Steve Pepper

The typology of binominal lexemes

Noun-noun compounds and their functional equivalents
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)
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>Table of Tables</td>
<td>xvii</td>
</tr>
<tr>
<td>Preface</td>
<td>xix</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xxi</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xxiii</td>
</tr>
<tr>
<td>Typographical and naming conventions</td>
<td>xxv</td>
</tr>
<tr>
<td>1 Towards a comparative concept</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Topic Maps and associative thought</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 Nominal compounding in Nizaa</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Binominal lexemes as a comparative concept</td>
<td>7</td>
</tr>
<tr>
<td>1.2.1 The limitations of compounding</td>
<td>7</td>
</tr>
<tr>
<td>1.2.2 The potential of binominal lexemes</td>
<td>8</td>
</tr>
<tr>
<td>1.2.3 An onomasiological perspective</td>
<td>10</td>
</tr>
<tr>
<td>1.2.4 Defining the object of study</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Design of the study</td>
<td>15</td>
</tr>
<tr>
<td>1.3.1 An empirical, data-driven approach</td>
<td>15</td>
</tr>
<tr>
<td>1.3.2 Research questions</td>
<td>15</td>
</tr>
<tr>
<td>1.4 Theoretical framework</td>
<td>16</td>
</tr>
<tr>
<td>1.5 Structure of this work</td>
<td>17</td>
</tr>
<tr>
<td>2 On the shoulders of giants</td>
<td>21</td>
</tr>
<tr>
<td>2.1 Compounding</td>
<td>22</td>
</tr>
<tr>
<td>2.1.1 Bauer (2001)</td>
<td>22</td>
</tr>
<tr>
<td>2.1.2 Arnaud (2004)</td>
<td>29</td>
</tr>
<tr>
<td>2.1.3 Morbo/Comp (2004-2006)</td>
<td>32</td>
</tr>
<tr>
<td>2.1.4 Scalise &amp; Bisetto (2009)</td>
<td>34</td>
</tr>
<tr>
<td>2.1.5 Guevara &amp; Scalise (2009)</td>
<td>37</td>
</tr>
</tbody>
</table>
2.2 Word-formation

2.2.1 Aikhenvald (2007) 40
2.2.2 Štekauer, Valera & Körtvélyessy (2012) 43

2.3 Prefiguring binominals 47

2.3.1 Levi (1978) – Complex nominals 48
2.3.2 Rainer (2013) – Relational adjectives, etc. 49
2.3.3 Bauer & Tarasova (2013) – Adnominal modification 49

2.4 Morphological complexity 51

2.4.1 Haspelmath & Tadmor (2009) – WOLD 52
2.4.2 Urban (2012) 54

2.5 Chapter summary 59

3 Meanings, languages and data 61

3.1 The meaning list 61

3.1.1 Initial extraction of 159 meanings 62
3.1.2 Evaluation against five languages 65
3.1.3 Expansion to 201 meanings 69
3.1.4 Reduction to 100 meanings 69
3.1.5 Overall evaluation 80

3.2 The language sample 84

3.2.1 Types of sample 84
3.2.2 The present sample 87
3.2.3 Overall evaluation 89

3.3 Data collection 91

3.3.1 Online database 91
3.3.2 Questionnaires 94
3.3.3 Dictionaries 95
3.3.4 Data clean-up 96

3.4 Data annotation 98

3.4.1 Identifying binominals 98
3.4.2 Determining the head 105
3.4.3 Defining the construction 111

3.5 Chapter summary 113
Table of Contents

4 Formal classification 115
  4.1 Theoretical prerequisites 116
    4.1.1 Koptjevkaja-Tamm (2002; 2003) 116
    4.1.2 Croft (2003) 119
  4.2 A non-hierarchical alternative 122
  4.3 Binominal types 126
    4.3.1 No additional marker: cmp, der and cls 126
    4.3.2 One additional marker: prp, gen, adj and con 134
    4.3.3 Two additional markers: dbl 139
  4.4 Unattested types 139
    4.4.1 Head-marking correlate of prp (prn) 140
    4.4.2 Head-marking correlate of adj (nml) 142
  4.5 Gradience 142
    4.5.1 Univerbation cmp ➔ simplex 143
    4.5.2 Affixoids cmp ➔ der 143
    4.5.3 Bound nouns cmp ➔ cls 144
    4.5.4 Linking elements gen ➔ cmp 144
    4.5.5 Adpositions or case affixes? prp ➔ gen 145
    4.5.6 Inflection or derivation? gen ~ adj 146
    4.5.7 Head replacement adj ➔ der 147
    4.5.8 Morpheme loss dbl ➔ con 147
  4.6 Chapter summary 148

5 Semantic relations 149
  5.1 Theoretical prerequisites 150
    5.1.1 Background 151
    5.1.2 Nominal modification (Bauer & Tarasova 2013) 154
    5.1.3 Metonymy and derivation (Janda 2011) 159
  5.2 An integrated approach to classification 161
    5.2.1 A low-level classification (Bourque 2014) 161
    5.2.2 A high-level classification (Hatcher 1960) 164
    5.2.3 Extending Hatcher’s classification (Arnaud 2016) 170
## 5.3 Classifying binominals
- 5.3.1 The Bourquifier
- 5.3.2 Bourque 2.0
- 5.3.3 The Hatcher-Bourque classification

## 5.4 Chapter summary

## 6 Patterns in the data

### 6.1 Data description
- 6.1.1 Data description by language
- 6.1.2 Data description by meaning

### 6.2 Formal patterning
- 6.2.1 Distribution of binominals by type
- 6.2.2 Intralingual competition
- 6.2.3 Genetic and areal patterning

### 6.3 Semantic patterning
- 6.3.1 Low-level semantic relations
- 6.3.2 High-level semantic relations
- 6.3.3 Form and semantics

### 6.4 Exploratory statistics
- 6.4.1 Mining for associations
- 6.4.2 Formal types and functions
- 6.4.3 Motivational grids
- 6.4.4 Degree of marking

### 6.5 Chapter summary

## 7 Typological investigations

### 7.1 Word order typology
- 7.1.1 Constituent order
- 7.1.2 Consistency
- 7.1.3 No basic order

### 7.2 The two-paths hypothesis
- 7.2.1 Evidence from Nizaa and binominals
- 7.2.2 Salience of head and modifier
- 7.2.3 Salience of elaboration sites

## Chapter summary
7.2.4 Species-framing vs. attribute-framing 273

7.3 A model of associative relations 276

7.3.1 Metonymic relations (Peirsman & Geeraerts 2006) 277

7.3.2 Cognitive relations (Koch and Blank) 283

7.3.3 Topic Maps: roles and granularity 285

7.4 Chapter summary 290

8 Conclusion 291

8.1 Summary 291

8.2 Contribution to science 292

8.3 Further research 294

8.4 Envoie 294

References 297

Appendices 321

A. Languages 323

B. Sources 327

C. Meanings 333

D. Constructions (summary) 335

Africa 335

(Afro-Asiatic) Berber 335

(Afro-Asiatic) Chadic 335

(Afro-Asiatic) Cushitic 336

(Afro-Asiatic) Semitic 337

(Atlantic-Congo) Adamawa-Ubangi 337

(Atlantic-Congo) Bantoid 338

(Atlantic-Congo) Gur 338

(Atlantic-Congo) Kwa 338

(Atlantic-Congo) Mande 338

(Atlantic-Congo) North-Central Atlantic 339

(Nilo-Saharan) Nilotic 339

(Nilo-Saharan) Saharan 339
<table>
<thead>
<tr>
<th>Region</th>
<th>Subregion</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurasia</td>
<td>(Altaic) Tungusic</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>(Altaic) Turkic</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>Basque</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>Dravidian</td>
<td>341</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Baltic</td>
<td>341</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Celtic</td>
<td>341</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Germanic</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Greek</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Indo-Aryan</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Romance</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>(Indo-European) Slavic</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>(Japonic) Japanese</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>Koreanic</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>(Nakh-Daghestanian) Avar-Andic-Tsezic</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>(Nakh-Daghestanian) Lezgic</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>Nuclear Torricelli</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>(Uralic) Finnic</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>(Uralic) Hungarian</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>(Uralic) Mari</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>(Uralic) Saami</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>Yeniseian</td>
<td>348</td>
</tr>
<tr>
<td>Oceania/SE Asia</td>
<td>(Austro-Asiatic) Aslian</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>(Austro-Asiatic) Vietic</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>(Austronesian) Formosan</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>(Austronesian) Greater Barito</td>
<td>349</td>
</tr>
<tr>
<td></td>
<td>(Austronesian) Greater Central Philippine</td>
<td>349</td>
</tr>
<tr>
<td></td>
<td>(Austronesian) Malayo-Sumbawan</td>
<td>349</td>
</tr>
<tr>
<td></td>
<td>(Austronesian) Oceanic</td>
<td>349</td>
</tr>
<tr>
<td></td>
<td>Hmong-Mien</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>(Sino-Tibetan) Bodic</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>(Sino-Tibetan) Macro-Bai</td>
<td>350</td>
</tr>
</tbody>
</table>
Table of Contents

(Sino-Tibetan) Sinitic 350
(Tai-Kadai) Kam-Tai 350

PNG/Australia 351
Gunwinyguan 351
Nuclear Torricelli 351
(Pama-Nyungan) Desert Nyungic 351
(Pama-Nyungan) Paman 351
West Bomberai 351

North America 352
Athabaskan-Eyak-Tlingit 352
(Eskimo-Aleut) Yupik 352
(Mayan) Core Mayan 352
(Otomanguean) Otopamean 353
Seri 353
(Tupian) Tupi-Guarani 354
(Uto-Aztecan) Southern Uto-Aztecan 354

South America 354
Araucanian 354
(Cariban) Guianan 354
Chibchan 355
Huitotoan 355
Matacoan 355
Nadahup 355
(Quechuan) Quechua II 355
Ticuna-Yuri 355

Pidgins/Creoles 356
(Pidgins & Creoles) English-based 356
(Pidgins & Creoles) French-based 356

E. Binominal data set 357
Africa 357
Eurasia 364
Oceania/SE Asia 379
PNG/Australia 384
The typology of binominal lexemes

North America 385
South America 387
Pidgins/Creoles 389

F. Database structures 391
   word (w) and binominal (nn) 391
   language (l) 392
   meaning (m) 392
   semantic relation (s) 392

G. Questionnaire 393

H. Additional tables 395
   Morbo/Comp database structure 395
   Štekauer et al.’s language sample 396
   Alphabetical list of the initial 201 meanings 397
   List of 201 meanings (binominality order) 399
   Lost constructions (84 meaning sample) 401

