

(Indo-European) Iranian

Western Farsi (PES): 103W / 47B / 9C ► con:22 jxt:6 adj:6 cmp:5 dbl:5 der:3

Possessives: con strategy. “The suffix known as the ezāfe… is used to connect two nouns in certain relationships… The first such relationship can be called the possessive structure” (Mace 2003: 44).

ketābhā.ye šāgerd [books.EZ pupil] ‘the pupil’s books’
xānē.ye modir [house.EZ director] ‘the director’s house’
The typology and semantics of binominal lexemes

Binominals: Strong preference for con. Several other strategies (and orders) also occur.

22 con Head.EZ Mod: čerk.e guš [dirt.EZ ear] EARWAX
5 jxt Head Mod: rāh āhan [way iron] RAILWAY
5 adj Head Mod: asiyāb bād.i [mill wind.ADJZ] WINDMILL
5 dbl Head.EZ Mod: surāx.e bin.i [hole.EZ nose.ADJZ] NOSTRIL
4 cmp Mod.Head: boz.qāle [goat.???] KID
3 der Base.AGT: kuze.gar [earthen_bottle.AGT] POTTER
1 cmp Head Mod: ney.šekar [cane.sugar] SUGAR CANE
1 jxt Mod Head: korre asb [young horse] FOAL OR COLT
1 adj Mod.ADJZ Head: rang.in kamān [colour.ADJZ arc] RAINBOW

► Recruitment of anchoring strategy: mostly

(Indo-European) Romance

French (FRA): 102W / 49B / 10C ► prp:34 der:9 adj:4 jxt:1 cmp:1


le manteau de Patrick [DET coat PREP Patrick] ‘Patrick’s coat’
les cahiers des élèves [DET books PREP:DEF pupils] ‘the pupil’s books’
le cadastre de la Marie [DET land_register PREP DEF town_hall] ‘the land register in/of the Town Hall]

Binominals: Strong preference for prp.

17 prp Head PREP Mod: chemin de fer [road of iron] RAILWAY
14 prp Head PREP2 Mod: moulin à vent [mill to wind] WINDMILL
4 der Base.DIM: mamel.on [breast.DIM] NIPPLE OR TEAT
4 der Base.NMLZ: pot.ier [pot.NMLZ] POTTER
4 adj Head Mod: voie lactée [way milk:ADJZ] MILKY WAY
2 prp Head PREP DEF Mod: lobe de l’oreille [lobe of DEF:ear] EARLOBE
1 jxt Base.F: sorcièr.e [sorcerer.F] SORCERER OR WITCH
1 jxt Head Mod: mot-clé [word-key] KEYWORD
1 prp Head PREP3 Mod: arc-en-ciel [bow-in-sky] RAINBOW
1 cmp Mod.Head: omo.plate [shoulder.plate] SHOULDERBLADE

► Recruitment of anchoring strategy: mostly

Italian (ITA): 112W / 38B / 9C ► prp:20 der:9 adj:5 cmp:3 jxt:1

Possessives: prp strategy. “In Italian, you indicate the person to whom things belong by using di with the person involved; there is no equivalent of the English possessive form ‘Franco’s car’, ‘Anna’s motorbike’ (Proudfoot & Cardo 2013: 121).

il maglione di Alessandro [DET sweater PREP Alessandro] ‘Alessandro’s sweater’
la macchina di mio cugino [DET car PREP POSS:1SG cousin] ‘my cousin’s car’
le ciabatte dei bambini [DET flip-flops PREP:DEF children] ‘the children’s flip-flops’

**Binominals:** Strong preference for **prp**, usually in the indefinite form.

# Numbers

13 **prp** Head PREP Mod: cucina da campo [kitchen from camp] COOKHOUSE
7 **prp** Head PREP:DET Mod: dito del piede [finger of DET foot] TOE
5 **der** Base.NMLZ: ceram.ista [ceramics.NMLZ] POTTER
5 **adj** Head Mod.ADJZ: via latt.ea [way milk.ADJZ] MILKY WAY
3 **der** Base.DIM: capr.etto [goat.NMLZ] KID
2 **cmp** Head.Mod: capo.tribù [head.clan] CHIEFTAIN
1 **jxt** Head Mod: parola chiave [word key] KEYWORD
1 **cmp** Mod.Head: ferro.via [iron.way] RAILWAY

► Recruitment of anchoring strategy: **mostly**


**Possessives:** **dbl** strategy. “The following means of encoding [nominal phrase internal] possession are found in Romanian: (a) nouns and pronouns in the genitive case; (b) possessive adjectives (including the possessive affix); (c) the adnominal possessive clitic (traditionally known as the adnominal possessive dative); (d) the prepositions de and a; (e) the definite article” (Nicolae 2013: 335-349). Referencing Koptjevskaja-Tamm (2002), Nicolae states explicitly that type (d) (“de-structures”) may be considered “non-anchoring genitives” (p.347). Thus the only anchoring strategy with for nominal possessors is (a).

(a) cărțile profesorului [books.DEF professor.DEF] ‘the professor’s books’ (p.265)
(c) cartea.i [book:DEF=3SG] ‘his/her book’ (pronominal possessors only)
(d) mână de copil [hand PREP child] ‘the child’s hand’
(e) tata [father:DEF] ‘my father’ (kinship and social relation nouns only)

**Binominals:** Strong preference for **prp**: **adj** also common.

17 **prp** Head PREP Mod: pânză de păianjen [cloth of spider] SPIDER WEB
7 **der** Base.AGT: o(a)lă.ar [pot.AGT] POTTER
7 **adj** Head Mod.ADJZ: calea fer.ata [way iron.ADJZ] RAILWAY
2 **der** Base.F: nepo(a)t.ă [grandson/nephew.F] NIECE
2 **dbl** Head.DET Mod.DET.GEN: încheietura mân.ii [joint:DEF hand.DEF:GEN] WRIST
1 **der** Base.ABST: capitan.ie [captain.NMLZ] CHIEFTAIN
1 **der** Base.AGT.ABST: bucătă.ar.ie [piece_of_food.AGT.NMLZ] COOKHOUSE
1 **der** Base.DIM: mână.us.ă [hand.SUF.F] GLOVE
1 **der** Base.NMLZ: brăt.ără [arm.NMLZ] BRACELET
1 **der** Base.SUF: gălben.us [yellow.SUF] YOLK
1 **jxt** Head Mod: cuvînt cheie [word key] KEYWORD

► Recruitment of anchoring strategy: **never**
The typology and semantics of binominal lexemes

(Indo-European) Slavic

Croatian (HRV): 100W / 58B / 8C ► der:30 adj:22 prp:4 cmp:1 gen:1

Possessives: gen and adj strategies. “A possessor can be expressed as a genitive on the possessed phrase (a)… If a possessor is definite, singular, human (or animal) and expressed by one word, it forms an adjective instead of going into the genitive (b)” (Brown & Alt 2004: 67-77).

(a) *knjige Mark.a Marković.a* [books Marko.GEN Marković.GEN] ‘Marko Marković’s books’ (multiword human possessor)

   *rep mačk.e* [tail cat.GEN] ‘the tail of a cat’ (indefinite possessor)

(b) *Mark.ove knjige* [Marko.ADJZ books] ‘Marko’s books’ (single word human possessor)

   *mačk.in rep* [cat.ADJZ tail] ‘the cat’s tail’ (definite possessor)

Binominals: der and adj both common.

22 der Base.NMLZ: željez.n.ica [iron.ADJZ.NMLZ] RAILWAY

22 adj Mod.ADJZ Head: pauk.ova mreža [spider.ADJZ web] SPIDER WEB

5 der Base.AGT: lonč.ar [pot.AGT] POTTER

4 prp Head PREP Mod: *pumpa za bicikl* [pump for bicycle] BICYCLE PUMP

2 der Base.DIM: janje.tina [lamb.DIM] LAMB

1 gen Head Mod.GEN: *dlan ruk.e* [palm hand.GEN] PALM OF HAND

1 cmp Head Mod: *po.dne* [half.day] MIDDAY

► Recruitment of anchoring strategy: rarely

Czech (CES): 98W / 44B / 7C ► adj:22 der:19 prp:2 gen:1

Possessives: gen and adj strategies. “The genitive case can be used for many of the purposes that English of serves, including the marking of part-whole relationships, possession, etc… [Furthermore] possessive adjectives can be formed from person’s names and kinship terms using -ův for male possessors and -in for female possessors [and] from the names of animals using -í and its variants -cí and -čí’ (Janda & Townsend 2002: 68, 58).

*střecha dom.u* [roof house.GEN] ‘the roof of the house’

*žena soused.a* [wife neighbour.GEN] ‘the neighbour’s wife’

*Martin.ův* [Martin.ADJZ] ‘Martin’s’ *bratr.ův* [brother.ADJZ] ‘brother’s’

*Alen.in* [Alena.ADJZ] ‘Alena’s’ *matč.in* [mother.ADJZ] ‘Mother’s’

*krav.ř* [cow.ADJZ] ‘cow’s’ *prase.čí* [pig.ADJZ] ‘pig’s’

Binominals: adj and der both common.

21 adj Mod.ADJZ Head: nos.ní dírka [nose.ADJZ hole] NOSTRIL

13 der Base.NMLZ: želez.n.ice [iron.ADJZ.NMLZ] RAILWAY

4 der Base.DIM: kůz.ře [goat.DIM] KID

2 der Base.F: čaroděj.nice [wizard.F] SORCERER OR WITCH
Recruitment of anchoring strategy: sometimes

Lower Sorbian (DSB): 118W / 43B / 11C ► adj:26 der:11 gen:3 prp:2 cmp:1

Possessives: adj and gen strategies. “Both Upper and Lower Sorbian have a strong inclination to express possession by means not of an adnominal genitive but of an adjectival construction. The simplest form of this construction is the possessive adjective, derived from a noun by means of the suffix -owy (for masculines, including a-stems) or -iny (-yny) (for feminines) (a)... The use of the adnominal genitive (b) is also possible, but rarer. It either imparts the stylistic connotation of formality or it puts emphasis on the noun in the genitive” (Stone 1993: 671).

(a) nan.owy dom [father.ADJZ house] ‘father’s house’
(b) dom nan.a [house father.GEN] ‘father’s house’

Binominals: Overwhelming preference for adj.

Polish (POL): 104W / 54B / 11C ► der:28 adj:19 gen:3 prp:4

Possessives: gen strategy; adj also occurs. “A nearly defunct possessive adjective with a short Nsg. masc. form can be derived from male nouns and first names (a)... The Genitive is possibly the Polish case most diversified as to use. It is the most frequently occurring of the cases after the Nominative and Accusative. Its basic use is to express various meanings of ‘of’ in noun-to-noun relations, including most importantly the notion of possession broadly speaking (b)” (Swan 2002: 137,329).

(a) Michał.ów dom [Michael.ADJZ house] ‘Michael’s home’
(b) książka student.a [book student.GEN] ‘a/the student’s book’

Binominals: der and adj both common.
The typology and semantics of binominal lexemes

Russian (RUS): 103W / 43B / 10C ► adj:20 der:16 gen:4 cmp:2 prp:1

Possessives: gen and adj strategies. “Most nouns can be possessed. Possessors that are nouns are expressed in the genitive, and are placed after the possessed noun… A very old option for expressing possession for nouns that specify unique people – first names or nouns identifying familial roles – is possessive adjectives” (Timberlake 2004: 205-206).

dom брат.a dom brat.a [house brother.GEN] ‘(my) brother’s house’ (Wade 2011: 106)
член партии chlen parti.i [member party.GEN] ‘a member of the party’
крыша дома krysha dom.a [roof house.GEN] ‘the roof of the house’

Binominals: adj and der both common; gen rather rare.

Slovak (SLK): 100W / 57B / 10C ► der:33 adj:21 prp:2 gen:1

Possessives: gen and adj strategies. “To say ‘of’ you use the genitive case on its own, typically feminine -y, masculine/neuter -a. We could call it the ‘of’ form… Personal nouns can form possessives that are like ’s (e.g. ‘Anna’s’) in English. Males have possessive forms ending in -ov, females in -in. They work as adjectives… Only a single proper noun can form a possessive. Otherwise you just use the genitive case…” (Naughton 1997: 50,187).

Petr.ov dom [Peter.ADJZ house] ‘Peter’s house’
Binominals: adj and der both common; gen rare.

21 adj Mod.ADJZ Head: nos.ná dierka [nose.ADJZ hole] NOSTRIL
18 der Base.SUF: po.brež.ie [LOC.shore/waterside.SUF] SHORE
5 der Base.AGT: hrnč.iar [pot.AGT] POTTER
4 der Base.DIM: kozľa [goat:DIM] KID
2 der Base.LOC: ohn.isko [fire.LOC] FIREPLACE
2 der Base.NMLZ: ná.uš.nica [on.ear.NMLZ] EARRING
2 prp Head PREP Mod: pumpa na bicykel [pump for bicycle] BICYCLE PUMP
1 der Base.ADJZ.NLMZ: želez.n.ica [iron.ADJZ.NMLZ] RAILWAY
1 der Base.F: kráľ.ovná [king.F] QUEEN
1 gen Head Mod.GEN: kmeňstrom.u [trunk.tree.GEN] TREE TRUNK

► Recruitment of anchoring strategy: rarely

(Japonic) Japanese

Japanese (JPN): 130W / 82B / 4C ► cmp:68 gen:12 der:2

Possessives: gen strategy. “The particle no links two nouns into a noun phrase, in which the preceding noun modifies the following noun. The meaning varies depending on the relationship between the two nouns, such as possessor, content, location, or source” (McGloin et al. 2014: 38).

先生の本 sensei no hon [teacher POSTP book] ‘teacher’s book’

Binominals: Overwhelming preference for cmp: gen also occurs.

67 cmp Mod.Head: tetsu.dō [iron.road] RAILWAY
12 gen Mod.GEN Head: kumo.no.su [spider.GEN.web] SPIDER WEB
2 der Mod Head: gyo.fu [fish.man] FISHERMAN
1 cmp Head.Mod: mi.ki [body.tree] TREE TRUNK

► Recruitment of anchoring strategy: sometimes

Koreanic

Korean (KOR): 144W / 63B / 2C ► cmp:59 gen:4

Possessives: gen strategy; omission of the genitive particle results in a jxt strategy. “The genitive particle -ni is said to express possession... Possession, however, is only one of the roles that the genitive particle encodes. The range of roles or meanings that the genitive particle covers is, in fact, so wide that it is not incorrect to say that the function of the genitive particle is to mark the modifier-modified relationship between two noun phrases, the exact nature of which is to be determined in the light of the context of use or the general
knowledge of the world. Moreover, it is not uncommon to leave out the genitive particle and rely on the simple juxtaposition of two noun phrases” (Song 2005: 116).

*yenghi.uy chayk* [Yonghee.GEN book] ‘Yonghee’s book’

*Kim.paksa.uy atul* [Kim.PhD.GEN son] ‘Dr. Kim’s son’ (Sohn 1994: 174)

*hankwukin.uy sako.pangsik* [Korean.GEN thinking.way] ‘Korean way of thinking’ (ibid.)

**Binominals:** Overwhelming preference for **cmp**. Archaic genitive (gen) in -s- occasionally preserved.

- **cmp** Mod.Head: *chel.kil* [iron.road] RAILWAY
- **gen** Mod.GEN Head: *kho.s.kwumeng* [nose.GEN.hole] NOSTRIL

► Recruitment of anchoring strategy: grammaticalized

(Nakh-Daghestanian) Avar-Andic-Tsezic

**Bezhta (KAP):** 93W / 35B / 8C ► **gen:**26 cmp:**3 adj:**2 der:**2 jxt:**1

**Possessives:** **gen** strategy. “The 1st genitive in -s and the 2nd genitive in -la differ clearly: the 1st genitive puts a noun in the nominative, while the 2nd genitive puts it in any oblique case… No other functional difference between the 1st and 2nd genitives can be found and they will therefore be discussed as a single case. One type of the genitive’s determinant meaning is to be a general determinant… But the genitive may also have [causal, relative, possessive or correlative] meaning” (Kibrik & Testelets 2004: 232-233).

- *abo.s žämi* [father.GEN plate] ‘father’s plate’
- *bilo.s čärdäx* [house.GEN roof] ‘the house roof / roof of the house’
- *aga.s mot’o* [woman.GEN face] ‘a woman’s face’

**Binominals:** Strong preference for **gen**.

- **gen** Mod.GEN Head: *kil.o.s hino* [iron.OBL.GEN way] RAILWAY
- **cmp** Mod.Head: *šayt’an.mašina* [devil.machine] BICYCLE
- **adj** Mod.ATTR Head: *xidalaƛ.ko čür* [snot.ATTR scarf] HANDKERCHIEF OR RAG
- **der** Base.NMLZ: *ƛ’erec’.madi* [(onom).NMLZ] WATER PUMP
- **der** Base.PAUC: *cāⁿ.bo* [star.PAUC] MILKY WAY
- **jxt** Head Mod: *c’uddo c’emuc’* [red egg] YOLK
- **gen** Mod.ABL Head: *loba.ƛ’a.s beşiyołi* [midday.SUP.ABL meal] LUNCH
- **adj** Mod.ADJZ Head: *mucodaq t’ot’* [honey:ADJZ fly] BEE

► Recruitment of anchoring strategy: always

(Nakh-Daghestanian) Lezgic

**Archi (AQC):** 66W / 14B / 3C ► gen:**9 jxt:**3 adj:**2

**Possessives:** **gen** strategy. “The genitive (marker -n) has attributive and possessive meanings” (Kibrik 1994: 312-313).

*diya.n k’oč’o* [father.GEN cup] ‘father’s cup’
noL’.li.n harq [house.OBL.GEN roof] ‘the roof of the house’

Binominals: Preference for gen; jxt and adj also occur.

9 gen Mod.GEN Head: muč.li.n klan [nose.OBL.GEN hole] NOSTRIL
3 jxt Mod Head: řenne lo [female child] GIRL
2 adj Mod.ADJZ Head: ak:onnilu.t:ut kummul [in_the_morning.ADJZ food] BREAKFAST

► Recruitment of anchoring strategy: mostly

(Uralic) Finnic

Estonian (EST): 104W / 65B / 7C ► gen:32 cmp:25 der:6 jxt:1 adj:1
Possessives: gen strategy. “In the … examples, the words Kitzbergi, venna, and etenduse, which correspond to the English possessives (‘Kitzberg’s’, ‘brother’s’, ‘of the beginning’), are in the genitive case” (Oinas 1966). Thus, one binominal strategy (gen) is identical to the possessive construction; the other (cmp) involves reduced morphological complexity of the dependent.

Kitzbergi draama [Kitzberg:GEN drama] ‘Kitzberg’s drama’
venna talu [brother:GEN farm] ‘the brother’s farm’
etenduse algust [performance:GEN beginning] ‘the beginning of the performance’

Binominals: Both gen and cmp are common.

32 gen Mod.GEN Head: ämbliku.võrk [spider:GEN.web] SPIDER WEB
24 cmp Mod.Head: raud.tee [iron:NOM.way] RAILWAY
4 der Base.SUF: pea lik [head.DER] CHIEFTAIN
2 der Base.GEN.SUF: kuninga.mna [king:GEN.DER] QUEEN
1 cmp Head.Mod: kesk.päev [middle.day] MIDDAY
1 jxt Mod Head: wc.paber [wc.paper] TOILET PAPER
1 adj Mod.NMLZ Head: mesi.las.vaha [honey.NMLZ.wax] BEESWAX

► Recruitment of anchoring strategy: mostly

Finnish (FIN): 100W / 57B / 5C ► cmp:42 gen:11 der:3 jxt:1
Possessives: gen strategy. “The genitive singular ending is always -n, which is added to the inflectional stem… The genitive often marks the possessor, belonging to someone or something, or origin” (Karlsson 2013: 143;151).

Maj.n velje.n nimi [Maj.GEN brother.GEN name] ‘Maj’s brother’s name’
ihmise.n elämä [man.GEN life] ‘man’s life’
englanni.n kieli [England.GEN language] ‘the English language’

Binominals: Strong preferences for cmp; gen also common.

41 cmp Mod.Head: rauta.tie [iron.road] RAILWAY
11 gen Mod.GEN Head: hämähäki.n.verkko [spider.GEN.net] SPIDER WEB
The typology and semantics of binominal lexemes

2 der Base.NMLZ: isä.ntä [father.NMLZ] HOST
2 cmp Head.Mod: keski.päivä [middle.day] MIDDAY
1 der Base.F: kuninga.tar [king.F] QUEEN

► Recruitment of anchoring strategy: sometimes

(Uralic) Hungarian

Hungarian (HUN): 105W / 57B / 6C ► cmp:42 der:6 adj:6 gen:3

Possessives: dbl and con strategies. “In possessive constructions the possessor noun precedes the possessed noun. The possessor noun is inflected for the DATIVE case and the possessed noun receives a POSSESSIVE/PERSON suffix. The possessed noun/pronoun agrees with the possessor in person. If the possessed noun is plural, it receives the plural possessed version of the POSSESSIVE/PERSON suffix. The noun in the Dative is followed by the definite article a/az (a)… It is usual to omit both the dative suffix and the following definite article (never just one of them!) (b)” (Törkenczy 2005: 160).

(a) Péter.nek a könyv.e [Peter.DAT DEF book.3SG] ‘Peter’s book’
(b) Péter könyv.e [Peter book.3SG] ‘Peter’s book’

Binominals: Strong preference for cmp.

42 cmp Mod.Head: vas.út [iron.road] RAILWAY
6 adj Mod.ADJZ Head: észak.i fény [north.ADJZ light] ARCTIC LIGHTS
3 der Base.NMLZ: asztal.os [table.NMLZ] CARPENTER
3 gen Mod.PROP Head: kép.es.lap [picture.PROP.card] POSTCARD
2 der Base.PROP: fazék.as [pot.PROP] POTTER
1 der Base.F: király.né [king.F] QUEEN

► Recruitment of anchoring strategy: never

(Uralic) Mari

Western Mari (MRJ): 93W / 43B / 4C ► jxt:24 cmp:17 gen:2

Possessives: dbl strategy. “When two nouns form a possessive construction (e.g., the boy’s name, the capital of the country) the first element is the possessor word in the genitive followed by the possessed word marked with the possessive suffix of the third person singular or plural (Riese et al. 2010: 36).

Елу.н эргы же Yelu.i erfə.že [Yelu.GEN son.3SG] ‘Yelu’s son’
Серге.н юдар.жо Serge.i üdär.žo [Serge.GEN daughter.3SG] ‘Serge’s daughter’
Елу ден Серге.н йоча.шт Yelu dei Serge.i joča.št [Yelu and Serge.GEN child.3PL] ‘Yelu and Serge’s child’

Binominals: Overwhelming preference for jxt and cmp.

24 jxt Mod Head: kõrtni kornõ [iron road] RAILWAY
17 cmp Mod.Head: ner.raž [nose.?] NOSTRIL
Appendix D. Strategies and constructions

1 gen Mod.GEN Head: mǝnǝ.n sarǝ [egg.GEN yellow] YOLK
1 gen Mod.LAT Head: šand.eš pumaga [toilet.LAT paper] TOILET PAPER

► Recruitment of anchoring strategy: *never*

(Uralic) Saami

Kildin Sami (SJD): 87W / 35B / 4C ► jxt:29 gen:3 adj:3

Possessives: gen strategy. “Constituent order is for the most part head-final, including… head-finality in noun phrases with adjective, determiner, noun, and pronoun modifiers… Adnominal possessors are marked for genitive” (Rießler to appear).

muun aahkev [1SG:GEN grandchild:PL] ‘my grandchildren’

Binominals: Overwhelming preference for jxt.

26 jxt Mod Head: rūvv’t čuekas [iron road] RAILWAY
3 jxt Mod.Head.DIM: koass.a.all’k.a [goat.DIM.son.DIM] KID
3 adj Mod.ATTR.Head: jīnc.es’.pierrk [morning.ATTR.meal] BREAKFAST
3 gen Mod.GEN.Head: oaǝ.ʒ sājjm [spider.GEN net] SPIDER WEB

► Recruitment of anchoring strategy: *grammaticalized*

Yeniseian

Ket (KET): 70W / 29B / 4C ► cmp:14 gen:12 jxt:2 adj:1

Possessives: gen strategy. “The genitive is used to express all kinds of possession” (Vajda 2004: 23).

ām.d ógdèn [mother.GEN ears] ‘mother’s ears’

Binominals: cmp and gen both common.

14 cmp Mod.Head: ekj.qoˀt [thunder.path] RAINBOW
12 gen Mod.GEN.Head: oln.d qūk [nose.GEN hole] NOSTRIL
2 jxt Mod Head: Alba kàŋ [Alba hunting_trail] MILKY WAY
1 adj Mod.ADIZ Head: sol.tu taq.ol [gold.ADIZ finger.covering] GOLD RING

► Recruitment of anchoring strategy: *sometimes*

Oceania/SE Asia

(Austro-Asiatic) Aslian

Ceq Wong (CWG): 38W / 20B / 1C ► jxt:20

Possessives: jxt strategy. “Nominal modifiers typically occur in an associative construction where the nominal compound [NPHEAD NP] denotes a specific relationship between two entities. The associative construction is used to express a range of semantic relations, including possession, kin relations, part-whole, object material and so forth. Associatives
The typology and semantics of binominal lexemes

are distinguished from non-compositional compounds by the fact that each constituent retains its original meaning and may be independently modified” (Kruspe & al 2015: 447).


**Binominals:** Single strategy: jxt.

20 jxt  Head Mod: daraŋ m̄ɔ̃h [hole nose] NOSTRIL

- Recruitment of anchoring strategy: always

(Austro-Asiatic) Vietic

**Vietnamese (VIE): 86W / 51B / 2C ► jxt:51**

**Possessives:** prp strategy, but the preposition is often dropped, resulting in a jxt strategy. “Cúà is used to create the meaning of ownership or responsibility (it forms the genitive). The possessive phrase has this word order: object – cūa – owner… The use of cūa is in some cases optional and it is frequently omitted” (Healy 2003: 82). In Nguyên’s (1997: 162) terms, cūa is a preposition, one of a set of connectives which “express possession, means, direction, etc.”

mẹ cūa tôi [mother PREP 1SG] ‘my mother’
sách cūa tôi [book PREP 1SG] ‘my book(s)’
xē_dap cūa chí Hoa [bicycle PREP miss Hoa] ‘Hoa’s bicycle’

**Binominals:** jxt strategy.

43 jxt  Head Mod: duong xē lié [road train] RAILWAY
8 jxt  Mod Head: nè hoàng [female emperor] QUEEN

- Recruitment of anchoring strategy: sometimes

(Austronesian) Formosan

**Puyuma (PYU): 54W / 3B / 3C ► der:2 db:1**

**Possessives:** dbl strategy. Alienable possession is divided into “two types according to the encoding of the possessor. Basically, the possessor can either be coded as a pronoun (Nmkr) or as an NpObl [oblique-marked NP]… If the possessor is manifested as an NpObl, the possessor follows the possessum. Usually, the possessor is manifested twice; as a pronoun in the Nmkr slot, and as an NpObl following the possessum (a). Sometimes, a third-person possessor only occurs once as an NpObl (b); this is only possible when both the possessum and the possessor are indefinite” (Teng 2007: 143-144).

(a) tu=tiyal kana unan [3.PSR=belly DEF:OBL snake] ‘the snake’s belly’
    tu=walak kan kalikali [3.PSR=child SG.OBL Kalikali] ‘Kalikali’s child’
(b) Da sa’aD Da kawi [INDF.OBL branch INDF.OBL tree] ‘branches of trees’
**Binominals:** Binominals appear to be rare (most complex nominals contain action roots). The only non-derivational strategy exhibited in the data is **dbl**, but the data is too sparse to indicate whether this is the preferred strategy.

1. **dbl** 3SG.Head DEF Mod: *tu-biraʔ kana ʔanjila* [3SG-earlobe DEF:OBL ear] EARLOBE

1. **der** Mod.PERF.Head: *d<in>papal-an* [foot<PERF>-LOC] FOOTPRINT

1. **der** TMP.Base.LOC: *ka-ław-an* [TMP-lunch-LOC] MIDDAY

► Recruitment of anchoring strategy: **sometimes**

**(Austronesian) Greater Barito**

**Malagasy (PLT): 89W / 58B / 7C ► cmp:22 con:18 jxt:9 der:8 prp:1**

**Possessives:** **con** strategy. “If the genitive argument consists of a noun or noun phrase, a number of different scenarios have to be distinguished. In the case of definite noun phrases and personal names functioning as genitive arguments” =n’ny or =n’i are used, respectively (a, b). “If the genitive argument is an indefinite (common) noun phrase, the genitive marker is =n (c)” (Rasoloson & Rubino 2005: 468).

(a) *ny bôki.n’ny mpampaànatra* [DET book.GEN:DEF teacher] ‘the teacher’s book’

(b) *ny kiràro.n’i Fâly* [DET shoes.GEN Fâly] ‘Fâly’s shoes’

(c) *tràno.n’andriana* [house.GEN’nobleman] ‘a nobleman’s house’

**Binominals:** Preference for **cmp**; **con** also common.

22 **cmp** Head.Mod: *vàva.òrona* [mouth.nose] NOSTRIL

18 **con** Head.PER.Mod: *lala.m.by* [road.PER.iron] RAILWAY

9 **jxt** Head Mod: *mpanèfy tanimànga* [moulder/maker clay] POTTER

6 **der** AGT.Base: *mpan.dràfitra* [AGT.carpentry] CARPENTER

1 **prp** Head SOC.Mod: *vehivàvy manam.bàdy* [woman with.spouse] MARRIED WOMAN

1 **der** NMLZ.Base: *fi.lòha* [NMLZ.head] CHIEFTAIN

1 **der** NMLZ.Base.CIRC: *fa.mosavi.ana* [NMLZ.witchcraft.CIRC] MAGIC

► Recruitment of anchoring strategy: **sometimes**

**(Austronesian) Greater Central Philippine**


**Possessives:** **prp** strategy. “The possessive-ng-phrase construction is the most common Tagalog translation equivalent of English possessive modification constructions involving the possessive suffix -’s, or involving ‘of’” (Schachter & Otness 1972: 136).

*lapis ng bata* [pencil LNK child] ‘a/the child’s pencil’

*ang buntot ng aso* [DET tail POSS dog] ‘the dog’s tail’, ‘the tail of the dog’
Binominals: Strong preference for **prp**.¹

19 **prp** Head LK Mod: *daa.ng-bakal* [road.LK-iron] RAILWAY
5 **jxt** Head Mod: *bahay-gagamba* [house-spider] SPIDER WEB
3 **der** Base.LOC: *hapun.an* [afternoon.LOC] DINNER
2 **der** AGT.RED.Base: *mag.sa.saka* [AGT.RED.farming] FARMER
2 **prp** Head LOC Mod: *sipilyo sa ngipin* [brush LOC tooth] TOOTHBRUSH
2 **prp** Mod LK Head: *ging.to ng singsing* [gold.LK ring] GOLD RING
1 **der** CIRC.Base.CIRC: *ka.sangkap.an* [CIRC.belongings.CIRC] TOOL
1 **cmp** Head Mod: *manggaga.way* [worker.sorcery] SORCERER OR WITCH
1 **jxt** Mod Head: *kamay preno* [hand brake] HAND BRAKE

► Recruitment of anchoring strategy: *mostly*

(Austronesian) Malayo-Sumbawan

**Indonesian (IND): 108W / 50B / 4C ► jxt:48 der:2**

Possessives: **jxt** strategy. “A possessor follows the head word. It can be a noun or a pronoun” (Sneddon 1996: 144).

rumah Tomo [house Tomo] ‘Tomo’s house’
nama negeri itu [name country DET] ‘the name of that country’
mobil saya [car 1SG:POSS] ‘my car’

Binominals: Overwhelming preference for **jxt**.

47 **jxt** Head Mod: *jalan keréta api* [road carriage fire] RAILWAY
1 **der** AGT.Base: *peng.sihir* [AGT.sorcery] SORCERER OR WITCH
1 **jxt** AGT.Head Mod: *pem.buluh darah* [AGT.bamboo blood] VEIN OR ARTERY
1 **der** LOC.Base.CIRC: *per.api.an* [LOC.fire.CIRC] FIREPLACE

► Recruitment of anchoring strategy: *always*

¹ According to Schachter & Otanes (1972: 107ff.) there are two linkers, *ng*/Ø (used for “nominal compounds”) and *na/-ng* (used for “modification constructions”). These cannot always be distinguished on the basis of form. The choice between *na* and *-ng* is determined by the final consonant of the independent form of the element that precedes the linker. Furthermore, *ng* can be written as a separate word or as a suffix. These facts, along with the rather flexible word order, have two consequences. First, forms using *ng* could be classified as either **gen** or **prp**; for consistency with *na* (which is always written separately) the latter is used here. Second, nominal compounds without an overt marker are classified as **jxt** (or **cmp**), while those with the marker are classified as **prp**.
(Austronesian) Oceanic

Äiwoo (NFL): 48W / 14B / 4C ► jxt:5 prp:5 cls:3 cmp:1

Possessives: con and prn strategies. Äiwoo distinguishes between “direct possessives, where possessor indexing attaches directly to the possessed noun (a), and indirect possessives, where possessor indexing attaches instead to an independent morpheme, typically described as a possessive classifier (b)” (Næss forthc.).

(a) tumwä John [father.3MIN John] ‘John’s father’
   nyimä singedâ [hand.3MIN woman] ‘the woman’s hand’
(b) nabe na nubââ [bait POSS(food):3MIN shark] ‘shark bait’

Binominals: jxt and prp are both common.

5 jxt Head Mod: sikonya nugokä [waste ear] EARWAX
5 prp Head PREP Mod: talâu wä nuu pevaio [meal of place morning] BREAKFAST
3 cls Head BN:Mod: kio mi.sigiläi [hen BN:GNL.male] COCK/ROOSTER
1 cmp Head.Mod: nupo.lea [net.?spider] SPIDER WEB

► Recruitment of anchoring strategy: never

Hawaiian (HAW): 91W / 50B / 3C ► jxt:39 cmp:11

Possessives: prp strategy. “The possessive prepositions are a (or ā) and o (or ō), and kā and kō… The use of a and o is one of the most discussed, and most intriguing, of Polynesian problems.” Various dichotomies (including alienable/inalienable) have been proposed, but no terminology “is completely logical; there are always some usages that don’t seem to fit any scheme… Many possessed objects may take either a or o, usually but not always with different meanings” (Elbert & Pukui 1979: 136-139).

nā iwi a Pua [DEF:PL bone PREP Pua] ‘Pua’s bones’ (as the chicken bones she is eating)

Binominals: Overwhelming preference for jxt; cmp also common.

38 jxt Head Mod: `upena nananana [net spider] SPIDER WEB
11 cmp Head.Mod: ala.hao [path.iron] RAILWAY
1 jxt Mod Head: poho lima [hollow hand] PALM OF HAND

► Recruitment of anchoring strategy: never

Takia (TBC): 71W / 24B / 7C ► dbl:11 con:7 jxt:3 prp:2 gen:1

Possessives: con and prp strategies. “The possessor noun phrase is the first constituent of the possession construction. As in other Oceanic languages, a possessor suffix cross-references the possessor in both direct and indirect possession constructions… Most kin terms, and most nouns referring to body parts or to parts of wholes, are directly possessed (a)… Indirect possession: The possessive classifiers in Rigen Takia are sa- and a-. The latter co-varies in other dialects with ane- (occasionally kane-) (b)” (Ross 2002: 228-229).
The typology and semantics of binominal lexemes

(a)  

Madi nao.n [Madi face.3SG] ‘Madi’s face’

   ab ilo.n [inside house.3SG] ‘the inside of the house’

(b)  

Madi sa.n ab [Madi POSS.3SG house] ‘Madi’s house’

Binominals: Preference for dbl; con also common.

10 dbl Mod.3SG Head.3SG: ydu.n awa.n [nose.3SG mouth.3SG] NOSTRIL
7   con Mod Head.3SG: su mala.n [breast eye.3SG] NIPPLE OR TEAT
2   jxt Mod Head: ab mroun [house owner] HOST
2   prp Mod POSS.3SG Head: graian sa.n anay [evening POSS.3SG food] DINNER
1   jxt Head Mod: tamol sos [man Derris_root] WIDOWER
1   gen Mod.3SG Head: grane.n tatu [side.3SG bone] RIB
1   dbl Mod.3SG Head.INAL.3SG: pao.n tatuw.a.n [shoulder.3SG bone.INAL.3SG] COLLARBONE

► Recruitment of anchoring strategy: sometimes

Hmong-Mien

Hmong Daw (mww): 86W / 46B / 2C ► jxt:45 cmp:1

Possessives: jxt strategy. “Seules, les êtres (esprits, humains ou animaux) sont considérés comme ayant pouvoir de posséder; les choses ne le peuvent pas… Le possesseur précède le possédé, chacun étant accompagné de son classificateur, selon le modèle: Cl + Possesseur + CI + Possédé” (Mottin 1978: 46).

tus me_ nyuam lub  kaus_mom [CLF child CLF hat] ‘the child’s hat’

Binominals: Overwhelming preference for jxt.

45 jxt Head Mod: kev tsheb nqaj hlau [way car rail iron] RAILWAY
1   cmp Head.Mod: poj.niam [female.mother] MARRIED WOMAN

► Recruitment of anchoring strategy: always

(Sino-Tibetan) Bodic

Manange (nmm): 61W / 22B / 5C ► jxt:16 cmp:3 con:2 gen:1

Possessives: gen strategy (Hildebrandt 2004: 66,70).

ju=lʌ  mi [village=GEN people] ‘people of the village’
kola=lʌ kɔlɔ [child=GEN clothing] ‘the child’s dress’

Binominals: Overwhelming preference for jxt.

16 jxt Mod Head: 4pʰoli 3tsay [spider nest/bed] SPIDER WEB
3   cmp Mod.Head: Ina.kʰuw [nose.hole] NOSTRIL
1   con Head.LOC.Mod: toyko ri kʰeta [basement.LOC cattle] STABLE OR STALL
1   con Mod Head.DET: 2mre titi.ko [door side.DET] DOORPOST
1   gen Mod.GEN.Head: lano.lʌ.tsami [elder_sister.GEN.daughter] NIECE

430 Submitted version 2020-05-26
Recruitment of anchoring strategy: *grammaticalized*

**(Sino-Tibetan) Macro-Bai**

Caijia *(CAI)*: 66W / 31B / 1C ► cmp:31

Possessives: prn strategy.¹ Possessive constructions require a possessive classifier (Shan-shan Lü, p.c.).

\[ mɔ²¹ u²¹tsʰo²⁴ ni²⁵ hɔ⁵⁵ wu⁵⁵ \] [that person CLF POSS clothes] ‘that person’s clothes’

Binominals: cmp is the only strategy.

31 cmp Mod.Head: ɛi⁵⁵.hɔ³ [iron.road] RAILWAY

Recruitment of anchoring strategy: *grammaticalized*

**(Sino-Tibetan) Sinitic**

Mandarin Chinese *(CMN)*: 137W / 105B / 4C ► cmp:84 jxt:18 der:3

Possessives: prp strategy. “In Modern Standard Chinese… possessive meanings are formed with the enclitic *de*, with the constituent order ‘Possessor *de* Possessee’” (Luo 2013: 187).

\[ mǔqin de háizi \] [mother POSS child] ‘the mother’s child’

Binominals: Overwhelming preference for cmp.

84 cmp Mod.Head: bi².kong³ [nose.hole] NOSTRIL

15 jxt Mod Head: tie³.lu⁴ [iron.road] RAILWAY

3 der Base.SUF: ju¹.zi¹ [foal.SUF] FOAL OR COLT

3 cmp Mod.Head.SUF: shou³.wan⁴.zi¹ [hand.wrist.SUF] WRIST

Recruitment of anchoring strategy: *never*

**(Tai-Kadai) Kam-Tai**

Thai *(THA)*: 103W / 56B / 3C ► cmp:55 jxt:1

Possessives: prp and jxt strategies. “In general, a full possessive construction is represented by a sequence of head noun – possessive marker (PssM) – possessor (Pssr). In this pattern, the Possessor is in direct construction with the possessive marker \( khɔɔn \) ‘of’… Three possessive constructions are distinguished that involve the occurrence of the PssM; optional, obligatory and prohibited… The optional occurrence of the PssM is restricted to a common head noun and an animate possessor, the obligatory occurrence is restricted to a proper head noun, and the prohibition is restricted to a common head noun and a common, inanimate possessor” (Singnoi 2000: 147-150).

\[ sīa (khɔɔn) dèk \] [shirt POSS child] ‘a child’s shirt’

¹ Recall that prn denotes the analytic head-marking strategy that was not encountered amongst the binominal data (cf. §5.4.1).
The typology and semantics of binominal lexemes

phanraya (khɔ̌ɔn) Somchai [wife POSS Somchai] ‘Somchai’s wife’
làŋkhəa bān [roof house] ‘house’s roof’
? làŋkhəa khɔ̌ɔn bān [roof POSS house] ‘house’s roof’ (“unpreferable”)

Binominals: Overwhelming preference for cmp.
50 cmp Head.Mod: thaŋ.roифay [way.train] RAILWAY
5 cmp Mod.Head: kasëettra.kɔɔn [agriculture.doer] FARMER
1 jxt Head Mod: kradūk ĕiplără [bone fermented_fish_jar] COLLARBONE
► Recruitment of anchoring strategy: sometimes

NG/Australia

Gunwinyguan

Anindilyakwa (A0I): 47W / 16B / 6C ► con:10 der:5 jxt:1
Possessives: con strategy. “The suffix -lhangwa has a wide range of functions, from grammatical (marking an argument of the verb, glossed DATive), adnominal (indicating relations between NPs, glossed POSSessive), to semantic (indicating the semantic role of an adjunct nominal, glossed ABLative)...” (van Egmond 2012: 284-285).
martuviya.lhangwa wurri.yukwayuwa [emu.POSS 3a.children] ‘Emu’s children’
emeba amakalyuwakbv.lhangwa [song Bickerton.Island.POSS] ‘Bickerton Island’s song’

Binominals: Strong preference for con.
9 con CLF.INAL.Head Mod: e.me.dhvrра e.mindha [CLF(neut).INAL.hole
CLF(neut).nose] NOSTRIL.
2 der CLF.AL.Mod (Head): e.nv.ng.a.rrvrra [CLF(neut).M.AL.CLF(neut).wind]
BICYCLE PUMP
2 der CLF.INAL.Head (Mod): a.mv.ngina [CLF(neut).INAL.joint] WRIST
1 der CLF.INAL.Head (Mod): n.env.m.alhvka [3SG:M.M.INAL.foot] FOOTPRINT
1 jxt Head Mod: a.mukwa a.ngura [CLF(neut).source CLF(neut).fire] FIREPLACE
1 con Mod CLF.INAL.Head: yu.kudhukudha yi.nv.ma.dhvlvrра [CLF(masc).chest
CLF(masc).M.INAL.bone] COLLARBONE
► Recruitment of anchoring strategy: mostly

Morehead-Wasur

Warta Thuntai (GNT): 31W / 16B / 3C ► jxt:9 cmp:7
Possessives: gen strategy. “Possession is indicated through the use of the genitive case marker -an” (Quinn 2014: 20).
nju ngam.an ngauw.an menk [1SG:GEN mother.GEN father.GEN house] ‘my mother’s
father’s house’

Binominals: jxt and cmp both common.
Appendix D. Strategies and constructions

- jxt  Mod Head: karipa roroku [river bank] SHORE
- cmp  Mod.Head: them.yewe [nose.hole] NOSTRIL
- jxt  Head Mod: than thebra [half day] MIDDAY

► Recruitment of anchoring strategy: *grammaticalized*

### Nuclear Torricelli

**Srenge (LSR): 40W / 20B / 3C ► jxt:19 cmp:1**

**Possessives:** jxt and prp strategies with possessor-possessum order. There is a tendency for the [N+N] construction to signal inalienable meaning. (Lea Brown, p.c.).

- *nimbo talm* [tree branch] ‘the branch of the tree’
- *yimbo kakawio* [dog young_of_animal] ‘puppies’
- *eimu nendi.Ø sa* (name PREP.F:SG village) ‘the name of the village’

**Binominals:** Overwhelming preference for jxt.

18  jxt  Mod Head: mupə tengə [nose hole] NOSTRIL
- jxt  Mod Head: REDUP: *ala talmtalm* [leg/foot branch:RED] TOE
- cmp  Mod.Head: bili.wi [spirit?.skin] HANDKERCHIEF OR RAG

► Recruitment of anchoring strategy: *always*

**Walman (VAN): 48W / 18B / 1C ► jxt:18**

**Possessives:** jxt and prn strategy. “Walman has three constructions, one in which the noun for the possessor precedes the noun for the possessed (a), a second in which this order is reversed, where the possessed noun precedes the possessor (b), and a third in which a genitive form of the pronoun agreeing in gender and number with the possessor occurs between the possessed and possessor nouns (c)” (Lea Brown, p.c.).

- (a) *mis chei* [cat tail] ‘cat’s tail’
  - *wapon chu* [older_brother wife] ‘older brother’s wife’
- (b) *nakol nyiki* [house woman] ‘the women’s house’
  - *lasi Kasiel* [name Kasiel] ‘Kasiel’s name’
- (c) *wolu wru Kasiel* [little_sister GEN:F Kasiel] ‘Kasiel’s little sister’
  - *chalien wri Akur* [land GEN.3PL Akur] ‘land of the Akur people’

**Binominals:** jxt strategy only.

18  jxt  Mod Head: nyamayki ra [nose hole] NOSTRIL

► Recruitment of anchoring strategy: *always*
The typology and semantics of binominal lexemes

(Pama-Nyungan) Desert Nyungic

Gurindji (GUE): 53W / 8B / 5C ► jxt:4 der:3 gen:1

Possessives: No information available.

Binominals: Clear preference for jxt.

- jxt Mod Head: jitji jarriny [nose hole] NOSTRIL
- der Base.INS: warlu.waji [fire.INS] FIREPLACE
- der Base.SUF: pirnti.yi [side.SUF] RIB
- der Base.VBLZ.AGT: wuyurrurn.karra.aji [fishing_line.ACT.AGT] FISHERMAN
- gen Mod.DAT Head: yawarta.wu marru [horse.DAT house] STABLE OR STALL

► Recruitment of anchoring strategy: (no data)

(Pama-Nyungan) Paman

Wik-Mungkan (WIM): 44W / 25B / 3C ► jxt:24 gen:1

Possessives: gen strategy. The case suffix -antam denotes a source or possessor (Kilham et al. 1986: 412-415).

- puk Martha.antam [child Martha.POSS] ‘Martha’s child’
- kangk banana.antam [leaf banana.POSS] ‘leaves from the banana tree’

Binominals: Overwhelming preference for jxt.

- jxt Mod Head: kaa' uuyan [nose hole] NOSTRIL
- jxt Head Mod: puk wuut [child old_man] BOY
- gen Head Mod.DAT: yuk mee'.akana [tree/thing eye.DAT] SPECTACLES/GLASSES

► Recruitment of anchoring strategy: grammaticalized

Walio

Tuwari (TWW): 21W / 8B / 1C ► cmp:8

Possessives: gen strategy: the possessor is marked with a possessive suffix -we. There is no alienability distinction (Sylvain Loiseau, p.c.).

Binominals: cmp only.

- Andrew.we o [Andrew.POSS house] Andrew’s house

► Recruitment of anchoring strategy: grammaticalized

West Bomberai

Kalamang (KGV): 90W / 57B / 6C ► jxt:26 con:14 cmp:8 gen:6 adj:2 prp:1

Possessives: con strategy (Eline Visser, p.c.).

- Malik kewe.un [Malik house.3POSS] ‘Malik’s house’
Appendix D. Strategies and constructions

kewe anggas.un [house door.3POSS] ‘the door of the house’
tumun bal.un [child dog.3POSS] ‘the child’s dog’

Binominals: Strong preference for jxt, but with varying word order; con also common.

16 jxt Head Mod: uriap kereta api [street carriage fire] RAILWAY
14 con Mod Head.3POSS: bustang pos.un [nose hole.3POSS] NOSTRIL
11 jxt Mod Head: kalis tanggir [rain ??] RAINBOW
 8 cmp Mod Head: os.ket [sand.above?] SHORE
 6 gen Head Mod.POSS: sapu tan.kin [broom hand.POSS] HANDKERCHIEF OR RAG
 2 adj Head Mod.REL: sontum war.ten [person sorcery.ADZ] SORCERER OR WITCH
  ► Recruitment of anchoring strategy: sometimes

North America

Athabaskan-Eyak-Tlingit

Navajo (NAV): 77W / 27B / 5C ► cmp:17 con:9 prp:1

Possessives: con strategy. “If a simple possessive relationship between two nouns is expressed, the possessive prefix is bi-” (3SG/PL) (Reichard 1951: 86).

‘awéé’ bamá (< bi-má) [baby 3SG.mother] ‘the baby’s mother, it is the baby’s mother’
xästiti be. esdzáq’ [man 3SG.wife] ‘the man’s wife, she is the man’s wife’
tciiči bi.djàad [car 3SG.wheel] ‘the car wheel, it is an automobile wheel’
Kii bi.léécháq [Kii 3SG.dog] ‘Kee’s dog’ (Goossen 1995: 99)
Bilagáana bi.łaad [White_man 3SG.language] ‘English language’ (Goossen 1995: 2)

Binominals: Preference for cmp; con also common. “The consonants /s/, /sh/, and /ʟ/ sometimes appear between the elements of compound nouns, joining a noun with a following component” (Young & Morgan 1980:1-6).

16 cmp Mod.Head: tó.bąąh [water.edge] SHORE
 9 con Mod 3SG.Head: béésh bi.tiin [iron 3SG.path] RAILWAY
 1 cmp Mod.LIG.Head: nák’ee.sh.to’ [eye_area.LIG.water] TEAR
 1 prp Mod INS Head: tsé bee na’ní’á [stone with_it bridge] STONE BRIDGE
  ► Recruitment of anchoring strategy: sometimes

(Eskimo-Aleut) Yupik

Central Yupik (ESU): 70W / 21B / 15C ► der:18 dbl:3

Possessives: dbl strategy. “A relative-case NP, being the dependent in an attributive construction with another NP as the head, refers to a possessor … or a point of reference; the latter nominal, i.e. possessum, must have a third person (possessor) suffix inflection, cross-referencing in number with the former” (Miyaoka 2012: 739).

arna.m atr.a [woman.REL name.POSS:ABS] ‘the woman’s name’

**Binominals:** Overwhelmingly preference for *der*, which covers a large range “postbases” whose meanings vary in specificity from “thing of/pertaining to N” (TAQ2) to “device associated with N” (CUUN). The only non-derivational pattern is *dbl*.

3 *der* Base.TAQ2: *ataku.taq* [evening.TAQ2] DINNER

3 *dbl* Mod.REL Head.POSS:ABS: *imarpi.im ceń.ii* [sea.REL shore.POSS:ABS] SHORE

2 *der* Base.AQ3: *arnar.aq* [woman.AQ3] GIRL

2 *der* Base.LLEQ1: *keni.legg*[fire.LLEQ1] FIREPLACE

1 *der* Base.AR(AQ): *cuk.ar(aq)* [person.AR(AQ)] TOE

1 *der* Base.CENGAQ: *quia.cenga* [waist.CENGAQ] BEE

1 *der* Base.CUUN: *anuqe.ssuun* [wind.CUUN] WINDMILL

1 *der* Base.ILITAQ: *taynarner.ilita* [wrist.ILITAQ] BRACELET

1 *der* Base.IRIN: *pingay.irin* [three.IRIN] WEDNESDAY

1 *der* Base.LEK: *emu.legg* [mother’s milk/breast.LEK] NIPPLE OR TEAT

1 *der* Base.QLIQ: *elaqliq* [outside?.QLIQ] NEIGHBOUR

1 *der* Base.QUQ: *epu.quq* [shaft.QUQ] TREE TRUNK

1 *der* Base.UAQ: *pacig.ua* [gills.UAQ] NOSTRIL

1 *der* Base.YAGAQ: *kusngi.yaga* [reindeer/sheep.YAGAQ] LAMB

1 *der* Base.YAQ: *agalr.yaq* [ridgepole.YAQ] RAINBOW

► Recruitment of anchoring strategy: *sometimes*

(Mayan) Core Mayan

**Kekchí (KEK): 101W / 40B / 14C ► con:22 jxt:10 der:5 dbl:2 adj:1**

**Possessives:** *con* strategy. “Ownership (the genitive) is expressed by the prefixing of an ergative pronoun [that refers to the owner N₂] to the object owned” (Stone 1976: 69). Note: The sources use different orthographies.

*li š.bo:ls liš Flora* [DET 3ERG.purse DET Flora] ‘Flora’s purse’

*li r.aq li kwi:nq* [DET 3ERG.pig DET man] ‘the man’s pig’

*li š.tib’-el a:a* [DET 3ERG.meat pig] ‘the pig’s meat’

**Binominals:** Strong preference for *con*: *jxt* also common.

20 *con* 3ERG.Head Mod: *x.k’ot xik* [3ERG.excrement ear] EARWAX

8 *jxt* Head Mod: *sa’ u’aj* [belly nose] NOSTRIL

1 *con* 3ERG.Head M Mod: *x.kem aj am* [3ERG.web M spider] SPIDER WEB

1 *dbl* 3ERG.Head Mod.DER: *x.na’ ixaq.ill* [3ERG.mother woman.DER] MOTHER-IN-LAW (OF A MAN)

1 *con* 3ERG.Head.SUF Mod: *x.q’an.al mol* [3ERG.yellow.SUF egg] YOLK

1 *der* AGT Base.AGT: *aj k’al.om* [AGT cornfield.AGT] FARMER

1 *der* Base.ANTIP.NMLZ: *awas.in.el* [secret.ANTIP.NMLZ] SORCERER OR WITCH
Appendix D. Strategies and constructions

1 **der** Base.DER: *jolom.il* [head.DER] CHIEFTAIN
1 **der** Base.INS.NMLZ: *k'uibb'.leb'.aal* [hearthstone.INS.NMLZ] COOKHOUSE
1 **der** Base.NMLZ: *k'oteb'.aal* [shit.NMLZ] TOILET
1 **dbl** Head.NPOS 3ERG.Mod: *b'aq.el x.b'een_tel* [bone.NPOS 3ERG.upper_arm] SHOULDERBLADE

1 **jxt** Head.UNPOSS Mod: *b'aq.el jolom* [bone.UNPOSS head] SKULL
1 **jxt** Mod Head: *asuukr utz'a jl* [sugar?] SUGAR CANE
1 **adj** Mod ADJZ Head: *k'im.al kab'l* [straw.ADJZ house] THATCH

► Recruitment of anchoring strategy: *mostly*

**Zinacantán Tzotzil (TZO):** 64W / 29B / 3C ► gen:15 jxt:14

Possessives: con and **dbl** strategies. “A possessed phrase has the form: Object Possessed + Possessor. The possessor follows the possessum, whose possessive prefix cross-indexes its possessor. When the possessor is another noun – another person, for example – the prefix is *s- ‘his/her’*” (Haviland 2001:§4.1).

s.k'u` `antz [3SG.blouse/clothing woman] ‘(a) woman’s blouse’
s.na li Xun=e [3SG.house DEF John=DEF] ‘John’s house’ ¹

Binominals: Strong preference for **gen; jxt** also common.

15 **gen** Head Mod.INAL: *niʔ chu'.ul* [nose breast.INAL] NIPPLE OR TEAT
12 **jxt** Head Mod: *be sim* [road mucous] NOSTRIL
2 **jxt** Mod Head: *shokan na* [side house] NEIGHBOUR

► Recruitment of anchoring strategy: *never*

(Otomanguean) Otopamean

**Querétaro Otomi (OTQ):** 134W / 35B / 4C ► cmp:22 jxt:12 prp:1

Possessives: **prn** strategy. “La FN que codifica al Poseedor se yuxtapone después de la FN principal que refiere a lo Poseído, pero esta FN no recibe una marca gramatical de posesión. La marca de la relación de posesión entre las dos FFNN se sitúa en la FN núcleo, que expresa lo Poseído” (Palancar 2009: 117).

*nór 'bede 'nar jö'i* [DEF:SG:3POS story IND:SG person] ‘the story of a lord’

Binominals: Preference for **cmp; jxt** also common.

21 **cmp** Head Mod: *oku.xiñu* [hole.nose] NOSTRIL
12 **jxt** Head Mod: *'behñä ndä* [woman king] QUEEN
1 **prp** Head DEF Mod: *hai ya mboxita* [land DEF:PL ancestor] NATIVE COUNTRY
1 **cmp** Mod Head: *do.xymo* [stone.bowl] SKULL

► Recruitment of anchoring strategy: *grammaticalized*

¹ The meaning of *li* is not clear from the source. It is glossed (once) as ‘of’ but referred to as a definite article.
The typology and semantics of binominal lexemes

Seri

**Seri (SEI): 61W / 16B / 6C ► dbl:6 jxt:5 con:3 gen:2**

**Possessives:** con or dbl strategy. “Possessor phrases precede the noun they modify… The noun must be a possessed noun — a body part noun, a kinship term, other personal item, or deverbal noun (object-oriented or action/oblique-oriented type). The possessor phrase therefore has a variety of semantic roles with respect to the possessed noun. With a body part noun, it represents the ‘whole’ (a). With a kinship term, it represents the person who is the point of reference (b). With a personal item, it represents the owner (c)” (Marlett 1981: 256-257).

(a) xazoŋ i.lít [puma 3POSS.head] ‘(a) puma’s head’
(b) zacaam cop ata cop [young_woman DET 3P:mother DET] ‘the young woman’s mother’
(c) zizquisii° çtam ticop yaaco cop [child male DEM 3P:house DET] ‘that boy’s house’

**Binominals:** Preference for dbl strategy; con and jxt also common. In all the examples of dbl the modifier is a body part and thus inalienable; the dbl strategy can thus be regarded as a variant of the con strategy.

5 dbl POSS.Mod POSS.Head: Ø.yanopj i.t [3POSS.fist 3POSS.base] Wrist
4 jxt Head Mod: quihehe cmaam [chief_person female] QUEEN
3 con Mod POSS.Head: pnaal ii.me [honey_bee 3POSS.abode] BEEHIVE
1 dbl ABS.Mod POSS.Head: ha.mt i.ten [ABS.breast 3POSS.opening/mouth] NIPPLE OR TEAT
1 jxt Mod Head: xepe poosj [sea line] FISHING LINE
2 gen POSS.Mod Head: i.to.cams [3POSS.eye.??] EYEBROW

► Recruitment of anchoring strategy: sometimes

Siouan

**Ho-Chunk (WIN): 61W / 20B / 1C ► jxt:14 cmp:3 der:3**

**Possessives:** con and dbl strategies. “The only way to express a possessive relation between two nominals in a noun phrase in a narrow sense is juxtaposition… The whole possessive noun phrase needs to be specified by a determiner, i.e. the definite article, a demonstrative pronoun, or the indefinite article. The determiner in this position controls the reference of the whole expression. If there is a definite article following the possessor (b), then it is the possession of a specific definite possessor. If the indefinite article follows the possessor (c), it is the possession of an indefinite or unspecific possessor” (Helmbrecht 2003: 12-13)

(a) hinį́k hisja.rá [woman face.DEF] ‘the woman’s face’
(b) hinį́k.rá hisja.rá [woman.DEF face.DEF] ‘the face of the (specific) woman’
(c) hinį́k.izá hisja.ra [woman.INDEF face.DEF] ‘the face of a woman’
Binominals: **jxt** strategy.

20 **jxt** Mod.Head: *maqs naqgu* [metal road] RAILWAY

► Recruitment of anchoring strategy: *never*

(Tupian) Tupi-Guaraní

**Mbyá Guarani (Gun)**: 61W / 33B / 7C ► **jxt**:16 cmp:7 gen:4 der:3 con:3

**Possessives**: **con** strategy. The prefix *r*- (allomorph *Ø*) occur in possessive constructions identifying the possessed element, whenever the possessor (which can be a clitic pronoun or a nominal) is explicitly expressed (Martins 2003: 38-39).\(^1\)

- *kivi r.âi* [tooth POSS.jaguar] ‘jaguar’s tooth’ (“dente da onça”)
- *xe=r.âí* [1SG=POSS.dente] ‘my tooth’ (“meu dente”)
- *mitâ Ø.po* [child POSS.hand] ‘child’s hand’ (“mão da criança”)
- *xe=Ø-po* [lSG=POSS.hand] ‘my hand’ (“minha mão”)

**Binominals**: Strong preference for **jxt**; **cmp** also common. The **con** strategy employed by three binominals (e.g. COLLARBONE) uses a temporal suffix rather than a possessive prefix and is thus quite different from the possessive strategy.

13 **jxt** Mod Head: *yy rembe* [water edge] SHORE
7 **cmp** Mod.Head: *py.apê* [foot.nail] TOE
4 **gen** 3.Mod Head: *h.uvixa kunha* [3.leader woman] QUEEN
3 **jxt** Mod Head.DIM: *kavara ra’y.î* [goat son.DIM] KID
3 **con** Mod Head.PST: *axi’y kâ.gue* [shoulder bone.PST] COLLARBONE
2 **der** Base.DIM: *kunha.î* [woman.DIM] GIRL
1 **der** 3.Mod ABL.Head: *h.exa re.gua* [3.eye ABL.NMLZ] SPECTACLES/GLASSES

► Recruitment of anchoring strategy: **grammaticalized**

---

\(^1\) “O prefixo relacional {r-} divide as raízes verbais e nominais em duas grandes classes morfológicas *r*- *Ø*, que, no caso das raízes nominais, englobam os vocábulos inalienáveis no Mbyá. Essas classes, além de carregar informações sintáticas e morfológicas, parece apresentar alguma motivação fonológica, quando, na maioria dos casos, o alomorfe *r*- é unido a radicais que apresentam vocal no seu segmento inicial, e o alomorfe *Ø* une-se a vocábulos iniciados por consoante. Contudo, a motivação fonológica parece ser parcial, pois enquanto vocábulos, como akə ‘galho’, são prefixados com o alomorfe *r*-; outros como o seu homônimo -akə ‘cabeça’ é unido ao alomorfe *Ø*.

Esses prefixos ocorrem em construções possessivas identificando o elemento possuído, sempre que o possuidor, que pode ser tanto um pronom clítico quanto um nominal, vier imediatamente expresso na locução” (Martins 2003: 38-39).
The typology and semantics of binominal lexemes

(Uto-Aztecan) Southern Uto-Aztecan

Yaqui (YAQ): 92W / 20B / 6C ► jxt:15 cmp:3 der:2

Possessives: gen and jxt strategies “A noun that becomes a genitive through Relativization takes on Dependency marking… Dependency marking is also present in genitive constructions that refer to inalienable possessions (b)” (Lindenfeld 1973: 55-56,79).

(a) peo.ta ačai [Peter.DEP father] ‘Peter’s father’
   itom pare.ta kar [our priest.DEP house] ‘our priest’s house’
   Maria.ta sewa.m [Mary.DEP sewa.PL] ‘Mary’s flowers’
(b) in ačai.ta čonim [my father.DEP hair] ‘my father’s hair’

“Il existe plusieurs restrictions à l’usage du suffixe -ta comme marqueur génitif. La première renvoie à la relation entre les deux nominaux qui doit être une relation de possession. Toute autre relation (holonymie, hyponymie par exemple) utilise d’autres ressources… Les autres contraintes sont liées aux caractéristiques du possesseur: ce doit être un possesseur animé et singulier. En effet, à nouveau, si le possesseur est au pluriel, le suffixe -ta ne peut apparaître et il en va de même s’il renvoie à une entité inanimée”1 (Fernández & Gonzalez 2008: 76).

tótoi.ta kába.m [chicken.DEP egg.PL] ‘the chicken’s eggs’
tótoi.m kába.m [chicken.PL egg.PL] ‘the chickens’ eggs’
puéta yábem [door key] ‘the key of the door’

Binominals: Strong preference for jxt; cmp also common.

1  jxt  Mod Head: yeka wojo’oria [nose hole] NOSTRIL
3  cmp  Mod.Head: pipim.koba [breast.head] NIPPLE OR TEAT
2  jxt  DIM Base: ili jamut [DIM woman] GIRL
1  der  Base.LOC: sisi’iwoo.chi [iron.LOC] TOOL
1  der  Base.PSS.LIG.NMLZ: jo’a.ak.a.me [home.PSS.LE.NMLZ] HOST
1  jxt  Mod Head.APPL: mam betala.riam [hand plain.APPL] PALM OF HAND

▶ Recruitment of anchoring strategy: sometimes

1 /There are several restrictions on using the -ta suffix as genitive marker. The first concerns the relationship between the two nominals which must be a possession relation. Any other relationship (e.g. holonymy, hyponymy) uses other resources… The other constraints are related to the characteristics of the possessor: it must be animate and singular… if the possessor is plural, the suffix -ta cannot appear and likewise if it refers to an inanimate entity/ (my translation).
South America

Araucanian

Mapudungun (ARN): 81W / 25B / 6C ► jxt:17 cmp:4 der:3 gen:1

Possessives: prn strategy. “A complex noun phrase contains more than one noun. The relation between the nouns is either subordinative or coordinative. There are various types of subordinative complex noun phrases: possessive, partitive and genitive… The noun phrase which refers to the possessor precedes the noun phrase which indicates the possessed. A possessive pronoun which corefers to the possessor noun phrase stands in between (a)… Genitive constructions are preferred in order to refer to an object which belongs to, forms part of or is connected to something else (b)” (Smeets 2008: 133-136).

(a) chaw ñi wenüy [father POSS3 friend] ‘father’s friend’
   tūifa-chi kawellu ñi pilun [this-ADJZ horse POSS3 ear] ‘this horse’s ear’
   kolū tren ñi chofer [red train POSS3 chauffeur] ‘the engine-driver of the red train’

(b) namun mesa [leg table] ‘table-leg’
   tūifa-chi pilun kawellu [this-ADJZ ear horse] ‘this horse ear’

Binominals: Strong preference for jxt.

14 jxt  Head Mod: wechoz yu [hole nose] NOSTRIL
4  cmp  Mod.Head: tren.rüpi [train.way] RAILWAY
3  jxt  Mod Head: malle ñawe [father’s_brother daughter] NIECE
2  der  Base.NMLZ: ruka.fe [house.NMLZ] CARPENTER
1  der  Base.LOC: kütiral.we [fire.LOC] FIREPLACE
1  gen  Mod.ESS.NONF Head: füta.nge.n zomo [husband.ESS.NONF woman] MARRIED WOMAN

► Recruitment of anchoring strategy: grammaticalized

Arawakan

Trinitario (TRN): 88W / 25B / 10C ► cls:12 con:4 cmp:4 dbl:3 der:2

Possessives: con or prn strategy. A pronominal possessor is expressed “with a personal prefix. If one wants to specify the identity of the possessor with another noun phrase (NP), the two noun phrases are combined in the following way: NP_POSSESSUM NP_POSSESSOR… The possessed noun must be marked as possessed (a). If it can not carry a prefix, the relational noun ye’e is used (b)” (Rose 2015: 86-87; my translation).

(a) to ta.kunara’i to koje [DET 3NH.reflect DET moon] ‘the reflection of the moon’
   to s.jora psena ’seno [to 3F.wound DET woman] ‘the wound of that woman’

(b) to ma.ye’e paku ma mópern [DET 3M.RELN dog DET boy] ‘the boy’s dog’
Binominals: Strong preference for cls strategy; con, cmp and jxt also occur.

9 cls Base.CLF: mopo.ji [bee_related.CLF(shapeless)] BEESWAX
3 cls Base.DER.CLF: tapaj.ro.gi [door.DER.CLF(trunk)] DOORPOST
4 con DET.Head DET.Mod: ta.nuk(u).’i -chene [3NH.mouth.CLF(round) -breast] NIPPLE OR TEAT
2 cmp DET.Head.Mod: -sira.giño [blockage.ear] EARWAX
2 jxt Head Mod: chiwa.gira ’moyo [goat.DIM child] KID
2 cmp Mod.Head: -siri.peno [nose.house] NOSTRIL
1 der Base.AGT: tyuraj(i).eru [mud.AGT] POTTER
1 der Base.DIM: kwoy.gira [horse.DIM] FOAL OR COLT
1 cmp Head.Mod: -jiwo.cho.no’u [hair.CLF(board).front] EYEBROW

► Recruitment of anchoring strategy: sometimes

(Cariban) Guianan

Galibi Carib (CAR): 65W / 20B / 7C ► jxt:6 con:6 der:4 cmp:2 gen:1 dbl:1

Possessives: con strategy. “The most common way to create a possessive form is by suffixation with -ry… a possessive form of a noun is preceded by either a pronominal prefix or a noun unit indicating the possessor… A few dozen nouns use a possessive suffix -ty instead of -ry” (Courtz 2008: 56-57). Note: The sources describe different dialects and use different orthographies based on different phonological analyses (cf. Courtz 2008: 14-16).

X aina.ry [X hand.POSS] ‘hand of X, X’s hand’
X kurijara.ry [X boat.POSS] ‘boat of X, X’s boat’
X py.ty [X wife.POSS] ‘wife of X, X’s wife’

Binominals: con and jxt strategies are both common.

6 jxt Mod Head: moyowai etaweini [spider web] SPIDER WEB
6 con Mod Head.POSS: manati poti.li [breast tip.POSS] NIPPLE OR TEAT
3 der Base.DIM: kapala.menpo [sheep.DIM] LAMB
2 cmp Mod.Head: amek.unti [wrist.base] BRACELET
1 der Base.DEV: kamisa.npo [piece_of_tissue.DEV] HANDKERCHIEF OR RAG
1 gen Mod.POSS Head: -pana.li weti [ear.POSS dirtiness] EARWAX
1 dbl Mod.POSS Head.POSS: emo.li sakila.li [nose.POSS aperture.POSS] NOSTRIL

► Recruitment of anchoring strategy: sometimes

Chibchan

Cabécar (CJP): 81W / 33B / 2C ► jxt:26 con:7

Possessives: jxt strategy. “The possessive noun phrase is formed by juxtaposition of possessor and possessed in that order. The possessor phrase can be formed by a noun or a pronoun and can itself be complex, e.g. a possessive noun phrase… Cabécar does not structurally distinguish between different kinds of possessive relations, juxtaposition being
the only structural means to indicate a possessive relationship in the nominal domain. The possessive construction can be lexicalized resulting in a compound…” (Verhoeven 2012: 165-166).

José mĩng [José mother] ‘José’s mother’
ju kõ [house door] ‘door (of the house)’

**Binominals:** Strong preference for **jxt**.

26 **jxt**  Mod Head: kõbākā ńqalg [train road] RAILWAY

7 **con**  Mod Head.SPEC: kukā ńqē [ear excrement:SPEC] EARWAX

▶ Recruitment of anchoring strategy: *mostly*

Harakmbut

**Harakmbut (AMR): 76W / 28B / 27C ► cmp:17 cls:4 gen:3 jxt:2 der:2**

**Possessives:** **gen** strategy. “In Harakmbut the syntactic relation of attributive possession is reflected by dependent marking: (pro)nouns denoting the possessor are marked for genitive case; the possessum is unmarked... The order is that of possessor – possesum. Free and bound nouns show distinct morphosyntactic behaviour. Free possessed nouns, like ēnq in (a), use the pattern in which the possessor and possesum form two distinct phonological words; stressed syllable nuclei are underlined. Bound possessed nouns have two patterns at their disposal: (i) the default pattern shared with free nouns (b), and (ii) a pattern exclusively available to bound nouns (c)” (Van linden forthcoming).