Index of names 405
Index of languages 409
Index of subjects 415
Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The TAO model</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>The anatomy of an association</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>The five basic onomasiological types</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Extent of compounding in 13 languages (based on Arnaud 2004)</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Classifying compounds (Bisetto &amp; Scalise 2005)</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>Classifying compounds (Scalise &amp; Bisetto 2009)</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>Scales of preference in compounding (Guevara &amp; Scalise 2009)</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>Extract from questionnaire (Štekauer et al 2012)</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>Pepper as a loanword in WOLD</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>SQL query used to extract meanings from WOLD</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>View of the table <strong>meaning</strong> in its original form</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>View of the table <strong>meaning</strong> showing frequencies</td>
<td>71</td>
</tr>
<tr>
<td>13</td>
<td>Simplified view of the table <strong>word</strong> showing various constructions</td>
<td>71</td>
</tr>
<tr>
<td>14</td>
<td>Number of constructions as a function of number of meanings</td>
<td>72</td>
</tr>
<tr>
<td>15</td>
<td>Cross-linguistic representation of meanings</td>
<td>82</td>
</tr>
<tr>
<td>16</td>
<td>Distribution of meanings across semantic fields</td>
<td>83</td>
</tr>
<tr>
<td>17</td>
<td>Distribution of meanings across semantic types</td>
<td>83</td>
</tr>
<tr>
<td>18</td>
<td>Areal distribution of language sample</td>
<td>88</td>
</tr>
<tr>
<td>19</td>
<td>Extract of the data after import</td>
<td>93</td>
</tr>
<tr>
<td>20</td>
<td>Subset of the Polish data</td>
<td>100</td>
</tr>
<tr>
<td>21</td>
<td>Givón’s scale of temporal stability</td>
<td>102</td>
</tr>
<tr>
<td>43</td>
<td>Subset of the Polish binominal data showing constructions</td>
<td>112</td>
</tr>
<tr>
<td>23</td>
<td>Major structural types of PNPs in Europe</td>
<td>116</td>
</tr>
<tr>
<td>24</td>
<td>Formal classification of binominal lexemes</td>
<td>124</td>
</tr>
<tr>
<td>25</td>
<td>Degree of fusion (Bybee 1985)</td>
<td>124</td>
</tr>
<tr>
<td>26</td>
<td>Formal classification showing gradient phenomena</td>
<td>143</td>
</tr>
<tr>
<td>27</td>
<td>Hatcher’s reworking of Jespersen’s classification</td>
<td>166</td>
</tr>
<tr>
<td>28</td>
<td>Summary of Hatcher 2.0</td>
<td>182</td>
</tr>
</tbody>
</table>
Figure 29: Bourque’s template for POSSESSION .......................................................... 185
Figure 30: The Bourquifier – FAMILY ESTATE .......................................................... 186
Figure 31: The Bourquifier – BEESWAX ................................................................. 187
Figure 32: The Bourquifier – OAK TREE ................................................................. 195
Figure 33: The Bourquifier – BEAR CUB ................................................................. 195
Figure 34: The Bourquifier – SUNBURN ................................................................. 197
Figure 35: Boathouse words .................................................................................... 198
Figure 36: The Bourquifier v 2.0 ............................................................................. 201
Figure 37: Distribution of data by language and meaning ........................................ 204
Figure 38: Basic data summary (by language) .......................................................... 205
Figure 39: Morphological complexity (A/S) by language ........................................ 207
Figure 40: Binominal frequency (NN/S) by language ............................................... 208
Figure 41: Basic data summary (by meaning) ........................................................... 210
Figure 42: Vocabulary sizes across semantic fields .................................................. 211
Figure 43: Morphological complexity (A/S) by meaning .......................................... 212
Figure 44: Binominal frequency (NN/S) by meaning ................................................. 213
Figure 45: Binominal frequency (NN/A) by meaning ................................................. 214
Figure 46: Number of binominals by type ............................................................... 214
Figure 47: Areal distribution of binominal types ...................................................... 215
Figure 48: Distribution of binominal types across 99 languages ................................ 216
Figure 49: Binominal types in terms of language preferences .................................... 218
Figure 50: Geographical distribution of binominal type preferences ....................... 218
Figure 51: Intralingual competition ........................................................................ 219
Figure 52: Correlations between binominal types ................................................. 220
Figure 53: The Indo-European spectrogram ............................................................ 221
Figure 54: The Indo-European scatter plot ............................................................... 222
Figure 55: Overall frequency of low-level semantic relations ................................. 224
Figure 56: Number of languages that exhibit a particular relation ......................... 224
Figure 57: Number of meanings that exhibit a particular relation ............................ 224
Figure 58: Number of binominal types that exhibit each relation ............................ 224
Figure 59: How low-level relations vary across semantic fields ............................. 226
Figure 60: Number of binominals, languages, meanings and ftypes by htype ....... 227
Figure 61: How high-level relations vary across semantic fields ............................. 227
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>Low-level semantic relations across binominal types</td>
<td>228</td>
</tr>
<tr>
<td>63</td>
<td>High-level semantic relations across binominal types</td>
<td>229</td>
</tr>
<tr>
<td>64</td>
<td>Residuals for cells in the Polish contingency table</td>
<td>241</td>
</tr>
<tr>
<td>65</td>
<td>Non-normal distribution of marking across languages</td>
<td>244</td>
</tr>
<tr>
<td>66</td>
<td>Normal distribution of marking across meanings</td>
<td>245</td>
</tr>
<tr>
<td>67</td>
<td>Constituent order by binominal type</td>
<td>250</td>
</tr>
<tr>
<td>68</td>
<td>Constituent order (cmp) by area</td>
<td>251</td>
</tr>
<tr>
<td>69</td>
<td>Distribution of right-headed and left-headed binominals</td>
<td>253</td>
</tr>
<tr>
<td>70</td>
<td>The Bourquifier – WINDMILL</td>
<td>261</td>
</tr>
<tr>
<td>71</td>
<td>Conceptual variability of head and modifier</td>
<td>267</td>
</tr>
<tr>
<td>72</td>
<td>Scatterplot of heads and modifiers</td>
<td>269</td>
</tr>
<tr>
<td>73</td>
<td>Heatmap of heads and modifiers by semantic type</td>
<td>270</td>
</tr>
<tr>
<td>74</td>
<td>Species- vs. attribute-framing</td>
<td>275</td>
</tr>
<tr>
<td>75</td>
<td>Semantic relations as a hierarchy</td>
<td>277</td>
</tr>
<tr>
<td>76</td>
<td>The anatomy of an association</td>
<td>285</td>
</tr>
<tr>
<td>77</td>
<td>Associative relations as a hierarchy</td>
<td>289</td>
</tr>
</tbody>
</table>
# Table of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Language sample (Bauer 2001)</td>
<td>22</td>
</tr>
<tr>
<td>Table 2</td>
<td>Order of noun-adjective and noun-modifier (Bauer 2001)</td>
<td>25</td>
</tr>
<tr>
<td>Table 3</td>
<td>Noun-adjective tetrachoric tables (Bauer 2001; Pepper 2015)</td>
<td>26</td>
</tr>
<tr>
<td>Table 4</td>
<td>Order of noun-possessor and noun-modifier (Bauer 2001)</td>
<td>26</td>
</tr>
<tr>
<td>Table 5</td>
<td>Noun-possessor tetrachoric tables (Bauer 2001; Pepper 2015)</td>
<td>27</td>
</tr>
<tr>
<td>Table 6</td>
<td>Consistency of head-modifier ordering (Bauer 2001)</td>
<td>27</td>
</tr>
<tr>
<td>Table 7</td>
<td>Consistency of head-modifier ordering (Pepper 2015)</td>
<td>28</td>
</tr>
<tr>
<td>Table 8</td>
<td>Languages covered in Arnaud (2004)</td>
<td>30</td>
</tr>
<tr>
<td>Table 9</td>
<td>Onomasiological cross-linguistic comparison (Arnaud 2004)</td>
<td>31</td>
</tr>
<tr>
<td>Table 10</td>
<td>Languages represented in Morbo/Comp</td>
<td>33</td>
</tr>
<tr>
<td>Table 11</td>
<td>Morphological classification (Štekauer et al. 2012)</td>
<td>44</td>
</tr>
<tr>
<td>Table 12</td>
<td>Examples of WOLD meanings</td>
<td>53</td>
</tr>
<tr>
<td>Table 13</td>
<td>Constructions yielded by NOSTRIL</td>
<td>72</td>
</tr>
<tr>
<td>Table 14</td>
<td>Additional constructions yielded by EARLOBE</td>
<td>73</td>
</tr>
<tr>
<td>Table 15</td>
<td>Constructions lost with a sample of 84 meanings (extract)</td>
<td>74</td>
</tr>
<tr>
<td>Table 16</td>
<td>Meanings represented by lost constructions (84 meaning sample)</td>
<td>75</td>
</tr>
<tr>
<td>Table 17</td>
<td>Final set of 100 meanings</td>
<td>77</td>
</tr>
<tr>
<td>Table 18</td>
<td>Complete list of languages in the database (99)</td>
<td>89</td>
</tr>
<tr>
<td>Table 19</td>
<td>Genetic distribution of languages</td>
<td>90</td>
</tr>
<tr>
<td>Table 20</td>
<td>Areal distribution of languages</td>
<td>90</td>
</tr>
<tr>
<td>Table 21</td>
<td>Preliminary classification of structural types</td>
<td>99</td>
</tr>
<tr>
<td>Table 22</td>
<td>Types of PNP and binominal</td>
<td>118</td>
</tr>
<tr>
<td>Table 23</td>
<td>Reordered binominal types</td>
<td>123</td>
</tr>
<tr>
<td>Table 24</td>
<td>Levi’s (1978) classification applied to six binominal types</td>
<td>157</td>
</tr>
<tr>
<td>Table 25</td>
<td>Bourque’s semantic relations</td>
<td>163</td>
</tr>
<tr>
<td>Table 26</td>
<td>Variety of word-formation types – professions</td>
<td>189</td>
</tr>
<tr>
<td>Table 27</td>
<td>Summary of changes to Bourque’s system</td>
<td>199</td>
</tr>
<tr>
<td>Table 28</td>
<td>Hatcher-Bourque classification of semantic relations</td>
<td>200</td>
</tr>
</tbody>
</table>
Table 29: Overall database statistics 203
Table 30: Preferred binominal types by language 217
Table 31: Summary of high- and low-level semantic relations 223
Table 32: Most frequent low-level semantic relations 225
Table 33: Some basic association rules 234
Table 34: Koch’s motivational grid 239
Table 35: Motivational grid for Polish 239
Table 36: Chi-squared and Fisher tests of ftype2 x htype 242
Table 37: Number of markers per binominal type 243
Table 38: Constituent order pace Morbo/Comp 250
Table 39: Languages with significant degree of mixed order 255
Table 40: Motivational grid for Mapudungun 257
Table 41: Motivational grid for Kalamang 257
Table 42: Motivational grid for Greek 257
Table 43: Motivational grid for Barain 258
Table 44: Motivational grid for Ticuna 258
Table 45: Semantic relations in Nizaa 259
Table 46: Motivational grid for Kalamang with row sums 263
Table 47: Competition between cmpR and conR in Kalamang 264
Table 48: Conceptual variation in heads and modifiers 266
Table 49: Salience of elaboration sites (by relation) 271
Table 50: Salience of elaboration sites (by role) 272
Table 51: Components that elaborate the concept HOUSE 273
Table 52: Peirsman & Geeraerts’ metonymical patterns 278
Table 53: Bourquifier template for PRODUCTION 279
Table 54: Metonymic patterns and semantic relations 280
Table 55: Blank’s ten cognitive relations 283
Table 56: Association types in the Italian Opera Topic Map 287
Table 57: Morbo/Comp database structure 395
Table 58: Language sample (Štekauer, Valera & Körtvélyessy 2012) 396
Table 59: The original 201 meanings (alphabetical order) 397
Table 60: The original 201 meanings (“binominality” order) 399
Table 61: Constructions lost with a sample of 84 meanings 401
Preface
Acknowledgements

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 forcing me, at long last, to empty my head onto paper. 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.