(a) ndoʔ.ød.nāq [1SG.GEN mother] ‘my mother’
(b) ndoʔ.ød wa.nq.a.po [1SG.GEN NMLZ.CLF(fruit).CLF(round)] ‘my belly’
(c) arakmbut-ød-ndik [people-GEN-name] ‘native word’ (‘name of the people’)

**Binominals:** Overwhelming preference for **cmp**. Constructions do not distinguish between independent and bound nouns, and consequently the NPOS prefix is disregarded. (Note that this prefix also functions as a nominalizer and is consequently glossed NMLZ in Van linden 2019.) **cls** is only used for binominals in which the head is denoted by a classifier on its own. When this occurs together with a spatial linker, the resulting “in between” word is classified as **cls**; otherwise such linkers give rise to binominals of type **gen** rather than **prp** because they are not independent forms.

18 **cmp**  Mod.Head: wāʔ.œ.wē [NPOS.nose.liquid] NOSTRIL

2 **cls**  Base.CLF: šērā.po [honey.CLF(round)] BEE

2 **der**  Base.DIM: wettonē.sipo [woman.DIM] GIRL

2 **cls**  Base.SPAT.CLF: wa.kmbere.ku.to.pa [NPOS.forehead.CLF(edge?).SPAT(down).CLF(ROD)] EYEBROW

2 **gen**  Mod.SPAT.Head: wa.mbaʔ.ta.meh(.po) [NPOS.hand.SPAT(base).hump.CLF(round)] WRIST

1 **cmp**  Head.LOC.Mod: nō.po.te.menoe [vital_centre.CLF(round).LOC.day] MIDDAY

1 **gen**  Mod.GEN Head: amiko.en kutamah [foreigner.GEN bag] BACKPACK
Huitotoan

Murui Huitoto (HUU): 48W / 18B / 4C ► cls:11 jxt:4 con:3

Possessives: jxt and prn strategies. “Murui lacks the distinction between alienable and inalienable possession. The most frequent marking of possession involves a simple juxta-position of words within the NP that requires the Possessor (R) - Possessed (D) order (a)... Verbless possessive constructions involve the R (and, often, the D) followed by the connective ie that refers to the R anaphorically (b)” (Wojtylak 2017: 240).

(a) *Lusio yoe.fai* [Lucio metal.CLF] ‘Lucio’s machete’
   *Maria int* [Maria husband] ‘Maria’s husband’
   *konago oma.kai* [lizard tail.CLF] ‘the tail of the lizard’

(b) *[moo ie] jafai.ki* [father CON breath.CLF] ‘father’s spirit (belonging to the father)’

Binominals: Strong preference for cls.

| 10 cls | Base.CLF: *defo* [nose:CLF(cavity)] NOSTRIL |
|  4 jxt | Mod Head: *iye fiue* [river mouth/edge] SHORE |
|  3 con | Base ANA.CLF: *jefo i.goï* [ear:CLF(cavity) ANA.CLF(leather)] EARLOBE |
|  1 cls | Base.CLF.CLF: *taizi.ko.ño* [heel.CLF(cover).CLF(FEM)] ANKLE |

Recruitment of anchoring strategy: sometimes

Matacoan

Wichi (MZH): 82W / 36B / 10C ► jxt:17 cmp:9 der:9 con:1

Possessives: jxt strategy. “La relación poseedor-poseído también puede expresarse mediante la yuxtaposición de dos sustantivos, y el núcleo de la construcción posesiva está siempre a la derecha de la frase” (Nercesian 2014: 162).

*atsinha w’et* [woman house] ‘the woman’s house’

Binominals: Preference for jxt; cmp also common.

|  9 jxt | INAL.Mod Head: *to.ch’ute’ lhits’i* [INAL.ear wax] EARWAX |
|  8 jxt | Mod Head: *tewuk lhip* [river part] SHORE |
|  5 cmp | INAL(Mod.Head: *tot.kwe.w’u* [INAL.hand.neck] WRIST |
|  4 cmp | Mod.Head: *si.jhile* [side.bone] RIB |
|  4 der | INAL.Base.LOC: *to.nhes.pe* [INAL.nose.LOC] NOSTRIL |
|  2 der | Base.AGT: *sapatu.wu* [shoe.AGT] SHOEMAKER |
|  2 der | Base.LOC: *y’amekw.hi* [excrement.LOC] TOILET |
|  1 der | INAL.Base.AGT: *to.lhokwe.wu* [INAL.container.AGT] POTTER |

Recruitment of anchoring strategy: sometimes

1 /The possessor-possessed relationship can also be expressed through the juxtaposition of two nouns, and the head of the possessive construction is always to the right of the phrase/ (my translation).
Appendix D. Strategies and constructions

1. con Mod Head.LOC: *hep lhetek.ch’u* [hut head.LOC] THATCH
   - Recruitment of anchoring strategy: *mostly*

Nadahup

**Hupdë (JUP): 61W / 28B / 3C ▶ jxt:17 cmp:11**

Possessives: jxt and prp strategies. “Most Hup nouns are non-inherently (i.e., non-obligatorily) possessed… For nouns of this type, their possession by another entity requires an additional morphological indicator of the possessive relationship, defining the relationship as alienable… In contrast, an inalienable relationship is indicated by the simple juxtaposition of possessor and possessed” (Epps 2008: 224-237).

*pedú nh cug’èt* [Pedro POSS book] ‘Pedro’s book’ [alienable]
*tāʔāy.ʔíp* [woman.father] ‘the woman’s father’ [inalienable]

Binominals: Preference for jxt; cmp also common.

17 jxt Mod Head: *toj m̄y* [nose hole/house] NOSTRIL
10 cmp Mod.Head: *yɔʔm.ʔh* [power.male] CHIEFTAIN
1 cmp Head.Mod: *tsp.g’æt* [house.leaf] THATCH
   - Recruitment of anchoring strategy: *mostly*

(Quechuan) Quechua II

**Imbabura Quechua (QVI): 71W / 18B / 2C ▶ jxt:18**

Possessives: gen strategy. “The possessor is marked by the morpheme -paj, also used for benefactives… There is no marking for possessed nominals in IQ. The possessive suffixes found in non-northern Quechua have been lost in Ecuador” (Cole 1985: 115).

*Juzi.paj wa* [José.POSS wife] ‘José’s wife’
*chay runa.paj jatun llama* [that man.POSS big sheep] ‘that man’s big sheep’ (p.76)

Binominals: jxt strategy only.

17 jxt Mod Head: *singa utuju* [nose hole] NOSTRIL
1 jxt Head Mod: *chaupi puncha* [half day] MIDDAY
   - Recruitment of anchoring strategy: *grammaticalized*
The typology and semantics of binominal lexemes

Ticuna-Yuri

**Ticuna (TCA): 73W / 20B / 5C ► cmp:17 gen:3**

Possessives: gen, cmp and prp strategies. The usual structures are **Mod.GEN Head** (“if the possessum is a full independent noun”) (a), **Mod.Head** (“if the possessum is a bound noun”) (b) or **Head CON Mod.GEN** (“especially if the possessor is a long, complex NP”) (c) (Denis Bertet, p.c.).

(a) ygr̃.ârũ yōrã [tamarin.GEN master] ‘the master of the tamarin monkeys’
(b) kõr̃-gâ [white_person-language] ‘Spanish’
(c) yōrã yã ygr̃.ârũ [master CON tamarin.GEN] ‘the master of the tamarin monkeys’

**Binominals:** Strong preference for cmp. Some (bound) nouns are obligatorily possessed and prefixed with ná- (Anderson 1962: 23-24):

12 cmp UNPOSS.Head.Mod: ná. ’mã.ruũ [UNPOSS.hole.nose] NOSTRIL
5 cmp Mod.Head: pàwú. chiâũ [spider.web] SPIDER WEB
1 gen Head.Mod.GEN: âněrũ òrò.ârũ [ring gold.GEN] GOLD RING
1 gen Head.Mod.PURP: pè.  ì  e. rũũ [cover.hand/finger.PURP] GLOVE
1 gen UNPOSS.Head.Mod.GEN: nà. ãp̃ã̃ ’kòtũ.ârũ [UNPOSS.roof palm_species.GEN] THATCH

► Recruitment of anchoring strategy: **sometimes**

Pidgins/Creoles

(Pidgins & Creoles) English-based

**Saramaccan (SRM): 65W / 32B / 4C ► cmp:25 jxt:4 der:2 prp:1**

Possessives: jxt and prp strategies, which tendentially represent an alienability distinction (McWhorter & Good 2012: 80-82).

Djéfì m’má [Jeff Mom] ‘Jeff’s Mom’
dágũ mamã [dog mother] ‘dog’s mother’
dì kɔkɔnì fútu [DEF rabbit foot] ‘the rabbit’s foot’
dì tã fa u Rohit [DEF table POSS Rohit] ‘Rohit’s table’
dì bûku (f)u Ámba [DEF book POSS Amba] ‘Amba’s book’

**Binominals:** Overwhelming preference for cmp.

25 cmp Mod.Head: nûsû. bãaku [nose.hole] NOSTRIL
4 jxt Mod.Head: talán fútu [train foot] RAILWAY
2 der Base.AGT: koósu.ma [skirt.AGT] GIRL
1 prp Head PREP Mod: ñìnga u fútu [finger of foot] TOE

► Recruitment of anchoring strategy: **sometimes**
Possessives: jxt strategy. “Le SN ou N déterminatif correspond aux divers compléments de détermination qui, en français, se joignent au nom à l’aide d’une préposition, surtout à ou de (qui, tous les deux, ne subsistent pas en créole), et, comme en français, ils indiquent des rapports très variés entre le nom principal et le nom déterminatif, p.ex… possession”¹ (Bollée 1977: 46).

lerwa zãs Zwif [king people Jews] ‘the king of the Jews’
zistwar Sesel [history Seychelles] ‘the history of the Seychelles’
lagel sa sak [mouth his bag] ‘the opening of his bag’

Binominals: jxt strategy only.

21 jxt Head Mod: trou nennen [hole nose] NOSTRL

► Recruitment of anchoring strategy: always

¹ /The determinative NP or N corresponds to the various complements of determination which, in French, are joined to the noun by means of a preposition, especially à or de (neither of which remain in Creole), and, as in French, they indicate a very varied relationship between the principal name and the determinative name, e.g. … possession/ (my translation).
E. Binominal data set

This appendix lists every binominal lexeme in the database, along with its gloss and meaning, ordered (for ease of comparison with Appendix D) by area, genus and language. Within each language constructions are ordered alphabetically. Glosses are as given by the contributor, with minor adjustments for consistency and in order to accord with the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel 2015).

Africa

(Afro-Asiatic) Berber

Tarifit (79): 96W / 18B / 4C
Head Mod (1)
Friinu manu [brake hand] HAND BRAKE
Head Mod:STC (3)
Tahr’us’t umazżun [ring? ear:STC] EARLOBE
ta’qqa ufas [palm hand:STC] PALM OF HAND
tisi ufas [bottom hand:STC] PALM OF HAND
Head PREP Mod:STC (13)
Apq’ab n wzrf’rw [bag of back:STC] BACKPACK
’Pumpa n baszklit [pump of bike:STC] BICYCLE PUMP
Tahr’us’t umazżun [ring? ear:STC] EARLOBE
Sfuwa n tít’t [back of eye:STC] EYELID
Tdınnt n wáy [ring of gold:STC] GOLD RING
Azqon n mhá [middle of day:STC] MIDDAY
Ta’qqa ufas [palm hand:STC] PALM OF HAND
tisi ufas [bottom hand:STC] PALM OF HAND
Head REL ASS Mod (7)
Tii de n bonte [eat:INF REL:SG:F ASS morning] BREAKFAST
Tii de n buso [eat:INF REL:SG:F ASS noon] LUNCH
Assu ge n kee [bone REL:3SG:M ASS head] SKULL
Geva ge n gargar [home REL:3SG:M ASS spider] SPIDER WEB
Assu ge n dikjia [bone REL:3SG:M ASS waist] SPINE
Tii de n bodo [eat:INF REL:SG:F ASS night] SUPPER

Hausa (35): 100W / 43B / 4C
Base.ABST (1)
Mádáu [sorcerer.ABST] MAGIC
Base.F (1)
Sáráu.niyya [king.F] QUEEN
Head.LK Mod (40)
Kákki.n zámáu [wax:LK bee] BEESWAX
dóókki.n kárfe [horse:LK metal] BICYCLE
Abí.n hánnùu [thing:LK hand] BRACELET
Káshì.n káññálá [bone:LK shoulder] COLLARBONE
Abí.n dááre [meal:LK night] DINNER
Iccé.n kóññá [wood:LK door] DOORPOST
Fláátá.f kánné [skin:LK ear] EARLOBE
Yá.n kánné [children:LK ear] EARRING
dáádu.a kánné [dirty:LK ear] EARWAX
Gáāshì.n idóo [hair:LK eye] EYELASH
Fláátá.r idóo [skin:LK eye] EYELID
Hárshè.n wáddá [tongue:LK fire] FLAME
Bírì.ki.n hànnu [brake:LK hand] HAND BRAKE
Já.ka.n hànnu [bag:LK hand] HANDBAG
Yá.r ákwdýáya [daughter:LK goat] KID
Dá.ki.n hànnu [son:LK goat] KID
Dá.n númá [son:LK sheep] LAMB
Yá.r tükúndá [daughter:LK ewe] LAMB
Lámáb.b r móótúu [number:LK car] LICENSE PLATE
Abí.n n fánàa [meal:LK day] LUNCH
Ákkuá.ti.n wáyikku [box:LK mail] MAIL BOX
Yá.r wáu [daughter:LK elder_brother] NIECE
Báá.shì.n mááuma [mouth:LK breast] NIPPLE OR TEAT
Hebrew (37): 100W / 44B / 6C
Head Mod (1) tik gav [bag back] BACKPACK
Head PREP.DEF Mod (1) kelet be. tan [arc in:DEF,sky] RAINBOW
Head.STC DEF Mod (2) tenux ha.zeen [lobe:STC DEF,ear] EARLOBE sevil he.salav [path:STC DEF,milk] MILKY WAY
Head.STC MOD ADJ (1) or ot arki.isim [light:STC arctic.ADJ] ARCTIC LIGHTS

Maltese (61): 96W / 32B / 8C
Base.AGT (1) skarp.an [shoe.AGT] SHOEMAKER
DEF Head Mod (1) il-kėlna mużieh [DEF-word key] KEYWORD
DEF Head.PREP Mod (1) Ir-Tiq ta' Sant’Anna [DEF-way of Saint_Anne] MILKY WAY
Head DEF Mod (2) tebuqt il-għajn [half:DEF,eye] EYELID xatt il-buqar [shore:DEF,sea] SEA

Head PREP Mod (3) kaf ta’ triba [head of tribe] CHIEF TAIN ikla ta’ nafsinhar [meal of midday] LUNCH saba ta’ tiq [digit of foot] TOE
Head Mod (2) nofs.inhir [half:day] MIDDAY qaws.ally [bow:god] RAINBOW

(App) Adamawa-Ubangi
Kam (48): 69W / 27B / 3C

(Atlantic-Congo) Bantoid
Swahili (90): 104W / 52B / 5C
CL Base (4) mtoto [CL:child] BOY

2020-05-26 Submitted version 451

Appendix E. Binominal data set
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>Base.AGT (1)</th>
<th>Chakali (15): 51W / 25B / 5C</th>
</tr>
</thead>
<tbody>
<tr>
<td>jüg nááb.ó.i</td>
<td>[house food.POSS.ASS] COOKHOUSE</td>
</tr>
<tr>
<td>laccès wá-wó.i</td>
<td>[tongue fire.POSS.ASS] FLAME</td>
</tr>
<tr>
<td>mún dyër.ó.i</td>
<td>[child goat.POSS.ASS] KID</td>
</tr>
<tr>
<td>mún tmbón.ó.i</td>
<td>[child sheep.POSS.ASS] LAMB</td>
</tr>
<tr>
<td>katféf náám.ó.i</td>
<td>[finger mother.POSS.ASS] THUMB</td>
</tr>
<tr>
<td>dyër dyëm.ó.i</td>
<td>[road blood.POSS.ASS] VEIN OR ARTERY</td>
</tr>
</tbody>
</table>

Head Mod (1) sängäti tēëngä [sun middle] MIDDAY

(Atlantic-Congo) Gur

Baa (56): 90W / 47B / 5C

Head L.K.Mod (5) dékin à.vil [fowl L.k.male] COCK/ROOSTER
mibón à.nár [ox L.k.female] DAIRY COW
li râ,pil [ring L.k.gold?] GOLD RING
bisin râ.nâ.it [horse L.k.female] MARE
pô à.tûn [cane L.k.sweet] SUGAR CANE

Head Mod (1) ge’tëngâ [day Yanga] WEDNESDAY

Head Mod.AL (1) lâ,ba.lî [land.father.AL] NATIVE COUNTRY

Head Mod.INAL (14) kakût.nâ.vi [bone.foot.INAL] ANKLE
wâwâ.nji.vi [bag.back.INAL] BACKPACK
ní.kâ.nâ.vi [thing ##.hand.INAL] BRACELET
kakût.mîv [bone.throat.INAL] COLLAR/NECK
pâ.nî.vi [mouth.eye.INAL] EARlobe
pën.nâ.vi [mark.eye.INAL] EYEBROW
són.nâ.vi [hair.eye.INAL] EYELASH
vâ.kâ.mî [child:AL.brother.INAL] NIECE
pâ,jî.vi [mouth.milk.INAL] NIPPLE OR TEAT
nî.nâ.vi [##.arm.INAL] PALM OF HAND
ge,kâ.kô.vî [##.shoulder.INAL] SHOULDERS/SHOULDER
ribî.nî.vi [covers.eye.INAL] SPECTACLES/GLASSES
kakût.nji.vi [bone.back.INAL] SPINE
kâ.nâ.vi [##.arm.INAL] WRIST

Head Mod (26) gî’t,ttâvî [room.bee] BEEHIVE
biss.in.mê [horse.iron] BICYCLE
và.nîvâl [child/small.male] BOY
pâ.kê [mouth.road] DOORPOST
pâ,ttî [thing:PL.ear] EARING
ri.mû.nî [PL.wax.ear] EARRING
bêl.nî [skin.eye] EYELID
ê,gbîn [rope.hook] FISHING LINE
dêm.lê [tongue.fire] FLAME
và.bisin [child/small.horse] FOAL OR COLT
rînà.nâsâ [PL.leg.person] FOOTPRINT
và.nât [child/small.girl] GIRL
kpo.nâkwên [piece.clothes] HANDKERCHIEF OR RAG
ghân.kô [hôle.medicine] HOSPITAL
và.mijdâ [child:goat] KID
gît.tàkàdà [room.book] MAIL BOX
nât.mûvâl [woman.man] MARRIED WOMAN
ghân.gô [hole.nose] NOSE
krâ.ki.sâ [road.tree.outside] RAILWAY
pâ.môn [mouth.water] SHORE
kût.gî [bone.head] SKULL
ghî.mgî [room.cattle] STABLE OR STALL
gwâ.làhâdi [day.Lahadi] SUNDAY
û.ri.nî [bag.PL.thing] TOOLBOX
klî.sâ [tree.outside] TRAIN
sôh.ki [##.tree] TREE TRUNK

Wawa (105): 27W / 13B / 4C

Head Mod (3) mún nômbar [child man] BOY
mûn ngûvë [child woman] GIRL
dâñgâi ngûvë [horse woman] MARE

Head Mod.LOC.ASS (1) nàdôb.t.v.i.l [fufu ear.LOC.ASS] EARWAX

Head Mod.POSS.ASS (8)

<table>
<thead>
<tr>
<th>Base.AGT (1)</th>
<th>Base.AT (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nàdôlkoor [shoe.make]</td>
<td>SHOE/SHOEMAKER</td>
</tr>
</tbody>
</table>

452
Submitted version 2020-05-26
Appendix E. Binomial data set

Base.F (2)

n3nu [cow.F:ADLT] DAIRY COW
pe.lor [sheep.F: YNG] LAMB

Base.M (2)

zm.bal [fowl.M:ADLT] COCK/ROOSTER
böm.belee [goat.M: YNG] KID

Head.Mod (3)

bi.nubaal [child.man] BOY
bi.nuñaŋ [child.woman] GIRL
nebi.kukawal [finger.??] THUMB

Mod.Head (17)

náñ.ʃoŋ [leg.joint??] ANKLE
náñ.sí [leg.eye] ANKLE
sii.kuro [eye.??] EYEBROW
sii.ʃoŋ [eye.hair] EYELASH
sii.ʃəp [eye.animal_skin] EYELID
diŋ.toŋ [fire.flame] FLAME
náñ.pul [leg.flat] FOOTPRINT
il.nóŋ [breast.mouth] NIPPLE OR TEAT
mu.bou [nose.hole] NOSTRIIL
nepel.ʃatgin [hand.stomach] PALM OF HAND
kuŋ.dabbi [rib.cage,log.part] RIB
moŋ.nóŋ [sea/big_river.mouth] SHORE
gantal.baam [back.middle] SPINE
sii.nu [eye.water] TEAR
náñ.bíl [leg.seed] TEAR
bu.báal [funeral.man] WIDOWER
ne.ʃif [arm.joint??] WRIST

(Atlantic-Congo) Kwa

Mamara Senoufo (64): 81W / 28B / 2C

Base.AGT (2)

kuŋ.ʃoŋ [village.AGT] CHIEFTAIN
guŋaŋ.ʃoŋ [witchcraft.AGT] SORCERER OR WITCH

Mod.Head (26)

pilèki.fu.ni [night.meal.CL] DINNER
nikí.tumm.ɔn.ni [ear.iron.CL] EARRING
nikí.fí [ear.excrement.CL] EARWAX
kuŋ.ʃoŋgoŋi.fí.ni [eye.edge.hair.CL] EYEBROW
kuŋ.fí [eye.hair.CL] EYELASH
kuŋ.ʃe.ʃe.yí [eye.skin.CL] EYELID
kuŋ.ʃa.lógy [fire.??CL] FLAME
kuŋ.ʃoŋgoŋi.[arm.mark] FOOTPRINT
kuŋ.ʃe.ʃe.yí [hand.brake] HAND BRAKE
dìg3ɔɔɛ́.sò [doctor.house] HOSPITAL
dìm.ʃi [speech.core.CL] KEYWORD
càmà.nu.ni [day.meal.CL] LUNCH
fu.ɔɔɔí [horse.female.CL] MARE
guví.ɔɔ [marriage.woman] MARRIED WOMAN
mu.na.ʃe.yí [nose.hole.CL] NOSE
sàmà.cö [king.woman] QUEEN
tumùn.ko.ni [iron.road.CL] RAILWAY
ki.ʃoŋ.ni [God.knife.CL] RAINBOW
kuŋ.póɔɔɔí.ʃi [head.back.CL] SKULL
sàmà.ʃe.yí [back.Log.CL] SPINE
jukari.kace.ni [sugar.cane.CL] SUGAR CANE
pilèki.fu.ni [night.meal.CL] SUPPER
juŋ.ɔɔɔí [eye.water.CL] TEAR
juŋ.ko.ni [blood.road.CL] VEIN OR ARTERY
yɔɔɔí.ʃoŋgoŋi [water.pump] WATER PUMP
kuŋ.karo.ku [hand.junction] WRIST

(Atlantic-Congo) Mande

Bambara (8): 81W / 39B / 6C

Base.SUF.DIM (1)

mísó.mání [woman.DIM] GIRL

Head.Mod (1)

cé.ʃiilaŋ [man widow] WIDOWER

Head.Mod.SUF (1)

dìlẹ̀n.jùn.bá [finger.head.AUG] THUMB

Mod POSS Head (1)

Ala ká miru [God POSS knife] RAINBOW

Mod.Head (25)

xáŋ.ɔ [foot.back] ANKLE
káŋ.ʃoŋgoŋi [back.pack] BACKPACK
dí.ʃe [honey.son] BEE
nìɡ.ʃó [iron.horse] BICYCLE
kuŋ.ʃoŋ [ear.excrement] EARWAX
nì.ʃi [eye.hair] EYEBROW
nì.ʃi [eye.hair] EYELASH
nì.ʃoŋ [eye.skin] EYELID
tásumə.yelən [fire.light] FLAME
sò.ʃe [horse.son] FOAL OR COLT
bóló.ní [hand.mark] FOOTPRINT
dìg3ɔɔɛ́.sò [doctor.house] HOSPITAL
bá.den [goat.kid] KID
sàɡa.dén [sheep.kid] LAMB
mómbí.plákí [car.plate] LICENSE PLATE
sò.mísó [horse.woman] MARE
ítle.cmáwu [sun.middle] MIDDAY
bíl:dí [mother_s.parents.kid] NIECE
nú.ʃi [nose.hole] NOSTRIIL
ité.kí [hand.interior] PALM OF HAND
kuŋ.ʃa.lógy [day.meal.CL] LUNCH
pí.na.dùbálén [eye.LOC.mirror] SPECTACLES/GLASSES
kuŋ.ʃá.lá [back.road] RAILWAY
sí.lu.ʃa.ní [night.meal] SUPPER
jú.lu.pò [water.LOC.pump] WATER PUMP

(Atlantic-Congo) North-Central Atlantic

Bandial (9): 72W / 16B / 3C

CLF.Base (3)

jí.ʃiilí [CLF.horse] FOAL OR COLT
jí.jàmèn [CLF.god] KID
a.sú [CLF.sorcery] SORCERER OR WITCH

Head CON Mod (4)

(íṣan) sa kúbújúm [rice AGR:CON morning] BREAKFAST
(íṣan) sa kúrúsa [rice AGR:CON evening] DINNER
(íṣan) sa tíjànu [rice AGR:CON afternoon] LUNCH
(íṣan) sa kúrúsa [rice AGR:CON evening] SUPPER

Head Mod (9)

jí.ʃo.ŋílì [CLF.young CLF.horse] FOAL OR COLT
bá.ṣe ƙà.ŋen [CLF.bag CLF.hand] HANDBAG
jí.ʃo.jàmèn [CLF.young CLF.god] KID
jí.ʃo.jàmèn ɛ.mándì [CLF.young CLF.god CLF.AGR:CON] LAMB
a.ʃe.ŋe [CLF.child CLF.same_s: sibling] NIECE
bà.ʃe.ʃe.ŋe [CLF.palm CLF.hand] PALM OF HAND
ku.ʃá [CLF.knife CLF.god/sky] RAINBOW
kú.ʃá [CLF.bone CLF.shoulder] SHOULDERBLADE
ʃá.ní [CLF.digit CLF.foot] TOE
The typology and semantics of binominal lexemes

tinir suol [iron path] RAILWAY
tas ku:rb [stone bridge] STONE BRIDGE

Mod Head.3SG (22)
djake:bi:t not.a [Yukaghir fire.3SG] ARCTIC LIGHTS
ispria nya:ta [bee nest.3SG] BEEHIVE
tinir at nasu:k.a [iron horse pump.3SG] BICYCLE PUMP
tinir u.h.a [iron master.3SG] BLACKSMITH
mas u.h.a [wood master.3SG] CARPENTER
a:n sipay.a [door cheek.3SG] DOORPOST
kalgu:z eminye.e [ear earlobe.3SG] EARLOBE
kalgu:z payoho.to [ear dranduff.3SG] EARWAX
ataq suol.a [leg path.3SG] FOOTPRINT
il:i: toormah.a [hand brake.3SG] HAND BRAKE
puo:ta ja:hi.g.a [mail box.3SG] MAIL BOX
kin orto.to [day middle.3SG] MIDDAY
zal:la ni:s:ge [sky seam.3SG] MILKY WAY
murun yayah.a [nose hole.3SG] NOSTRIL
oyoyo usugy.a [side bone.3SG] RIB
sarik yapatayay.a [shoulder flat.3SG] SHOULDERBLADE
tōbō iquoy.e [head bone.3SG] SKULL
oy yut sitim.e [spider thread.3SG] SPIDER WEB
yar:y u: ta [eye water.3SG] TEAR
u: nasu:k.a [water pump.3SG] WATER PUMP
simi't arayah.a [egg yellow.3SG] YOLK
simi't nuyah.a [egg colostum.3SG] YOLK

Mod.ADIZ Head (1)
tu:letu:ny kuma:yi [toilet.ADIZ paper] TOILET PAPER

Mod.PROP Head (1)
saz.ar:da:z trostnik [sugar.PROP cane] SUGAR CANE

Basque

Basque (26): 102W / 55B / 6C

Base.LOC (2)
erla.tegi [bee:LOC] BEEHIVE
be.koki [eye:LOC] EYEBROW

Base:AGT (4)
errement.ari [tool:AGT] BLACKSMITH
ostal.ari [host:AGT] HOST
buczin.lari [mud:AGT] POTTER
zapat.ari [shoe:AGT] SHOE MAKER

Base:SUF (1)
gox.ar [morning.SUF] BREAKFAST

Mod Head (15)
iper-argi [north-light] ARCTIC LIGHTS
bickar zorro [back purse] BACKPACK
erle argirari [bee wax] BEESWAX
esne behi [milk cow] DAIRY COW
ate zango [door stilt] DOORPOST
belarri-tobula [ear-lobe] EARLOBE
guko-hitx [key-word] KEYWORD
arrau-gurpit [paddle-wheel] PADDLE WHEEL
posta-txartel [post.card] POSTCARD
itsas bazter [sea corner] SHORE
sorbula-heza [shoulder-bone] SHOULDERBLADE
azukre kainabera [sugar cane] SUGAR CANE
tresna-kutxa [tool-box] TOOLBOX
zuhaiti. enbor [tree trunk] TREE TRUNK
ur pumpa [water pump] WATER PUMP

Mod.ADIZ Head (6)
urre.zko eratzan [gold. ADIZ ring] GOLD RING
eskulo baleta [hand. ADIZ brake] HAND BRAKE
eskuo poltsu [hand. ADIZ bag] HANDBAG
harri.zko zubi [stone.ADIZ bridge] STONE BRIDGE
laita:ko teitalu [straw. ADIZ roof] THATCH
komune:ko paper [toilet.ADIZ paper] TOILET PAPER

Mod.Head (27)
su:arot [fire.carpenter] BLACKSMITH
oil.arra [chicken.male] COCK/ROOSTER
lepa:aztu [neck.ring] COLLARBONE
suk:alde [fire zone] COOKHOUSE
bet:ile [eye.hair] EYELASH
bet:azul [eye.skin] EYE-LID
su:pazar [fire.corner] FIREPLACE
zaldi:kume [horse.child] FOAL OR COLT
esku:larra [hand.fur] GLOVE
ontx.une [goat.child] KID
ar.kume [sheep.child] LAMB
post:ontzi [post.container] MAIL BOX
egun.erdi [day.half] MIDDAY
esne.bide [milk.way] MILKY WAY
titl.buru [seat.head] NIPPLE OR TEAT
sudur:tyulo [nose.hole] NOSTRIL
esku:barra [hand.inside] PALM OF HAND
bardin.bide [iron.way] RAILWAY
tren.bide [train.way] RAILWAY
ortz.adar [sky.horn] RAINBOW
bar:ezar [head.bone] SKULL
bikar:rear [back.bone] SPINE
er:pura [finger.head] THUMB
be:hat [bottom.finger] TOE
mahats.ondo [grape.tree] VINE
hai:errotza [wind.mill] WINDMILL
esku:mutur [hand.end] WRIST

Dravidian

Malayalam (60): 100W / 46B / 5C

Head.Mod (1)
kanai.kil [part.foot] ANKLE

Mod Head (4)
saikkil pump [bicycle pump] BICYCLE PUMP
prabh:la bhak:san [morning food] BREAKFAST
svarn:na motsirn [gold ring] GOLD RING
kalu p:ala:n [stone bridge] STONE BRIDGE

Mod GEN Head (1)
att.in.kutti [goat.GEN child] LAMB

Mod Head (31)
ti:n.liccu [honey.fly] BEE
ti:nccu melku [bee.wax] BEESWAX
kalutt.ello [neck.bone] COLLARBONE
kus.pili [eye.lash] EYELASH
kus.pola [eye.lid] EYELID
i:jyala [fire.flame] FLAME
kal.pajo [foot.mark] FOOTPRINT
pen:ku:ti [female.child] GIRL
kai:sanji [hand.bag] HANDBAG
ucca bhak:saun [midday.food] LUNCH
pen.kutira [female.horse] MARE
ki:ra:pati:n [milky.way] MILKY WAY
am:miyi yam'ma [aunt.mother] MOTHER-IN-LAW (OF A MAN)
tula:ca:kram [paddle.wheel] PADDLE WHEEL
ti:sunj.pata [fire.vehicle.way] RAILWAY
ma:la:illo [rain.bow] RAINBOW
v:iri ello:]. [bone] RIB
tol:palaka [shoulder.blade] SHOULDERBLADE
tula:yo:ti [head.shell] SKULL
cilanta:vala [spider.web] SPIDER WEB
kutira:pari [horse.ROW] STABLE OR STALL
ähl ay:dar [sun.day] SUNDAY
ku:n:uso [eye.water] TEAR
tula:viraj [mother.finger] THUMB
käl viraj [leg.finger] TOE
mätra:pura [urine.house] TOILET
fi:vant [fire.vehicle] TRAIN

456
Submitted version 2020-05-26
### The typology and semantics of binominal lexemes

| Mod.Head (2) | blus.turgis [flea.market] FLEA MARKET  
son.kaulis [side.bone] RIB |
|---|---|
| Mod.L.E.Head (5) | rakt.i.kaulis [key.LE.bone] COLLARBONE  
rakt.a.zodis [key.LE.word] KEYWORD  
vaiv.o.rykšė [Vaiva.LE.cane] RAINBOW  
vor.a.tinklus [spider.LE.web] SPIDER WEB  
cukr.a.nendre [sugar.LE.red] SUGAR CANE |
| (Indo-European) Celtic |
| Irish (29): 100W / 40B / 6C |
| Base.AGT (3) | femail.óir [farm.AGT] FARMER  
iasc.aire [fish.AGT] FISHERMAN  
greas.al [embroidery.AGT] SHOE MAKER |
| Base.DIM (5) | tinte.din [fires.DIM] FIREPLACE  
caille.in [veil.DIM] GIRL  
mot.óig [mitt.DIM] GLOVE  
slinn.din [sate.hile.DIM] SHOULDER BLADE  
oróig.ig [sledge.hammer.DIM] THUMB |
| Base.NMLZ (1) | drauchacht [magician.NMLZ] MAGIC |
| Base.SUF (1) | roth.ar [wheel.SUF] BICYCLE |
| Head Mod.Gen (27) | mod.la.droma [bag back:GEN] BACKPACK  
ceir blench [waex:GEN] BEEF  
saor.adhrmid [craftsman wood:GEN] CARPENTER  
bó bánnte [cow milk:GEN] DAIRY COW  
bun na cluaise [bottom DET ear:GEN] EARlobe  
fáinne cluaise [ring ear:GEN] EARRING  
cesir cluaise [waex ear:GEN] EARwax  
caip.in súile [cap.DIM eye:GEN] EYELlid  
lorg.coise [trace foot:GEN] FOOTPRINT  
fáinne óir [ring gold:GEN] GOLD RING  
molna [lámh] [bag hand:GEN] HANDBAG  
boasa poist [box post:GEN] MAIL BOX  
meán lae [middle day:GEN] MIDDay  
bealach na bó finne [way DET cow:GEN fair:GEN] MILKY WAY |
| Head Mod. (3) | míthair cèile [mother companion/spouse:GEN] MOTHER-IN-LAW (OF A MAN)  
tír dichuir [land heritage:GEN] NATIVE COUNTRY  
poll xáná [hole/pool nose:GEN] NOSTRIL  
roth liàin [wheel trowel/blade:GEN] PADDLE WHEEL  
cárta poist [card post:GEN] POSTCARD  
bogha básít [bow rain:GEN] RAINBOW  
droichead liá [bridge stone:GEN] STONE BRIDGE  
de Dómhaigh [day Sunday:GEN] SUNDAY  
pálpéar leithris [paper toilet:GEN] TOILET PAPER  
boasa artís [box tools:GEN] TOOLBOX  
suabh fiacra [brush teeth:GEN] TOOTHBRUSH  
de Céadaoin [day Wednesday:GEN] WEDNESDAY  
meall eanna gaoithe [mill wind:GEN] WINDMILL |
| Welsh (19): 106W / 63B / 10C |
| Base.AGT (4) | ffairn.wr [farm.AGT] FARMER  
pysgol.wr [fishes.AGT] FISHERMAN  
gweseti.wr [guest.AGT] HOST  
crochan.ydd [cauldron.AGT] POTTER  

---

| Base.F (1) | brenin.es [king.F] QUEEN |
| Base.NMLZ (1) | pen.aeth [chief.NMLZ] CHIEFTAIN |
| Base.SUF (1) | cwnnedwg [district.SUF] NEIGHBOUR |

---

### Head DET Mod (5) | goleuuni ‘r ysgwedd [light DET north] ARCTIC LIGHTS  
pont yr ysgwedd [bridge DET shoulder] COLLARBONE  
tor (γ) llaw [palm DET hand] PALM OF HAND  
bwa ‘r Drindod [bow DET trinity] RAINBOW  
assywrn y ceft [bone DET back] SPINE |

---

### Head Mod (40) | bag ysgwedd [bag back] BACKPACK  
gwenny nél [bees honey] BEE  
cwch gwynen [hive bees] HIVE  
ciw xwennyw [waex bees] BEEWAX  
pwpmp beic [pump bike] BICYCLE PUMP  
saer coed [carpenter wood] CARPENTER  
post drws [post door] DOORPOST  
lledd clust [lappet ear] EARlobe  
tlw clust [jewel ear] EARRING  
cwyr clust [waex ear] EARwax  
blew amran [hairs eyelid] EYELASH  
tli tân [place fire] FIREPLACE  
lein bysgota [line fishing] FISHING LINE  
ffair petheuach [market bits_and_pieces] FLEA MARKET  
öl troed [mark foot] FOOTPRINT  
modryn awr [ring gold] GOLD RING  
brèc llaw [brack hand] HAND BRAKE  
bag llaw [bag hand] HANDBAG  
cadach poked [rag pocket] HANDKICKER OR RAG  
wyn gafr [kid goat] KID  
phéil rhft [plate number] LICENSE PLATE  
boisc llwybrau [box letters] MAIL BOX  
canol dydd [middle day] MIDDay  
caer arianrhod [fort Arianrhod] MILKWy WAY  
caer gwydioni [fort Gwydion] MILKWy WAY  
olwyn bado [wheel paddle] PADDLE WHEEL  
cerdyn post [card post] POSTCARD  
gwe pryf cop [web insect spider] SPIDER WEB  
pont garreg [bridge stone] STONE BRIDGE  
cansen sivuy [cane sugar] SUGAR CANE  
dydd sul [day sun] SUNDAY  
to gwellt [roof straw] TATCH  
bys bawd [finger thumb] THUMB  
bys troed [finger foot] TOE  
bhwch offer [box tools] TOOLBOX  
brov dannedd [brush teeth] TOOTHBRUSH  
pwpmp dêr [pump water] WATER PUMP  
dydd merchwr [day Mercury] WEDNESDAY  
gŵr gweddol [man widow] WIDOWER  
melin wint [mill wind] WINDMILL |

---

### Head Mod.AdJz (2) | bious laeth.og [cow milk. ADJZ] DAIRY COW  
lwybr llwyth.og [cow milk. ADJZ] MILK Wy WAY |

---

### Head Prep Mod (1) | papur lle cweiche [paper for toilet] TOILET PAPER |

---

### Head.Mod (2) | bôn.cyff [base.trunk] TREE TRUNK  
melyn.yi [yellow.egg] YOLK |

---

### Head.Mod (6) | clust.bws [ear.jewel] EARRING  
yshyd.fý [hospital.house] HOSPITAL  
allwedd.gair [key.word] KEYWORD  
mam.gwlad [mother.country] NATIVE COUNTRY  
rheid.flrroll [railroad] RAILWAY  
gwin.gwyl [wine.trees] VINE |
Appendix E. Binomial data set

English (23): 105W / 54B / 8C

Base.AGT (3)
farmer [farm.AGT] FARMER
potter [pott.AGT] POTTER
widower [widow.M] WIDOWER

Head.Mod (1)
midday [middle.day] MIDDAY

Mod Head (13)
bicycle pump [bicyc.pump] BICYCLE PUMP
flea market [flea market] FLEA MARKET
gold ring [gold ring] GOLD RING
hand brake [hand brake] HAND BRAKE
handkerchief [hand.kerchief] HANDKERCHIEF OR RAG
number plate [number plate] LICENSE PLATE
mail box [mail box] MAIL BOX
paddle wheel [paddle wheel] PADDLE WHEEL
stone bridge [stone bridge] STONE BRIDGE
sugar cane [sugar cane] SUGAR CANE
toilet paper [toilet paper] TOILET PAPER
tree trunk [tree trunk] TREE TRUNK
water pump [water pump] WATER PUMP

Mod.ADJZ Head (2)
northern lights [north.ADJZ lights] ARCTIC LIGHTS
milky way [milk.ADJZ way] MILKY WAY

Mod.Head (25)
backpack [back.pack] BACKPACK
collarbone [collar.bone] COLLARBONE
doorpost [door.post] DOORPOST
carloe [carloe] EARLOBE
doorpost [door.post] DOORPOST
earring [ear.ring] EARRING
earring [ear.ring] EARRING
eyelash [eye.lash] EYELASH
eyebrow [eye.brow] EYEBROW
eyelid [eye.lid] EYELID
fireplace [fire.place] FIREPLACE
footprint [foot.print] FOOTPRINT
handbag [hand.bag] HANDBAG
herdsman [herd.POSS.man] HERDSMAN
key [key] KEYWORD
kitchen [kitchen] KITCHEN
nostril [nose.hole] NOSTRIL

German (20): 131W / 75B / 8C

Base.AGT (3)
fisch.er [fish.AGT] FISHERMAN
öpf.er [pot.AGT] POTTER
sauber.er [magic.AGT] SORCERER OR WITCH

Base.DIM (4)
knöch.el [bone.DIM] ANKLE
müd.chen [maid.DIM] GIRL
gelb.lein [goat.DIM] KID
zick.lein [goat.DIM] KID

Base.F (1)
könig.in [king.F] QUEEN

Base.NMLZ (2)
tisch.ler [table.NMLZ] CARPENTER
schrey.in [cupboard.NMLZ] CARPENTER
The typology and semantics of binominal lexemes

**Head.Mod (2)**
mit.tag [middle.day] MIDDAY
mitt.woch [middle.week] WEDNESDAY

**Mod.ADJZ Head (1)**
pol.ar.lichter [pole.ADJZ lights] ARCTIC LIGHTS

**Mod.Head (51)**
nord.lichter [north.lights] ARCTIC LIGHTS
ruck.sack [back.sack] BACKPACK
luft.pumpe [air.pump] BICYCLE PUMP
fahrrad.pumpe [bicycle.pump] BICYCLE PUMP
arm.band [arm.band] BRACELET
zimmer.mann [timmerman] CARPENTER
schlüssel.bein [key.bone] COLLARBONE
milch.kuh [milk.cow] DAIRY COW
abend.brot [evening.bread] DINNER
abend.essen [evening.meal] DINNER
tür.pfosten [door.post] DOORPOST
ohr.lipp.chen [ear.lobe] EARLOBE
ohr.ring [ear.ring] EARRING
hand.gelenk [hand.joint] WRIST
wind.mühle [wind.mill] WINDMILL
zahn.bürste [toothbrush] TOOTHBRUSH
werkzeug.kasten [tool.box] TOOLBOX
werkzeug.kasten [tool.box] TOOLBOX

**Head.Mod (70): 100W / 57B / 3C**

**Head.Mod (1)**
mid. dag [middle.day] MIDDAY

**Mod.Head (51)**
nord.lys [north.lights] ARCTIC LIGHTS
rygg.sekk [back.back] BACKPACK
bi.kube [bee.cube] BEEHIVE
bi.voks [bee.wax] BEEWAX
sykkel.pumpe [bicycle.pump] BICYCLE PUMP
arm.bånd [arm.band] BRACELET
krage.bein [collar.bone] COLLARBONE
eld.hus [fire.house] COOKHOUSE
dor.stolpe [door.post] DOORPOST
dø opp [ear.??] EARLOBE
dø dob [ear.button] EARRING
dø.voks [ear.wax] EARWAX
eye.vipp [eye.??] EYELASH
eye.lok [eye.lid] EYELID
ild.sted [fire.place] FIREPLACE
løpp.marked [face.market] FLEA MARKET
fot.spor [foot.trace] FOOTPRINT
gull.ring [gold.ring] GOLD RING
hånd.bremse [hand.brake] HAND BRAKE
hånd.veske [hand.bag] HANDBAG
lomme.tørkle [pocket.kerchief] HANDKERCHIEF OR RAG
nøkkel.ord [key.word] KEYWORD
bil.kort [car.sign] LICENSE PLATE
brev.kasse [letter.box] MAIL BOX
melke.vei.en [milk.way] MILK WAY
hjem.land [home.country] NATIVE COUNTRY
bryst.vorte [breast.wart] NIPPLE OR TEAT
nes.ebor [nose.hole] NOSTRIL
skovle.hjul [shovel.wheel] PADDLE WHEEL
hånd.flate [hand.surface] PALM OF HAND
post.kort [post.card] POSTCARD
jern.bane [iron.way] RAILWAY
regn.bue [rain.bow] RAINBOW
rib.bein [rib.bone] RIB

**Mod.LE.H (5)**
melk.e.ku [milk.LE.cow] DAIRY COW
eye.n.bryn [eye.LE.brow] EYEBROW
kved.s.mat [evening.LE.food] SUPPER
øns.dag [ Odin.LE.day] WEDNESDAY
egg.e.plomme [egg.LE.yolk] YOLK

**Old High German (31): 63W / 23B / 5C**

**Base.AGT (5)**
hirt.i [herd.AGT] HERDSMAN
hufan.âri [pot.AGT] POTTER
zoubar.âri [magic.AGT] SORCERER OR WITCH

**Base.DIM (1)**
zickin [goat.DIM] KID
Appendix E. Binominal data set

Base.F (1)
kunig.in [king.SUF] QUEEN

Head.Mod (2)
mitti(t).tag [middle.day] MIDDAY
mitta.wehha [middle.week] WEDNESDAY

Mod.Head (16)
bli.kar [bee.container] BEEHIVE
zimbar.man [timber.man] CARPENTER
turi.stud[a]l [door.post] DOORPOST
ör.lappa [ear.cloth] EARLOBE
ör.smero [ear.wax] EARWAX
akkar.man [field.man] FARMER
fiur.stat [fire.place] FIREPLACE
fueo.spor [foot.trace] FOOTPRINT
hant.suwo[hand.shoe] GLOVE
regan.bogyo [rain.bow] RAINBOW
hirni.skala [brain.bowl] SKULL
spinne.webbi [spider.fabric] SPIDER WEB
ruggi.bein [back.spine] SPINE
sunnûn.tag [sun.day] SUNDAY
ābund.maos [evening.meal] SUPPER

(Indo-European) Greek

Greek (22): 100W / 30B / 7C

Base.DIM (1)
katisik.aki [goat.DIM] KID

Base.SUF (4)
proi.ino [morning.SUF] BREAKFAST
vi.far.īd[a] [eyelid.SUF] EYELASH
mesimer.iano [noon.SUF] LUNCH
vrās.ino [night.SUF] SUPPER

Head.DEF:GEN Mod.GEN (3)
kolona tis porta.s [post DEF:GEN door.GEN] DOORPOST
lovos tu afit.ou [lobe DEF:GEN ear.GEN] EARLOBE
adlia ner.ur [wax DEF:GEN ear.GEN] EARWAX

Head Mod (1)
leksi-kliō [word-key] KEYWORD

Head Mod.GEN (9)
ayelada yalakotos [cow.dairy:GEN] DAIRY COW
spasino heri.ou [brake hand:GEN] HAND BRAKE
tsanta heiro.s [bag hand:GEN] HANDBAG
pinakida kikloforia.s [sign traffic:GEN] LICENSE PLATE
istos araxan.s [web spider:GEN] SPIDER WEB
charti iyia.s [paper hygiene:GEN] TOILET PAPER
kut erilino. [box tool:GEN] TOOLBOX
kormos bentru.s [trunk tree:GEN] TREE TRUNK
adlia ner.ur [pump water:GEN] WATER PUMP

Mod.ADJ Head (3)
vorio selas [north:ADJ lights] ARCTIC LIGHTS
uron.to tokso [sky:ADJZ bow] RAINBOW
petr.ιn pefire [stone:ADJZ bridge] BRIDGE

Mod.LE.Head (9)
ki λ.o.kokalo [collar:LE.bone] COLLARBONE
ik.o.desopot [house:LE.lord] HOST
yrammat.o.kivotto [letters:LE.box] MAIL BOX
sióhr.o.bromos [iron:LE.road] RAILWAY
om.o.platt [shoulder:LE.blade] SHOULDERBLADE
zachar.o.kalamo [sugar:LE.cane] SUGAR CANE
achir.o.ukpei [straw:LE.cover] THATCH
okol.o.vurtas [tooth:LE.brush] TOOTHBRUSH
anom.o.milos [wind:LE.wind] WINDMILL

(Indo-European) Indo-Aryan

Assamese (7): 87W / 34B / 4C

Head Mod (1)
gu.gys [body.tree] TREE TRUNK

Mod Head (4)
mou.makhi [honey.beet] BEE
sor bozar [thief market] FLEA MARKET
xuc koxub [index word] KEYWORD
dak siti [postal letter] POSTCARD

Mod.Gen.Gen (17)
aiinkel:or hawa [cycle:GEN gas] BICYCLE PUMP
ratipua.r ahbar [morning:GEN food] BREAKFAST
rat:r ahar [night:GEN food] DINNER
duar:or khuta [door:GEN post] DOORPOST
ek:or loti [ear:GEN lobe] EARLOBE
kan:or mof [ear:GEN dirt] EARWAX
soka.r nom [eye:GEN hair] EYELASH
soka:r poza [eye:GEN lid] EYELID
ghora.r powali [horse:GEN calf] FOAL OR COLT
bhori.r sap [foot print] FOOTPRINT
son:or angutili [gold:GEN ring] GOLD RING
sithi.r bakis [letter:GEN box] MAIL BOX
nij:or dex [self:GEN country] NATIVE COUNTRY
nak:or phuta [nose:GEN hole] NOSE
hat:r tolava [hand:GEN backside] PALM OF HAND
hat:or burha.anguli [hand:GEN old.finger] THUMB
bhori.r burha.anguli [leg:GEN old.finger] TOE

Head Mod (12)
mou:xak [bee.hive] BEEHIVE
kath.mistrī [wood.carpenters] CARPENTER
kan phuli [ear:GEN lobe] EARRING
hat.moca [hand.sock] GLOVE
sikitsa.lov [treatment.place] HOSPITAL
mugya.por[license plate/certificate] LICENSE PLATE
ram.dhena [Lord_Rama.bow] RAINBOW
kami.Μ [chest:bone] BIK
sil.sako [stone:bridge] BRIDGE
suka.pani [eye:water] TEAR
nutra.lov [urine.place] TOILET
budh.bar [Mercury.day] WEDNESDAY

Hindi (38): 95W / 30B / 4C

Base.AGT (1)
lohā.ṛ [gold.GEN ring] GOLD RING

Mod Gen Head (10)
ṛaṭ kā khāṇā [evening GEN food] DINNER
ghore kā bacca [horse GEN child] FOAL OR COLT
pair kā cinhi)' [foot GEN sign] FOOTPRINT
sone kā angūti [gold GEN ring] GOLD RING
dopahar kā khāṇā [noon GEN food] LUNCH
mukrī kā jāl [spider GEN web] SPIDER WEB
ṛaṭ kā khāṇā [night GEN food] SUPPER
pāṃvī kā angūṭhā [foot GEN finger] TOE
saic kā kāgā [purification GEN paper] TOILET PAPER
dāṃt kā bāī [tooth GEN brush] TOOTHBRUSH

Mod Head (6)
kabārī bādzū [scrap_merchant market] FLEA MARKET
arugvā pēr [license plate] LICENSE PLATE
jālū-tōnā [magic-charm] MAGIC
dīk-baksā [post-box] MAIL BOX
jal pām [water pump] WATER PUMP
pavan cakkī [wind mill] WINDMILL

Mod.Head (13)
madhu.mom [honey.wax] BEESWAX
karna.phil [ear.flower] EARRING
karna.mol [ear.dirt] EARWAX
ākāś.γ [sky.Ganges] MILKY WAY
The typology and semantics of binominal lexemes

mātr. bhumī [mother.earth] NATIVE COUNTRY
bhāṭi.jī [brother.daughter] NIECE
bhādmā],jī [sister.daughter] NIECE
post.kārd [post.card] POSTCARD
indra.dhanus [Indra.bow] RAINBOW
ghar.sil [horse.hall] STABLE OR STALL
sau.cälāy [purification.area] TOILET
yatra.peti [instrument.box] TOOLBOX
rel.gārī [rail.car] TRAIN

Nepali (71): 66W / 11B / 2C
Mod.Gen Head (6)
kukhārā.ko bāhā [hen.Gen cock] COCK/ROOSTER
sun.ko ānūth i [gold.Gen ring] GOLD RING
dud.ko munto [milk.Gen head] NIPPLE OR TEAT
mākurā.ko jālo dudh.ko mun [milk.Gen head] NIPPLE OR TEAT
kuze.gar āhan.gar [earthen_bottle.AGT] IRON
Base.AGT (3)
Mod.GEN Head (6)
kuze.gar āhan.gar [earthen_bottle.AGT] IRON
Base.F (1)
kuze.gar āhan.gar [earthen_bottle.AGT] IRON

Selice Romani (80): 88W / 8B / 4C
Base.ABST (1)
čohan.ipe [witch/sorcerer.ABST] MAGIC
Mod.Gen Head (2)
ča.j.ori [Gypsy_child.F.DIM] GIRL
Base.F (1)
kirā.čikha [king.F] QUEEN

(Indo-European) Iranian
Western Farsi (74): 103W / 47B / 9C
Base.AGR (3)
āhan.gar [iron.AGR] BLACKSMITH
kuze.gar [earthen_bottle.AGT] POTTER
jādu.gar [magic.AGR] SORCERER OR WITCH
Head Mod (5)
kalane kelādī [word key] KEYWORD
somāre māšin [number car] LICENSE PLATE
mīdār zan [mother woman] MOTHER-IN-LAW (OF A MAN)
rāh āhan [way iron] RAILWAY
ūr ankabat [cord spider] SPIDER WEB

(Indo-European) Romance
French (28): 102W / 49B / 10C
Base.DIM (4)
bruc.elet [arm.DIM] BRACELET
chev.eau [goat.DIM] KID
mamelon [breast.DEF] NIPPLE OR TEAT
poing.e [fist.DIM] WRIST
Base.F (1)
sorcer.e [sorcerer.F] SORCERER OR WITCH
Base.NMLZ (4)
farm.ier [farm.NMLZ] FARMER
pêch.eur [fish.NMLZ] FISHERMAN
pot.ier [pot.NMLZ] POTTER
ecurier [horsemanship.NMLZ] STABLE OR STALL

Head Mod (1)
mot-clé [word-key] KEYWORD
Head Mod.ADJZ (4)
aurore boréale [light north.ADJZ] ARCTIC LIGHTS
voie lactée [way milk:ADJZ] MILKY WAY
carte post.ale [card post.ADJZ] POSTCARD
colonne vertébrale [column vertebra.ADJZ] SPINE

Head Prep DEF Mod (2)
lorbe de l'oreille [lobe of DEF:ear] EARLOBE
paume de la main [palm of the hand] PALM OF HAND

Head Prep Mod (17)
ruche d'abeille.s [hive of:bee.PL] BEEHIVE
cire d'abeille [wax:bee] BEESWAX
chef de clan [chief of clan] CHIEFTAIN
cuisine de châtrier [kitchen of worksite] COOKHOUSE
jambage de porte [jamb of door] DOORPOST
boucle d'oreille [ring of:ear] EARRING
cire d'oreille [wax:ear] EARWAX
ligne de pêche [line of fishing] FISHING LINE
Appendix E. Binomial data set

dito del piede [finger of DET foot] TOE
cassetta degli attrezzi [box of DET tools] TOOLBOX
tronco dell'albero [trunk of DET:tree] TREE TRUNK
pompa dell'acqua [pump of DET:water] WATER PUMP

Head.MOD (2)
capo.tribù [head.clan] CHIEFTAIN
mezzo.giorno [half.day] MIDDAY

Mod.Head (1)
ferro.via [iron.way] RAILWAY

Romanian (81): 135W / 41B / 11C

Base.ABST (1)
capitan.ie [captain.NMLZ] CHIEFTAIN

Base.AGT (7)
fier.ar [iron.AGT] BLACKSMITH
potoa(o)vă.r [hoof.AGT] BLACKSMITH
usă.o.r [door.SUF] DOORPOST
pară.an [earth.AGT] FARMER
păst.or [pasture.AGT] HERDSMAN
o(a)lă.ar [pot.AGT] POTTER
cizmă.ar [boots.AGT] SHOEMAKER

Base.AGT.ABST (1)
bucătă.ar.ie [piece_of_food.AGT:NMLZ] COOKHOUSE

Base.DIM (1)
mână.uş.ă [hand.SUF:F] GLOVE

Base.F (2)
socră.ă [father_in_law.F] MOTHER-IN-LAW (OF A MAN)
nepo(a)lă.ă [grandson/nephew.F] NIECE

Base.NMLZ (1)
brăz.ară [arm.NMLZ] BRACELET

Base.SUF (1)
gălb.ens [yellow.SUF] YOLK

Head Mod (1)
cuvânt cheie [word key] KEYWORD

Head.MOD.ADJZ (7)
lumină.polară [light polar] ARCTIC LIGHTS
cutie poștală [box post.ADJZ] MAIL BOX
calea lactee [way milk:ADJZ] MILKY WAY
carte poștală [card postal] POSTCARD
calea ferată [way iron:ADJZ] RAILWAY
coloană vertebrală [column vertebral] SPINE
hârtie igien.ica [paper hygiene.ADJZ] TOILET PAPER

Head.PREP Mod (17)
pompa de bicicletă [pump of bicycle] BICYCLE PUMP
cuja de lapte [cow of milk] DAIRY COW
sfărăi de undătă [thread of fishing_rod] FISHING LINE
urmac de picior [trace of foot] FOOTPRINT
inel de aur [ring of gold] GOLD RING
frână de mânt [brake of hand] HAND BRAKE
guăntă de mână [bag of hand] HAND BAG
rotă cu zu布置 [wheel with paddle] PADDLE WHEEL
pânză de păianjen [cloth of spider] SPIDER WEB
pod de piatra [bridge of stone] STONE BRIDGE
trestie de zahăr [reed of sugar] SUGAR CANE
acoperiş de paie/stuf [roof of straw/reed] THATCH
trusă de scule [kit of tools] TOOLBOX
periău de dinţi [brush of tooth] TOOTHBRUSH
vîţa de vie [creepers_plant of vine/vineyard] VINE
pompă de apă [pump of water] WATER PUMP
moară de vant [mill of wind] WINDMILL

Head.DET.MOD.DET.GEN (2)
lopățica umăr.u.LUI [shoulder.DEF.DEF.GEN] SHOULDERBLADE

inochietaura mâin.ii [joint:DEF hand.DEF.GEN] WRIST

2020-05-26 Submitted version 463

Czech (13): 98W / 44B / 7C
Base.DIM (4)
větrár [man.DIM] BOY
kabelka [purse.DIM] HANDBAG
kůží [goat.DIM] KID
lopaťka [shovel.DIM] SHOULDERBLADE
Base.F (2)
čarodějnice [wizard.F] SORCERER OR WITCH
střed a [middle.F] WEDNESDAY
Base.NMLZ (13)
kovář [iron.NMLZ] BLACKSMITH
večeře e [evening.NMLZ] DINNER
náušnice [ear.on.NMLZ] EARRING
farmář [farm.NMLZ] FARMER
rybář [fish.NMLZ] FISHERMAN
ručička [hand.NMLZ] GLOVE
kuželovník [pocket.NMLZ] HANDKERCHIEF OR RAG
nemocnice [illness.NMLZ] HOSPITAL
hrnčíř [pot.NMLZ] POTTER
královna [king.GEN.NMLZ?] QUEEN
pavučená [spider.NMLZ] SPIDER WEB
zástěra [after.fist.NMLZ] WRIST
Head PREP Mod (2)
pumpička na kolo [pump on bicycle] BICYCLE PUMP
kartáček na zuby [brush.DIM on teeth] TOOTHBRUSH
Head.LE.Mod (1)
půl e.dne [half.LE.day] MIDDAY
Mod.ADZ Head (21)
polárni záře [polar.ADZ lights] ARCTIC LIGHTS
včela i il [bee.ADZ hive] BEEHIVE
včela i vosk [bee.ADZ wax] BEESWAX
klíč i kost [key.ADZ bone] COLLARBONE
dvě díl sloupek [door.ADZ post] DOORPOST
uši dílalček [ear.ADZ lobe] EARLOBE
uši díla [ear.ADZ wax] EARWAX
uši dílasy [ear.ADZ lash] EYELASH
blešní trh [flea.ADZ market] FLEA MARKET
zlatý prsten [gold.ADZ ring] GOLD RING
ruční brada [hand.ADZ bristle] HAND BRAKE
klíčové slovo [key.ADZ word] KEYWORD
použitá skrínka [mail.ADZ box] MAIL BOX
milčí dára [milky.ADZ track] MILKY WAY
rod.ní říše [clan.ADZ land] NATIVE COUNTRY
nos.ní díra [nose.ADZ hole] NOSTRIL
vod.ní kolo [water.ADZ wheel] PADDLE WHEEL
kamen.ní most [stone.ADZ bridge] STONE BRIDGE
červená tráva [sugar.ADZ cane] SUGAR CANE
toaletter.ní papír [toilet.ADZ paper] TOILET PAPER
větrné mlýny [wind.ADZ mill] WINDMILL
Mod.ADZ Head.DIM (1)
oc.ní víčko [eye.ADZ lid.DIM] EYELID

Lower Sorbian (21): 118W / 43B / 11C
Base.ADZ.NMLZ (1)
zejl.nica [iron.NMLZ] BLACKSMITH
railway
Base.AGT (3)
rybár [fish.AGT] FISHERMAN
góści inar [guest.AGT] HOST
gajnar ['pot.AGT] POTTER
Base.F (1)
królowa [king.F] QUEEN
Base.LOC (1)
wogniwo [fire.LOC] FIREPLACE
pel.thong [be.container] BEEHIVE
mil.cho [honey.wax] BEESWAX
kongki.phemphu [air.pump] BICYCLE PUMP
taycang.cang' [smithy.professional] BLACKSMITH
namca.ai [man.child] BOY
sanay.ai [man.child] BOY
achim.pap [morning.meal] BREAKFAST
achim.siksa [morning.meal] BREAKFAST
cang.talk [male.animal.chicken] COCK/ROOSTER
su.thalik [male.animal.chicken] COCK/ROOSTER
cec.so [milk.cow] DAIRY COW
cenyek.siksa [evening.meal] DINNER
cenyek.pap [evening.meal] DINNER
mun.selcwa [door.support_pillar] DOORPOST
kwu.keli [ear.hook/hanger] EARRING
sok.nwunsep [inside.eyebrow] EYELASH
nwan.kkepul [eye.layer] EYELID
nongcang.kwanliin [farm.manager] FARMER
pyek.nanlo [wall.stove] FIREPLACE
nakksi.kkwun [fishing.professional] FISHERMAN
pal.kko'ch [fire.flower] FLAME
pyeluk.sicang [flea.market] FLEA MARKET
pal.calwuk [person.wife] FOOTPRINT
kyveicip.ai [girl.deroq.child] GIRL
yeca.ai [woman.child] GIRL
cang.kap [palm_of_hand.box/armour] GLOVE
kum.panci [gold.ring] GOLD RING
son.swukan [hand.towel] HANDKERCHIEF OR RAG
saykk.iyemso [baby.goat] KID
saykk.yang [baby.sheep] LAMB
catong.ya.penphapan [car.number_plate] LICENSE PLATE
penho.phan [number.board] LICENSE PLATE
am.mal [female.animal.horse] MARE
si.emen [husband.side.mother] MOTHER-IN-LAW (OF A MAN)
ivus.saram [neighbour.person] NEIGHBOUR
chokha.tul [niece/nephew.daughter] NIECE
cec.klokci [breast.knob] NIPPLE OR TEAT
son.patuk [hand.floor/ground] PALM OF HAND
kalim.yepsa [picture.postcard] POSTCARD
chel.kit [iron.road] RAILWAY
kalpi.pye [rib.bone] RIB
ekkay.pye [shoulder.bone] SHOULDERBLADE
kemi.cwul [spider.string] SPIDER WEB
kemi.cip [spider.house] SPIDER WEB
tung.pye [back.bone] SPINE
vol.tari [stone.bridge] STONE BRIDGE
sathang.susu [sugar.sorghum] SUGAR CANE
cenyek.siksa [evening.meal] SUPPER
nwan.mal [eye.water] TEAR
choka.cibung [thatch.roof] THATCH
pal.kalak [foot.digit] TOE
yencang.kwey [tool.box] TOOLBOX
yencang.thong [tool.box] TOOLBOX
tengkoul.sikul [vine.plant] VINE
mul.phemphu [water.pump] WATER PUMP
son.mok [hand.neck] WRIST
phal.mok [arm.neck] WRIST

(Nakh-Daghestanian) Avar-Andic-Tsezic

Bezhta (47): 93W / 35B / 8C

Base.NMLZ (1)

Base.PAUC (1)

Head Mod (1)

c’uddo c’emuc’ [red.egg] YOLK

Mod.ABL Head (1)

juba.λ’a.s beiyoli [midday.SUP.ABL meal] LUNCH
The typology and semantics of binominal lexemes

Mod.ADJZ Head (1)

Mod.ATTR Head (2)

Mod.Gen Head (25)

Mod.Head (3)

Mod.Head (32)

Archí (5): 66W / 14B / 3C

Mod Head (3)

Nakh-Daghestanian Lezgic

Archi (5): 66W / 14B / 3C

Mod Head (3)

(Nakh-Daghestanian) Lezgic

Archí (5): 66W / 14B / 3C

Mod Head (3)

Mod.Gen Head (9)

Uralic Finnish

Estonian (24): 104W / 65B / 7C

Base.Gen.SUF (2)

Base.Gen.SUF (4)

pea.lik [head:DER] CHEFTAIN
talu.nik [farm:DER] FARMER
sör.m.ik [finger:DER] GLOVE

Head.Mod (1)

Head.Mod (1)

Head.Mod (24)

Estonian (24): 104W / 65B / 7C

Base.Gen.SUF (2)

Base.Gen.SUF (4)

mesi.lane [honey:Nom.DER] BEE
(Uralic) Mari

Western Mari (62): 93W / 43B / 4C

Mod Head (24)

ir kuckš [morning meal] BREAKFAST
vadš kuckš [evening meal] DINNER
amaša kūšik [door post] DOORPOST
pūšmi mēčaš [ear end] EARLOBE
pūšši kši [ear wax] EARWAX
tš vac [fire fireplace] FIREPLACE
āngōr sirtš [river thread] FISHING LINE
tš lāšım [fire flame] FLAME
sőrt'n sārgā [gold ring] GOLD RING
kid tormoc [hand brake] HAND BRAKE
kēss iš [goat baby_animal] KID
kēčâval kuckš [midday meal] LUNCH
kek kombš kornš [bird goose road] MILK WAY
kid lapa [hand pad] PALM OF HAND
korošk mastor [pot master] POTTER
kērt nornš [iron road] RAILWAY
kem ērgūš [boot shoemaker] SHOEMAKER
innimサー [horse barn] STABLE OR STALL
ki kēver [stone bridge] STONE BRIDGE
saxar trostnik [cargo cane] SUGAR CANE
wadš kuckš [evening meal] SUPPER
šēš.m šādš koropła [make.NMLZ thing box] TOOLBOX
pušangš [tree trunk] TREE TRUNK
vinograd padšrēngš [grape vine] VINE

Mod.Gen Head (1)
mūnš.n sarš [egg.GEN yellow] YOLK

Mod Head (17)
kid.soň [hand.?] BRACELET
kelde.lu [seine.bone] COLLABORNE
pūšši tāngā [ear.coin] EARRING
sēncūxal [eye.brow] EYEBROW
sēncūxūn [eye.hair] EYELASH
sēncūkomšš [eye.cover] EYELID
ner raž [nose.?] NOSTRIL
ōrdšš.lu [side.bone] RIB
pulšš. savala [shoulder.spoon] SHOULDERBLADE
vał.karka [head.cloak] SKULL
āngōremšš vad [spider.cauldron] SPIDER WEB
utp.nūš [back.core] SPINE
sēncūxūk [eye.water] TEAR
jāl varnā [foot.finger] TOE
vār šār [blood.?] VEIN OR ARTERY
vār keč [blood.?].day] WEDNESDAY
kid.piān [hand.?] WRIST

Mod.Lat Head (1)
šand.eš pumaga [toilet.LAT paper] TOILET PAPER

(Uralic) Saami

Kildin Sami (86): 87W / 35B / 4C

Mod Head (26)

siellk vuss [back sack] BACKPACK
vezzviss-pneiss’ [wasp-nest] BEEHIVE
vezzviss-tuře’ [wasp-tar] BEESWAX
vuell’k-taxx’ [shoulder-bone] COLLABORNE
piejv-vearr [day-soup] DINNER
pielij-kied’k [ear-stone] EARLOBE
cīl’ m-rūmās [eye-brim] EYELID
kōll’s surmus [gold ring] GOLD RING
kid vuss [hand bag] HANDBAG
nosor-nipp ex’ [nose-kirsch] HANDKERCHIEF OR RAG
piejv-pierk [day-meal] LUNCH
pọltu čibk’ [post box] MAIL BOX
nimless jèb’ e’s [female horse] MARE
piejv-kēsk [day-middle] MIDDAY
tāss=tōll [star fire] MILK WAY
rūv’nt tückš [iron road] RAILWAY
fīr’mes’ jākkš [thunder.bow] RAINBOW
jērr’ t-ācc’ [flank-bone] RIB
vuelj-čālp’ [head-cap] SKULL
kied’d’most [stone bridge] STONE BRIDGE
saxar kērbā [sugar cane] SUGAR CANE
jāl’k-čiehp [foot-toe] TOE
vēr-rūn’ [blood-thread] VEIN OR ARTERY
vinn’ mūr’ [vine-tree] VINE
līess-kāll’ es’ [widow-old_man] WIDOWER
kīdt-lālp’ [handsole] WRIST

Mod.Head.Dim (3)
jēb’ es’ all’ k.a [horse.son.DIM] FOAL OR COLT
kouss.a all’ k’a [goat.DIM.son.DIM] KID
lānm’ pes’ all’ k’a [sheep.son.DIM] LAMB

Mod.Attr.Head (3)
jīnc. es’ pierrk [morning.ATTR.meal] BREAKFAST
mājiʃ’.es’ liʃʃkm [milk.ATTR cow] DAIRY COW
jēk’ es’ pierrk [evening.ATTR.meal] SUPPER

Mod.Gen.Head (3)
līl’ nūxcem’ [fire.GEN.tongue] FLAME
masiša nūm’ er [car.GEN number] LICENSE PLATE
oye s sājjm [spider.GEN.net] SPIDER WEB

Yeniseian

Ket (50): 70W / 29B / 4C

Mod Head (2)
Alba käy [Alba hunting_trail] MILK WAY
ho q.d.oks bu’ŋ [excrement.POSS.stick place] TOILET