Three eminent scholars, without whose work mine would have been impossible, are Bill Croft, Martin Haspelmath and Ron Langacker. I take my hat off to all three 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 the depth 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 Sonia Cristofaro, Martin Haspelmath (again), Pavol Štekauer (again) and Kjell-Magne Yri. In addition, Bård Uri Jensen devoted several hours to checking my statistics.

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. They are, in alphabetical order:


In Appendix B, Sources, I acknowledge the same contributors in the context of the language(s) they assisted me with.

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.
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>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>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>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>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>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INS</td>
<td>instrumental</td>
</tr>
<tr>
<td>IPPOS</td>
<td>inalienable possession</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>NMLZ</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NONF</td>
<td>nonfinite</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>PROP</td>
<td>proprietive</td>
</tr>
<tr>
<td>PROX</td>
<td>proximate</td>
</tr>
<tr>
<td>PST</td>
<td>past</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>SG</td>
<td>singular</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
</tr>
<tr>
<td>SGLT</td>
<td>singulative</td>
</tr>
<tr>
<td>SPEC</td>
<td>specific</td>
</tr>
<tr>
<td>STC</td>
<td>construct state</td>
</tr>
<tr>
<td>SUF</td>
<td>suffix</td>
</tr>
<tr>
<td>SUP</td>
<td>superessive</td>
</tr>
<tr>
<td>TAQ2</td>
<td>thing of/pertaining to N (874)</td>
</tr>
<tr>
<td>TMP</td>
<td>temporal</td>
</tr>
</tbody>
</table>
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.
- **NOR** – 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 301) and by name in the Index of Languages (page 387). To find the ISO code for a language in the database, use either the index or Appendix B (page 305). 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 (2017).
1 Towards a comparative concept

The aim 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.¹

The most familiar mechanism serving this purpose is noun-noun compounding, as in Ger. *eisen.bahn* [iron.way] RAILWAY,² but there are many others, including Fr. prepositional phrases (*chemin de fer* [road of iron] RAILWAY), Rus. relational adjectives (*želez.naja doroga* [iron.ADJZ road] RAILWAY), and more. Informally, the object of study can be thought of as *noun-noun compounds and their functional equivalents*.

Starting from a set of 100 meanings,³ and using data from the World Loanword Database (WOLD), supplemented by data collected specifically for this project, I describe both the formal variation and some of the functional-semantic features exhibited by such “binominals” (as I call them for short) across 100 languages, and their correlations with genetic, areal and typological features.

The theoretical framework behind the choice of comparative concept, and also the explanatory part of the study, is that of Cognitive Linguistics, in particular Cognitive Grammar (Langacker 1987; 1991) and Radical Construction Grammar (Croft 2001). However, the descriptive part of the study is framed in terms of Basic Linguistic Theory (Dixon 2010a; 2010b; 2012) and its findings should therefore be amenable to linguists of all theoretical persuasions.

While the primary goal of the project is to chart the formal and semantic diversity of binominal lexemes, the longer term interest is in understanding how this aspect of language reflects the associative nature of human thought; this topic will come to the fore from Chapter 5 onward.

---

¹ A more precise definition is developed in §1.2.4.
² See page xxiii regarding typographical conventions.
³ The term ‘meaning’ as used here reflects the usage in WOLD, where it denotes a concept, such as RAILWAY, that is not language-specific. 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 §7.2.3.) 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 was 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.

Figure 1: The TAO model

Figure 2: The anatomy of an association
(after Pepper 2010a)
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 both represent relationships of various kinds; associations of different arities (unary, binary, ternary) resemble clauses of different valencies (intransitive, transitive and 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 article, *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 both based on associative relations between entities and 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 7.

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, he 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,” wrote 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 nouns of a compound that resembles an association (such as ‘com-

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. (5) 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). (6) 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.
posed 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 question I posed was: could an understanding of the nature of the relationships inherent in Nizaa compounds help explain the presence of two types of compound 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 7 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 buh [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 cúŋ [person tree] ‘carpenter’ (purpose or occupation) and cam ɲunnam [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, exploiting what Langacker (1993) calls our “reference point ability”; the other, in left-headed compounds, via a superordinate concept, employing 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 subtype 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 existing hypotheses regarding universals of compounding, and where possible to propose new generalizations” (from the original project description). Phenomena to be studied included “formal marking, head position (and its correlation with constituent 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). 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 Ger. Regierungschef ‘head of government’ is not a “true” compound because it contains a linking element, -s-. Other linguists dispute whether a construction involving prepositions, such as Fr. chemin de fer, is a compound (some might admit them as compound phrases or prepositional compounds, Bauer 2001:705), and no-one, to my knowledge, has ever entertained the idea that Rus. želez.naja doroga might be a compound.

Now, what interests me, as is surely by now apparent, 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 Eng. railway, Ger. Eisenbahn and Nor. jernbane. Furthermore, English solar energy and Czech sluneční energie [sun:ADJZ energy] are just as relevant as Nor. solenergi [sun:energy] and Ger. Sonnenenergie [sun:LE:energy], since they involve the same semantic relation, ð, (“from” or “produced by”), between the same two concepts, SUN and ENERGY, to denote the same target concept, SOLAR ENERGY, and thus presumably involve the same underlying associative processes. All of

---

¹ Paolo has since informed me that he was deliberately being provocative, but the point stands.
these examples can be reduced to three basic schemas: NN, N PREP N and N ADJZ N. What they have in common (in addition to their function as naming units) is that their major constituents represent two nominal concepts (RAIL, IRON, WAY; SUN, ENERGY; GOVERNMENT, CHIEF) and that  is unstated (or underspecified). This realization offered a convenient way out of the above-mentioned definitional impasse: I could simply use a cognitive-functional comparative concept instead of the formal concept of noun-noun compound. So if we now 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 could thus be characterized as a binominal naming strategy and my project would become, in informal terms, a cross-linguistic study of noun-noun compounds and their functional equivalents, and 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 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 (Štekauer 1998:165, fn.2).

I experimented with the term ‘binominal naming unit’ for a while, but it felt too unwieldy, and so I opted for binominal lexeme instead.¹

My definition was a good first approximation, but several issues remained. I had known all along that I wanted to exclude coordinative compounds (such as Hmong Daw *zaub-mov* [vegetable-rice] FOOD or Vietnamese *bố mẹ* [father mother] PARENTS from my study, since their typology has been described by Wälchli (2005). This was easily done 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 was restricted to *determinative* noun-noun compounds and their functional equivalents. Another, more serious, problem was that I wanted to exclude synthetic compounds like *truck driver* in which the head is a deverbal noun. There were two reasons for this. First of all, they are inherently less interesting in terms of their semantic relations, because the relation is stated explicitly: a truck driver is an Agent who drives trucks; they correspond to the typed association ‘composed by’ in Figure 2. Secondly, it seemed likely that the presence of the verbal element would likely involve a different set of properties, related to argument structure, which again would complicate the typology unnecessarily.

### 1.2.3 An onomasiological perspective

The answer to this dilemma presented itself when I came across 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). 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; a complex mark consists of a determined (or Actional) constituent (M_A) and a determining (i.e. Qualifying) constituent (M_Q).

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

---

1 Unfortunately, much of the literature is in Czech or Slovak and inaccessible to many linguists.
2 This is *pace* Štekauer. I would argue that forms like *italia* porto-lettere [carry-letters] POSTMAN, can be analysed as consisting of the determining and determined constituents of the onomasiological mark without the onomasiological base, and I would define it as Onomasiological Type 6, one in which the onomasiological base is not present.
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 models, 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+AGT_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 mentioned above (page 10) 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 must be included. So, too, must noun classifier constructions, such as Bora tuú,heja [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 101 ff).

### 1.2.4 Defining the object of study

Štekauer’s model of onomasiological types provided a satisfying rationale 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 it seemed advisable to define my object of study in more theoretically neutral terms. The definition given in (1) on page 9 was thus in need of further refinement. The first issue is that the term ‘nominal constituent’ includes deverbal nominalizations, such as *driver* in *truck driver*, which we now know are Onomasiological Type 1. Simply replacing ‘nominal constituent’ with ‘noun’ does not solve the problem, and in any case, ‘noun’ is a language-specific descriptive category and therefore unsuitable for cross-linguistic comparison, as many typologists have pointed out (e.g. Croft 2001, Haspelmath 2012). To avoid what Croft terms “methodological opportunism”, I considered adopting Haspelmath’s term ‘thing-root’: “a root that denotes a physical object (animate or inanimate)”. This clearly excludes *driver*, since *DRIVE* is an ‘action-root’, “a root that denotes a volitional action” in Haspelmath’s terms.¹ However, Haspelmath makes a distinction between roots and affixes; thus, the term ‘thing-root’ does not include nominalizing affixes of the železnica type. 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 propose the terms ‘thing-affix’, ‘action-affix’ and ‘property-affix’, defining the first of which as “an affix that denotes a physical object (animate or inanimate)”. Then, since roots and affixes are both morphs, I propose the superordinate concept of ‘thing-morph’ to cover both (2).

---

¹ Other linguists use other terms for semantic, cognitive or onomasiological types or classes: Croft has objects, actions and properties; Langacker has thing and relationship (processual or non-processual); Dokulil has substance, action, quality and concomitant circumstance, etc. I have chosen to adopt Haspelmath’s terminology (thing, action, property) because it denotes linguistic items rather than conceptual classes, even though it is based on the latter.
(2) **thing-morph**
   
   *a morph that profiles a thing (prototypically a physical object, animate or inanimate)*

The definition in (2) involves two significant changes relative to Haspelmath’s definition of thing-root (over and above replacing root with morph): firstly, in using the verb ‘profile’ instead of ‘denote’, I embrace the terminology of Cognitive Grammar (the difference between the two can be ignored for present purposes); secondly, I allow for the inclusion of non-prototypical thing-morphs that profile more abstract entities than physical objects. Having 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 underspecified) 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 what will become a major concern from Chapter 5 onwards.

Binominal lexemes are called binominals for short and I use the term binominal construction to refer to schemas that individual binominals instantiate, such as Mod Head for typical Germanic “root” compounds and Head PREP Mod for prepositional compounds. Note that the term binominal in my sense does not include constructions that do not name a generic concept, even when it occurs in the term used by other linguists (4).

(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 following list of binominals from around the world (5), all of which mean RAILWAY (unless otherwise stated), illustrates some of the variety covered by my comparative concept and offers a taste of things to come:

(5) a. root compounds: German *eisen.bahn* [iron.track]
    b. compounds with linking elements: Plains Cree *pīwāpisk.o.mēskanaw* [iron.CON.road]
    c. prepositional compounds: French *chemin de fer* [way of iron]
    d. relational compounds: Russian *želez.naja doroga* [iron.ADJZ road]
    e. genitive lexemes: Bezhta *kil.o.s hino* [iron.OBL.GEN way]
    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: Takia *ŋdu.n awa.n* [nose.3SG mouth.3SG] NOSTRIL
    j. classifier constructions: Murui Huitoto *ui.tiraї* [eye.CL(hair)] EYELASH

As noted above, synthetic compounds are out of scope; so too are NVN constructions such as Vietnamese *bữ a ăn sáng* [meal eat morning] BREAKFAST, in which the determined element of the onomasiological mark is also present; contrast this with Kildin Sami *īnc.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 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 Design of the study

1.3.1 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, in which he compares the extent of 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. Together they inspired the design of the present study, which is a detailed analysis of binominal lexemes representing 100 meanings across 100 languages, drawn from approximately 10,000 data points. More precise details of how the meanings and languages were selected, how the data was gathered, and how it was analysed are given in Chapter 3.