Mod.Adjz Head (1)
sol tu təq.ol [gold.ADJZ finger.covering] GOLD RING

Mod.Gen Head (12)
more.d bo’k [sea.GEN fire] ARCTIC LIGHTS
dēs.d’a’d [eye.GEN.bone] EYEBROW
dēs.d qār [eye.GEN hair] EYELASH
dēs.d ipol [eye.GEN skin] EYELID
bo’k.d.ot [fire.GEN.??] FIREPLACE
Dōy da ko’ [Doh.GEN trail] MILK WAY
Alba da käy [Alba.GEN hunting_trail] MILK WAY
ma’m.d kūb [breast.GEN.tap] NIPPLE OR TEAT
obm.d qūk [nose.GEN hole] NOSTRIL
sēs.d ba’ŋ [river.GEN land] SHORE
elim.d unay [spider.GEN.net] SPIDER WEB
dēs.d il’ [eye.GEN.water] TEAR

Mod.Head (14)
līk dāl’ [male.child] BOY
iil.tāb [arm.ring] BRACELET
okde kāb [ear.point/end] EARLOBE
bul sej [leg.place] FOOTPRINT
bul qūk [leg.way] FOOTPRINT
qūn dāl’ [woman.child] GIRL
hāj kōn [female.horse] MARE
qūm am [grandmother,mother] MOTHER-IN-LAW (OF A MAN)
qun qiv [khan.woman] QUEEN
qub qo’ [thunder.path] RAINBOW
dō.n’ol’ [brain.covering] SKULL
se’n qua’s [reindeer.PL.tent] STABLE OR STALL
sāł’əŋ [blood.rop] VEIN OR ARTERY
l’a’n’ə’ [wrist.bone] WRIST
### Oceania/SE Asia

#### (Austronesian) Aslian

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Height</th>
<th>Bust</th>
<th>Hip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceq Wong (18)</td>
<td>38W / 2B / 1C</td>
<td>158</td>
<td>86</td>
<td>70</td>
</tr>
</tbody>
</table>

#### (Austronesian) Formosan

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Height</th>
<th>Bust</th>
<th>Hip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puyuma (77)</td>
<td>54W / 3B / 3C</td>
<td>162</td>
<td>89</td>
<td>75</td>
</tr>
</tbody>
</table>

#### (Austronesian) Greater Barito

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Height</th>
<th>Bust</th>
<th>Hip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malagasy (75)</td>
<td>89W / 58B / 7C</td>
<td>173</td>
<td>89</td>
<td>76</td>
</tr>
</tbody>
</table>

---

### Appendix E: Binominal data set

<table>
<thead>
<tr>
<th>Class</th>
<th>Datatype</th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree</td>
<td>trunk</td>
<td>García</td>
<td>García</td>
</tr>
<tr>
<td>Vine</td>
<td>grape</td>
<td>García</td>
<td>García</td>
</tr>
<tr>
<td>Foot</td>
<td>hand</td>
<td>García</td>
<td>García</td>
</tr>
<tr>
<td>Wrist</td>
<td>García</td>
<td>García</td>
<td>García</td>
</tr>
</tbody>
</table>

---

2020-05-26 Submitted version 471
The typology and semantics of binominal lexemes

Tagalog (94): 92W / 36B / 9C
AGT.RED.Base (2)
mag.su.saka [AGT.RED.farming] FARMER
mang ing isda [AGT.RED.fish] FISHERMAN

Head Mod (19)
panghangin ng bisikletu [wind? LK bicycle] BICYCLE PUMP
bata.ng lakal [child.LK male] BOY
tulapug ng mata [eyelid LK eye] EYELID
bakas ng paa [remains LK foot] FOOTPRINT
bata.ng babae [child.LK female] GIRL
bata.ng kambing [child.LK goat] KID
bata.ng tupa [child.LK sheep] LAMB
babai.ng kahayo [female.LK horse] MARE
pulo.ng-bituin [island.LK star] MILKY WAY
bihaya.ng babae [parent.in_law.LK female] MOTHER-IN-LAW

butas ng ilong [hole LK nose] NOSTRIL
palad ng kamay [palm LK hand] PALM OF HAND
manggagawa ng palayok [worker LK pot] POTTER
laan.ng-bukal [road.LK-iron] RAILWAY
tabi.ng-dagat [edge.LK-sea] SHORE
tulay na bato [bridge LK stone] STONE BRIDGE
daliri ng paa [digit LK foot] TOE
bomba ng tubig [pump LK water] WATER PUMP
pula ng ilog [red LK egg] YOLK

Head LOC Mod (2)
salamin sa mata [mirror LOC eye] SPECTACLES/GLASSES
sipilyo sa nggin [brush LOC tooth] TOOTHBRUSH

Head Mod (5)
bahay-gaganba [house-spider] SPIDER WEB
bato tulay [stone bridge] STONE BRIDGE

Head Mod (1)
manggagawa [worker] SORCERER OR WITCH

Mod Head (1)
kamay preno [hand brake] HAND BRAKE

Mod LK Head (2)
ginto.ng sising [gold LK ring] GOLD RING
kasal na babae [wedding LK woman] MARRIED WOMAN

(Austronesian) Malayo-Sumbawan

Indonesian (42): 108W / 50B / 4C
AGT.Base (1)
peng.sihir [AGT.sorcery] SORCERER OR WITCH
AGT.Head Mod (1)
pem.buluh darah [AGT.bamboo blood] VEIN OR ARTERY

Head Mod (47)
mata kaki [eye foot] ANKLE
tukang besi [craftsman iron] BLACKSMITH
tukang kaya [craftsman wood] CARPENTER
kepala adat [head custom] CHIEFTAIN
kepala suku [head tribe] CHIEFTAIN
ayam jago [chicken rooster] COCK/ROOSTER
tulong selangka [bone collarbone] COLLARBONE
tiang pintu [post door] DOORPOST
tahi kaping [excrement ear] EARWAX
kotoran telinga [dirtiness ear] EARWAX
alis (mata) [eyebrow eye] EYEBROW
bulu mata [body_hair eye] EYELASH
kelepak mata [sheath eye] EYELID
tali kai [cord hook] FISHING LINE
lidah api [tongue fire] FLAME
nyala [API] [flame fire] FLAME
sarung tangan [cover hand] GLOVE
sapu tangan [broom hand] HANDKERCHIEF OR RAG
kain lap [cloth rag] HANDKERCHIEF OR RAG
tuan rumah [master house] HOST
plat (mobili) [plate car] LICENSE PLATE
kotak pos [box post] MAIL BOX
tengah hari [middle day] MIDDAY
tanah-air [land-water] NATIVE COUNTRY
labang hidung [hole nose] NOSTRIL
rongga hidung [cavity nose] NOSTRIL
tapak/kelepak tangan [palm hand] PALM OF HAND
karto pos [card mail] POSTCARD
tukang tembikai [craftsman earthenware] POTTER
jalan kereta api [road carriage fire] RAILWAY
tukang sibir [craftsman magic] SORCERER OR WITCH
kaca mata [glass eye] SPECTACLES/GLASSES

dayak belakang [back] SPINE

(Austronesian) Greater Central Philippine

vâtana.bâzo [trunk/body.tree/wood] TREE TRUNK
vôa.lôbôka [fruit.grape] VINE
hâtôka.tâmûna [neck.hand] WRIST

Head.PER.Mod (18)
fêh.y.n.tâmûna [tying/knot.PER.hand] BRACELET
tuola.m.parâv [bone.PER.charm] COLLARBONE
tâi.n.sôfîna [dirt shit.PER.ear] EARWAX
tûd.y.n.frì thu [line/rope.Fishhook] FISHING LINE
dia.n.tôngétta [print.PER.foot] FOOTPRINT
gu.n.tâmûna [glove.PER.hand] GLOVE
tampo.n.trão [owner/lord.PER.house] HOST
nimero.n.automobîla [number.PER.car] LICENSE PLATE
tûnî.y.n.ràçâna [land.PER.ancestor] NATIVE COUNTRY
lala.m.by [road.PER.iron] RAILWAY
âtms.bi.n.andriaminântra [knife.big.PER.God] RAINBOW
tuiltona.n.tèhezànâa [bone.PER.rib_side] RIB
kàvàna.n.lôbu [shell/coral.PER.head] SKULL
tràño.n.bàl[a [house.PER.spider] SPIDER WEB
hâzo.n.lâmòsina [wood/tree.PER.back] SPINE
tràño.n.ômybî [house.PER.bovine] STABLE OR STALL
fiara.n.latamby [vehicle.PER.railway] TRAIN
lîlâna.n.rra [road/path.PER.blood] VEIN OR ARTERY

NMILZ.Base (1)
fl.ôhu [NMILZ.head] CHIEFTAIN

NMILZ.Base.CIRC (1)
fa.mosavi.anu [NMILZ.witchcraft.CIRC] MAGIC

fa.sangkap.an [CIRC.belongings.CIRC] TOOL

Head LK Mod (19)
panghangin ng bisikletu [wind? LK bicycle] BICYCLE PUMP
bata.ng lakal [child.LK male] BOY
tulapug ng mata [eyelid LK eye] EYELID
bakas ng paa [remains LK foot] FOOTPRINT
bata.ng babae [child.LK female] GIRL
bata.ng kambing [child.LK goat] KID
bata.ng tupa [child.LK sheep] LAMB
babai.ng kahayo [female.LK horse] MARE
pulo.ng-bituin [island.LK star] MILKY WAY
bihaya.ng babae [parent.in_law.LK female] MOTHER-IN-LAW

butas ng ilong [hole LK nose] NOSTRIL
palad ng kamay [palm LK hand] PALM OF HAND
manggagawa ng palayok [worker LK pot] POTTER
laan.ng-bukal [road.LK-iron] RAILWAY
tabi.ng-dagat [edge.LK-sea] SHORE
tulay na bato [bridge LK stone] STONE BRIDGE
daliri ng paa [digit LK foot] TOE
bomba ng tubig [pump LK water] WATER PUMP
pula ng ilog [red LK egg] YOLK

Head LOC Mod (2)
salamin sa mata [mirror LOC eye] SPECTACLES/GLASSES
sipilyo sa nggin [brush LOC tooth] TOOTHBRUSH

Head Mod (5)
bahay-gaganba [house-spider] SPIDER WEB
bato tulay [stone bridge] STONE BRIDGE

Head Mod (1)
manggagawa [worker] SORCERER OR WITCH

Mod Head (1)
kamay preno [hand brake] HAND BRAKE

Mod LK Head (2)
ginto.ng sising [gold LK ring] GOLD RING
kasal na babae [wedding LK woman] MARRIED WOMAN

(Loc) Base.CIRC (1)

per.api.an [LOC.fire.CIRC] FIREPLACE
The typology and semantics of binominal lexemes

NG/Australia

Gunwinyguan

Anindilyakwa (4): 47W / 16B / 6C

CLF.INAL.Head (Mod) (1)
n.env.m.allhoka [3SG:M:INAL:foot] FOOTPRINT

CLF.AL.Mod (Head) (2)
pump
e.ev.ng.m.enba [CLF(neut).M:AL:CLF(veg).eye] SPECTACLES/GLASSES

CLF.INAL.Head (Mod) (2)
a.mv.regina [CLF(neut).INAL:joint] WRIST

CLF.INAL.Head Mod (9)
ma.ma.yama.m.enba [CLF(veg).INAL:body_hair]
CLF(neut).eye) EYEBROW
ma.ma.yama.m.enba [CLF(veg).INAL:body_hair]
CLF(veg).eye) EYELASH
ma.ma.ukulya.m.enba [CLF(veg).skin:CLF(veg).eye]
EYELID
a.ma.muwaa.nge.nema [CLF(neut).INAL:round
CLF(neut).breast) PALM OF HAND

Head Mod (4)

Head Mod:REDUP (1)
a.ma.talimala [leg/foot branch:REDUP] TOE

Mod.Head (1)
bili.woi [spirit?.skin] HANDKERCHIEF OR RAG

Warta Thuntai (30): 31W / 16B / 3C

Head Mod (1)
thent.thebra [half day] MIDDAY

Mod HEAD (8)
wongan.nangai [person child] BOY
si.batu [eye ??] EYEBROW
si.yowi [eye hair] EYELASH
si.pampam [eye leaf:REDUP] EYELID
men kar [fire place] FIREPLACE
men.thathamin [fire ??] FLAME
 tambi.kapkap [hand foot] PALM OF HAND
karipa.rrorki [river bank] SHORE

Mod.Head (7)
konjo.men [oven:house] COOKHOUSE
charve.per [door:tree] DOORPOST
them.yew [nose:hole] NOSTRIL
tha.bar [rib:bone] RIB
ndaka.nyan [magic:man] SORCERER OR WITCH
nggambu.chuk [head:digit] THUMB
weh.men [poo:house] TOILET

Nuclear Torricelli

Srenge (59): 40W / 20B / 3C

Mod Head (18)
ala.yolta [leg/foot eye] ANKLE
bilibina.wusoro [bee testicles] BEEHIVE
ape.aka [stomach:mouth] BREAKFAST
karkara.buna [child male_non_human] COCK/ROOSTER
ape.aka [stomach:mouth] DINNER
yungolo.pio [ear:crecenta] EARWAX
yolta.wals [eye hair:fur/feather] EYEBROW
yolta.io [eye skin] EYELID
ala.mleko [leg/foot scar/scab] FOOTPRINT
srenge.suka [child:water] GIRL
ape.aka [stomach:mouth] LUNCH
srenge.suka [child:water] MARRIED WOMAN
nupa.teng [nose:hole] NOSTRIL
wando.bningi [head:bone] SKULL
billa.wiro [spirit initiated:man] SORCERER OR WITCH
ape.aka [stomach:mouth] SUPPER
nanga.yotim [hand:thumb/big:toe] THUMB
nanga.awara [hand:knot] WRIST

Mod.Head:REDUP (1)
ala.talimala [leg/foot branch:REDUP] TOE

Mod.Head (1)
bili.woi [spirit?.skin] HANDKERCHIEF OR RAG

Walman (100): 48W / 18B / 1C

Mod Head (18)
kayal.chkuel [foot eye] ANKLE
ngel.kranoo.mukan [chicken male] COCK/ROOSTER
oputo.nyuku.eu [thing:food] DINNER
mkuel.kal [ear:leaf] EARLOBE
mkuel.nkyar [ear:wax] EARWAX
chkuel.okul [eye fur] EYEBROW
chkuel.nukul [eye:small_piece] EYELASH
chkuel.nukul [eye:small_piece] EYELID
oputo.nyuku.eu [thing:food] LUNCH
mamul.roune [parent_of_wife old:human] MOTHER-IN-LAW
(OF A MAN)
nyamayki.ra [nose:hole] NOSTRIL
nganu.oto [sun:flat_leaf] RAINBOW
woruen.annungko [head:bone] SKULL
oputo.nyuku.eu [thing:food] SUPPER
motu.chuto [finger:digit female] THUMB
kayal.motu [foot:finger/digit] TOE
wi.pariki [hand:bent_part?] WRIST
ngal.meten.oltun [(bird seed/egg) ??] YOLK

(Pama-Nyungan) Desert Nyungic

Gurindji (32): 53W / 8B / 5C

Base.INS (32)
warlu.waji [fire:INS] FIREPLACE

Base.SUF (1)
pirrit.yi [side:SUF] RIB

Base.VBL.Z.AGT (1)
wayarrurn.karra.aji [fishing_line.ACT.AGT] FISHERMAN

Mod Head (4)
jitji.jarriny [nose:hole] NOSTRIL
wartarn.majuu [hand:stomach] PALM OF HAND
wartarn.nganay [hand:mother] THUMB
jimana.nganay [foot:mother] TOE

Mod.DAT Head (1)
yawarra.wu.marra [horse:DAT house] stable OR STALL
(Pama-Nyungan) Paman

Wik-Mungkan (102): 44W / 25B / 3C

Head Mod (2)
puk wuut [child old_man] BOY
aak puul [place fathers_father] NATIVE COUNTRY

Head Mod.DAT (1)
yuk mee.čakana [tree[thing eye].DAT] SPECTACLES/GLASSES

Mod Head (22)
tha’ kont [foot prawn] ANKLE
man kaunch [neck bone] COLLABONET
kon maq [ear stickiness] EARWAX
mee’ paam [eye palm] EYEBROW
mee’ pench [eye body_hair] EYELASH
thum nganth [fire light] FLAME
tha’ pemp [foot track] FOOTPRINT
puk wanch [child woman] GIRL
piny-kench [fathers_sibling-mother_in-law] MOTHER-IN-LAW
(OF A MAN)
thuat mee’ [breast eye] NIPPLE OR TEAT
kaa’ uuyan [eye.?] EYEBROW
ma’ un [hand front] PALM OF HAND
ma’ paam [hand palm] PALM OF HAND
ngangk ek [heart shell] SHOULDERBLADE
kuchek kaunch [head bone] SKULL
ma’ mangk [hand small_of_back] SORCERER OR WITCH
pikpungk kaunch [back bone] SPINE
mee’ kam [eye juice] TEAR
tha’ puq [foot child] TOE
yuk think [tree lower_back] TREE TRUNK
ma’ kongk [hand nuts] WRIST
ma’ ochangan [hand round_shell] WRIST

Walió

Tuwarí (98): 21W / 8B / 1C

Mod.Head (8)
hele.tiile [leg.joint] ANKLE
nomo.čanese [eye.?] EYEBROW
nomof.e [eyelid.hair] EYELASH
nomo.fe [eye.?] EYELID
saaluo.hon [nosehole] NOSTRIL
u.pase [fresh_water.sand] SHORE
hele.ne [leg.finger] TOE
iane.pe [arm.?] WRIST

West Bomberai

Kalamang (51): 90W / 57B / 6C

Head Mod (16)
pompa sepedu [pump bike] BICYCLE PUMP
tuman caunam [child male] BOY
tukan ror [worker wood] CARPENTER
kepala suku [head tribe] CHEFTEAN
kokok caunam [chicken male] COCK/ROOSTER
sontum amdir [person garden] FARMER
tuman pas [child female] GIRL
tanggarargar maraun [ring gold] GOLD RING
muap yuom:nawa:ririn [food midday] LUNCH
laharang pas [horse face] MARE
ketan.un pas [inlaw.1SG.POSS female] MOTHER-IN-LAW (OF A MAN)
leit pas [king female] QUEEN
uriap kereta api [street carriage fire] RAILWAY
jarir puseer [web spider] SPIDER WEB
los (wame) var [bridge PROX stone] STONE BRIDGE
pompa air [pump water] WATER PUMP

Head Mod.POSS (6)
muap goosaun.kin [food evening.POSS] DINNER

nka war.kin [line fish.POSS] FISHING LINE
tas tan.kin [bag hand.POSS] HANDBAG
sapu tan.kin [broom hand.POSS] HANDKERCHIEF OR RAG
kotak surat.kin [box letter.POSS] MAIL BOX
muap goosaun.kin [food evening.POSS] DINNER

Mod Head.3POSS (14)
akaar kang.un [chest bone.3POSS] COLLABONET
unggas ror.un [door wood.3POSS] DOORPOST
kelkam elao.un [ear under.3POSS] EARLOBE
kolkan kit.un [ear ear.3POSS] EARWAX
kanggir pul.un [eye skin.3POSS] EYEBROW
kanggir pul.un [eye skin.3POSS] EYELID
din tompat.un [fire place.3POSS] FIREPLACE
kuda tumun.un [horse child.3POSS] FOAL OR COLT
kewe tura.un [house old.person.3POSS] HOST
domba tumun.un [sheep child.3POSS] LAMB
mobil pelat.un [car plate.3POSS] LICENSE PLATE
am bel.un [breast .3POSS] NIPPLE OR TEAT
bustang pos.un [nose hole.3POSS] NOSTRIL
tan el.un [hand under.3POSS] PALM OF HAND

Mod.Head (8)
kor.kom [foot.?] FOOTPRINT
kir.kang [side.bone] RIB
os.ker [sand.above.?] SHORE
bekiem.kang [shoulder.bone] SHOULDERBLADE
som.kang-kang [person.bone.—RED] SKULL
suol.kang [back bone] SPINE
peri.tam [water.drops] TEAR
kor.parok [foot.?] TOE

North America

Athanaskan-Eyak-Tlingit

Navajo (66): 77W / 27B / 5C

3SG.Mod.Head (1)
‘a.ké.ts’iin [3SG.foot.bone] ANKLE

Mod 3SG.Head (9)
tsis’ná bi.ghan [bee 3SG.home] BEEHIVE
tsis’ná bi.jeel [bee 3SG.resin] BEESWAX
náa’ohóóhai bi.ká’ii [chicken 3SG.male.NFE] COCK/ROOSTER
lóó’ bi_hahdleehe bi.tl’óól [fishing_rod 3SG.rope] FISHING LINE
béeso bi.zís [money 3SG.bag] HANDBAG
béesh bi.tiin [iron 3SG.path] RAILWAY
na’ashí’í bi.tl’óól [spider 3SG.string] SPIDER WEB
lí’ bi.ghan [horse 3SG.home] STABLE OR STALL
tsin bi.isin [tree 3SG.handle] TREE TRUNK

Mod INS Head (1)
tsé bee na’ní’á [stone with_it bridge] STONE BRIDGE
The typology and semantics of binominal lexemes

Mod.Head (15)

Kekch (49): 101W / 40B / 14C

3ERG.Head M Mod (1)
x.kem aj om [3ERG.web M spider] SPIDER WEB

3ERG.Head Mod (20)
x.map oq [3ERG.palm foot] ANKLE
r.oochok kab’ [3ERG.house bee] BEEHIVE
x.che’, el kab’ [3ERG.tree.DER bee] BEEHIVE
x.na’aj xam [3ERG.place fire] COOKHOUSE
x.k’ot sîk [3ERG.excrement ear] EARWAX
r.ismat u [3ERG.hair surface] EYELASH
r.îx u [3ERG.back surface] EYELID
r.na’aj xam [3ERG.place fire] FIREPLACE
r.al kawaw [3ERG.offspring_of_female horse] FOAL OR COLT
x.tu’ u [3ERG.covering hand] GLOVE
r.al yu’k [3ERG.offspring goat] KID
x.na’ kawaw [3ERG.mother horse] MARE
x.rab’in itz’in [3ERG.daughter_of_man younger_brother] NIECE
r.al chaq’ na’ [3ERG.son mature.mother] NIECE
x.rab’ in kax [3ERG.daughter_of_man elder_brother] NIECE
r.alal ikan [3ERG.offspring_of_man uncle] NIECE
r.u’uj tu’ [3ERG.point/nose breast] NIPPLE OR TEAT
x.na’aj wakax [3ERG.place cow] STABLE OR STALL
x.ya’al u [3ERG.liquid surface] TEAR
x.u’uj oq [3ERG.place nose] TOE
3ERG.Head Mod.DER (1)
x.na’ isaq.îl [3ERG.mother woman.DER] MOTHER-IN-LAW
(OF A MAN)
3ERG.Head.SUF Mod (1)
x.q’an.al mol [3ERG.yellow.SUF egg] YOLK
AGT Base.AGT (1)
aj k’al om [AGT cornfield.AGT] FARMER
Base.ANTIP.NMLZ (1)
awas.in.el [secret.ANTIP.NMLZ] SORCERER OR WITCH
Base.DER (1)
jolom.îl [head.DER] CHIEFTAIN
Base.INS.NMLZ (1)
k’ub’.leb’aal [hearthstone.INS.NMLZ] COOKHOUSE
Base.NMLZ (1)
k’oteb’aal [shit.NMLZ] TOILET

Head Mod (8)
kux oq [joint leg] ANKLE
b’aqal ch’îch’ [corncob iron] BICYCLE
tzo’ kaxlan [male chicken] COCK/ROOSTER
sa’u’u’ [belly nose] NOSTRIL
sa’uq’ [belly hand] PALM OF HAND
starjet postal [card postal] POSTCARD
na’uq’m [mother hand] THUMB
kux uq’m [neck hand] WRIST

Head.NPOS 3ERG.Mod (1)
b’uq.el x.b’een tel [bone.NPOS 3ERG.upper_arm]
SHOULDERBLADE

Head.UNPOSS Mod (1)
b’uq.el jolom [bone.UNPOSS head] SKULL

Mod Head (1)
assuik rts’ajl [sugar?] SUGAR CANE

Mod.ADJ Head (1)
k’im.al kab’l [straw.ADJ house] THATCH

Zinacantán Tzotzil (99): 64W / 29B / 3C

Head Mod (12)
kachimpa pom [pipe incense] BEE

Submitted version 2020-05-26
Chanul pom [bug incense] BEE
ton pom [stone incense] BEE
mishik' pom [bellly_button incense] BEE
na pom [house incense] BEEHIVE
sera 'ruha-pom [wor honey-incense] BEESWAX
'at k'ok' [penis fire] FLAME
'ot k'ak' [middle sun/day] MIDDAY
be sin [road mucous] NOSTRIL
me' k'inbal [mother drizzle] RAINBOW
hol 'amimu [head soul] SKULL
nen sar [mirror eye] SPECTACLES/GLASSES

Head Mod.INAL (15)
nuk' 't.o.k. [neck foot.INAL] ANKLE
'at bal bag. [chaplet/rosary bone.INAL] COLLARBONE
tak' 'in chik'in. [metal ear.INAL] EARRING
shik' sat'il [wing/fin eye.INAL] EYELASH
'au 't.o.k. [jag foot.INAL] FOOTPRINT
na k'o. b.o. [house hand.INAL] GLOVE
ni' chu 't.a [nose breast.INAL] NIPPLE OR TEAT
pachomal k'o. b.o. [basket hand.INAL] PALM OF HAND
'at k'o. b.o. [inside hand.INAL] PALM OF HAND
ch'il. te'il. [wooden_frame.tree.INAL] KIB
ch'ich pat'il. [thorn/spine/wound] BACK.INAL
spine
me' k'o. b.o. [mother hand.INAL] THUMB
ni' t.o.k. [nose foot.INAL] TOE
be ch'il ch'il. [road bleed.INAL] VEIN OR ARTERY
nuk' 't.o.k. [neck hand.INAL] WRIST

Mod Head (2)
shokan na [side house] NEIGHBOUR
me' anal vinik [poverty/misery/grief man] WIDOWER

( Otomanguean Otopamean )

Quer 退 taro Otomi (73): 134W / 35B / 4C

Head DEF Mod (12)
sera kolmenä [wax bee] BEESWAX
ndo'yo tihu [bone breast] COLLARBONE
'mai goxhi [post/pole door] DOORPOST
t'yu t'oda [little silly] KID
ky t'udaxi [little nanny] KID
lo'lo' t'unxi [little nanny] KID
made mpaa [middle day] MIDDAY
'bebi' n'dä [woman king] QUEEN
'het'e n'hojo [sorcerer/witch man] SORCERER OR WITCH
'het'e bebi' [sorcerer/witch woman] SORCERER OR WITCH
'me'le paxa [roof straw] THATCH
gna nua na [house bath] TOILET

Head.Mod (21)
fan'i bjüa [horse.iron] BICYCLE
dede.oni [cock.chicken] COCK/ROOSTER
ndai.igua [beginning/base/ear] EARLOBE
xi 'da [body_hair.eye] EYELASH
xi 'da [body_hair.eye] EYELID
nt'o. 'ye [sheath hand] GLOVE
tsu. fin [female.horse] MARE
do. ba [stone.milk] NIPPLE OR TEAT
ok'xi. ulha [hole.nose] NOSTRIL
nengi.dehe [bank/shore/water] SHORE
bänj. dehe [bank/shore/edge.water] SHORE
tsu. sene [female.wizard/sorcerer] SORCERER OR WITCH
ndo. sene [male.sorcerer/witch] SORCERER OR WITCH
hh. deq [mirror.eye] SPECTACLES/GLASSES
na. myeh [??/spider] SPIDER WEB
hu. xathu [path/road.back] SPINE
gnu. yo [house.livestock] STABLE OR STALL
gnu. meji [house.animal/cattle] STABLE OR STALL
ngi. dy [juice.eye] TEAR

wa. caa [foot.tree] TREE TRUNK
'ënu. ji [path/road.blood] VEIN OR ARTERY

Head Head (1)
do. yono [stone.bowl] SKULL

Seri

Seri (84): 61W / 16B / 6C

ABS.Mod.Poss.Head (1)
ha. nut. teen [ABS.breast 3POSS.opening/mouth] NIPPLE OR TEAT

Head Head (4)
zizpussil cet [child male] BOY
toottar cet [chicken male] COCK/ROOSTER
cuyu caaun [horse female] MARE
dzhehe caaun [chief_person female] QUEEN

Mod Head (1)
sepe poof [sea line] FISHING LINE

Mod Poss.Head (3)
pe. na. me [3POSS.bee 3POSS.abebe] BEEHIVE
sepe. teel [sea 3POSS.edge] SHORE
hehe. i. [plant 3POSS.base] TREE TRUNK

POSS.Mod.Head (2)
'to. cauns [3POSS.eye.??] EYEBROW
a. queect caaun [3POSS.parent_in_law female] MOTHER-IN-LAW (OF A MAN)

POSS.Mod.Poss.Head (5)
'to. u. i. pq [3POSS.eye.PL 3POSS.??] EYELASH
'to. u. naa [3POSS.eye 3POSS.skin] EYELID
'to. u. pq [3POSS.eye 3POSS.back] EYELID
'to. i. u. [3POSS.head 3POSS.bone] SKULL
yanao. i. [3POSS.foot 3POSS.base] WRIST

Siouan

Ho-Chunk (103): 61W / 20B / 1C

Head Mod (20)
si. kaq [foot.sinew] ANKLE
bijinec huætæk [shoulder bone] COLLARBONE
cexa hj [?? hair] EYEBROW
hijaasu hj [eye hair] EYELASH
hijaasu hj [face/eye skin] EYELID
šju, xetæ, nj [horse.DIM] FOAL OR COLT
hinya, jk [woman.DIM] GIRL
cuasga, njk [sheep.DIM] LAMB
šju, xetæ hiy [horse woman] MARE
hæpp kisæk [day half] MIDDAY
waas. paa [breast.nose/head] NIPPLE OR TEAT
puq̱æ huq̱æ [nose/snot hole] NOSTRIL
nuq̱æ huq̱æ [metal road] RAILWAY
nuq̱æ huætæ [head bone] SKULL
nuq̱æ hi̱jaasu [metal eye] SPECTACLES/GLASSES
nuq̱æ huætæ [back bone] SPINE
haramfæ hæpp [cross day] SUNDAY
hoxjæ wær [evening meal] SUPPER
hi̱ja nj [Face water] TEAR
peeæ waa [fire boat] TRAIN

(Tupian) Tupi-Guarani

MbyGU Guarani 원 (33): 61W / 33B / 7C

3.Mod ABL.Head (1)
h. exa re. gua [3.eye ABL.NMLZ] SPECTACLES/GLASSES

3.Mod Head (4)
h. wisa kunha [3.leader woman] QUEEN
h. exa. y [3.eye.water] TEAR
The typology and semantics of binominal lexemes

South America

Araucanian

Mapudungun (6): 81W / 25B / 6C

Base.LOC (1)
ku'tral.we [fire.LOC] FIREPLACE

Base.NML.Z (2)
ruka.ve [house.NMLZ] CARPENTER
challwa.ve [fish.NMLZ] FISHERMAN

Head Mod (14)
we.chen wentr [new.person man] BOY
lafikétr pilig [sea ear] EARLOBE
férá pilun [dirt ear] EARWAX
ngen ruka [owner house] HOST
zomo kawella [female horse] MARE
rangü antí [middle day] MIDDAY
longkó moyo [head breast] NIPPLE OR TEAT
wechou ya [hole nose] NOSTRIL
pílay kuwi [palm hand] PALM OF HAND
fítra changull kuwi [big finger/toe hand] THUMB
changull jamug [finger/toe foot] TOE
chill mollfi [branched blood] VEIN OR ARTERY
troy kuwi [joint hand] WRIST
choy kuraam [yellow egg] YOLK

Mod Head (3)
malle nawe [father's brother daughter] NIECE
afas foki [grape plant] VINE
jantu wenteru [widow man] WIDOWER

Mod.ESS.NONF Head (1)
fútš.nge.n zomo [husband.ESS.NONF woman] MARRIED

Mod.Head (4)
milla.íwíl.kuwi [gold.surround hand] GOLD RING
kuwi,pírínka [hand.brake hand] HAND BRAKE
tren.ripú [train.way] RAILWAY
kura.kuykuy [stone.bridge] STONE BRIDGE

Yaqi (106): 92W / 20B / 6C

Base.LOC (1)
sísí 'woo.chí [iron.LOC] TOOL

Base.POSS.LIG.NML.Z (1)
jo'a.k.a.me [home.POSS.LE.NMLZ] HOST

DIM Base (2)
ili kaba'i [DIM horse] FOAL OR Colt
ili jamut [DIM woman] GIRL

Mod Head (12)
mama ma 'urá [bee house] BEESWAX
pueta moyoa [door edge] DOORPOST
naka bwiw [ear secretion] EARWAX
kaba i jamut [horse female] MARE
bwiw tooos [hand nest] NATIVE.COUNTRY
yeka wojo orú [nose hole] NOSTRIL
bave moyoa [sea edge] SHORE
koba ota [head bone] SKULL
joo'o ota [back bone] SPINE
bwe'wokk wusia [big feet.finger] TOE
ojo wìi'i [blood thread] VEIN OR ARTERY
kaba jiapsi [egg heart] YOLK

Mod Head.APPL (1)
mam betala.riam [hand plain.APPL] PALM OF HAND

Mod.Head (3)
mam.booosam [hands.bags] GLOVE
pipim.koba [breast.head] NIPPLE OR TEAT
mam.pusiam [hand.finger] THUMB

(UTO-Aztec) Southern UTO-Aztecan

Trinitario (96): 88W / 25B / 10C

3SG.Head Mod (3)
ta.choe.ro.pi.choka [3NH.lip.DER.CLF(rope) ear] EARLOBE
ita.muk(i).i -chene [3NH.mouth.CLF(round) -breast] NIPPLE
or TEAT
ita.chípamo -yeeruupa [3NH.box tool] TOOLBOX

3SG.Head.Mod (1)
ta.sii.peno [3NH.nose.house] FIREPLACE

Base.AGT (1)
tyuraj(i).eru [med.AGT] POTTER

Base.CLF (9)
mopo.si [bee_related.CLF(round)] BEE
mopo.ji [bee_related.CLF(shapeless)] BEESWAX
-uqi.mo [eye.CLF(fabric)] EYELID
e(s)an(e).ti.m(o)re [field.NPOS.CLF(FAN)] FARMER
ta.pi.gi [3NH.point(?).CLF(trunk)] FLAME
-muí.pewú [clothes.CLF(hand)] GLOVE
-jek.pevú [interior.CLF(hand)] PALM OF HAND
-mits.gí [3NH.CLF(trunk)] SPINE
-it(e).ne.re.pi [blood.PSD.NZR?.CLF(snake)] VEIN OR ARTERY

Base.DER.CLF (3)
tapaj.ro.gí [door.DER.CLF(trunk)] DOORPOST
-lyped.re.ku [foot.DER.CLF(inside)] FOOTPRINT
-pow.ro.cho [wing.DER.CLF(board)] SHOULDERBLADE

Base.DIM (1)
kwoy.gíra [horse.DIM] FOAL OR Colt
DET.Head.Mod (1)
-sira.giho [blockage.ear] EARWAX

Head Mod (3)
tig.wa.re oro [?.finger.NPOS gold] GOLD RING
chiwa.gira 'moyo [goat.DIM child] KID
'moyo 'yesa [child sheep] LAMB

Head.Mod (1)
-jivo.cho.no'I [hair.CLF(board).front] EYEBROW

Mod.Head (2)
-siri.penO [nose.house] NOSTRIL
smatu.penO [spider.house] SPIDER WEB

(Cariban) Guianan

Galibi Carib (12): 65W / 20B / 7C
Base.DEV (1)
kamisa.npo [piece_of_tissue.DEV] HANDKERCHIEF OR RAG

Base.DIM (3)
kawale.menpo [horse.DIM] FOAL OR COLT
kapitilia.menpo [goat.DIM] KID
kapala.menpo [sheep.DIM] LAMB

Mod Head (6)
wana wetI [bee excrement] BEESWAX
wato aponI [fire place] FIREPLACE
ime norI [son grandmother] MOTHER-IN-LAW (OF A MAN)
moyowai etaweni [spider web] SPIDER WEB
aina yumI [hand father] THUMB
weve untI [tree base] TREE TRUNK

Mod.Head.POSS (6)
pana sepeI LI [ear.lobe.POSS] EARLOBE
manat potI LI [breast tip.POSS] NIPPLE OR TEAT
aina la.LI [hand flat] PALM OF HAND
upapo kawai.yI [head calabash.POSS] SKULL
pupu sikIli.LI [foot little.end.POSS] TOE
i'mo kani.LI [egg yolk] YOLK

Mod.Head (2)
anekuntI [wrist.base] BRACELET
enulu.pipo [eye.skin] EYELID

Mod.POSS Head (1)
-pana.wetI [ear.POSS dirtiness] EARWAX

Mod.POSS Head.POSS (1)
emo.li sakila.LI [nose.POSS aperture.POSS] NOSTRIL

Chibchan

Cab 退 car (14): 81W / 33B / 2C
Mod Head (26)
lā tālik [foot spur] ANKLE
bukalā ju [bee house] BEEHIVE
bula ñgI [bee excrement] BEEESWAX
iškoro sīgyir [chicken male] COCK/ROOSTER
wā kā [face hair] EYEBROW
wāblā kā [eye hair] EYELASH
wā kūa [eye skin] EYELID
shkgālīnālī [smoke road] FIREPLACE
bikolō kicha [hook liana] FISHING LINE
yōko wā [fire fruit/ball] FLAME
kalwā yaba [horse child] FOAL OR COLT
yaba alākīzI [child woman] GIRL
due yaba [goat child] KID
oveja yaba [sheep child] LAMB
kalwā alākī [horse female] MARE
kāwā mākū [sun half] MIDDAY
jāyī yakI [man mother_in_law] MOTHER-IN-LAW (OF A MAN)
tsa' batI [teat tip] NIPPLE OR TEAT
jula kūI [hand underside] PALM OF HAND
kōhākā ḳālgI [train road] RAILWAY
wā rīā [eye liquid] TEAR
julāskā mīkō [finger grandmother] THUMB
tebēli yakaI [knife mass] TOOL
pi kīchā [blood liana] VEIN OR ARTERY
jula kiū ngā [hand underside] WRIST
iškoro sīgā bala [chicken egg] INTERIOR YOLK

Mod Head.MOD.7 (4)
kuā ḳē [ear excrement] SPECKLE
bak chichē [shoulder bone] SHOULDERBLADE
tāgū chichē [head bone] SKULL
kalwā juI [horse head] STABLE OR STALL
hā juI [excrement house] TOILET
kal wākūchī [tree trunk] TREE TRUNK
ava kalI [grape tree] VINE

Harakmbut

Harakmbut (3): 76W / 28B / 27C
Base.CLF (2)
aymōro.po [honey.CLF(round)] BEE
sērā.po [honey.CLF(round)] BEE

Base.DIM (2)
wambo.sipo [youngster.DIM] BOY
wettonē.sipo [woman.DIM] GIRL

Base.SPAT.CLF (2)
wam.ka.mbe.to.pa [NPOS.forehead.CLF(edge?).SPAT(down).CLF(rod)] EYEBROW
wam.ka.ki.ti.mba [NPOS.eye.CLF(edge?).SPAT(on).CLF(rod)] EYELID

Head.LOC.Mod (1)
no.pō.te.menoe [vital centre.CLF(round).LOC.day] MIDDAY

Mod.Head (18)
wā.i'ko.pe [NPOS.foot.shape_half_month.CLF(disc)] ANKLE
aymōro.ōn [on.peak] BEEHIVE
akvikai.mbag.mpe [door.lip.edge.CLF(disc)] DOORPOST
wam.pe.sodn'I [NPOS.ear.flappy_object] EARLOBE
wam.pe.otkusō [NPOS.ear.viscose_substance] EARWAX
wam.ko.ki.mba.wih [NPOS.eye.CLF(edge?).SPAT(on).CLF(hand)] HAIR EYELASH
wam.ka.kupign [NPOS.breast.tip] NIPPLE OR TEAT
wā.o.wē [NPOS.nose.liquid] NOSTRIL
wam.ka.mbe.nek [NPOS.hand.belly] PALM OF HAND
wam.ka.kumbogn [NPOS.eye.CLF(edge?).SPAT(on).CLF(rod)] DOORPOST
wam.ka.ki.ti.mba [NPOS.eye.CLF(edge?).SPAT(on).CLF(rod)] EYEBROW
wam.ka.ki.ti.mba [NPOS.eye.CLF(edge?).SPAT(on).CLF(rod)] EYEBROW

Mod.SPAT.Head (2)
wam.ko.tu.wē [NPOS.eye.SPAT(base).liquid] TEAR
wam.ka.ta.me(tI) [NPOS.hand.SPAT(base).hump.CLF(round)] WRIST

Huitotoan

Murui Huitoto (41): 48W / 16B / 4C
Base.ANA.CLF (3)
jefo.gi [ear.CLF(animal) ANA.CLF(leather)] EARLOBE
onojI i.ko [hand.ANA.CLF(cover)] GLOVE
The typology and semantics of binominal lexemes

Wichi (65): 82V / 36B / 10C
Head.AGT (2)
tshos.yo [animals.AGT] HERDSMAN
sapatu.wu [shoe.AGT] SHOEAKER

Matacoan

Imbabura Quechua (78): 71W / 18B / 2C
Head Mod (1)
chaupi puncha [half day] MIDDAY

(Quechuan) Quechua II

Ticuna (92): 73W / 20B / 5C
Head.Mod.GEN (1)
ānëruit ořo.ãrã [ring gold.GEN] GOLD RING
Appendix E. Binominal data set

(Pidgins & Creoles) French-based

Seychelles Creole (17): 107W / 21B / 1C

Head Mod (21)
banan lalimyé aritik [PL light arctic] ARCTIC LIGHTS
mous dimynel [fly honey] BEE
kaka zorey [feces ear] EARWAX
tapo lizye [skin eye] EYELID
met lakour [master home] HOST
plak nimo [plate number] LICENSE PLATE
bout tey [end breast] NIPPLE OR TEAT
trous nenén [hole nose] NOSTRIL
plak lanmen [plate hand] PALM OF HAND
kart postal [card postal] POSTCARD
bor-lameron [edge-sea] SHORE
plak zepol [plate shoulder] SHOULDERBLADE
bënonm dibwa [man woods] SORCERER OR WITCH
akaz bib [house spider] SPIDER WEBSHE
lezo leren [bone kidney] SPINE
kolonn vertebraal [string spinal] SPINE
bwa fey [roof straw] THATCH
pous lipye [thumb foot] TOE
ledwa lipye [finger foot] TOE
pye rezen [tree grape] VINE
zonn dizef [yellow egg] YOLK

(Pidgins & Creoles) English-based

Saramaccan (89): 65W / 32B / 4C

Base.AGT (2)
ka:nsa.mta [loincloth.AGT] BOY
ko:sma [skirt.AGT] GIRL

Head PREP Mod (1)
finga u fût [finger of foot] TOE

Mod Head (4)
góutu andél [gold ring] GOLD RING
deiwe pásì [dairy path] MILKY WAY
talán fût [train foot] RAILWAY
tanda boso [tooth brush] TOOTHBRUSH

Mod.Head (25)
hôni.wîso [bee.house] BEEHIVE
hôni.káki [bee.excrement] BEESWAX
wônt.ganu [man.chicken] COCK/ROOSTER
dôs.pàsì [door.tree] DOORPOST
jësi.bùka [ear.mouth] EARLOBE
This appendix contains a summary of the database structures. Each table is shown here in the form of the corresponding data frame in R. The source code (available from https://dataverse.no/dataverse/trolling) shows how the latter are constructed from the original tables exported from Microsoft Excel.

### word (w) and binominal (nn)

The data frame `w` contains the complete data set, including information merged from the data frame `l` (see below); `nn` is a subset of `w` (minus two variables not needed for the data analysis), consisting of binominal data only. The structure of `nn` is the more instructive, so it is shown here. Note that four of the last five variables are merged from the data frames `m` and `s` (see below), and that the last variable is generated by concatenating `ftype` and `headPos`. Most variables are self-explanatory, given the name and examples. Thus, `ftype` represents the eight-way formal classification developed in Chapter 5, with values such as `cmp`, `prp`, `gen`, etc. The two semantic types, `semType` and `semTypeH`, represent my revision of the system proposed by Hatcher and the original, respectively (see page 81).

```r
> str(nn)
'data.frame': 3556 obs. of 24 variables:
  $ meaning: Factor w/ 100 levels "ankle","arctic lights",...
  $ stype : Factor w/ 27 levels "TAX","TAX2","COOR",...
  $ language: Factor w/ 99 levels "Äiwoo","Akkadian",...
  $ word  : chr  "mata kaki"  "'akéts'iin"  "kla¨´ ti li'k"  "kifundo cha mguu" 
  $ gloss : chr  "[eye foot]"  "'a.ké.ts'iin [3SG.foot.bone]"  "foot spur" ...
  $ construction: chr  "Head Mod"  "3SG.Mod.Head"  "Mod Head"  "Head CON Mod" ...
  $ headPos: Factor w/ 2 levels "L","R": 1 2 2 2 1 2 2 2 2 2 ...
  $ head  : chr  "eye"  "bone"  "spur"  "joint" ...
  $ mod   : chr  "foot"  "foot"  "foot"  "foot" ...
  $ ftype : Factor w/ 8 levels "cmp","prp","gen",...
  $ area2 : Factor w/ 7 levels "A","E","O",...
  $ area  : Factor w/ 7 levels "A","E","O","C",...
  $ iso639: Factor w/ 99 levels "akk","amh","aoi",...
  $ family: Factor w/ 37 levels "Afro-Asiatic",...
  $ genus : Factor w/ 66 levels "Adamawa-Ubangi",...
  $ latitude: chr  "-7.33458"  "36.2106"  "9.67031"  "-8.25605" 
  $ longitude: chr  "109.716"  "-110.082"  "-83.4102"  "37.624" 
  $ area  : Factor w/ 7 levels "Africa","Eurasia",...
  $ htype : Factor w/ 5 levels "MisH","HinM",...
  $ semField: Factor w/ 16 levels "Agriculture and vegetation",...
  $ semType: Factor w/ 7 levels "person","animal",...
  $ semTypeH: Factor w/ 6 levels "person","animal",...
  $ ftype2 : Factor w/ 16 levels "adjL","adjR",...
```

language (l)

The data frame l contains information relating to each of the languages in the sample, including some administrative data not included here. The variable language2 is contains abbreviated language names used for reasons of space in some figures in Chapter 7 (e.g. “Seych. Creole”).

```r
> str(l)
de慰料枃： 99 倒的 11 變異
$ language : Factor w/ 99 levels "Aiwoo","Akkadian",...: 1 2 3 4 5 6 7 8 9 10 ...
$ language2: Factor w/ 99 levels "Aiwoo","Akkadian",...: 1 2 3 4 5 6 7 8 9 10 ...
$ area : Factor w/ 7 levels "A","E","O","G",...: 3 1 1 4 2 2 1 1 1 1 ...
$ iso639 : Factor w/ 99 levels "akk","amh","aoi",...: 64 1 2 3 4 6 53 7 8 9 ...
$ glottocode : Factor w/ 99 levels "akka1240","amha1245",...: 7 1 2 3 4 6 48 8 9 ...
$ family : Factor w/ 37 levels "Afro-Asiatic",...: 7 1 1 13 22 16 5 5 5 1 ...
$ genus : Factor w/ 66 levels "Adamawa-Ubangi",...: 64 1 2 3 4 6 48 8 9 ...
$ latitude : chr  "-10.2302" "33.1000" "11.7082" ...
$ longitude : chr  "166.21" "44.1000" ...
$ sample : Factor w/ 2 levels "P","W": 1 1 1 1 2 1 1 1 1 1 ...
$ area2 : Factor w/ 7 levels "Africa","Eurasia",...: 3 1 1 4 2 2 1 1 1 1 ...
```

meaning (m)

The data frame m contains information relating to the 100 meanings. See above for the difference between semType and semTypeH.

```r
> str(m)
de慰料枃： 100 倒的 4 變異
$ meaning : Factor w/ 100 levels "ankle","arctic lights",...: 1 2 3 4 5 6 7 8 9 ...
$ semField: Factor w/ 16 levels "Agriculture and vegetation",...: 12 14 8 2 2 2 8 ...
$ semType : Factor w/ 7 levels "person","animal",...: 4 5 6 2 5 5 7 7 ...
$ semTypeH: Factor w/ 7 levels "person","animal",...: 3 3 2 3 3 3 3 1 ...
```

semantic relation (s)

The data frame s contains information relating to the two classifications of semantic relations (cf. Table 31 on page 238), with “B” and “H” standing for Bourque and Hatcher, respectively. Somewhat inconsistently, stype contains the codes for the Bourque2 classification and htype those for the Hatcher2 classification. The variable atype maps these two systems to the three classic associative relations identified by Aristotle.

```r
> str(s)
de慰料枃： 27 倒的 7 變異
$ B2 : Factor w/ 16 levels "Cause","Composition",...: 13 13 4 11 15 3 3 8 ...
$ stype : Factor w/ 27 levels "CAUS","CAUS2",...: 21 22 7 18 25 5 6 13 14 11 ...
$ atype : Factor w/ 3 levels "caus","cont",...: 3 3 3 3 2 2 2 2 2 ...
$ H2icon : Factor w/ 5 levels "M≈H","M←H","M→H",...: 1 1 1 4 5 4 4 5 5 ...
$ htype : Factor w/ 5 levels "MisH","HinM",...: 1 1 1 3 2 3 2 2 ...
$ B2template : chr  "(an) M is a kind of H" "(an) H is a kind of M" ...
$ B2example : chr  "oak tree" "bear cub" "boy king" "lion ant" ...
```
G. Questionnaire

**Instructions** (formatted for A4 or Letter)

Contact: pepper.steve@gmail.com
Slide show: SLE 2016 presentation
Description: Project description

Dear Contributor,

Thank you for volunteering to supply data for my PhD project on the typology of binominal lexemes. You will of course be credited for your work. Please fill out the **Data sheet** after carefully reading the following instructions. Contact me if you have any questions.

**DATABASE FIELDS**

**Meaning (ENG)**

This is the meaning to be translated. Also given in Russian, Spanish and French.

**Translation equivalent**

1. For each meaning give the canonical **translation equivalent** (TE) using the Latin script (or IPA); if no equivalent exists, leave the field blank
2. Choose the **most common** translation equivalent
3. If two translation equivalents are equally common, supply either one but prefer one that is **analysable** to than one that is mono-morphemic

**TE (non-Latin script)**

4. For non-Latin writing systems, provide the word in the **native script** (see the RUSSIAN example)

**Gloss (complex words only)**

5. For TEs consisting of more than one morpheme provide a **gloss**.
6. If the TE does not contain any **polymorphemic words**, simply supply the gloss: e.g. for FRENCH 'railway' (chemin de fer) enter "**way of iron**"
7. For TEs in which **one or more words are polymorphemic**, repeat the translation with word-internal morpheme breaks indicated by a **period**, and add the gloss in **square brackets**: e.g. for GERMAN 'railway' (Eisenbahn) enter "**eisen.bahn [iron.way]**"
8. Only words that are **synchronically analysable** should be glossed
9. Use recommended abbreviations from the **Leipzig Glossing Rules** wherever possible, except:
10. Use a **colon** instead of a period when a single object-language element is rendered by several metalanguage elements (see the BEZHTA example)
11. Use a **colon** instead of a period when a single object-language element is rendered by several metalanguage elements (see the BEZHTA example)

**Notes**

12. Put any comments regarding the source of loans, calques, etc. in this column

**EXAMPLES**

<table>
<thead>
<tr>
<th>Meaning (ENG)</th>
<th>Translation equivalent</th>
<th>TE (non-Latin script)</th>
<th>Gloss (complex words only)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>railway</td>
<td>kil.os hinu</td>
<td>iron.OBL:GEN way</td>
<td></td>
<td>BEZHTA</td>
</tr>
<tr>
<td>railway</td>
<td>chemin de fer</td>
<td>way of iron</td>
<td></td>
<td>FRENCH</td>
</tr>
<tr>
<td>railway</td>
<td>eisenbahn</td>
<td>eisen.bahn</td>
<td>[iron.way]</td>
<td>GERMAN</td>
</tr>
<tr>
<td>railway</td>
<td>železnaja doroga</td>
<td>железная дорога</td>
<td>[iron.ADJZ road]</td>
<td>RUSSIAN</td>
</tr>
<tr>
<td>railway</td>
<td>reli</td>
<td>-</td>
<td></td>
<td>SWAHLI</td>
</tr>
<tr>
<td>railway</td>
<td>železnica</td>
<td>želez.n.ica</td>
<td>[iron.ADJZ.NMLZ]</td>
<td>SLOVAK</td>
</tr>
</tbody>
</table>

Many thanks for your help,
Steve
<table>
<thead>
<tr>
<th>RUS</th>
<th>SPA</th>
<th>FRA</th>
<th>Meaning (ENG)</th>
<th>Translation equivalent</th>
<th>TE (non-Latin script)</th>
<th>Gloss (complex words only)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>лодька</td>
<td>tobillo</td>
<td>cheville</td>
<td>ankle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>северное сияние</td>
<td>aurora boreal</td>
<td>aurore boréale</td>
<td>arctic lights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>рюкзак</td>
<td>mochila</td>
<td>sac à dos</td>
<td>backpack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>велосипед</td>
<td>bicicleta</td>
<td>vélo</td>
<td>bicycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>насос</td>
<td>bomba de bicicleta</td>
<td>pompe à vélo</td>
<td>bicycle pump</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>пчела</td>
<td>abeja</td>
<td>abeille</td>
<td>bee</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>улей</td>
<td>colmena</td>
<td>ruche d'abeilles</td>
<td>beehive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>волосянка</td>
<td>clavícula</td>
<td>clavicule</td>
<td>collarbone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ужин</td>
<td>cena</td>
<td>dîner</td>
<td>dinner</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>вождь</td>
<td>cacique</td>
<td>chef de clan</td>
<td>chieftain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>пчеловод</td>
<td>пчелиный улей</td>
<td>abeille ruche</td>
<td>apiary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>камень</td>
<td>камень</td>
<td>rocher</td>
<td>rock</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
This appendix contains various tables that would have taken up too much space in the body of the text.

Morbo/Comp database structure

Cf. §2.1.3 on page 32. The structural description is adapted from Guevara et al. (2006) for readability and to conform with the subset of the database kindly made available to me by Sergio Scalise. This table only shows fields for two constituents but the database was designed to accommodate more, as in EFL δeka.pend.a.silavos [ten.five.LE.syllable] ‘fifteen-syllable metre’. The second “linking element” field was also used to record the POSS:3SG suffix of the Turkish izafet construction, as in kaldırım mühendi.si [pavement engineer.POSS:3SG] ‘unemployed person’.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value set</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>2-letter code</td>
<td>FI (FIN)</td>
</tr>
<tr>
<td>compound</td>
<td>orthographic form (Latin script)</td>
<td>kuusi+puu</td>
</tr>
<tr>
<td>category</td>
<td>N, V, A, P, Adv, etc.</td>
<td>N</td>
</tr>
<tr>
<td>structure</td>
<td>[N+N], [N+A], [V+N], etc.</td>
<td>[N+N]</td>
</tr>
<tr>
<td>classification</td>
<td>SUB</td>
<td>CRD</td>
</tr>
<tr>
<td>syntactic head</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>semantic head</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>1st const (C₁)</td>
<td>root form</td>
<td>kuusı</td>
</tr>
<tr>
<td>category of C₁</td>
<td>N, V, A, P, Adv, etc.</td>
<td>N</td>
</tr>
<tr>
<td>1st linking element</td>
<td>additive/subtractive morpheme</td>
<td>-</td>
</tr>
<tr>
<td>2nd const (C₂)</td>
<td>root form</td>
<td>puu</td>
</tr>
<tr>
<td>category of C₂</td>
<td>N, V, A, P, Adv, etc.</td>
<td>N</td>
</tr>
<tr>
<td>2nd linking element</td>
<td>additive/subtractive morpheme</td>
<td>-</td>
</tr>
<tr>
<td>plural marking</td>
<td>C₁</td>
<td>C₂</td>
</tr>
<tr>
<td>gender</td>
<td>m</td>
<td>f</td>
</tr>
<tr>
<td>gloss</td>
<td>English gloss (C₁+C₂=COMP)</td>
<td>spruce+tree = spruce</td>
</tr>
</tbody>
</table>

*Table 66: Morbo/Comp database structure*
**Štekauer et al’s language sample**

Cf. §2.2.2 on page 44. The genetic groupings given here are from Glottolog, not WALS. The 15 languages listed under the heading Eurasia 2 were excluded from the study sample.

<table>
<thead>
<tr>
<th>Africa (A) 14</th>
<th>Eurasia (E) 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dangaleat DAA (Chadic)</td>
<td>Udihe UDE (Tungusic)</td>
</tr>
<tr>
<td>Hausa HAU (Chadic)</td>
<td>Tatar TAT (Turric)</td>
</tr>
<tr>
<td>Amharic AMH (Semitic)</td>
<td>Malayalam MAL (Dravidian)</td>
</tr>
<tr>
<td>Hebrew HEB (Semitic)</td>
<td>Tamil TAM (Dravidian)</td>
</tr>
<tr>
<td>Ganda LUG (Bantoid)</td>
<td>Telugu TEL (Dravidian)</td>
</tr>
<tr>
<td>Swahili SWH (Bantoid)</td>
<td>Breton BRE (Celtic)</td>
</tr>
<tr>
<td>Zulu ZUL (Bantoid)</td>
<td>English ENG (Germanic)</td>
</tr>
<tr>
<td>Jola-Fonyi DYO (Central Atlantic)</td>
<td>Greek ELL (Greek)</td>
</tr>
<tr>
<td>Yoruba YOR (Defoid)</td>
<td>Marathi MAR (Indo-Aryan)</td>
</tr>
<tr>
<td>Konni KMA (Gur)</td>
<td>Spanish SPA (Romance)</td>
</tr>
<tr>
<td>Ga GAA (Kwa)</td>
<td>Slovak SLK (Slavic)</td>
</tr>
<tr>
<td>Kua TYU (Non-Khoech)</td>
<td>Japanese JPN (Japanese)</td>
</tr>
<tr>
<td>Datooga TCC (Nilotic)</td>
<td>Georgian KAT (Kartvelian)</td>
</tr>
<tr>
<td>Luo LUO (Nilotic)</td>
<td>Estonian EST (Finnic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>North America (N) 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slave SCS (Athapaskan-Eyak-Tlingit)</td>
</tr>
<tr>
<td>Kalaallisut KAL (Eskimo-Aleut)</td>
</tr>
<tr>
<td>Zinacantán Tzotzil TZO (Core Mayan)</td>
</tr>
<tr>
<td>Clallam CLM (Salishan)</td>
</tr>
<tr>
<td>Lakota LKT (Siouan)</td>
</tr>
<tr>
<td>Totonac TKU (Totonacan)</td>
</tr>
<tr>
<td>Pipil PPL (Southern Uto-Aztec)</td>
</tr>
<tr>
<td>Kwakialt KWK (Wakashan)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>South America (S) 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maipure MAIP1246 (Arawakan)</td>
</tr>
<tr>
<td>Ja'garu JQR (Aymara)</td>
</tr>
<tr>
<td>Wichí MZH (Matacoan)</td>
</tr>
<tr>
<td>Movima MZP (Movima)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southeast Asia &amp; Oceania (O) 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vietnamese VIE (Vietic)</td>
</tr>
<tr>
<td>Anejom ATY (Eastern Malayo-Polynesian)</td>
</tr>
<tr>
<td>Kumak NEE (Eastern Malayo-Polynesian)</td>
</tr>
<tr>
<td>Maori MRI (Eastern Malayo-Polynesian)</td>
</tr>
<tr>
<td>Indonesian IND (Malayo-Sumbawan)</td>
</tr>
<tr>
<td>Iloko ILO (Northern Luzon)</td>
</tr>
<tr>
<td>Karao KYJ (Northern Luzon)</td>
</tr>
<tr>
<td>Tibetan BOD (Bodic)</td>
</tr>
<tr>
<td>Mandarin Chinese CMN (Sinitic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eurasia 2 (E) 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afrikaans AFR (Germanic)</td>
</tr>
<tr>
<td>Dutch NLD (Germanic)</td>
</tr>
<tr>
<td>German DEU (Germanic)</td>
</tr>
<tr>
<td>Swedish SWE (Germanic)</td>
</tr>
<tr>
<td>Trinidadian Creole English TRF (Germanic)</td>
</tr>
<tr>
<td>Hindi HIN (Indo-Aryan)</td>
</tr>
<tr>
<td>Catalan CAT (Romance)</td>
</tr>
<tr>
<td>French FRA (Romance)</td>
</tr>
<tr>
<td>Italian ITA (Romance)</td>
</tr>
<tr>
<td>Portuguese POR (Romance)</td>
</tr>
<tr>
<td>Romanian RON (Romance)</td>
</tr>
<tr>
<td>Belarusian BEL (Slavic)</td>
</tr>
<tr>
<td>Russian RUS (Slavic)</td>
</tr>
<tr>
<td>Serbian SRP (Slavic)</td>
</tr>
<tr>
<td>Ukrainian UKR (Slavic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Australia / New Guinea (G) 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bardi BCI (Nyulnyulan)</td>
</tr>
<tr>
<td>Kalkutunga KTG (Galgadungic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pidgins &amp; Creoles (P) 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amele AEU (Madang)</td>
</tr>
</tbody>
</table>

*Table 67: Language sample (Štekauer, Valera & Körtvélyessy 2012)*
Alphabetical list of the initial 201 meanings

Cf. §3.1.3 and §3.1.4 on page 70ff.

This table lists the full set of 201 meanings in alphabetical order, together with the frequency of their occurrence as binominals in the 50 language sample.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>afternoon</td>
<td>12</td>
</tr>
<tr>
<td>airplane</td>
<td>9</td>
</tr>
<tr>
<td>ancestors</td>
<td>8</td>
</tr>
<tr>
<td>anger</td>
<td>2</td>
</tr>
<tr>
<td>ankle</td>
<td>36</td>
</tr>
<tr>
<td>anxiety</td>
<td>4</td>
</tr>
<tr>
<td>arctic lights</td>
<td>65</td>
</tr>
<tr>
<td>armpit</td>
<td>14</td>
</tr>
<tr>
<td>arson</td>
<td>9</td>
</tr>
<tr>
<td>baby</td>
<td>11</td>
</tr>
<tr>
<td>backpack</td>
<td>26</td>
</tr>
<tr>
<td>bad luck</td>
<td>4</td>
</tr>
<tr>
<td>bark</td>
<td>17</td>
</tr>
<tr>
<td>bee</td>
<td>21</td>
</tr>
<tr>
<td>beehive</td>
<td>46</td>
</tr>
<tr>
<td>beeswax</td>
<td>58</td>
</tr>
<tr>
<td>beggar</td>
<td>9</td>
</tr>
<tr>
<td>beginning</td>
<td>2</td>
</tr>
<tr>
<td>bicycle</td>
<td>21</td>
</tr>
<tr>
<td>bicycle pump</td>
<td>33</td>
</tr>
<tr>
<td>birth certificate</td>
<td>11</td>
</tr>
<tr>
<td>blacksmith</td>
<td>22</td>
</tr>
<tr>
<td>bow tie</td>
<td>20</td>
</tr>
<tr>
<td>boy</td>
<td>28</td>
</tr>
<tr>
<td>bracelet</td>
<td>34</td>
</tr>
<tr>
<td>breakfast</td>
<td>25</td>
</tr>
<tr>
<td>bruise</td>
<td>2</td>
</tr>
<tr>
<td>calf</td>
<td>19</td>
</tr>
<tr>
<td>capital city</td>
<td>16</td>
</tr>
<tr>
<td>captive</td>
<td>13</td>
</tr>
<tr>
<td>carpenter</td>
<td>31</td>
</tr>
<tr>
<td>cave</td>
<td>15</td>
</tr>
<tr>
<td>chieftain</td>
<td>21</td>
</tr>
<tr>
<td>cock/rooster</td>
<td>33</td>
</tr>
<tr>
<td>collarbone</td>
<td>50</td>
</tr>
<tr>
<td>cookhouse</td>
<td>22</td>
</tr>
<tr>
<td>cousin</td>
<td>23</td>
</tr>
<tr>
<td>crowd</td>
<td>5</td>
</tr>
<tr>
<td>dairy cow</td>
<td>28</td>
</tr>
<tr>
<td>darkness</td>
<td>0</td>
</tr>
<tr>
<td>dawn</td>
<td>3</td>
</tr>
<tr>
<td>deceit</td>
<td>4</td>
</tr>
<tr>
<td>defeat</td>
<td>0</td>
</tr>
<tr>
<td>defendant</td>
<td>7</td>
</tr>
<tr>
<td>descendants</td>
<td>9</td>
</tr>
<tr>
<td>dinner</td>
<td>42</td>
</tr>
<tr>
<td>disease</td>
<td>4</td>
</tr>
<tr>
<td>divorce</td>
<td>0</td>
</tr>
<tr>
<td>doorpost</td>
<td>46</td>
</tr>
<tr>
<td>drink</td>
<td>2</td>
</tr>
<tr>
<td>earlobe</td>
<td>71</td>
</tr>
<tr>
<td>earring</td>
<td>37</td>
</tr>
<tr>
<td>earwax</td>
<td>65</td>
</tr>
<tr>
<td>earthquake</td>
<td>2</td>
</tr>
<tr>
<td>east</td>
<td>4</td>
</tr>
<tr>
<td>election</td>
<td>2</td>
</tr>
<tr>
<td>end</td>
<td>5</td>
</tr>
<tr>
<td>envy</td>
<td>6</td>
</tr>
<tr>
<td>eyebrow</td>
<td>29</td>
</tr>
<tr>
<td>eyelash</td>
<td>42</td>
</tr>
<tr>
<td>eyelid</td>
<td>56</td>
</tr>
<tr>
<td>farmer</td>
<td>36</td>
</tr>
<tr>
<td>fireplace</td>
<td>37</td>
</tr>
<tr>
<td>firewood</td>
<td>19</td>
</tr>
<tr>
<td>fisherman</td>
<td>33</td>
</tr>
<tr>
<td>fishing line</td>
<td>33</td>
</tr>
<tr>
<td>flame</td>
<td>30</td>
</tr>
<tr>
<td>flea market</td>
<td>28</td>
</tr>
<tr>
<td>foal or colt</td>
<td>29</td>
</tr>
<tr>
<td>food</td>
<td>2</td>
</tr>
<tr>
<td>footprint</td>
<td>40</td>
</tr>
<tr>
<td>freeman</td>
<td>3</td>
</tr>
<tr>
<td>girl</td>
<td>32</td>
</tr>
<tr>
<td>glove</td>
<td>54</td>
</tr>
<tr>
<td>gold ring</td>
<td>46</td>
</tr>
<tr>
<td>grief</td>
<td>3</td>
</tr>
<tr>
<td>guard</td>
<td>12</td>
</tr>
<tr>
<td>hand brake</td>
<td>29</td>
</tr>
<tr>
<td>handbag</td>
<td>30</td>
</tr>
<tr>
<td>handkerchief</td>
<td>25</td>
</tr>
<tr>
<td>herdsman</td>
<td>22</td>
</tr>
<tr>
<td>horseshoe</td>
<td>22</td>
</tr>
<tr>
<td>hospital</td>
<td>21</td>
</tr>
<tr>
<td>host</td>
<td>40</td>
</tr>
<tr>
<td>hummingbird</td>
<td>8</td>
</tr>
<tr>
<td>idea</td>
<td>2</td>
</tr>
<tr>
<td>intention</td>
<td>9</td>
</tr>
<tr>
<td>itch</td>
<td>0</td>
</tr>
<tr>
<td>judgment</td>
<td>7</td>
</tr>
<tr>
<td>keyword</td>
<td>25</td>
</tr>
<tr>
<td>kid</td>
<td>41</td>
</tr>
<tr>
<td>lamb</td>
<td>29</td>
</tr>
<tr>
<td>license plate</td>
<td>60</td>
</tr>
<tr>
<td>lipstick</td>
<td>20</td>
</tr>
<tr>
<td>lunch</td>
<td>27</td>
</tr>
<tr>
<td>magic</td>
<td>19</td>
</tr>
<tr>
<td>mail box</td>
<td>35</td>
</tr>
<tr>
<td>mare</td>
<td>39</td>
</tr>
<tr>
<td>married man</td>
<td>10</td>
</tr>
<tr>
<td>married woman</td>
<td>20</td>
</tr>
<tr>
<td>meal</td>
<td>5</td>
</tr>
<tr>
<td>meeting house</td>
<td>15</td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>merchant</th>
<th>15</th>
<th>servant</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>midday</td>
<td>39</td>
<td>shoemaker</td>
<td>23</td>
</tr>
<tr>
<td>milky way</td>
<td>46</td>
<td>shore</td>
<td>30</td>
</tr>
<tr>
<td>mistake</td>
<td>2</td>
<td>shoulderblade</td>
<td>33</td>
</tr>
<tr>
<td>mother-in-law</td>
<td>29</td>
<td>sibling</td>
<td>10</td>
</tr>
<tr>
<td>murder</td>
<td>4</td>
<td>skull</td>
<td>45</td>
</tr>
<tr>
<td>native country</td>
<td>46</td>
<td>sorcerer or witch</td>
<td>35</td>
</tr>
<tr>
<td>necklace</td>
<td>17</td>
<td>spectacles/glasses</td>
<td>31</td>
</tr>
<tr>
<td>neighbour</td>
<td>18</td>
<td>spider web</td>
<td>63</td>
</tr>
<tr>
<td>niece</td>
<td>51</td>
<td>spine</td>
<td>52</td>
</tr>
<tr>
<td>nipple or teat</td>
<td>50</td>
<td>spring or well</td>
<td>11</td>
</tr>
<tr>
<td>nostril</td>
<td>77</td>
<td>stable or stall</td>
<td>29</td>
</tr>
<tr>
<td>nurse</td>
<td>6</td>
<td>stepdaughter</td>
<td>13</td>
</tr>
<tr>
<td>nut</td>
<td>11</td>
<td>stepfather</td>
<td>16</td>
</tr>
<tr>
<td>old woman</td>
<td>5</td>
<td>stepmother</td>
<td>15</td>
</tr>
<tr>
<td>older sister</td>
<td>6</td>
<td>stepson</td>
<td>6</td>
</tr>
<tr>
<td>paddle wheel</td>
<td>26</td>
<td>stone bridge</td>
<td>40</td>
</tr>
<tr>
<td>palm of hand</td>
<td>68</td>
<td>stranger</td>
<td>2</td>
</tr>
<tr>
<td>parents</td>
<td>2</td>
<td>sugar cane</td>
<td>22</td>
</tr>
<tr>
<td>perjury</td>
<td>6</td>
<td>Sunday</td>
<td>56</td>
</tr>
<tr>
<td>pestle</td>
<td>12</td>
<td>supper</td>
<td>40</td>
</tr>
<tr>
<td>pity</td>
<td>2</td>
<td>swelling</td>
<td>0</td>
</tr>
<tr>
<td>plaintiff</td>
<td>3</td>
<td>tailor</td>
<td>7</td>
</tr>
<tr>
<td>postage stamp</td>
<td>24</td>
<td>teacher</td>
<td>4</td>
</tr>
<tr>
<td>postcard</td>
<td>55</td>
<td>tear</td>
<td>31</td>
</tr>
<tr>
<td>potter</td>
<td>51</td>
<td>thatch</td>
<td>39</td>
</tr>
<tr>
<td>praise</td>
<td>0</td>
<td>thief</td>
<td>5</td>
</tr>
<tr>
<td>prostitute</td>
<td>11</td>
<td>thumb</td>
<td>32</td>
</tr>
<tr>
<td>pupil</td>
<td>9</td>
<td>toe</td>
<td>52</td>
</tr>
<tr>
<td>quarrel</td>
<td>0</td>
<td>toilet</td>
<td>17</td>
</tr>
<tr>
<td>queen</td>
<td>53</td>
<td>toilet paper</td>
<td>33</td>
</tr>
<tr>
<td>railway</td>
<td>39</td>
<td>tool</td>
<td>21</td>
</tr>
<tr>
<td>rainbow</td>
<td>33</td>
<td>toolbox</td>
<td>33</td>
</tr>
<tr>
<td>rape</td>
<td>2</td>
<td>toothbrush</td>
<td>36</td>
</tr>
<tr>
<td>razor</td>
<td>8</td>
<td>train</td>
<td>31</td>
</tr>
<tr>
<td>remains</td>
<td>0</td>
<td>tree trunk</td>
<td>42</td>
</tr>
<tr>
<td>rib</td>
<td>26</td>
<td>twins</td>
<td>2</td>
</tr>
<tr>
<td>roof</td>
<td>24</td>
<td>vein or artery</td>
<td>27</td>
</tr>
<tr>
<td>school</td>
<td>3</td>
<td>victory</td>
<td>0</td>
</tr>
<tr>
<td>screwdriver</td>
<td>4</td>
<td>vine</td>
<td>36</td>
</tr>
<tr>
<td>sculptor</td>
<td>21</td>
<td>water pump</td>
<td>29</td>
</tr>
<tr>
<td>waterfall</td>
<td>7</td>
<td>weapons</td>
<td>12</td>
</tr>
<tr>
<td>weapons</td>
<td>12</td>
<td>wedding</td>
<td>11</td>
</tr>
<tr>
<td>Wednesday</td>
<td>30</td>
<td>west</td>
<td>2</td>
</tr>
<tr>
<td>west</td>
<td>2</td>
<td>whetstone</td>
<td>9</td>
</tr>
<tr>
<td>whirlpool</td>
<td>15</td>
<td>widow</td>
<td>13</td>
</tr>
<tr>
<td>widow</td>
<td>13</td>
<td>widower</td>
<td>21</td>
</tr>
<tr>
<td>windmill</td>
<td>32</td>
<td>womb</td>
<td>22</td>
</tr>
<tr>
<td>womb</td>
<td>22</td>
<td>wrist</td>
<td>43</td>
</tr>
<tr>
<td>young man</td>
<td>6</td>
<td>yolk</td>
<td>10</td>
</tr>
<tr>
<td>young woman</td>
<td>8</td>
<td>young woman</td>
<td>6</td>
</tr>
<tr>
<td>younger brother</td>
<td>10</td>
<td>younger sister</td>
<td>14</td>
</tr>
<tr>
<td>younger sister</td>
<td>14</td>
<td>younger sister</td>
<td>10</td>
</tr>
</tbody>
</table>
List of 201 meanings (binominality order)

Cf. §3.1.3 and §3.1.4 on page 70ff.