1.3.2 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 its exploratory, data-driven nature:
• What is the extent and diversity of binominal word-formation in the world’s languages?
• How can binominals be classified typologically, in terms of both form and meaning?
• 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?

More precise research questions were expected to emerge from the data as the study progressed, as indeed they did. They are to be found in Chapters 6 and 7.

1.4 Theoretical framework

As mentioned previously, the theoretical framework underlying the choice of comparative concept and the explanatory part of the study are Cognitive Grammar (Langacker 1987; 1991) and Radical Construction Grammar (Croft 2001; 2003). Some of its central ideas are expressed by Langacker as follows:

Lexicon, morphology, and syntax form a continuum, divided only arbitrarily into discrete “components.” Everything along this continuum is fully describable as assemblies of symbolic structures. A symbolic structure is specifically defined as the pairing between a semantic structure and a phonological structure (its semantic and phonological poles). This has certain consequences. First, grammar is not distinct from semantics but incorporates it as one pole. Second, the elements of grammatical description are not special, irreducible primitives, but reduce to form-meaning pairings. Finally, every valid grammatical construct should be meaningful (Langacker 2005: 104).

Binominals are ideal for exploring this continuum: as naming units they are all part of the lexicon, but while some are ‘syntactic’ (e.g. *chemin de fer*), some are ‘morphological’ (e.g. *železnica*) and others (e.g. *Eisenbahn*) are “an embarrassment” that form “a subsystem of grammar somewhat distinct from ordinary syntax and morphology” (Jackendoff 2009; 2010). However, while the theoretical framework is firmly that of Cognitive Linguistics, the descriptive framework is that of Basic Linguistic Theory (Dixon 2010a) and should thus be amenable to linguists of all theoretical persuasions. The study belongs to the field of typology, but not to either morphological typology or syntactic typology. I situate it myself within lexical typology, despite the fact that it is broader in many respects than most studies within that field. 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 step in such a direction.

1.5 Structure of this work
Song (2007:9) describes the 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

I largely follow this scheme in the structure of this work. The first stage has been covered in this chapter and will be further elucidated in the literature review that follows (Chapter 2, *On the shoulders of giants*). There I discuss previous work that is most generally relevant to the present study, in particular my choice of comparative concept. This includes cross-linguistic studies of word-formation (especially compounding), studies that prefigure the concept of binominal lexemes, and large-scale, typological studies that take a broadly onomasiological approach. (Other work that pertains to more specific topics is discussed in the relevant context, especially in chapters 4, 5 and 7).

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 the challenges posed by each of them (§3.3). The final section covers the pre-theoretic data annotation (“coding”) that preceded the development of the formal taxonomy (§3.4).

In Chapter 4, *Formal classification*, I develop my first classification (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, based on a kind of semantic map. I also consider the issue of gradience, including how it can be captured in a visual representation.
In Chapter 5, *Semantic relations*, I develop a second classification, based on the unstated (or underspecified) relation between the two major constituents of a binominal. Again I start out with theoretical prerequisites, this time invoking Arnaud, Bauer and Tarasova, Bourque, Hatcher 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 integrate the two into a single whole: the Hatcher-Bourque classification.

In Chapter 6, *Patterns in the data*, I seek to identify the statistical patterns in the data and formulate the kinds of generalization appropriate to Song’s fourth stage. The chapter starts out (§6.1) with a general description of the data based on the pre-theoretic annotation described in Chapter 3. This is followed by three different types of analysis: one based on the eight-way formal typology (§6.2), one on the Hatcher-Bourque classification (§6.3), and one using more advanced methods to investigate the intersection between the two (§6.4).

Those findings which I find most interesting are discussed further in Chapter 7, under the heading *Typological investigations*. The topics under consideration are word order typology (§7.1), my two-paths hypothesis (§7.2), and the broader issue of associative relations (§7.3). Where possible, given the exploratory nature of the study, I attempt to put forward explanations of the kind Song was looking for in the fifth and final stage of his model, and once again I force the linguist reader out of her comfort zone in order to show how Topic Maps can enrich the discussion.

Finally, in Chapter 8, *Conclusion*, I offer a brief summary, discuss the contribution to science of the present work, and indicate 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 299).

B A list of every source of data and grammatical information for each of the languages in the sample, ordered by language name (page 305).

C The list of meanings, how they are categorised and various statistics associated with them (page 311).

D A list of binominal constructions, ordered by area, genus and language, with an example of each (page 313).
Chapter 1. Towards a comparative concept

E  The complete binominal data set (page 335).
F  A summary of the database structure (page 369).
G  The questionnaire sent to contributors (page 369).
H  Various tables that were too large for the main body of the work (page 373).

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.
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 in turn 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 Štékauer, Valera & Körtvélyessy (2012). In §2.3 Prefiguring binominals I present three studies that in one way or another prefigure 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 4, Formal classification, Koptjevskaja-Tamm’s (2002; 2003) and Croft’s (2003) typologies of possessive constructions; in Chapter 5, 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 6, Patterns in the data, Koch’s (2001) idea of motivational grids, and in Chapter 7, Typological investigations, 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 1). The sample comprises six languages from each of Dryer’s 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 HeB (Semitic)</td>
<td>Yimas YEE (Lower Sepik-Ramu)</td>
</tr>
<tr>
<td>Tswana TSN (Bantoid)</td>
<td>Kobon KPW (Madang)</td>
</tr>
<tr>
<td>Yoruba YOR (Defoid)</td>
<td>Siroi SSD (Madang)</td>
</tr>
<tr>
<td>Ewe EWE (Kwa)</td>
<td>Waskia WSK (Madang)</td>
</tr>
<tr>
<td>Turkana TUV (Niloitic)</td>
<td>Mara MEC (Mangarrayi-Maran)</td>
</tr>
<tr>
<td>Kanuri KNC (Saharan)</td>
<td>Arabana ARD (Karnic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eurasia (E)</th>
<th>North America (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abkhaz-Adyge AĐY (NW Caucasian)</td>
<td>Kalaallisut KAL (Eskimo-Aleut)</td>
</tr>
<tr>
<td>Chukchi ČKT (Chukotko-Kamchatkan)</td>
<td>Kiowa KIO (Kiowa-Tanoan)</td>
</tr>
<tr>
<td>Tamil TAM (Dravidian)</td>
<td>Tz’utujil TŻI (Mayan)</td>
</tr>
<tr>
<td>Danish DAN (Germanic)</td>
<td>Dakota DAK (Siouan)</td>
</tr>
<tr>
<td>Basque EUS (Basque)</td>
<td>Takelma TKM (Takelma)</td>
</tr>
<tr>
<td>Finnish FIN (Finnic)</td>
<td>Shoshone SHH (Northern Uto-Aztecan)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southeast Asia &amp; Oceania (O)</th>
<th>South America (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khmer KHM (Khmeric)</td>
<td>Paumari PAD (Arawan)</td>
</tr>
<tr>
<td>Vietnamesevenience (Vietic)</td>
<td>Hixkaryána HIX (Parukotoan)</td>
</tr>
<tr>
<td>Maori MRI (Eastern Malayo-Polynesian)</td>
<td>Cayubaba СВ (Cayubaba)</td>
</tr>
<tr>
<td>Batak Toba BBC (NW Sumatra-Barrier Isl)</td>
<td>Pirahã MYP (Pirahã)</td>
</tr>
<tr>
<td>Yue Chinese YUE (Sinitic)</td>
<td>Imbabura Highl. Quichua QVI (Quechua II)</td>
</tr>
<tr>
<td>Thai THA (Kam-Tai)</td>
<td>Paraguayan Guaraní GUG (Tupi-Guarani)</td>
</tr>
</tbody>
</table>

Table 1: Language sample (Bauer 2001)

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).
Bauer starts out by developing a definition of compound (6); this is only intended to present a “focal notion” of the way in which the term compound is used in the paper, since it is acknowledged 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 the order of elements, compound types, 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:

- **tatpurusa** (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*;
- **avvayībhāva** compounds¹ 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.
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 possessor (e.g. Takelma *p!iyin sge’xabâ:* [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 Onomasological Types 1 and 3, respectively.

2 I return to this issue in Chapter 5, *Semantic relations*.
Chapter 2. On the shoulders of giants

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 2 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 2: 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 2003: 56). This becomes very clear if the data are represented in the form of a tetrachoric table, as in Table 3a. From this we can derive the implicational universal Adj-N \(\not\rightarrow\) Mod-N (that is, adjective-noun order implies modifier-head noun order in compounds). Furthermore, it can be concluded that Mod-N (i.e. right-headedness) is the dominant order cross-linguistically, and that 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 3b). 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 §6.3, 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.

\[
\begin{array}{c|c|c}
    & N-Mod & Mod-N \\
\hline
N-Adj & 10 & 11 \\
Adj-N & 1 & 9 \\
\end{array}
\]  

(a) Bauer 2001

\[
\begin{array}{c|c|c}
    & N-Mod & Mod-N \\
\hline
N- Adj & 9 & 13 \\
Adj-N & 1 & 7 \\
\end{array}
\]  

(b) Pepper 2015

Table 3: 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 4, grouped by geographical area). Again, representing the data as a tetrachoric table (Table 5a) reveals a distribution from which it is possible to derive the implicational universal \( \text{Poss-N} \supset \text{Mod-N} \), with \( \text{Mod-N} \) (right-headedness) again emerging as dominant. In this case, however, only the two harmonic correlations (\( \text{N-Poss & N-Mod} \) and \( \text{Poss-N & Mod-N} \)) can really be said to be frequent. And while the \( \text{Poss-N & N-Mod} \) can be characterized as “extremely rare”, the other disharmonic pattern, \( \text{N-Poss & Mod-N} \), is also rather infrequent. The data thus tend toward a biconditional universal of the type \( \text{Poss-N} \equiv \text{Mod-N} \).

\[
\begin{array}{c|c|c|c|c|c|c|c}
    & A & E & O & G & N & S & \text{Total} \\
\hline
\text{N-Poss & N-Mod} & 3 & 0 & 4 & 0 & 1 & 0 & 8 \\
\text{N-Poss & Mod-N} & 1 & 0 & 0 & 2 & 0 & 1 & 3 \\
\text{Poss-N & N-Mod} & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
\text{Poss-N & Mod-N} & 1 & 5 & 1 & 4 & 1 & 2 & 14 \\
\text{insufficient data} & 1 & 1 & 1 & 2 & 2 & 3 & 10 \\
\end{array}
\]

Table 4: Order of noun-possessor and noun-modifier (Bauer 2001)

In my replication study, this tendency turned into an exceptionless pattern (Table 5b). 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 6]. 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 6]. 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>Consistent ordering</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Inconsistent ordering</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Unclear or missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6: Consistency of head-modifier ordering (Bauer 2001)

Bauer’s figures can be compared with those obtained in my replication study. In Table 6, 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 the row labelled sum. Numbers in the row labelled mixed correspond to Bauer’s ‘inconsistent ordering’ and those in the row labelled n/a (not applicable) correspond to Bauer’s ‘unclear or missing’. Since the replication study only investigated nominal compounds, the numbers in Table 7 correspond to those in parentheses in Table 6. The relevant comparison is thus between the numbers in parentheses in both tables (shown in boldface).
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 concluded 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” (Nguyễn 1997: 72, 77). Đinh (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 mention nominal compounds explicitly. However, he does state that a “noun in adjectival (modifying) function may be integrated into the NP” and to follow the nuclear noun.
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.