This table lists the full set of 201 meanings in binominality order, i.e. the frequency (in percent) of the meaning’s occurrence as a binominal in the 50 language sample.

Table 69: The original 201 meanings (“binominality” order)

<table>
<thead>
<tr>
<th></th>
<th>Meaning</th>
<th>Frequency</th>
<th></th>
<th>Meaning</th>
<th>Frequency</th>
<th></th>
<th>Meaning</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nostril</td>
<td>77</td>
<td>32</td>
<td>footprint</td>
<td>40</td>
<td>63</td>
<td>handbag</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>earlobe</td>
<td>71</td>
<td>33</td>
<td>stone bridge</td>
<td>40</td>
<td>64</td>
<td>flame</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>palm of hand</td>
<td>68</td>
<td>34</td>
<td>supper</td>
<td>40</td>
<td>65</td>
<td>Wednesday</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>arctic lights</td>
<td>65</td>
<td>35</td>
<td>midday</td>
<td>39</td>
<td>66</td>
<td>shore</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>earwax</td>
<td>65</td>
<td>36</td>
<td>railway</td>
<td>39</td>
<td>67</td>
<td>hand brake</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>spider web</td>
<td>63</td>
<td>37</td>
<td>thatch</td>
<td>39</td>
<td>68</td>
<td>water pump</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>license plate</td>
<td>60</td>
<td>38</td>
<td>mare</td>
<td>39</td>
<td>69</td>
<td>eyebrow</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>beeswax</td>
<td>58</td>
<td>39</td>
<td>fireplace</td>
<td>37</td>
<td>70</td>
<td>lamb</td>
<td>29</td>
</tr>
<tr>
<td>9</td>
<td>eyelid</td>
<td>56</td>
<td>40</td>
<td>earring</td>
<td>37</td>
<td>71</td>
<td>mother-in-law</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>Sunday</td>
<td>56</td>
<td>41</td>
<td>ankle</td>
<td>36</td>
<td>72</td>
<td>stable or stall</td>
<td>29</td>
</tr>
<tr>
<td>11</td>
<td>postcard</td>
<td>55</td>
<td>42</td>
<td>farmer</td>
<td>36</td>
<td>73</td>
<td>foal or colt</td>
<td>29</td>
</tr>
<tr>
<td>12</td>
<td>glove</td>
<td>54</td>
<td>43</td>
<td>toothbrush</td>
<td>36</td>
<td>74</td>
<td>dairy cow</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>queen</td>
<td>53</td>
<td>44</td>
<td>vine</td>
<td>36</td>
<td>75</td>
<td>flea market</td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>spine</td>
<td>52</td>
<td>45</td>
<td>mail box</td>
<td>35</td>
<td>76</td>
<td>boy</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>toe</td>
<td>52</td>
<td>46</td>
<td>sorcerer or witch</td>
<td>35</td>
<td>77</td>
<td>lunch</td>
<td>27</td>
</tr>
<tr>
<td>16</td>
<td>potter</td>
<td>51</td>
<td>47</td>
<td>bracelet</td>
<td>34</td>
<td>78</td>
<td>vein or artery</td>
<td>27</td>
</tr>
<tr>
<td>17</td>
<td>niece</td>
<td>51</td>
<td>48</td>
<td>bicycle pump</td>
<td>33</td>
<td>79</td>
<td>backpack</td>
<td>26</td>
</tr>
<tr>
<td>18</td>
<td>nipple or teat</td>
<td>50</td>
<td>49</td>
<td>cock/rooster</td>
<td>33</td>
<td>80</td>
<td>paddle wheel</td>
<td>26</td>
</tr>
<tr>
<td>19</td>
<td>collarbone</td>
<td>50</td>
<td>50</td>
<td>rainbow</td>
<td>33</td>
<td>81</td>
<td>rib</td>
<td>26</td>
</tr>
<tr>
<td>20</td>
<td>milky way</td>
<td>46</td>
<td>51</td>
<td>toilet paper</td>
<td>33</td>
<td>82</td>
<td>breakfast</td>
<td>25</td>
</tr>
<tr>
<td>21</td>
<td>beehive</td>
<td>46</td>
<td>52</td>
<td>toolbox</td>
<td>33</td>
<td>83</td>
<td>handkerchief</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>doorpost</td>
<td>46</td>
<td>53</td>
<td>fisherman</td>
<td>33</td>
<td>84</td>
<td>keyword</td>
<td>25</td>
</tr>
<tr>
<td>23</td>
<td>gold ring</td>
<td>46</td>
<td>54</td>
<td>fishing line</td>
<td>33</td>
<td>85</td>
<td>postage stamp</td>
<td>24</td>
</tr>
<tr>
<td>24</td>
<td>native country</td>
<td>46</td>
<td>55</td>
<td>shoulderblade</td>
<td>33</td>
<td>86</td>
<td>roof</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>skull</td>
<td>45</td>
<td>56</td>
<td>thumb</td>
<td>32</td>
<td>87</td>
<td>cousin</td>
<td>23</td>
</tr>
<tr>
<td>26</td>
<td>wrist</td>
<td>43</td>
<td>57</td>
<td>windmill</td>
<td>32</td>
<td>88</td>
<td>shoemaker</td>
<td>23</td>
</tr>
<tr>
<td>27</td>
<td>dinner</td>
<td>42</td>
<td>58</td>
<td>girl</td>
<td>32</td>
<td>89</td>
<td>herdsman</td>
<td>22</td>
</tr>
<tr>
<td>28</td>
<td>eyelash</td>
<td>42</td>
<td>59</td>
<td>train</td>
<td>31</td>
<td>90</td>
<td>horseshoe</td>
<td>22</td>
</tr>
<tr>
<td>29</td>
<td>tree trunk</td>
<td>42</td>
<td>60</td>
<td>carpenter</td>
<td>31</td>
<td>91</td>
<td>blacksmith</td>
<td>22</td>
</tr>
<tr>
<td>30</td>
<td>kid</td>
<td>41</td>
<td>61</td>
<td>spectacles/glasses</td>
<td>31</td>
<td>92</td>
<td>cookhouse</td>
<td>22</td>
</tr>
<tr>
<td>31</td>
<td>host</td>
<td>40</td>
<td>62</td>
<td>tear</td>
<td>31</td>
<td>93</td>
<td>womb</td>
<td>22</td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

<table>
<thead>
<tr>
<th>#</th>
<th>First term</th>
<th>Frequency</th>
<th>Second term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>sugar cane</td>
<td>22</td>
<td>prostitute</td>
<td>11</td>
</tr>
<tr>
<td>95</td>
<td>tool</td>
<td>21</td>
<td>spring or well</td>
<td>11</td>
</tr>
<tr>
<td>96</td>
<td>widower</td>
<td>21</td>
<td>servant</td>
<td>11</td>
</tr>
<tr>
<td>97</td>
<td>hospital</td>
<td>21</td>
<td>nut</td>
<td>11</td>
</tr>
<tr>
<td>98</td>
<td>chieftain</td>
<td>21</td>
<td>wedding</td>
<td>11</td>
</tr>
<tr>
<td>99</td>
<td>bicycle</td>
<td>21</td>
<td>married man</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>sculptor</td>
<td>21</td>
<td>sibling</td>
<td>10</td>
</tr>
<tr>
<td>101</td>
<td>bee</td>
<td>21</td>
<td>younger brother</td>
<td>10</td>
</tr>
<tr>
<td>102</td>
<td>lipstick</td>
<td>20</td>
<td>yolk</td>
<td>10</td>
</tr>
<tr>
<td>103</td>
<td>married woman</td>
<td>20</td>
<td>beggar</td>
<td>9</td>
</tr>
<tr>
<td>104</td>
<td>bow tie</td>
<td>20</td>
<td>airplane</td>
<td>9</td>
</tr>
<tr>
<td>105</td>
<td>magic</td>
<td>19</td>
<td>descendants</td>
<td>9</td>
</tr>
<tr>
<td>106</td>
<td>firewood</td>
<td>19</td>
<td>pupil</td>
<td>9</td>
</tr>
<tr>
<td>107</td>
<td>calf</td>
<td>19</td>
<td>intention</td>
<td>9</td>
</tr>
<tr>
<td>108</td>
<td>neighbour</td>
<td>18</td>
<td>arison</td>
<td>9</td>
</tr>
<tr>
<td>109</td>
<td>necklace</td>
<td>17</td>
<td>whetstone</td>
<td>9</td>
</tr>
<tr>
<td>110</td>
<td>toilet</td>
<td>17</td>
<td>young woman</td>
<td>8</td>
</tr>
<tr>
<td>111</td>
<td>bark</td>
<td>17</td>
<td>hummingbird</td>
<td>8</td>
</tr>
<tr>
<td>112</td>
<td>stepfather</td>
<td>16</td>
<td>ancestors</td>
<td>8</td>
</tr>
<tr>
<td>113</td>
<td>capital city</td>
<td>16</td>
<td>razor</td>
<td>8</td>
</tr>
<tr>
<td>114</td>
<td>meeting house</td>
<td>15</td>
<td>waterfall</td>
<td>7</td>
</tr>
<tr>
<td>115</td>
<td>whirlpool</td>
<td>15</td>
<td>tailor</td>
<td>7</td>
</tr>
<tr>
<td>116</td>
<td>cave</td>
<td>15</td>
<td>judgment</td>
<td>7</td>
</tr>
<tr>
<td>117</td>
<td>merchant</td>
<td>15</td>
<td>defendant</td>
<td>7</td>
</tr>
<tr>
<td>118</td>
<td>stepmother</td>
<td>15</td>
<td>young man</td>
<td>6</td>
</tr>
<tr>
<td>119</td>
<td>armpit</td>
<td>14</td>
<td>stepson</td>
<td>6</td>
</tr>
<tr>
<td>120</td>
<td>younger sister</td>
<td>14</td>
<td>envy</td>
<td>6</td>
</tr>
<tr>
<td>121</td>
<td>widow</td>
<td>13</td>
<td>nurse</td>
<td>6</td>
</tr>
<tr>
<td>122</td>
<td>stepdaughter</td>
<td>13</td>
<td>older sister</td>
<td>6</td>
</tr>
<tr>
<td>123</td>
<td>captive</td>
<td>13</td>
<td>perjury</td>
<td>6</td>
</tr>
<tr>
<td>124</td>
<td>guard</td>
<td>12</td>
<td>end</td>
<td>5</td>
</tr>
<tr>
<td>125</td>
<td>pestle</td>
<td>12</td>
<td>old woman</td>
<td>5</td>
</tr>
<tr>
<td>126</td>
<td>afternoon</td>
<td>12</td>
<td>meal</td>
<td>5</td>
</tr>
<tr>
<td>127</td>
<td>weapons</td>
<td>12</td>
<td>thief</td>
<td>5</td>
</tr>
<tr>
<td>128</td>
<td>birth certificate</td>
<td>11</td>
<td>crowd</td>
<td>5</td>
</tr>
<tr>
<td>129</td>
<td>baby</td>
<td>11</td>
<td>anxiety</td>
<td>4</td>
</tr>
</tbody>
</table>
Lost constructions (84 meaning sample)

Cf. §3.1.4 on page 70ff.

This table contains the complete list of constructions that would be lost from the 50-language data set if the number of meanings were to be reduced from the original 201 to 84. There are 55 in all. A subset of this table is reproduced in the text of §3.1.4 as Table 16 (page 75). Constructions marked with a dagger (†) did not make it into the final data set based on 100 meanings.

*Table 70: Constructions lost with a sample of 84 meanings*

**Bezhta**

<table>
<thead>
<tr>
<th>Base</th>
<th>Head Mod</th>
<th>Mod.ADJZ Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>† Base.SUP.ATTR</td>
<td>māhā.ā.kō [door_frame.SUP.ATTR]</td>
<td>SERVANT</td>
</tr>
<tr>
<td>† Base.SUP.ATTR</td>
<td>c’uddo c’emuc’ [red egg]</td>
<td>YOLK</td>
</tr>
<tr>
<td>† Mod.ADJZ Head</td>
<td>nucodaq t’ot’ [honey:ADJZ fly]</td>
<td>BEE</td>
</tr>
</tbody>
</table>

**Dutch**

<table>
<thead>
<tr>
<th>Base</th>
<th>Mod Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base.M</td>
<td>weduwn.aar [widow.M]</td>
</tr>
</tbody>
</table>

**Hausa**

**Hindi**

<table>
<thead>
<tr>
<th>Base</th>
<th>Mod.ADJZ Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>† Base.ABST</td>
<td>mūnāafū.ncīi [unfaith.ABST]</td>
</tr>
<tr>
<td>† Base.ABST</td>
<td>mā.i.tāa [sorcerer.ABST]</td>
</tr>
</tbody>
</table>

**Indonesian**

| † Mod Head | satpam < satuan pengamanan [security unit] | GUARD |
| † Mod Head | mata air [water eye] | SPRING OR WELL |

**Irish**

<table>
<thead>
<tr>
<th>Base</th>
<th>Head Mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base.NMLZ</td>
<td>draí.acht [magician.NMLZ]</td>
</tr>
<tr>
<td>† Head Mod</td>
<td>tráth.nóna [occasion.noon]</td>
</tr>
</tbody>
</table>

**Kanuri**

**Ket**

| Head Mod.ADJZ | kwôngá nyiýá.à [man marriage.ADJZ] | MARRIED MAN |
| Head Mod.ADJZ | kámú nyiýá.à [woman marriage.ADJZ] | MARRIED WOMAN |

**Ket**

**Lithuanian**

| † Base.ABST | žmogžud.ystė [murderer.ABST] | MURDER |
| Base.COLL | kaim.ynas [village.COLL] | NEIGHBOUR |

**Lower Sorbian**

| † Base.ABST | lut.osć [sorrow.ABST] | PITY |
| † Base.DIM | gōle.tko [child.DIM] | BABY |
| † Base.INST | šrub.owak [screw.INST] | SCREWDRIVER |
| † Base.VBLZ.AGT | wik.owa.ř [market.VBLZ.AGT] | MERCHANT |
### The typology and semantics of binominal lexemes

#### Malagasy

<table>
<thead>
<tr>
<th><strong>Head SOC.Mod</strong></th>
<th>lehilàhy manaN.vády [man with.spouse] MARRIED MAN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMLZ.Base</strong></td>
<td>vehivàvy manaN.vády [woman with.spouse] MARRIED WOMAN</td>
</tr>
<tr>
<td><strong>NMLZ.Base.CIRC</strong></td>
<td>fi.lòha [NMLZ.head] CHIEFTAIN</td>
</tr>
<tr>
<td><strong>fanambakàna &lt; faN.ambàka.ana</strong></td>
<td>[NMLZ.deceit.CIRC] DECEIT</td>
</tr>
<tr>
<td><strong>fisfidìàmana &lt; fi.fidy.(an)ana</strong></td>
<td>[NMLZ.choice.CIRC] ELECTION</td>
</tr>
<tr>
<td><strong>fialònana &lt; fi.àlona.ana</strong></td>
<td>[NMLZ.jealousy.CIRC] ENVY OR JEALOUSY</td>
</tr>
<tr>
<td><strong>fikasàna &lt; fi.kàsa.ana</strong></td>
<td>[NMLZ.witchcraft.CIRC] MAGIC</td>
</tr>
<tr>
<td></td>
<td>fifidiànana &lt; fi.fìdy.(an)ana [NMLZ.choice.CIRC] ELECTION</td>
</tr>
<tr>
<td></td>
<td>fialònana &lt; fi.àlona.ana [NMLZ.jealousy.CIRC] ENVY OR JEALOUSY</td>
</tr>
<tr>
<td></td>
<td>fiadìana &lt; fi.àdy.ana [NMLZ.fight.CIRC] WEAPONS</td>
</tr>
</tbody>
</table>

#### Mapudungun

<table>
<thead>
<tr>
<th><strong>Mod.VBLZ.AGT Head</strong></th>
<th>kure.nge.n wentru [wife.ESS.NON:F3 man] MARRIED MAN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mod.VBLZ.LOC Head</strong></td>
<td>füta.nge.n zomo [husband.ESS.NON:F3 woman] MARRIED WOMAN</td>
</tr>
<tr>
<td><strong>Old High German</strong></td>
<td>wingka lawen.tu.we ruka [medicine.VBLZ.LOC house] HOSPITAL</td>
</tr>
<tr>
<td><strong>† Base.DIM</strong></td>
<td>kindi.lîn [child.DIM] BABY</td>
</tr>
<tr>
<td><strong>Oroqen</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Base.REC</strong></td>
<td>amo.rok [stool.REC] TOILET</td>
</tr>
<tr>
<td><strong>† Base.VBLZ.AGT</strong></td>
<td>mayma.la.ri [business.VBLZ.AGT] MERCHANT</td>
</tr>
<tr>
<td><strong>Otomi</strong></td>
<td></td>
</tr>
<tr>
<td><strong>† Head Mod Lim</strong></td>
<td>jwàdä bàtsi.tho [brother child.LIM] YOUNGER BROTHER</td>
</tr>
<tr>
<td></td>
<td>nju bàtsi.tho [sister child.LIM] YOUNGER SISTER</td>
</tr>
<tr>
<td><strong>Polish</strong></td>
<td></td>
</tr>
<tr>
<td><strong>† Base.AUG</strong></td>
<td>maco.cha [mother.AUG] STEPMOTHER</td>
</tr>
<tr>
<td><strong>Kekchí</strong></td>
<td></td>
</tr>
<tr>
<td><strong>† AGT Base</strong></td>
<td>aj limoox [AGT alm] BeggAR</td>
</tr>
<tr>
<td></td>
<td>aj tz’alam [AGT prison] CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td><strong>Base.DER</strong></td>
<td>jolom.îl [head.DER] CHIEFTAIN</td>
</tr>
<tr>
<td><strong>Base.INST.NMLZ</strong></td>
<td>k’ub leb’.aal [hearthstone.INST.NMLZ] COOKHOUSE</td>
</tr>
<tr>
<td></td>
<td>ch’ut.leb’.aal [group.INST.NMLZ] MEETING HOUSE</td>
</tr>
<tr>
<td><strong>Base.PL.NMLZ</strong></td>
<td>k’ot.îb’.aal [shit.NMLZ] TOILET</td>
</tr>
<tr>
<td><strong>† Base.SUF</strong></td>
<td>sum.la.jik [mate.DER.SUF] WEDDING</td>
</tr>
<tr>
<td><strong>opaque</strong></td>
<td>b’aqlaq ch’îich’ [corncob iron] BICYCLE</td>
</tr>
<tr>
<td><strong>Romanian</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Base.ABST</strong></td>
<td>capitan.ie [captain.NMLZ] CHIEFTAIN</td>
</tr>
<tr>
<td></td>
<td>cumun.ie [crown.SUF] WEDDING</td>
</tr>
<tr>
<td><strong>Base.AGT.ABST</strong></td>
<td>bucătă.ar.ie [piece_of_food.AGT.NMLZ] COOKHOUSE</td>
</tr>
<tr>
<td><strong>Base.VBLZ.AGT.ABST</strong></td>
<td>vrăj.i.tor.ie [magic.VBLZ.AGT.ABST] MAGIC</td>
</tr>
<tr>
<td><strong>Selice Romani</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Base.ABST</strong></td>
<td>čohan.îpe [witch/sorcerer.ABST] MAGIC</td>
</tr>
<tr>
<td><strong>† Mod.ADJZ Head</strong></td>
<td>phivli džuvli [widow.ADJZ woman] WIDOW</td>
</tr>
<tr>
<td></td>
<td>özvédni džuvli [widow.ADJZ woman] WIDOW</td>
</tr>
</tbody>
</table>
### Seychelles Creole

<table>
<thead>
<tr>
<th>Base.AGT</th>
<th>prizon.nyen [prison.AGT] CAPTIVE OR PRISONER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bazar.dye [market.AGT] MERCHANT</td>
</tr>
</tbody>
</table>

### Takia

<table>
<thead>
<tr>
<th>Head Mod</th>
<th>tamol sos [man Derris_root] WIDOWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Mod LOC</td>
<td>nanuk swa te [child breast at] BABY</td>
</tr>
<tr>
<td>Head Mod.3SG COM PFV</td>
<td>tamol iwo.n da ya [man spouse.3SG COM PFV] MARRIED MAN</td>
</tr>
<tr>
<td></td>
<td>pein iwo.n da ya [woman spouse.3SG COM PFV] MARRIED WOMAN</td>
</tr>
</tbody>
</table>

### Welsh

<table>
<thead>
<tr>
<th>Base.NMLZ</th>
<th>pen.aeth [chief.NMLZ] CHIEFTAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base.SUF</td>
<td>carchar.or [prison.SUF] CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td></td>
<td>cwmwd.og [district.SUF] NEIGHBOUR</td>
</tr>
</tbody>
</table>

### Wichí

<table>
<thead>
<tr>
<th>Base.AGT</th>
<th>tshotoy.wu [animals.AGT] HERDSMAN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tiena.wu [market.AGT] MERCHANT</td>
</tr>
<tr>
<td></td>
<td>sapatu.wu [shoe.AGT] SHOEMAKER</td>
</tr>
<tr>
<td>Base.LOC</td>
<td>kanu.hi [needle.LOC] SUGAR CANE</td>
</tr>
<tr>
<td></td>
<td>y’amekw.hi [excrement.LOC] TOILET</td>
</tr>
</tbody>
</table>

### Yakut

<table>
<thead>
<tr>
<th>Base.NMLZ</th>
<th>bah.ilik [bas.NMLZ] CHIEFTAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yos.poγ [room.NMLZ] COOKHOUSE</td>
</tr>
<tr>
<td>† Base.VR.NMLZ</td>
<td>uot.taː.hin [fire.VR.NMLZ] ARSON</td>
</tr>
<tr>
<td>† Base.VR.REFL.NMLZ</td>
<td>sanaː.ryaː.hin [thought.VR.NMLZ] GRIEF</td>
</tr>
</tbody>
</table>

### Yaqui

<table>
<thead>
<tr>
<th>Base.LOC</th>
<th>sisi’iwoo.chi [iron.LOC] TOOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>† Base.NMLZ</td>
<td>ko ´oko.a [pain.NMLZ] DISEASE</td>
</tr>
<tr>
<td>† Mod:PL Head</td>
<td>ju<del>ubi.wa.me [RDP</del>wife.PASS.NMLZ] WEDDING</td>
</tr>
<tr>
<td></td>
<td>waim asoa [sister/brother:PL daughter] STEPDoughter</td>
</tr>
<tr>
<td></td>
<td>waim marai [sister/brother:PL father] STEPFather</td>
</tr>
<tr>
<td></td>
<td>waim achai [sister/brother:PL mother] STEPMOTHER</td>
</tr>
</tbody>
</table>

### Zinacantán Tzotzil

<table>
<thead>
<tr>
<th>Mod Head</th>
<th>shokan na [side house] NEIGHBOUR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hmulavi ?antz [sinner woman] PROSTITUTE</td>
</tr>
<tr>
<td></td>
<td>me?anal ?antz [poverty/misery/grief woman] WIDOW</td>
</tr>
<tr>
<td></td>
<td>me?anal vinik [poverty/misery/grief man] WIDOWER</td>
</tr>
<tr>
<td>opaque</td>
<td>k’op ?οʔon [word/argument heart] ANXIETY</td>
</tr>
<tr>
<td></td>
<td>kachimpa pom [pipe incense] BEE</td>
</tr>
<tr>
<td></td>
<td>mishik’ pom [belly_button incense] BEE</td>
</tr>
<tr>
<td></td>
<td>ton pom [stone incense] BEE</td>
</tr>
</tbody>
</table>
Name index

Abraham, Roy Clive, 388
Adams, Valerie, 185, 198, 226
Adelaar, Alexander, 390
Ahlqvist, Anders, 389
Ahmad, Mustapha, 388
Aikhenvald, Alexandra Y., 21, 41, 42, 43, 45, 61, 149, 150, 269, 299
Akdikmen, Resuhi, 68, 392
Allen, Keith, 149
Allen, Margaret Reece, 185
Ambrazas, Vytautas, 390
Anderson, Doris G., 113, 392, 446
Arcodia, Giorgio Francesco, 124
Arends, Jacques, 391
Argus, Reili, xxii, 388
Aristotle, 193, 209, 215, 281, 317, 325, 486
Aronoff, Mark, 9
Artiagoitia, Xabier, 387
Asher, R. E., 390, 408
Askel, John Ole, 390
Aronoff, Mark, 9
Aristotle, 193, 209, 215, 281, 317, 325, 486
Arnaud, Pierre J.L., xxii, 21, 29, 30, 31, 32, 61, 67, 82, 88, 188, 198, 203, 204, 205, 206, 207, 209, 211, 214, 215, 216, 217, 224, 331
Aronoff, Mark, 9
Baker, Mark C., 390
Bakker, Dik, 57, 85, 87, 88, 99, 391
Bao, Sokhna, xxii, 392
Bareggi, Cristina, 389
Barker, Chris, 311
Baron, Irène, 226, 269
Bartels, Hauke, xxii, 390
Belzėja, Z., 390
Bergsten, Nils, 184
Benners Lee, Tim, 4
Bertet, Denis, xxii, 392, 446
Bisetto, Antonietta, 7, 21, 33, 34, 35, 36, 38, 297, 337
Blank, Andreas, 193, 219, 318, 324, 325, 326, 327, 332, 334
Bloomfield, Leonard, 34, 60
Bollée, Annegren, 391, 447
Booij, Geert, 9, 34, 164, 337, 388
Booth, Trudie Maria, 388
Borger, Hagit, 389
Borg, Albert, 390
Bourdage, Mario, 193
Brekke, Herbert E., 185
Brindle, Jonathan, xxii, 388, 403
The typology and semantics of binominal lexemes

Brown, Cecil H., xxii, 38, 165, 390, 391, 392, 418, 433
Brown, Lea, xxii, 38, 165, 390, 391, 392, 418, 433
Bush, Vannevar, 4, 5
Butnariu, Cristina, 188
Bybee, Joan L., 144
Carlson, Robert, 403
Carr, Charles Telford, 184, 187, 188
Cauquelin, Josiane, 391
Ceccagno, Antonella, 33, 210, 216, 264, 390
Cetnarowska, Bożena, 261, 390
Chihuaicura, Antonio, xxii, 390
Chumakina, Marina, 38, 387
Cicin-Sain, Višnja, xxii, 388
Cole, Peter, 389, 445
Comrie, Bernard, xxii, 55, 55, 87, 95, 387, 448
Corne, Chris, 391
Coulibaly, Sekou, xxii, 387, 390, 403
Courtz, Henk, 388, 442
Cowan, Marion M., 392
Coyos, Jean-Baptiste, 387
Craig, James Patrick, 389
Creissels, Denis, 134, 144, 161
Cyffer, Norbert, 389
Dinh, Dien, 28, 392
Diouf, Jean-Léopold, 392
Dixon, R. M. W., 106, 144, 158, 269, 299, 388, 412
Dokulil, Miloš, 10
Dombrowsky-Hahn, Klaudia, 403
Don, Jan, 388
Donaldson, Bruce C., 388, 411
Donohue, Mark, xxii, 86, 391
Dooley, Bob, xxii, 390
Downing, Pamela, 185, 195, 221
Doyle, Aidan, 389, 410
Dressler, Wolfgang U., 175
Dryer, Matthew S., 8, 34, 57, 59, 86, 88, 89, 256, 383
Dumestre, Gérard, 387
Echols, John M., 389
Egmond, Marie-Elaine van, 387, 432
Eiesland, Eli-Anne, 184, 188, 223
Elbert, Samuel H., 389, 429
Elšík, Viktor, 391
Engelbart, Doug, 4
England, Nora C., 139, 320, 423
Epps, Patience, xxii, 389, 445
Evans, H. Meurig, 68
Faarlund, Jan Terje, 390
Fabri, Ray, 390, 400
Fal, Arame, 392
Fang Ching Teng, Stacy, xxii, 391
Fellbaum, Christiane, 305
Fengxiang Li, 185, 391
Fernández, Zarina Estrada, 392, 440
Foolen, Ad, 13
Foroodi-Nejad, Farzaneh, 392
Fortescue, Michael D., 29
Fragkopoulou, Katerina, xxii, 388
Gagné, Christina L., 175
Garai, Koldo, xxii, 387
Georg, Stefan, 390
Giurgea, Ion, 391
Glinert, Lewis, 389, 400
Göksel, Aslı, 392, 407
Goldberg, Adele E., 337
Golden, Anne, 390, 413
Golluscio, Lucía, 390
González Campos, Guillermo, xxii
Good, Jeff, xxii, 391, 446
Goossen, Irvy W., 390, 435
Goswami, G. C., 387, 414
Grant, Anthony, 388
Greenberg, Joseph H., 87, 175, 255
Grimm, Jacob, 184
Grossmann, Maria, 391
Grubišić, Vinko, 388
Guevara, Emiliano, 7, 16, 21, 32, 33, 34, 37, 38, 39, 40, 183, 256, 337, 489
Güller, Ashi, xxii, 392
Hacohen, Pius ten, 184, 185
Hagen Kaldhol, Nina, xxii, 391
Hammarström, Harald, xxv, 86
Harmelink, Bryan, 279
Harms, Robert T., 388
Hartmann, Iren, xxii, 389
Hasegawa, Yoko, 69, 389
Haviland, John, 392, 437
Healy, Dana, 392, 426
Heath, Jeffrey, 28
Heimbach, Ernest E., 389
Helmbrecht, Johannes, 389, 438
Hernaiz, Rodrigo, xxii, 387
Hildebrandt, Kristine A., xxiii, 194, 388, 418
Hinds, John, 69, 389
Holton, David, 388, 413
Holzinger, Daniel, 391, 415
Hutchison, John P., 389, 406
Iwasaki, Shoichi, 392
Jackendoff, Ray, 15, 183, 187, 188, 196, 198, 221, 223, 231, 337
Jacobson, Steven A., xxiii, 194, 388
Jaggar, Philip J., 388
Jakobson, Roman, 193, 325
Janda, Laura A., xxi, 18, 21, 193, 194, 209, 216, 252, 319, 321, 348, 388, 418
Jayaraj, Acharya, 390, 415
Jensen, Bård Uri, xxi
Jespersen, Otto, 49, 186, 187, 188, 198, 199, 201, 202, 221, 348
Johnston, Michael, 185
Jones, ʻŌiwi Parker, 389, 392, 410, 411
Kachru, Yamuna, 389
Kageyama, Taro, 69, 389
Karlsson, Fred, 388, 423
Kawachi, Kazuhiro, xxii, 390, 391, 398
Keenan, Edward L., 87
Khalilov, Madzhid, xxii, 387
Kibrik, Aleksandr E., 16, 157, 387, 422
Kiefer, Ferenc, 157
Kilham, Christine, 392, 434
King, Gareth, 392
Kirtchuk-Halevi, Pablo I., 389
Kittilä, Seppo, 326
Koch, Peter, 9, 57, 58, 193, 269, 280, 281, 282, 302, 318, 324, 325, 327, 331, 332, 345
Kornfilt, Jaklin, 68, 392, 407
Körtvélyessy, Lívia, 7, 10, 21, 44, 48, 88, 168, 490
Kossmann, Maarten, xxii, 156, 391, 395
Kövecses, Zoltan, 319
Krause, Anne, xxii, 388
Kruspe, Nicole, xxii, 388, 426
Lang, Margaret, 388, 416
Langacker, Ronald W., xxi, 6, 37, 115, 185, 296, 308, 337, 348
Lees, Robert B., 185
Lesage, Jakob, xxii, 389, 401
Leslau, Wolf, 387, 399
Levi, Judith N., 14, 21, 48, 49, 61, 185, 186, 187, 190, 195, 196, 217, 389
Levshina, Natalia, 274
Libben, Gary, 175, 184
Liberman, Marc, 234
Lieber, Rochelle, 7, 8, 24, 36, 113, 185, 297
Lindenfeld, Jacqueline, 392, 440
Lipkind, William, 389
Löbner, Sebastian, 311
Lødrup, Helge, xxi
Löhr, Doris, 389
Loiseau, Sylvain, xxii, 392, 434
Lovestrand, Joey, xxi, 267, 387, 396
Lovestrand, Joseph, xxii, 267, 387, 396
Lü, Shanshan, xxii, 388, 431
Lukas, Johannes, 389
Luo, Yongxian, 390, 431, 490
Lysvik, Julian, xxii, 389
MacAindir, Cormac, xxii, 389, 392
Mace, John, 392, 415
Mæhlum, Petter, xxii, 390
Mahieu, Marc-Antoine, 388
Mahootian, Shahrzad, 392
Maiden, Martin, 389
The typology and semantics of binominal lexemes