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 may be 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) would 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 8. 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).

/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 …/

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.
The typology 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</td>
<td>Khmer KHM (Khmeric)</td>
</tr>
<tr>
<td>(Berber)</td>
<td></td>
</tr>
<tr>
<td>Bambara BAM (Mande)</td>
<td>Kumak NEE (Eastern Malayo-Polynesian)</td>
</tr>
<tr>
<td></td>
<td>Angami Naga NJM (Kuki-Chin-Naga)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Eurasia (E)</td>
<td></td>
</tr>
<tr>
<td>Turkish TUR (Turkic)</td>
<td></td>
</tr>
<tr>
<td>Basque EUS (Basque)</td>
<td></td>
</tr>
<tr>
<td>Modern Armenian HYE (Armenian)</td>
<td></td>
</tr>
<tr>
<td>Welsh CYM (Celtic)</td>
<td></td>
</tr>
<tr>
<td>Udi UDI (Lezgic)</td>
<td></td>
</tr>
<tr>
<td>Hungarian HUN (Hungarian)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Pidgins &amp; Creoles (P)</td>
<td></td>
</tr>
<tr>
<td>Tok Pisin TPI (English-based</td>
<td></td>
</tr>
<tr>
<td>Creoles)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 8: Languages covered in Arnaud (2004)**

Arnaud’s results are reproduced below as Table 9 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 prompts the question of what languages that disfavour compounding do instead. In the case of French, the answer is well-known: complex concepts that are typically expressed through compounding in, say, English and German (and, to judge by Figure 4, Basque, Welsh and Cambodian as well), such as RAILWAY, are often expressed using a prepositional construction, as in *chemin de fer* [road of iron]. This prompted the central question of the present work: What are the functional equivalents of noun-noun compounds in the world’s languages?

| +   | composé | compound                  |
| (+) | très 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 |
| ×   | cranberry compound            | cranberry compound       |

*Legend for Table 9*
Chapter 2. On the shoulders of giants

<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>CRU</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></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>forehead</td>
<td>FRONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>eyebrow</td>
<td>SOURCIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</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></td>
<td></td>
<td></td>
<td></td>
<td>6</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></td>
<td></td>
<td></td>
<td></td>
<td>8</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></td>
<td></td>
<td></td>
<td></td>
<td>7</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></td>
<td></td>
<td></td>
<td></td>
<td>6</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></td>
<td></td>
<td></td>
<td></td>
<td>5</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></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>thumb</td>
<td>Pouce</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</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></td>
<td></td>
<td></td>
<td></td>
<td>3</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></td>
<td></td>
<td></td>
<td></td>
<td>5</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></td>
<td></td>
<td></td>
<td></td>
<td>4</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></td>
<td></td>
<td></td>
<td></td>
<td>3</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></td>
<td></td>
<td></td>
<td></td>
<td>8</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></td>
<td></td>
<td></td>
<td></td>
<td>3</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></td>
<td></td>
<td></td>
<td></td>
<td>10</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></td>
<td></td>
<td></td>
<td></td>
<td>5</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></td>
<td></td>
<td></td>
<td></td>
<td>8</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></td>
<td></td>
<td></td>
<td></td>
<td>3</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></td>
<td></td>
<td></td>
<td></td>
<td>3</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></td>
<td></td>
<td></td>
<td></td>
<td>9</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></td>
<td></td>
<td></td>
<td></td>
<td>9</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></td>
<td></td>
<td></td>
<td></td>
<td>7</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></td>
<td></td>
<td></td>
<td></td>
<td>5</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></td>
<td></td>
<td></td>
<td></td>
<td>4</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></td>
<td></td>
<td></td>
<td></td>
<td>7</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></td>
<td></td>
<td></td>
<td></td>
<td>2</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></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 9: Onomasiological cross-linguistic comparison (Arnaud 2004)
2.1.3 Morbo/Comp (2004-2006)

Morbo/Comp was an international research project on compounds, 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.sssmit.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 2006.¹ 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 (2/2009) containing papers on compounding in Russian, Chinese, Turkish, Finnish,

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).
Chapter 2. On the shoulders of giants

Swedish, Czech and Portuguese. To judge by the authorship of these papers, the principal participants in the project, besides Scalise himself, appear to have been 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 10, 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 10: Languages represented in Morbo/Comp

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>NLD</td>
</tr>
<tr>
<td>English</td>
<td>ENG</td>
</tr>
<tr>
<td>German</td>
<td>DEU</td>
</tr>
<tr>
<td>Norwegian</td>
<td>NOR</td>
</tr>
<tr>
<td>Swedish</td>
<td>SWE</td>
</tr>
<tr>
<td>Finnish</td>
<td>FIN</td>
</tr>
<tr>
<td>Hungarian</td>
<td>HUN</td>
</tr>
<tr>
<td>Turkish</td>
<td>TUR</td>
</tr>
<tr>
<td>Hebrew</td>
<td>HEB</td>
</tr>
<tr>
<td>Basque</td>
<td>EUS</td>
</tr>
<tr>
<td>Greek</td>
<td>ELL</td>
</tr>
<tr>
<td>Catalan</td>
<td>CAT</td>
</tr>
<tr>
<td>French</td>
<td>FRA</td>
</tr>
<tr>
<td>Italian</td>
<td>ITA</td>
</tr>
<tr>
<td>Latin</td>
<td>LAT</td>
</tr>
<tr>
<td>Spanish</td>
<td>SPA</td>
</tr>
<tr>
<td>Portuguese</td>
<td>POR</td>
</tr>
<tr>
<td>Belarusan</td>
<td>BEL</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>BUL</td>
</tr>
<tr>
<td>Polish</td>
<td>POL</td>
</tr>
<tr>
<td>Russian</td>
<td>RUS</td>
</tr>
<tr>
<td>Serbo-Croatian</td>
<td>HBS</td>
</tr>
<tr>
<td>Japanese</td>
<td>JPN</td>
</tr>
<tr>
<td>Korean</td>
<td>KOR</td>
</tr>
<tr>
<td>Mandarin</td>
<td>CMN</td>
</tr>
</tbody>
</table>

The structure of the database is shown in Table 57 (see page 373). 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. In fact only 4,000 or so of the 80,000 compounds were ever analysed in such detail.

---

\(^1\) Languages shown in italics in Error! Reference source not found. are not mentioned in Guevara & Scalise (2009).

\(^2\) The ISO 639-3 code point HBS now has the name South-Western Slavic.
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 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 ‘large 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 suggested that another approach was required in order to build a typologically well-balanced database for use in cross-linguistic comparison.

Despite these issues, the project produced a number of useful results. In the context of the present project, the two most important are the classification of compounds developed by Scalise and Bisetto (§2.1.4), and a discussion of 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 main problem with the last of these is 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, depending on “the presence or absence of a head constituent” (Figure 5).

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

*Figure 5: Classifying compounds (Bisetto & Scalise 2005)*

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, the class of subordinate compounds is subdivided into verbal-nexus and ground, ostensibly on the basis of whether or not they contain a deverbal head.\(^1\) Furthermore, the class of attributive compounds is 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 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.

Two things may be observed about the resulting classifications: Firstly, there is an inherent 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 may not

---

\(^1\) The exocentric *lavapiatti* ‘dishwasher’ < ‘wash’ + ‘dishes’ should, of course, be classified with *pickpocket* under verbal-nexus, not under ground. I assume this is a printing error.
even be present. 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).

![Diagram of compound classification]

Figure 6. Classifying compounds (Scalise & Bisetto 2009)

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 26 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 that the two paths of mental access may well be cognitive universals. Since they involve different sets of semantic relation (at least in Nizaa), I return to this issue in Chapters 6 and 7.

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, entitled 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 (2001) and Olsen (2000) 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 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 aside the problem of the definition of compound and look for general tendencies in the world’s languages. This leads them to identify the presence of a relation \( R \) between the constituents of a compound, which is not explicitly (phonetically) realized, as a key
defining feature. They then invoke canonical typology (Brown, Chumakina & Corbett 2013) and define the *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 the formal head and the semantic head, and the difficulty of determining the head in (i) coordinate compounds like It. *studente-lavoratore* ‘student-worker’, and (ii) verbal nexus compounds with the structure [V+N]N, such as It. *rompi.ghiaccio* [break-ice] ‘icebreaker’. They conclude (contra Haspelmath 2002) that (i) should be regarded as having two heads rather than none. As for (ii), the upshot appears to be that the form 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:

---


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. There is therefore no way of determining how close any specific compound is to the canon.
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.

As for the position of the head, Guevara and Scalise reject the right-hand head rule put forward by Williams (1981) and also the idea that the position of the head is a parameter that is fixed for any given language (Scalise 1994). Instead they “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 constituent (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\(^1\)

Data from these languages, comprising about 3,000 compounds, is analysed on the basis of five features defined in terms of the generic structure \([ X \mathcal{R} Y ] Z\):

\[
\begin{align*}
\text{a) } Z &= \text{ Output Category} \\
\text{b) } X \text{ and } Y &= \text{ Input Categories} \\
\text{c) } \mathcal{R} &= \text{ Relation between constituents (Classes)} \\
\text{d) } [X Y] &= \text{ Combination of Categories} \\
\text{e) } [X Y] Z &= \text{ Headedness}
\end{align*}
\]

Scales of preference are computed for each of these features to give the following results:

---

\(^1\) The authors acknowledge that Mandarin, Korean and Japanese are not genetically related.
The typology of binominal lexemes

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

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 typological balance, and the apparently unsystematic manner in which the data were collected, cast some doubt on the validity of the findings (see §2.1.3). 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; Takada 2008), but, as far as I am aware, there has not been a single cross-linguistic study of this kind of compound. That is the gap which the present work seeks to fill.
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 focuses on two types of distinction: the one based on the nature of the compound head (i.e. endocentric vs. exocentric vs. coordinate), the other on the word class of its constituents (i.e. 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 phrasal origins, such as Fr. *chemin de fer* and Rus. *železnaya doroga* (‘railway’). For Aikhenvald, as for most linguists, this process is not regarded as part of word-formation; if not, then where does it belong?

The 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.

Aikhenvald (2007) is thus 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

---

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-vrīhi 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.
any particular sample. Her purpose, as noted above, is simply to describe the
diversity of the domain rather than to formulate and explain typological generali-
zations. Nevertheless, certain generalizations are made, albeit at a high level and
without reference to evidence in the literature. 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 af-
 fixes” (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 com-pre-
hensive coverage of the features to be found cross-linguistically in the domain of
word-formation, rather than in any particular typological insights.

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 58 on page
374). The data is mostly based on questionnaires, but data for some languages was
acquired from published sources. The creation of a typological classification is a task
left to the reader. According to the authors, various sorts of typological classifica-
tions can be inferred from the data; these are

  primarily determined by the specific method of analysis, semasiological or onomasio-
  logical. 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 re-
lated 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 11). 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 (Sapirian) classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agglutinative languages (of various types)</td>
</tr>
<tr>
<td>Fusional (of various types)</td>
</tr>
<tr>
<td>Isolating (of various types)</td>
</tr>
<tr>
<td>Polysynthetic (of various types)</td>
</tr>
</tbody>
</table>

*Table 11: Morphological classification (Štekauer et al. 2012)*

The body of the questionnaire has three parts. Part I asks (1) which word-formation processes¹ 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 enquires about 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.

---

¹ The 20 processes comprise those listed by Aikhenvald in her formal classification of derivation processes (page 42) and combinations thereof, together with the other two major word-formation types, incorporation and compounding.
Chapter 2. On the shoulders of giants

(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 of phrasal lexemes, such as FRA chemin de fer and RUS železnaya daroga in the study: Apparently Štekauer and his colleagues follow Aikhenvald in excluding such constructions from the domain of word-formation.

The second part contains the cross-linguistic analysis and comprises five chapters. The first three of these adopt a semasiological perspective and investigate three kinds of formal 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 and subtractive processes, including back-formation). The fourth chapter takes an onomasiological approach and examines the formal mechanisms used to express 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 bulk of the first four analytical chapters consists of tables presenting the data collected using the questionnaire. Thus, for example, the first table in Chapter 3 lists the 50 languages (out of the basic sample of 55 languages) that make productive use compounding, along with their genetic affiliation (language family, not genus), morphological type and geographical area. Section 3.1.1.4 treats noun-noun compounds. It 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 fifth analytical chapter is quite different, consisting of a set of statistical analyses using the chi square test and multiple correspondence analysis (MCA). Both methods of analysis are applied to the twenty word-formation processes using the three independent variables: language family, morphological type, and basic word order. With the chi-square 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 they] 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 covering 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, used with a more fine-grained exploration of various subtypes of the four word-formation processes prefixation, suffixation, compounding and reduplication, produces inter alia the following results:

- **prefixation:** Sino-Tibetan, Kartvelian and Indo-European cluster together at one end of a gradient, while Dravidian and Uralic are at the other, with Australian and Japanese in between.
- **suffixation:** Khoisan, Siouan, Trans-New Guinea and Afro-Asiatic cluster at one end of a gradient, with Indo-European, Uralic, Kartvelian and Matacoan at
the other (and with clear separation between the first two and the last two of
the latter).

- **compounding**: Sino-Tibetan clusters with Indo-European as regards the pres-
ence of most types of compounding, while Totonacan clusters with Altaic and
Niger-Congo in terms of the absence of most types of compounding.

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 bring to light any really
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,
since “such a motivation or prediction would be a matter of speculation rather than
of clear linguistic facts”.

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.\(^1\) Unlike
Warren (1978), writing the same year on semantic relations, Levi is concerned with
both syntax and semantics, albeit from a Generative Semantics perspective that is
not relevant to the present study. The reason why the work is of interest here is that
it is based on the concept of ‘complex nominals’, a term covering three “partially

---

\(^1\) Earlier studies, including Mätzner (1860), Jespersen (1942) and Hatcher (1960) are discussed in
Chapter 5.
The typology of binominal lexemes

overlapping sets of expressions”, viz. (i) nominal compounds, (ii) nominalizations, and (iii) noun phrases with non-predicating adjectives (p. 1). On the basis of the examples provided, these can be defined more precisely as:

- compounds consisting of two root nouns (e.g. apple cake, windmill)
- deverbal nouns preceded by either a denominal adjective or a base noun (e.g. musical criticism, metal detection)
- 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.ADJZ.NMLZ] RAILWAY) or a thing-root and a noun classifier (Bora túú.heju [nose.CM(hole)] NOSTRIL).

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 linguistic orientation, but her typology of nine (“recoverably deletable”) predicates is relevant to the discussion of semantic relations in Chapter 5.

---

1 These categories are not disjunct, in that deverbal nouns qualified by denominal adjectives (e.g. musical criticism) belong to both types (2) and (3).
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. He concludes that they 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 competitors in question being (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 serves to validate it as a comparative concept.

2.3.3 Bauer & Tarasova (2013) – Adnominal modification

Bauer & Tarasova (2013) is a study of “the meaning relationships holding between the elements of endocentric nominal compounds in English.” Taking the relations inherent in Levi’s (1978) set of ‘recoverably deletable predicates’ as their starting point, the authors show that the same relations are implicit in a number of different constructions in which a noun is modified by another noun (‘adnominal nominal modification’). They conclude that these relations are not specific to compounding and must arise “from the nature of the modification”. This conclusion is of great interest in the context of the wider research question concerning the associative nature of human thought that underlies the present study and we will return to it in Chapter 5, together with the classification of semantic relations. What interests us here, however, are the kinds of construction considered by Bauer and Tarasova.
The typology of binominal lexemes

These are listed in (7) and illustrated using the items provided by the authors to exemplify Levi’s USE relation (\(N_2 \text{ USE } N_1\)):

(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)

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 “non-anchoring” relations. Thus, (7c) *car’s driver* and (7d) *driver of the car* do not qualify as binominals because they are not lexical items and therefore fall outside the scope of the present study.\(^1\) This does not mean, however, that these two constructions are outside the scope of the present study: Bauer & Tarasova provide three examples of prenominal possessives that clearly are naming units: *dog’s breakfast* ("a confused mess or mixture"),\(^2\) an instance of the HAVE relation, *ladies’ man* ("a man who is fond of, attentive to, and successful with women"),\(^3\) exemplifying the reversed HAVE relation, and *wolf’s bane* ("aconite; esp., a tall Eurasian plant (*Aconitum lycoctonum*) with showy, yellow flowers"),\(^37\) which embodies the FOR relation. As for postnominal possessives, none of Bauer & Tarasova’s examples are clearly naming units, presumably because this is a relatively marginal word-formation strategy in English, but a word like *man-of-war* ("a combatant warship of a recognized navy")\(^36\) shows that it does occur.

Thus all six of Bauer and Tarasova’s constructions can be accommodated by the notion of binominal, and the first four of them actually occur in the database. Neoclassical compounds and blends are not attested, presumably because they are not common enough to have been captured using the list of 100 meanings, 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 the list of 201 meanings used in the early phase of the project (see §3.1.3). These include the Hindi neoclassical form *par.deśī* [foreign_region. inhabitant] STRANGER,

---

\(^1\) On the other hand, *car driver* denotes a generic concept rather than a specific individual and is therefore a naming unit. However, it is also not a binominal since it contains an actional element (*drive*) and is thus OT1, not OT3.

\(^2\) https://www.merriam-webster.com/dictionary/

\(^3\) https://www.collinsdictionary.com/
which is “a Sanskrit word built like German Ausländer” (Claus-Peter Zoller, p.c.), and the Italian pomeriggio, which is “a sort of blend between pomeridiano ‘post-meridian’ and meriggio ‘midday (ancient)’” (Francesca Masini, p.c.). To summarize: Bauer and Tarasova’s category of “nominals that modify other nominals” is nearly co-extensive with that of binominal lexemes,\(^1\) provided the former is restricted to lexemes that have a naming function. Their study therefore serves, once again, to confirm the validity of binominal lexeme as a comparative concept.

### 2.4 Morphological complexity

This section brings together two rather different studies: the World Loanword Database (WOLD) and Matthias Urban’s (2012) work 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, proceed from meaning to form, and are concerned (each in their own way) with morphological complexity.

![Diagram](image.png)

**Figure 9: Pepper as a loanword in WOLD**

#### 2.4.1 Haspelmath & Tadmor (2009) – WOLD

The World Loanword Database (Haspelmath & Tadmor 2009a),\(^2\) is the web 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.

---

\(^1\) The pre-theoretical notion of binominal presented in Chapter 1 includes two types (denominal derivation and classifier constructions) that are not covered by Bauer and Tarasova. I discuss the status of neoclassical compounds as possible subtypes of derivation in §5.1.2.

\(^2\) [http://wold.cldf.org/](http://wold.cldf.org/)
The typology of binominal lexemes

The reason WOLD was particularly useful to me is that the data set on which it is based 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 88 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, Iraqiw, Japanese, Kanuri, Ket, Kildin Sami, Lower Sorbian, Malagasy, Manange, Mandarin Chinese, Mapudungun, Old High German, Oroqen, Q‘eqchi’, Querétaro Otomi, Romanian, Saramaccan, West Slovakian Romani (Selice Romani), Seychelles Creole, Swahili, Takia, Tarifit (Tarifiyt Berber), Thai, Vietnamese, Wichi, Yakut (Sakha), Yaqui, Zinacantán Tzotzil.¹

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 lan-

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 xxiii).
guages 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*. These meanings are grouped into 24 semantic fields and five broad semantic categories: *Noun, Verb, Adjective, Adverb 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 12.

<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 12: 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, analysability (with a morpheme gloss for analysable words), loanword status, age, etc. The morpheme glosses mostly (but not always) 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 fulfil 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 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 secondly, it provided an initial source of data for the project (see §3.3.1). The question of analysability played a key 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).

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, played an essential role in 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. The summary section 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 what the present work is about. 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, possibly correlating with the affiliation to a particular language family or a linguistic area?
- 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?

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 particularly gratifying to have been able to reuse the WOLD data, especially those parts that had been essentially discarded in the original project. This was only possible because of the original project editors’ commitment to the principles of Open Data, for which they are to be commended.
• 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 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 ŋgàŋ-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 patterns, Urban adopts an onomasiological approach and starts from a set of 160 nominal meanings, most of which are also found in WOLD. However, no explanation is given of the selection procedure employed. 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, EYLASH, EYELID, FINGERNAIL, HEADLAND, MILKY WAY, NOSTRILS, RAINBOW, RIVER BED, SATURDAY, SUNRISE, SUNSET, WATERFALL, WHIRLPOOL. This can serve as a rough measure of the frequency of binominal concepts in the 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 85), 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 provide an introduction to the major research questions and a brief discussion of 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 the classification of formal relations, which proceeds from a basic subdivision between 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 ǂnora.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) is monomorphemic and therefore classed as unanalysable, whereas (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 low-level subdivisions 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, which seeks to develop a more fine-grained formal typology (see Chapter 4).

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 applies a much more fine-grained semantic classification (see Chapter 5).

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 Nichol-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).”

Following this, Chapter 5 offers a quantitative evaluation of the data. Using typological parameters drawn from the World Atlas of Language Structures (Haspelmath et al. 2005), an attempt is made to provide an explanation for the variation in the number of analysable lexical items in the lexicon. 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, i.e. “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 contains much interesting material, but rather little in the way of easily discernible conclusions. It is clear that the present work 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 seeks to develop a more detailed classification of a subset of complex nominals: binominal lexemes.

2.5 Chapter summary

This chapter has discussed previous work of various types, 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 them all. 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 100 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 and dictionaries, which each presented its own challenges.

§3.4 concerns the process of data annotation (or coding) and some of the theoretical issues involved, including how to identify binominals (§3.4.1), how to determine the head of a binominal (§3.4.2), and how to define the construction (§3.4.3). Later stages in the coding of the data – cross-linguistic classification by formal type and semantic relation – are discussed in chapters 4 and 5, respectively.

Finally, §3.5 offers a short summary of the overall process.

3.1 The meaning list

Having decided upon a data-driven, onomasiological approach (see §1.3.1), 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. 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, Adv, 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 and 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 way.

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 53, 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.