Maksunova, Zoya, 390
Mallinson, Graham, 391
Marchand, Hans, 34, 105, 113, 185
Marino, Mary Carolyn, 389
Marlett, Stephen A., xxii, 391, 438
Marlett, Steve, xxii, 391, 438
Martin, Marieke, xxii, 392
Martins, Marci Fileti, 390, 439
Masini, Francesca, xxii, 13, 45, 52, 337, 389
Mathesius, Vilém, 9
Mathiassen, Terje, 390, 409
Matthews, P.H., 164
Mätzner, Eduard, 49, 184, 199, 201
McConvell, Patrick, 388
McGloin, Naomi Hanaoka, 389, 421
McGregor, Ronald Stuart, 389, 414
McGregor, William, 269
McWhorter, John, 391, 446
Meyer, Ronny, xxii
Michaelis, Susanne, 391
Miestamo, Matti, xxii, 87, 88, 99, 388
Mihatsch, Wiltrud, 13
Mistrík, Jozef, 391
Mitchell, Alice, xxii, 388, 405
Miyaoka, Osahito, 388, 435
Moldovan, Dan, 188
Möller Nwadigo, Mirjam, xxii, 387, 402
Montgomery, Christine Anne, 390, 405
Mottin, Jean, 389, 430
Mourigh, Khalid, xxii, 391
Mous, Maarten, 389
Munroe, Randall, 234
Næss, Áshild, xxii, 387, 429
Nagórko, Alicja, 147, 391
Nakov, Preslav, 188
Naughton, James, 388, 391, 420
Navickaitė-Klišauskienė, Agnė, xxii, 390
Nedjalkov, Igor, 391, 392, 406
Nefedov, Andrey, xxii, 390
Nelson, Ted, 4
Nercesian, Verónica, 392, 444
Newman, Paul, 161, 388, 396
Ngom, Fallou, 392, 404
Nguyễn, Dinh-Hoà, 28, 261, 392
Nichols, Johanna, 59, 86, 87, 138, 142
Nicolaes, Alexandru, 391, 417
Noailly, Michèle, 203, 206
Nordbustad, Frøydis, 389, 397
Novgorodov, Innokentij, xxii, 392
Ó Séaghdha, Diarmuid, 188
O’Brien, Richard J., 390, 405
Oinas, Felix J., 423
Olsen, Susan, 37
Osgarby, David, xxii, 392
Packard, Jerome L., 264
Palancar, Enrique, 391, 437
Parajuli, Krishna Prasad, xxii, 390
Parker, G. W., 97
Patgiri, Bipasha, xxii, 387
Peirsman, Yves, 21, 193, 318, 319, 320, 327
Pepper, Steve, 2, 3, 4, 5, 6, 10, 25, 26, 27, 28, 37, 53, 68, 69, 184, 188, 195, 216, 254, 266, 268, 274, 275, 276, 292, 295, 296, 327, 328, 329, 330, 331, 338, 344, 390, 514
Perkins, Revere D., 87
Pešková, Pavlína, xxii, 388
Petegham, Marleen Van, 391
Plank, Frans, 335
Pleshak, Polina, xxii, 392
Pohontsch, Anja, 390
Polomé, Edgar Charles, 391, 401
Praušinš, Dace, 390
Proudfoot, Anna, 389, 416
Pukui, Mary Kawena, 389, 429
Pustejovsky, James, 185, 312
Putten, Marijn van, 391
Quesada, Juan Diego, 387
Quinn, Kyla, xxii, 392, 432
R Core Team, xxv
Raghunathan, Arathi, xxii, 390
Rainer, Franz, 14, 21, 48, 50, 61, 192
Ralli, Angela, 165, 265, 388, 414
Rasoloson, Janie, 390, 427
Ratliff, Martha, xxii, 389
Ravindran, P.N., 390
Reichard, Gladys A., 390, 435
Renaud-Lescure, Odile, 388
Rendón, Jorge A. Gómez, 389
Richardson, James, 97
Riese, Timothy, 392, 424
Rießler, Michael, xxii, 390, 425
Rijk, Rudolf P. G. de, 387, 408
Rijkehoff, Jan, 57, 86, 87, 88, 99, 269
Rose, Françoise, xxii, 392, 441
Ross, Malcolm, xxii, 391, 429
Rottland, Franz, 388, 405
Rounds, Carol H., 389
Russell, Joan, 150
Ryder, Mary Ellen, 184, 185, 187, 217
Saade, Benjamin, xxii, 390
Sadock, Jerrold M., 29, 175
Saeed, John, 391, 399
Saltarelli, Mario, 387
Saulwick, Adam, 89
Saussure, Ferdinand de, 123
Scalise, Sergio, 7, 21, 32, 33, 34, 35, 36, 37, 38, 39, 40, 110, 111, 112, 144, 183, 210, 216, 256, 264, 297, 337, 389, 489
Schaaik, Gerjan van, 68, 407
Schachter, Paul, 42, 391, 427, 428
Schadeberg, Thilo, 391
Schäfer, Martin, 188
Schmidt, Christopher K., 389
Schuhmann, Roland, 391
Schulte, Kim, 391
Schuster-Šewc, Heinz, 390
Shadbolt, Nigel, 5
Sijs, Nicoline van der, 388
Singnoi, Unchalee, 392, 431
Smeets, Ineke, 107, 108, 390, 441
Smith, Benjamin C., 390, 391
Smyth, David, 392
Sneddon, James Neil, 389, 428
Snyder, William, 40
Søgaard, Anders, 185, 187, 203, 204
Sohn, Ho-min, 390, 422
Sokolova, Svetlana, xxii, 391
Song, Jae Jung, 17, 43, 44, 87, 390, 422
Spencer, Andrew, 45
Stachowski, Marek, 392, 408
Stassen, Leon, 131, 269
Štekauer, Pavol, xxi, xxii, 7, 8, 9, 10, 11, 21, 44, 45, 48, 61, 88, 113, 114, 116, 337, 391, 490
Štichauer, Pavel, 388
Stone, Gerald, 389, 390, 419, 436
Sultkala, Helena, 388
Suthiwan, Titima, xxii, 392
Sveen, Andreas, xxi
Swan, Oscar E., 391, 419
Szubert, Andrzej, 184, 188
Szymanek, Bogdan, 391
Takada, Hareo, 41
Tarasova, Elizaveta, 14, 18, 21, 48, 51, 52, 61, 187, 188, 190, 192, 216
Tauli, Valter, 388
Taylor, John R., 298
Tehranisa, Seyed Hassan, 140, 264, 392
Teng, Stacy Fang-Ching, 391, 426
Ternes, Elmar, 264
Theil Endresen, Rolf, 68, 391
Thorne, David, 392
Timberlake, Alan, 391, 420
Tompa, József, 389
Toquero, Luis Miguel, 184, 188
Törkenczy, Miklós, 424
Tosco, Mauro, xxii, 388, 397
Tratz, Stephen, 10, 188
Treis, Yvonne, xxii, 389, 398
Tuggy, David, 164, 184, 185, 337
Tulu, Gebrew, 388, 397
Turkina, E., 390
Tzoc, Juan, 389
Ullmann, Stephen, 123, 325
Ungnad, Arthur, 387, 399
Urban, Matthias, xxi, 21, 56, 57, 58, 59, 60, 61, 65, 70, 85, 112, 148, 324
Urbanik, Pavel, xxii, 391
Vajda, Edward J., xxii, 390, 425
Van de Velde, Mark, 391
Van Egmond, Marie-Elaine, xxii, 163, 387
Van linden, An, xxii, 165, 388, 443
Vanderwende, Lucy, 188, 198, 221
Vaovao, Andro, 97
Veenstra, Tonjes, 391
Vennemann, Theo, 210
Verhoeven, Elisabeth, 387, 443
Vervecken, Katrien Dora, 13
Vidal, Alejandra, 392
Villena, Belén, xxii, 390
Visser, Eline, xxii, 389, 434
Wade, Terence, 391, 420
Wälchli, Bernhard, 9, 241
Warren, Beatrice, 49, 185, 186, 195, 221, 328
Watson, Rachel, xxii, 150, 152, 387, 404
Whaley, Lindsay J., 391
The typology and semantics of binominal lexemes

Whittle, Ruth, 388, 412
Wichmann, Søren, 389
Wiebusch, Thekla, 390
Williams, David-Antoine, 234
Williams, Edwin, 39, 113
Wojtylak, Kasia, xxii, 390, 444
Wojtylak, Katarzyna I., xxii, 390, 444
Worthington, Martin, 250, 387
Wotango, Deginet, xxii, 389
Wu, Shuqiong, xxii, 390

Wurm, Stephen A., 387
Yeon, Jaehoon, 165, 390
Yip, Po-ching, 390
Young, Robert W., 390, 435
Yri, Kjell-Magne, xxi, xxii, 391
Zimmer, Stefan, 68
Zoller, Claus Peter, xxii, 52, 389
Zúñiga, Fernando, xxii, 254, 279, 326, 390
Zwicky, Arnold M., 40, 113
Language index

Äiwoo (nfl, Austronesian, Oceanic), xxv, 107, 151, 165, 171, 172, 221, 258, 275, 277, 278, 286, 384, 387, 429, 485, 486

Akkadian (akk, Afro-Asiatic, Semitic), 122, 159, 160, 172, 174, 177, 250, 276, 277, 286, 383, 387, 399, 485, 486

Amharic (amh, Afro-Asiatic, Semitic), 171, 172, 277, 286, 312, 313, 383, 387, 398, 399, 485, 486

Anindilyakwa (aoi, Gunwinyguan), 163, 172, 276, 277, 285, 286, 383, 387, 432, 485, 486

Arabela (arl, Zaparoan), 149

Arbore (arv, Afro-Asiatic, Cushitic), 168

Archi (aqc, Nakh-Daghestanian, Lezgic), 53, 55, 63, 73, 116, 157, 172, 277, 286, 383, 387, 422

Assamese (asm, Indo-European, Indo-Aryan), 129, 172, 178, 246, 277, 286, 383, 387, 414

Baa (kwb, Atlantic-Congo, Gur), 124, 147, 171, 172, 277, 286, 303, 384, 387, 402

Bambara (bam, Atlantic-Congo, Mande), 30, 147, 159, 172, 277, 286, 383, 387, 403

Bandial (bqj, Atlantic-Congo, North-Central Atlantic), 150, 151, 152, 172, 258, 277, 286, 383, 387, 404

Barain (bva, Afro-Asiatic, Chadic), 155, 158, 172, 263, 265, 266, 267, 277, 286, 315, 383, 387, 396

Basque (eus, Basque), 22, 30, 31, 33, 91, 172, 277, 286, 383, 387, 408

Bezhta (kap, Nakh-Daghestanian, Avar-Andic-Tsezic), 14, 49, 53, 57, 74, 75, 133, 142, 156, 157, 172, 189, 221, 277, 286, 384, 387, 422, 495

Bora (boa, Boran), 11, 14, 49, 149

Cabécar (cjp, Chibchan), 146, 172, 277, 286, 383, 387, 442

Cajia (cai, Sino-Tibetan, Macro-Bai), xxv, 146, 172, 173, 174, 230, 277, 286, 383, 388, 431

Cantonese (yue, Sino-Tibetan, Sinitic), 255

Central Yupik (esu, Eskimo-Aleut, Yupik), xxiii, 89, 98, 128, 159, 171, 172, 177, 194, 277, 286, 383, 388, 435

Ceq Wong (cwg, Austro-Asiatic, Aslian), 53, 73, 122, 146, 172, 174, 277, 383, 388, 425

Chakali (cli, Atlantic-Congo, Gur), 172, 277, 285, 383, 388, 403

Cora (crn, Uto-Aztecan), 115

Croatian (hrv, Indo-European, Slavic), 33, 39, 157, 172, 276, 277, 384, 388, 418

Czech (ces, Indo-European, Slavic), 8, 10, 33, 133, 142, 147, 150, 152, 157, 172, 177, 193, 194, 277, 286, 383, 388, 418

Datooga (tcc, Nilo-Saharan, Nilotic), 122, 124, 146, 172, 174, 249, 277, 385, 388, 405, 490

Dutch (nld, Indo-European, Germanic), 33, 39, 53, 73, 74, 75, 95, 164, 172, 262, 277, 285, 320, 385, 388, 411, 490, 495

English (eng, Indo-European, Germanic), xxv, 8, 15, 23, 24, 30, 31, 33, 38, 39, 41, 42, 49, 51, 52, 53, 57, 62, 66, 70, 73, 83, 95, 96, 97, 98, 104, 107, 109, 112, 113, 115, 123,
The typology and semantics of binominal lexemes


Estonian (est, Uralic, Finnic), 42, 157, 172, 198, 203, 226, 262, 277, 383, 388, 423, 440, 490

Fijian (fij, Austronesian, Oceanic), 58

Finnish (fin, Finnish, Uralic, Finnic), 22, 33, 172, 192, 222, 273, 383, 388, 423, 489, 490


Gawwada (gwd, Afro-Asiatic, Cushitic), 53, 72, 73, 147, 148, 150, 151, 172, 263, 277, 384, 388, 397

German (deu, Indo-European, Germanic), xxv, 1, 8, 14, 23, 30, 31, 33, 39, 42, 50, 52, 71, 72, 73, 74, 133, 142, 148, 164, 165, 166, 172, 180, 184, 188, 192, 194, 262, 277, 285, 320, 383, 388, 412, 490

Greek (ell, Indo-European, Greek), 33, 156, 165, 166, 172, 178, 199, 257, 261, 263, 265, 277, 320, 323, 383, 388, 413, 414, 489, 490

Gurindji (gue, Pama-Nyungan, Desert Nyungic), 172, 249, 277, 384, 388, 434, 439

Harakmbut (amr, Harakmbut), 91, 151, 153, 154, 165, 172, 258, 277, 383, 388, 443

Hausa (hau, Afro-Asiatic, Chadic), 53, 58, 73, 93, 94, 95, 147, 152, 158, 161, 172, 194, 276, 277, 384, 388, 396, 490, 495

Hawaiian (haw, Austronesian, Oceanic), 53, 73, 95, 147, 152, 172, 222, 226, 241, 277, 312, 313, 384, 389, 429

Hebrew (heb, Afro-Asiatic, Semitic), 14, 22, 24, 33, 97, 136, 140, 159, 172, 192, 194, 276, 277, 285, 384, 389, 400, 490

Hindi (hin, Indo-European, Indo-Aryan), 52, 71, 75, 77, 155, 167, 172, 178, 259, 277, 384, 389, 414, 490, 495

Hixkaryána (hix, Cariban), 28, 255

Hmong Daw (mww, Hmong-Mien), 9, 53, 72, 73, 172, 241, 277, 384, 389, 430

Ho-Chunk (win, Siouan), 71, 146, 172, 277, 385, 389, 438

Hungarian (hun, Uralic, Hungarian), 30, 31, 33, 50, 71, 73, 147, 157, 168, 172, 277, 384, 389, 424, 490

Hupdë (jup, Nadahup), 53, 73, 172, 241, 277, 384, 389, 445

Imbabura Quechua (qvi, Quechuan, Quechua II), 22, 73, 146, 172, 174, 177, 313, 385, 389, 445

Indonesian (ind, Austronesian, Malayo-Sumbawan), 53, 73, 95, 172, 262, 277, 384, 389, 428, 490, 495

Iraqw (irk, Afro-Asiatic, Cushitic), 53, 72, 73, 95, 98, 171, 172, 221, 277, 285, 384, 389, 397

Irish (gle, Indo-European, Celtic), 71, 73, 75, 77, 147, 156, 168, 172, 179, 180, 259, 262, 263, 277, 384, 389, 410, 495


Japanese (jpn, Japonic, Japanese), xxv, 24, 33, 39, 41, 49, 53, 67, 69, 73, 95,
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Subfamily</th>
<th>Genus</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamang</td>
<td>kgv</td>
<td>West Bomberai</td>
<td>108, 156, 159, 172, 263, 277, 285, 299, 300, 301, 302, 315, 384, 389, 434</td>
<td></td>
</tr>
<tr>
<td>Kam</td>
<td>kdx</td>
<td>Atlantic-Congo, Adamawa-Ubangi</td>
<td>22, 172, 277, 315, 384, 385, 389, 401, 431</td>
<td></td>
</tr>
<tr>
<td>Kambaata</td>
<td>ktb</td>
<td>Afro-Asiatic, Cushitic</td>
<td>107, 159, 172, 277, 384, 389, 398</td>
<td></td>
</tr>
<tr>
<td>Kanuri</td>
<td>knc</td>
<td>Nilo-Saharan, Saharan</td>
<td>22, 26, 28, 53, 72, 73, 77, 95, 128, 157, 172, 255, 263, 277, 285, 384, 389, 406, 495</td>
<td></td>
</tr>
<tr>
<td>Kekchi</td>
<td>kek</td>
<td>Mayan, Core Mayan</td>
<td>53, 73, 157, 159, 172, 277, 384, 389, 436, 496</td>
<td></td>
</tr>
<tr>
<td>Ket</td>
<td>ket</td>
<td>Yeniseian</td>
<td>53, 72, 73, 74, 157, 172, 222, 241, 246, 277, 285, 315, 384, 390, 425, 435, 490, 495</td>
<td></td>
</tr>
<tr>
<td>Khmer</td>
<td>khm</td>
<td>Austroasiatic, Khmeric</td>
<td>22, 24, 30, 82</td>
<td></td>
</tr>
<tr>
<td>Khoekhoe</td>
<td>naq</td>
<td>Khoe</td>
<td>58, 490</td>
<td></td>
</tr>
<tr>
<td>Kildin Sami</td>
<td>sjd</td>
<td>Uralic, Saami</td>
<td>13, 53, 58, 146, 156, 157, 172, 277, 315, 385, 390, 425</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>kor</td>
<td>Koreanic</td>
<td>33, 39, 122, 147, 165, 166, 172, 277, 301, 384, 390, 421, 422</td>
<td></td>
</tr>
<tr>
<td>Kupsabiny</td>
<td>kpz</td>
<td>Nilo-Saharan, Nilotic</td>
<td>155, 159, 172, 177, 277, 384, 390, 405</td>
<td></td>
</tr>
<tr>
<td>Latvian</td>
<td>lav</td>
<td>Indo-European, Baltic</td>
<td>172, 180, 277, 384, 390, 409</td>
<td></td>
</tr>
<tr>
<td>Lenakel</td>
<td>tnl</td>
<td>Austronesian, Oceanic</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Lithuanian</td>
<td>lit</td>
<td>Indo-European, Baltic</td>
<td>12, 42, 71, 74, 134, 156, 166, 172, 193, 243, 270, 272, 273, 277, 284, 390, 409, 495</td>
<td></td>
</tr>
<tr>
<td>Lower Sorbian</td>
<td>dsb</td>
<td>Indo-European, Slavic</td>
<td>53, 72, 73, 155, 156, 157, 171, 172, 194, 276, 277, 383, 390, 419, 495</td>
<td></td>
</tr>
<tr>
<td>Malagasy</td>
<td>plt</td>
<td>Austronesian, Greater Barito</td>
<td>49, 53, 73, 74, 75, 77, 96, 97, 98, 129, 142, 144, 147, 159, 172, 258, 277, 285, 315, 385, 390, 427, 496</td>
<td></td>
</tr>
<tr>
<td>Malayalam</td>
<td>mal</td>
<td>Dravidian</td>
<td>172, 277, 315, 384, 390, 408, 414, 490</td>
<td></td>
</tr>
<tr>
<td>Maltese</td>
<td>mlt</td>
<td>Afro-Asiatic, Semitic</td>
<td>155, 159, 160, 166, 168, 172, 276, 277, 384, 390, 400</td>
<td></td>
</tr>
<tr>
<td>Mamara Senoufo</td>
<td>myk</td>
<td>Atlantic-Congo, Kwa</td>
<td>127, 172, 384, 390, 403</td>
<td></td>
</tr>
<tr>
<td>Manange</td>
<td>nmm</td>
<td>Sino-Tibetan, Bodic</td>
<td>53, 73, 156, 172, 277, 385, 390, 430</td>
<td></td>
</tr>
<tr>
<td>Mapudungun</td>
<td>arn</td>
<td>Araucanian</td>
<td>53, 73, 77, 107, 147, 148, 172, 254, 263, 264, 265, 277, 279, 280, 286, 299, 383, 390, 441, 496</td>
<td></td>
</tr>
<tr>
<td>Marra</td>
<td>mec</td>
<td>Mangarrayi-Maran</td>
<td>22, 26, 28, 255</td>
<td></td>
</tr>
<tr>
<td>Mbyá Guaraní</td>
<td>gun</td>
<td>Tupian, Tupi-Guarani</td>
<td>172, 241, 277, 384, 390, 399, 439</td>
<td></td>
</tr>
<tr>
<td>Minangkabau</td>
<td>min</td>
<td>Austronesian, Malayic</td>
<td>149, 150</td>
<td></td>
</tr>
<tr>
<td>Nepali</td>
<td>npi</td>
<td>Indo-European, Indo-Aryan</td>
<td>167, 172, 178, 259, 277, 385, 390, 415</td>
<td></td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>kmr</td>
<td>Indo-European, Iranian</td>
<td>168, 273</td>
<td></td>
</tr>
<tr>
<td>Norwegian</td>
<td>nor</td>
<td>Indo-European, Germanic</td>
<td>8, 33, 39, 136, 172, 188, 2020-05-26 Submitted version 507</td>
<td></td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

223, 235, 262, 277, 301, 385, 390, 413
Oko (okoo1245, Atlantic-Congo, Oko-Eni-OSayen), 254, 264
Old High German (goh, Indo-European, Germanic), 53, 74, 172, 180, 384, 391, 413, 496
Oroqen (orh, Altaic, Tungusic), 54, 73, 159, 172, 250, 277, 315, 385, 391, 406, 407, 496
Portuguese (por, Indo-European, Romance), 33, 42, 490
Puyuma (pyu, Austronesian, Formosan), 89, 124, 171, 172, 177, 249, 277, 385, 391, 426
Querétaro Otomi (otq, Otopamean, Otopamean), 54, 127, 172, 194, 385, 391, 437
Rama (rma, Chibchan), 58, 414
Romanian (ron, Indo-European, Romance), 54, 95, 108, 172, 259, 263, 277, 385, 391, 417, 490, 496
Russian (rus, Indo-European, Slavic), xxv, 1, 8, 14, 33, 39, 95, 97, 128, 133, 142, 157, 162, 166, 167, 172, 189, 193, 194, 272, 273, 276, 277, 285, 385, 391, 420, 490
San Mateo del Mar Huave (huv, Huavean), 58
Saramaccan (srm, Pidgins & Creoles, English-based), 54, 73, 92, 146, 172, 277, 285, 315, 385, 391, 446
Selice Romani (rmc, Indo-European, Indo-Aryan), 54, 124, 171, 172, 177, 178, 249, 275, 277, 278, 385, 391, 415, 496
Seri (sei, Seri), 91, 159, 172, 250, 277, 285, 385, 391, 438
Seychelles Creole (crs, Pidgins & Creoles, French-based), 54, 73, 75, 77, 92, 127, 146, 172, 174, 383, 391, 447, 497
Sidamo (sid, Afro-Asiatic, Cushitic), 147, 156, 157, 172, 277, 285, 286, 385, 391, 398
Slovak (slk, Indo-European, Slavic), 10, 11, 14, 49, 157, 167, 172, 177, 189, 277, 385, 391, 420, 490
Somali (som, Afro-Asiatic, Cushitic), 159, 160, 172, 277, 385, 391, 399
Spanish (spa, Indo-European, Romance), 13, 112, 325
Sranen (srs, Nuclear Torricelli), 122, 172, 277, 384, 391, 433
Swahili (swh, Atlantic-Congo, Bantoid), 49, 54, 58, 72, 73, 94, 95, 124, 150, 151, 152, 153, 155, 168, 169, 172, 258, 277, 385, 391, 401, 490
Tagalog (tgl, Austronesian, Greater Central Philippine), 42, 152, 155, 172, 277, 286, 312, 313, 315, 385, 391, 427
Takia (tbc, Austronesian, Oceanic), 12, 54, 72, 73, 74, 77, 98, 112, 142, 155, 156, 159, 172, 192, 221, 250, 277, 385, 391, 429, 497
Taklema (tkm, extinct isolate), 22, 24
Tarifit (rif, Afro-Asiatic, Berber), xxv, 49, 54, 64, 74, 140, 156, 168, 169, 172, 277, 385, 391, 395
Thai (tha, Tai-Kadai, Kam-Tai), 22, 54, 73, 94, 97, 140, 147, 149, 172, 192, 277, 385, 392, 431
Ticuna (tca, Ticuna-Yuri), 91, 172, 277, 385, 392, 446
Toaripi (tqo, Eleman), 58
Trinitario (trn, Arawakan), 151, 153, 154, 172, 194, 258, 263, 277, 314, 385, 392, 441
Turkish (tur, Altaic, Turkic), 1, 14, 30, 31, 33, 66, 68, 95, 172, 272, 273, 277, 285, 315, 385, 392, 407, 489
Tuwari (tww, Walio), 122, 146, 172, 174, 249, 277, 385, 392, 434
Tz’utujil (tzj, Mayan, Core Mayan), 255
Walman (van, Nuclear Torricelli), 146, 172, 174, 277, 385, 392, 433
Warta Thuntai (gnt, Morehead-Wasur), 172, 277, 384, 392, 432
Wawa (www, Atlantic-Congo, Bantoid), 122, 172, 277, 315, 385, 392, 402
West Greenlandic (kal, Eskimo-Aleut, Inuit), 175, 255
Western Farsi (pes, Indo-European, Iranian), 146, 172, 263, 264, 277, 285, 286, 314, 315, 385, 392, 415
Western Mari (mrj, Uralic, Mari), 146, 156, 172, 277, 384, 392, 424
Wichi (mzh, Matacoan), 54, 72, 172, 277, 312, 313, 384, 392, 444, 490
Wik-Mungkan (wim, Pama-Nyungan, Paman), 172, 277, 285, 385, 392, 434
Wolof (wol, Atlantic-Congo, North-Central Atlantic), 172, 277, 285, 286, 385, 392, 404
Yakut (sah, Altaic, Turkic), 54, 72, 73, 146, 156, 157, 159, 172, 262, 277, 285, 385, 392, 408, 497
Yaqui (yaq, Uto-Aztecan, Southern Uto-Aztecan), 54, 73, 75, 76, 172, 222, 277, 285, 385, 392, 436, 440, 497
Yimas (yee, Lower Sepic Ramu), 22, 24
Zinacantán Tzotzil (tzo, Mayan, Core Mayan), 54, 73, 74, 75, 127, 172, 385, 392, 437, 490, 497
Subject index

<table>
<thead>
<tr>
<th>Term</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>action-affix</td>
<td>12, 101, 108, 109, 130</td>
</tr>
<tr>
<td>action-morph</td>
<td>102, 253</td>
</tr>
<tr>
<td>action-root</td>
<td>12, 49, 72, 101, 102, 107, 109, 112, 124, 130, 228</td>
</tr>
<tr>
<td>affixoid</td>
<td>104, 137, 164, 189</td>
</tr>
<tr>
<td>association</td>
<td>2, 3, 4, 5, 10, 21, 46, 48, 53, 56, 59, 60, 218, 226, 280, 282, 283, 284, 298, 325, 327, 328, 330</td>
</tr>
<tr>
<td>associative type</td>
<td>2, 3, 218, 328, 330</td>
</tr>
<tr>
<td>associative relation</td>
<td>1, 4, 5, 19, 51, 193, 194, 209, 215, 228, 317, 318, 319, 321, 324, 326, 327, 328, 331, 332, 334, 336, 339, 340, 486</td>
</tr>
<tr>
<td>binominal fingerprint</td>
<td>167, 169, 177, 180, 181, 259, 346</td>
</tr>
<tr>
<td>binominal lexeme (definition)</td>
<td>9, 13</td>
</tr>
<tr>
<td>blend</td>
<td>24, 52, 400</td>
</tr>
<tr>
<td>Bourque2 classification</td>
<td>224, 235, 236, 237, 238, 298, 486</td>
</tr>
<tr>
<td>classifier</td>
<td>xxiii, 11, 12, 13, 14, 49, 50, 52, 133, 134, 135, 140, 142, 143, 148, 149, 150, 151, 153, 154, 165, 210, 253, 429, 431, 443</td>
</tr>
<tr>
<td>CLDF (Cross-Linguistic Data Formats)</td>
<td>93</td>
</tr>
<tr>
<td>CLLD (Cross-Linguistic Linked Data)</td>
<td>326</td>
</tr>
<tr>
<td>Cognitive Grammar</td>
<td>6, 115, 184, 185, 187, 296</td>
</tr>
<tr>
<td>colour terms</td>
<td>12, 106</td>
</tr>
<tr>
<td>compound types</td>
<td></td>
</tr>
<tr>
<td>aggregative</td>
<td>23</td>
</tr>
<tr>
<td>appositional</td>
<td>199, 203, 207, 208, 209, 210, 211, 214, 215, 216, 236, 250, 252, 284, 295, 347</td>
</tr>
<tr>
<td>attributive</td>
<td>7, 35, 36, 40, 264</td>
</tr>
<tr>
<td>avyayibhava</td>
<td>23</td>
</tr>
<tr>
<td>coordinate</td>
<td>9, 23, 35, 36, 38, 40, 42, 101, 102, 109, 111, 112, 203, 209, 216, 241, 252, 257, 284, 297, 313, 441</td>
</tr>
<tr>
<td>copulative</td>
<td>23, 45, 208</td>
</tr>
<tr>
<td>determinative</td>
<td>7, 10, 23, 29, 198, 199, 202, 203, 216, 239, 332, 447</td>
</tr>
<tr>
<td>dvandva</td>
<td>23</td>
</tr>
<tr>
<td>endocentric</td>
<td>35, 36, 39, 41, 42, 51, 190</td>
</tr>
<tr>
<td>exocentric</td>
<td>23, 30, 35, 36, 39, 40, 42, 45, 112, 222</td>
</tr>
<tr>
<td>non-appositional</td>
<td>81, 84, 198, 202, 207, 209, 215, 216, 236, 331</td>
</tr>
<tr>
<td>possessive</td>
<td>132, 135, 268</td>
</tr>
<tr>
<td>prepositional</td>
<td>8, 13, 14, 214</td>
</tr>
<tr>
<td>root</td>
<td>8, 14, 24, 40</td>
</tr>
<tr>
<td>subordinate</td>
<td>7, 9, 35, 36, 38, 40, 41, 204, 264, 279, 280, 297, 300, 303, 441</td>
</tr>
<tr>
<td>synthetic</td>
<td>10, 11, 12, 13, 23, 24, 42, 43, 54, 132, 133, 141, 142, 257, 270, 332</td>
</tr>
<tr>
<td>tatpuruṣa</td>
<td>23, 24, 239</td>
</tr>
<tr>
<td>verbal-nexus</td>
<td>23, 35</td>
</tr>
<tr>
<td>contiguity</td>
<td>6, 59, 60, 193, 194, 209, 210, 215, 281, 317, 319, 325, 326, 327, 328, 331, 334</td>
</tr>
<tr>
<td>convenience sample</td>
<td>86, 87, 89</td>
</tr>
<tr>
<td>CSV (comma-separated values)</td>
<td>64, 93</td>
</tr>
<tr>
<td>elaboration site</td>
<td>115, 185, 295, 302, 308, 309, 310, 311, 312, 315, 328, 336, 342, 343</td>
</tr>
<tr>
<td>expressive binominal NPs</td>
<td>13</td>
</tr>
<tr>
<td>ezāfe</td>
<td>140, 415</td>
</tr>
<tr>
<td>flag</td>
<td>137, 138, 259</td>
</tr>
<tr>
<td>genetic grouping</td>
<td>86, 490</td>
</tr>
<tr>
<td>geographical area</td>
<td>25, 46, 59, 86, 88, 92, 170, 178, 258, 287, 288, 291, 344, 346</td>
</tr>
</tbody>
</table>
The typology and semantics of binominal lexemes

glotcode, xxv, 388, 485, 486
Glottolog, xxv, 22, 89, 383, 490
Hatcher2 classification, 237, 238, 254, 298, 486
ICM (Idealized Cognitive Model), 319
index, xxv, 6, 134, 135, 137, 138, 139, 140, 141, 156, 158, 161, 259, 273, 348, 499, 505, 511
intralingual competition, 169, 170, 175, 177, 181, 274
JSON (JavaScript Object Notation), 93
Leipzig Glossing Rules, xxv, 55, 95, 98, 448
linker, 42, 134, 135, 138, 139, 140, 141, 154, 155, 158, 160, 161, 166, 313, 396, 397, 428, 443
Lucca, 2, 329
Madame Butterfly, 2
metaphor, 36, 39, 209, 210
metonymy, 1, 21, 36, 39, 61, 193, 194, 209, 210, 318, 319, 321, 322, 323, 324, 326, 327, 328, 330, 332, 336, 348
Morbo/Comp project, 21, 32, 33, 34, 37, 39, 40, 61, 62, 63, 88, 256, 257, 258, 291, 489
motivational grid, 21, 281, 282, 292, 298, 299, 300, 324, 325, 331, 345, 348
nominal modification construction, 14, 48, 51, 188, 190, 253
anchoring, 15, 21, 181, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 292, 327, 332, 336, 344, 395, 417
non-anchoring, 21, 270, 271, 272, 395, 417
typifying, 270, 271, 273, 274, 332
occurrence, 2, 83, 98, 169, 177, 327, 395, 406, 431, 491, 493
onomasiological base, 10, 11, 66, 113, 114, 116
onomasiological mark, 10, 11, 13, 114
Onomasiological Types, 10, 11, 24, 51, 72, 102, 108, 114, 129, 222, 324, 332, 333
Pepper scale, 275, 276, 292, 338, 344
person index, 137, 138, 139, 158, 159, 161
phrasal lexeme, 45, 48, 58, 337
possessive noun phrase, 131, 132, 133, 134, 162, 268, 270, 274, 348, 407, 438, 442
Principal Components Analysis (PCA), 179
probability sample, 86, 87, 89, 91
property-affix, 12, 101, 108, 109, 130
property-morph, 130
property-root, 12, 84, 101, 109, 130, 216
Puccini, 2, 327, 329, 331
Radical Construction Grammar, 1, 14
RDF (Resource Description Framework), 5, 93, 326
reference point ability, 6, 37, 296, 299
relational adjective, 14, 48, 50, 106, 143, 157, 192, 273, 284
relator, 137, 138, 139, 154, 155, 158, 160
role type, 4
semantic category, 45, 46, 48, 54, 55, 63, 66, 81, 98
semantic field, 54, 55, 60, 77, 78, 79, 80, 84, 125, 127, 128, 239, 246, 247, 251, 252, 253, 346, 393
SGML (Standard Generalized Markup Language), 56, 218
similarity, 9, 59, 60, 186, 193, 197, 199, 209, 210, 211, 215, 216, 233, 237, 238, 247, 249, 250, 252, 274, 276, 278, 281, 292, 317, 322, 325, 326, 331, 332, 344, 347
subordination, 35, 37, 281, 297, 325
synset, 305, 308
TAO of Topic Maps, 2, 327
ting-affix, 12, 49, 101, 102, 108, 110, 111, 113, 116, 130, 137, 147, 149, 164, 189, 258, 335
ting-morph, 12, 13, 63, 101, 102, 108, 116, 130, 141, 145, 151, 183, 253, 285
thing-root, 12, 49, 84, 101, 102, 104, 106, 107, 108, 110, 111, 112, 113, 116, 124, 130, 137, 145, 146, 147, 149, 162, 164, 176, 189, 216, 222, 335

topic, 2, 4, 5, 18, 23, 61, 326, 327, 336, 347
topic map, 2, 5, 218, 329, 330
Topic Maps, xxi, 2, 4, 5, 7, 19, 56, 218, 318, 326, 329, 330
topic type, 2, 3, 218
Tosca, 2, 327, 329, 331
type binominals, 13

variety sample, 57, 86, 87, 88, 89, 99
WALS (World Atlas of Language Structures), 59, 86, 89, 218, 256, 326, 490
WOLD (World Loanword Database), xxi, 1, 53, 54, 55, 56, 57, 62, 63, 66, 67, 69, 70, 71, 77, 81, 84, 88, 89, 92, 93, 94, 95, 96, 97, 98, 99, 116, 122, 128, 164, 326, 387, 397
word order, 18, 25, 26, 27, 29, 46, 214, 255, 259, 264, 348, 402, 426, 428, 435
WordNet, 305, 348
Steve Pepper

The typology and semantics of binominal lexemes

Noun-noun compounds and their functional equivalents
Dissertation for the Degree of PhD 2019
Department of Linguistics and Scandinavian Studies
Faculty of Humanities  2020-05-26