(11)  
```sql
select p.id, p.name as meaning, count(distinct vs.id) as langs
from meaning as m,
    parameter as p,
    valueset as vs,
    value as v,
    counterpart as cp,
    word as w
where m.pk = p.pk and
    p.pk = vs.parameter_pk and
    v.valueset_pk = vs.pk and
    v.pk = cp.pk and
    cp.word_pk = w.pk and
    m.semantic_category = 'Noun' and
    substring(w.analyzability from 1 for 10) = 'analyzable'
  group by p.id, p.name
order by langs desc, meaning;
```

*Figure 10: SQL query used to extract meanings from WOLD*

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.¹

¹ The original query did not count the number of languages in which each meaning was represented by an analysable word, but rather the number of analysable words for each meaning. With this query, the meaning COUSIN appeared at the top of the list, with a count of 38, followed by NOSTRIL with a count of 37. The difference is due to the fact that some meanings have multiple counterparts in some languages. For example, COUSIN has six (analysable) counterparts in Tarifit, e.g. yagği.s n xači [daughter.3SG of maternal_aunt], all of which use the same construction Head.3SG PREP CON:Mod. As a result, undue weight is given to such meanings, as can be seen from the fact that under the revised query, the result of which is shown in (12), COUSIN only gets a count of 16.
The typology of binominal lexemes

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id</td>
<td>meaning</td>
</tr>
<tr>
<td>2</td>
<td>4-231</td>
<td>nostril</td>
</tr>
<tr>
<td>3</td>
<td>14-45</td>
<td>midday</td>
</tr>
<tr>
<td>4</td>
<td>4-213</td>
<td>eyelid</td>
</tr>
<tr>
<td>5</td>
<td>17-26</td>
<td>pupil</td>
</tr>
<tr>
<td>6</td>
<td>3-819</td>
<td>spider web</td>
</tr>
<tr>
<td>7</td>
<td>2-71</td>
<td>stepfather</td>
</tr>
<tr>
<td>8</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 11: View of the table meaning in its original form*

From this list, the top 159 meanings were selected. The cut-off point was set at meanings that are represented by complex words in 16 or more languages. This left space to balance the sample later, as described in §3.1.2 and §3.1.3, by adding further meanings in order to achieve the target of 200 meanings (originally chosen for commensurability with Urban 2012). A cut-off point of 17 languages would have resulted in 133 meanings and the need to add a further 67 during balancing, which was deemed too many; a cut-off point of 15, on the other hand, would have resulted in 191 meanings and space for only a further 9 to achieve balance, which was deemed too few. This resulted in a “starter set” of 159 meanings (12). The number of languages in which each form is represented by an analysable form (i.e. the value of langs) is 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, STEPDUGHTER, 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, PITY, 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,
FROM THE ENGLISH WORDS USED AS LABELS, TWO THINGS MAY BE OBSERVED. 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 THUS THE ENGLISH LANGUAGE COUNTERPARTS) ARE NOT BINOMINALS. SOME OF THESE ARE MONOMORPHEMIC, INDICATING THAT, WHILE SOME MEANINGS, E.G. PUPIL, CLEARLY HAVE ANALYSABLE COUNTERPARTS IN MANY LANGUAGES, THEY DO NOT DO SO IN ALL. OTHERS ARE ANALYSABLE BUT NOT BINOMINAL, SUCH AS SHOEMAKER, WHICH CONTAINS THE ACTIONAL ELEMENT MAKE AND IS THUS OF ONOMASILOGICAL TYPE 1 (SEE PAGE 11), AND BEGINNING, WHICH CONSISTS OF THE ONOMASILOGICAL BASE (THE NOMINALIZING AFFIX -ING) AND THE DETERMINED ELEMENT OF THE MARK (BEGIN), BUT NO DETERMINING ELEMENT, AND IS THUS OF ONOMASILOGICAL TYPE 2. MOREOVER IT CAN BE OBSERVED THAT WHILE SHOEMAKER PROBABLY HAS A rather high potential to be expressed as a binominal 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 constructions that are used across languages 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 construction to expect) were readily available. The languages chosen for this purpose were French, Welsh, .

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 .
The typology of binominal lexemes

Turkish, 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. For each language, every form representing one of the 159 meanings was analysed in order to determine the kind of construction it exhibited. The results are presented in the following subsections.

French

Based on the literature (Arnaud 2015; Bourque 2016), the following binominal 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. ADVJZ**, e.g. *incendie volontaire* [fire will. ADVJZ] 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).

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

*timbre-poste* POSTAGE STAMP (8), *oiseau-mouche* HUMMINGBIRD, *mot-clé* KEYWORD, *nœud papillon* BOW TIE, *papier toilette* TOILET PAPER

(14) **Head à Mod** (15x)


---

1 Meanings given in bold are represented in WOLD but are not among the top 159 produced by the SQL query. The number of languages in which the meaning is represented by a complex word 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 resulted in a further 10 examples of this type. Adding the two meanings in (15) then yielded three more examples.

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

 gyddfðorch, gyddfðadwyn NECKLACE (10), rheilffordd 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**

 başkent CAPITAL CITY

(17) **ModMATERIAL Head**

 Altın yüzük GOLD RING, taş köprü 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 patterns predicted in the literature (Pepper 2010b) would be revealed by the set of meanings. Those patterns were

- **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**

\[ n\text{\textit{i}}\text{\textit{ii} d\text{\textit{anni}}} [\text{person work}] \text{\textsc{Farmer}} (17); n\text{\textit{ii} s\text{\textit{inni}}} [\text{person darkness}] \text{\textsc{Cap\textsc{tive or Prisoner}} (23); k\text{\textit{ow} b\text{\textit{nn}}} [\text{bone head}] \text{\textsc{Skull}} (25) \]

(19) **Mod Head**

\[ n\text{\textit{y\text{\textit{u}r y\text{\textit{aa}}} [nose hole}] \text{\textsc{Nos\textsc{tril}} (35); t\text{\textit{w\text{\textit{a\text{\textit{a} y\text{\textit{a}}} [ear dirt}] \text{\textsc{Earwax}} (24); b\text{\textit{nn} k\text{\textit{ow}}} [head bone] \text{\textsc{Skull}} (25) \]

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

**Japanese JAP**

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 (2016:492 ff) proffered further examples (21).

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

\[ w\text{\textit{aki} n\text{\textit{o} sh\text{\textit{ita}}} [armpit \text{\textsc{Gen} under}] \text{\textsc{Armpit}} (11); k\text{\textit{i} n\text{\textit{o} mi} [tree\text{\textsc{Gen} fruit}] \text{\textsc{Nut}} (10); e\text{\textit{no} gu [painting\text{\textsc{Gen}?}] \text{\textsc{Paint}} (9); b\text{\textit{u\text{\textit{do} no ki} [grape \text{\textsc{Gen} tree}] \text{\textsc{Vine}} (14); t\text{\textit{osh\text{\textit{ish\textit{ita} no ky\text{\textit{\textit{odai}} [age:down \text{\textsc{Gen} sibling}] \text{\textsc{Younger Sibling}} (9); t\text{\textit{oshi\text{\textit{ue} no ky\text{\textit{\textit{odai}} [age:up \text{\textsc{Gen} sibling}] \text{\textsc{Older Sibling}} (9) \]

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

\[ k\text{\textit{o} no ha [tree\text{\textsc{Gen} leaf}] \text{\textsc{Leaf}}; h\text{\textit{i} no de [sun\text{\textsc{Gen} rise}] \text{\textsc{Sunrise}; k\text{\textit{um\text{\textit{o} no su} [spider\text{\textsc{Gen} web}] \text{\textsc{Spider Web}}; n\text{\textit{om\text{\textit{i} no iti} [flea\text{\textsc{Gen} market}] \text{\textsc{Flea Market}}; a\text{\textit{ma} no k\text{\textit{awa} [heaven\text{\textsc{Gen} river}] \text{\textsc{Milky Way}}; m\text{\textit{ago} no te [grandchild\text{\textsc{Gen} hand}] \text{\textsc{Back Scratcher}} \]

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 kum\textit{a 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.

\[ ^{1} \text{Theil’s word list contains two entries with the definition cr\text{\textit{ane} ‘skull’ (see Pepper 2016: 301).} \]
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 59) on page 375.

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
necessary to reduce the number of meanings drastically. By this time data had been collected and analysed for 50 languages.¹ In addition to the kind of basic information available in WOLD regarding analysability and morpheme gloss, each item had been annotated for its basic structural type (see Table 21 on page 99). From this it became clear that some meanings were much more “binominal” (that is, much more likely to be represented by binominal constructions) than others. For example, of the 52 words meaning NOSTRIL, 40 were binominal (most of them consisting of nouns meaning NOSE and HOLE), whereas none of those meaning VICTORY were (not surprisingly, many of the latter were derived from verbs with the actional meaning WIN). The set of meanings, in other words, was not optimal and could be pruned without significant reduction in the number and diversity of binominal constructions in the data. This was accomplished in several steps.

First, three columns were added to the spreadsheet meaning, as shown in Figure 12 (cf. Figure 11 on page 64).² Column D (words) shows the total number of words per meaning (including monomorphemic forms) in the expanded database of 50 languages; column E (NN) shows the number of binominals per meaning; and column F (freq) shows the frequency of binominals (NN / words) as a percentage. The table is sorted in descending order by column F. Observe that NOSTRIL is represented by 52 words, of which 40 (i.e. 77%) are binominal.³ The remaining 12, including Fr. narine and Sakha tani; are not (synchronously) analyzable. NOSTRIL is the “most binominal” of all the meanings (that is, it is the meaning that is most frequently represented as a binominal), and it is followed by EARLOBE, PALM OF HAND, ARCTIC LIGHTS, etc. The least binominal of the 201 meanings were DIVORCE, ITCH, PRAISE, QUARREL, REMAINS, SWELLING, VICTORY, none of which were represented by as much as a single binominal: Like VICTORY, most of them are derived from action-roots and are thus OT1 or OT2 and not OT3 (see page 11).

¹ French, German, Hindi, Hungarian, Irish, Italian, Lithuanian, Polish and Welsh, in addition to the 41 languages found in WOLD (see page 51).
² Recall that column C (langs) shows the number of languages in which the meaning is represented by an analysable word in WOLD. That information is no longer relevant at this stage.
³ There were 40 binominals that represented NOSTRIL in the data set of 50 languages, but two languages have two words for this concept and use the same construction for both: Indonesian lubang hidung [hole nose] and rongga hidung [cavity nose] (both Head Mod) and Swahili mwanzi wa pua [bamboo con nose] and tundu la pua [hole con nose] (both Head con Mod). These four data points thus give rise to just two construction types.
### Chapter 3. Meanings, languages and data

<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*

The complete ranking in order of “binominality”, that is, the likelihood of being represented as a binominal, is given in Appendix H (Table 60) on page 377. This ranking offers a principled method for trimming the list of meanings. A cut-off point could be chosen at random, for example, at 25, 50, 100 or 120 meanings, but it is important to know what the consequences of the 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 structural patterns. In order to make an optimal choice, a consequence analysis was performed using the results of the preliminary formal analysis described in §3.4. As explained there, every binominal was annotated with its construction, as illustrated 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>qhov 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>süwúlí kánzà.bè</td>
<td>[opening nose.GEN]</td>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>8</td>
<td>Ket</td>
<td>ołan.d qûk</td>
<td>[nose.GEN hole]</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>9</td>
<td>Lower Sorbian</td>
<td>nos.owa žërka</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>ńdu.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 ḥaya’yah.a</td>
<td>[nose hole.3SG]</td>
<td>Mod Head.3SG</td>
</tr>
</tbody>
</table>

*Figure 13: Simplified view of the table word showing various 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.

Figure 14: Number of constructions as a function of number of meanings

The graph shows the number of constructions found in the 50-language sample for various meaning sample sizes. 201 meanings yield 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).

(24) Archi Mod.GEN Head Ket Mod.GEN Head
    Ceq Wong Head Mod Lower Sorbian Mod.ADJZ Head
    Dutch Mod Head Malagasy Head Mod
    English Mod Head Manange Mod Head
    Galibi Carib Mod.POSS Head.POSS Mandarin Chinese Mod Head
    Gawwada BASE.FEM Mapudungun Head Mod
    German Mod.LE.Head Oroqen Mod.APOSS Head.POSS
    Gurinji Mod Head Otomi Head Mod
    Hausa Head.LK Mod Q’eqchi’ Head Mod
    Hawaiian Head Mod Saramaccan Mod Head
    Hmong Daw Head Mod Seychelles Creole Head Mod
    Hungarian Mod Head Swahili Head CON Mod
    Hupdë Mod Head Takia Mod.3SG Head.3SG
    Imababra Quechua Mod Head Thai Head Mod
    Indonesian Head Mod Vietnamese Head Mod
    Iraqw Head.CON Mod Wichi POSS.BASE.LOC
    Irish Head Mod.GEN Yakut Mod Head.3SG
    Japanese Mod Head Yaqui Mod Head
    Kanuri Head Mod.GEN Zinacantán Tzotzil Head Mod

Table 13: Constructions yielded by NOSTRIL
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 the word 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.

\[(25)\]

<table>
<thead>
<tr>
<th>Language</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>Mod Head.POSS</td>
</tr>
<tr>
<td>German</td>
<td>Mod.Head</td>
</tr>
<tr>
<td>Ket</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Kildin Saami</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>Old High German</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Polish</td>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>Takia</td>
<td>Mod Head.3SG</td>
</tr>
<tr>
<td>Tarifit</td>
<td>Head PREP CON.Mod</td>
</tr>
<tr>
<td>Welsh</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Wichí</td>
<td>POSS.Mod Head</td>
</tr>
</tbody>
</table>

Table 14: Additional constructions yielded by EARLOBE

As more meanings are added to the list, the number of constructions increases, 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 value of 25% or more (cf. Table 60 on page 377). The number of constructions that would be lost in such a scenario is 55. These are listed in their entirety in Table 61 on page 379, along with each individual binominal conforming to the pattern in question. For convenience, a subset is reproduced here as Table 15. (The significance of the dagger † is explained on page 76.)

The following observations may be made regarding Table 15 (and Table 61):

- 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.
### The typology of binominal lexemes

<table>
<thead>
<tr>
<th>Language</th>
<th>Base</th>
<th>Mod</th>
<th>Head</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bezhta</td>
<td>Base.SUP.ATTR</td>
<td>mähä.λ.'ä.kö</td>
<td>[door_frame.SUP.ATTR]</td>
<td>SERVANT</td>
</tr>
<tr>
<td></td>
<td>Head Mod</td>
<td>c’uddo c’emuc’</td>
<td>[red egg]</td>
<td>YOLK</td>
</tr>
<tr>
<td></td>
<td>Mod.ADJZ Head</td>
<td>nucodaq t’ot’</td>
<td>[honey:ADJZ fly]</td>
<td>BEE</td>
</tr>
<tr>
<td>Dutch</td>
<td>Base.M</td>
<td>weduw.naar</td>
<td>[widow.M]</td>
<td>WIDOWER</td>
</tr>
<tr>
<td>Hindi</td>
<td>Base.AGT</td>
<td>lohär</td>
<td>[iron.AGT]</td>
<td>BLACKSMITH</td>
</tr>
<tr>
<td></td>
<td>† Mod.ADJZ Head</td>
<td>havä.i jahäz</td>
<td>[air.ADJZ ship]</td>
<td>AIRPLANE</td>
</tr>
<tr>
<td>Irish</td>
<td>Base.NMLZ</td>
<td>draí.acht</td>
<td>[magician.NMLZ]</td>
<td>MAGIC</td>
</tr>
<tr>
<td></td>
<td>† Head Mod</td>
<td>tráth.nóna</td>
<td>[occasion.noon]</td>
<td>AFTERNOON</td>
</tr>
<tr>
<td>Malagasy</td>
<td>Head SOC.Mod</td>
<td>lehilahy manaN.vàdy</td>
<td>[man with.spouse]</td>
<td>MARRIED MAN</td>
</tr>
<tr>
<td></td>
<td>NMLZ.Base</td>
<td>fi.lòha</td>
<td>[NMLZ.head]</td>
<td>CHIEFTAIN</td>
</tr>
<tr>
<td></td>
<td>NMLZ.Base.CIRC</td>
<td>fanambakàna &lt; faN.ambàka.ana</td>
<td>[NMLZ.deceit.CIRC]</td>
<td>DECEIT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fifidiànanà &lt; fi.fidy.(an)ana</td>
<td>[NMLZ.choice.CIRC]</td>
<td>ELECTION</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fiadìnanà &lt; fi.àdy.ana</td>
<td>[NMLZ.fight.CIRC]</td>
<td>WEAPONS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>famosavàna &lt; faN.mosàvy.ana</td>
<td>[NMLZ.witchcraft.CIRC]</td>
<td>MAGIC</td>
</tr>
<tr>
<td>Seychelles Creole</td>
<td>† Base.AGT</td>
<td>prizon.nyen</td>
<td>[prison.AGT]</td>
<td>CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bazar.dye</td>
<td>[market.AGT]</td>
<td>MERCHANT</td>
</tr>
<tr>
<td>Wichí</td>
<td>Base.AGT</td>
<td>tshotoy.wu</td>
<td>[animals.AGT]</td>
<td>HERDSMAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tiena.wu</td>
<td>[market.AGT]</td>
<td>MERCHANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sapatu.wu</td>
<td>[shoe.AGT]</td>
<td>SHOEMAKER</td>
</tr>
<tr>
<td></td>
<td>Base.LOC</td>
<td>kanu.hi</td>
<td>[needle.LOC]</td>
<td>SUGAR CANE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y’amekw.hi</td>
<td>[excrement.LOC]</td>
<td>TOILET</td>
</tr>
<tr>
<td>Yaqui</td>
<td>Base.LOC</td>
<td>sisi’iwoo.chi</td>
<td>[iron.LOC]</td>
<td>TOOL</td>
</tr>
<tr>
<td></td>
<td>† Base.NMLZ</td>
<td>ko’oko.a</td>
<td>[pain.NMLZ]</td>
<td>DISEASE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ju-ubi.wa.me</td>
<td>[RDP-wife.PASS.NMLZ]</td>
<td>WEDDING</td>
</tr>
<tr>
<td></td>
<td>† Mod.PL Head</td>
<td>waim asoa</td>
<td>[sister/brother.PL daughter]</td>
<td>STEPdaughter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waim marai</td>
<td>[sister/brother.PL daughter]</td>
<td>STEPdaughter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waim achai</td>
<td>[sister/brother.PL father]</td>
<td>STEPFather</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waim maala</td>
<td>[sister/brother.PL mother]</td>
<td>STEPMOTHER</td>
</tr>
<tr>
<td>Zinacantán Tzotzil</td>
<td>Mod Head</td>
<td>shokan na</td>
<td>[side house]</td>
<td>NEIGHBOUR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hmulavi ʔantz</td>
<td>[sinner woman]</td>
<td>PROSTITUTE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>meʔanal ʔantz</td>
<td>[poverty/misery/grief woman]</td>
<td>WIDOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>meʔanal vinik</td>
<td>[poverty/misery/grief man]</td>
<td>WIDOWER</td>
</tr>
<tr>
<td></td>
<td>opaque</td>
<td>k’op ʔoʔon</td>
<td>[word/argument heart]</td>
<td>ANXIETY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kachimpa pom</td>
<td>[pipe incense]</td>
<td>BEE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mishik’ pom</td>
<td>[belly_button incense]</td>
<td>BEE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ton pom</td>
<td>[stone incense]</td>
<td>BEE</td>
</tr>
</tbody>
</table>

*Table 15: 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 Fr. ruche d’abeille.s [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.

There are 50 different meanings represented in Table 15. These are listed in Table 16 in (descending) order of the number of times they occur in Table 15 (column A). Those meanings that occur most frequently are CHIEFTAIN (5), MAGIC (5), BEE (4), MARRIED MAN (4), MARRIED WOMAN (4) and MERCHANT (4). The table on page 377 shows these six meanings ranked as numbers 98, 105, 101, 135, 103 and 117 respectively in terms of binominality. Thus, for these top six to be included, the size of the list 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 not contribute to any reduction in the number of lost constructions (since they are not listed in Table 15). Furthermore, the addition of both MARRIED MAN and MARRIED WOMAN is unnecessary, since they employ the same constructions wherever they occur in Table 16 (i.e. in Kanuri, Malagasy, Mapudungun and Takia). It is sufficient to add just one of them.

<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>
<tr>
<td>yolk</td>
<td>1</td>
<td>17</td>
</tr>
<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>13</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>intention</td>
<td>1</td>
<td>19</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>arson</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 16: Meanings represented by lost constructions (84 meaning sample)
From this it was clear that adding more meanings based solely on 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, but only if they occur in both Table 15 and Table 16. Accordingly, the 16 most binominal meanings (shown in boldface) were selected from Table 16 and added to the list of 84 that had resulted from applying a cut-off value of 25%. These all have a binominality value of at least 17% (column B of Table 16) and bring the number of meanings to a total of 100.

The final list of 100 meanings is shown in Table 17. 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 original data set based on 201 meanings. Those that were lost are marked with a dagger (†) in Table 15 (and Table 61). 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 arsenal of binominal constructions, that Irish exhibits the occasional head-initial compound, and that Seychelles Creole has a Base.AGT agentive derivation, etc. However, it could never be the goal of the present study to detect every single binominal construction in every language sampled. 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.

Meanings can be classified in two different ways. As mentioned in §2.4.1, the 1,460 meanings in WOLD are assigned to a set of 24 semantic fields (cf. page 53) and the same classification is retained in the present study for those 82 meanings that also occur in WOLD. For 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).

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 17) consisting of the seven types: Person; Animal; Concrete Object, Substance, Condition; Place; Time; Activity; and Miscellaneous Abstract Entities. This was suggested 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 adopt (and extend) the latter in the present study (see §5.2.2), it seemed reasonable to use her system of semantic types as well. However, as I show in the next section, this led to a very unbalanced result and therefore a revised scheme (labelled Semantic type in Table 17) was created instead.
<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>backpack</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 17: Final set of 100 meanings*
The typology 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>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 17: Final set of 100 meanings (cont.)
### Table 17: Final set of 100 meanings (cont.)

<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><strong>stone bridge</strong></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><strong>toilet paper</strong></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><strong>toolbox</strong></td>
<td>Modern world</td>
<td>advanced</td>
<td>concrete</td>
</tr>
<tr>
<td>90</td>
<td><strong>toothbrush</strong></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><strong>water pump</strong></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><strong>windmill</strong></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>
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 structural types. 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 in Figure 10): the words should be analysable (because every binominal must by definition be analysable) and the meaning they express 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 found in the most 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 structural types. This was accomplished by testing the initial set of 159 meanings against five languages for which information regarding types was easily available in the literature (§3.1.2), and then adding new meanings deemed 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, that list of 201 meanings was reduced to 100 meanings in such a way as to ensure that as much structural diversity as possible was retained (§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 general process could have been improved, but in retrospect 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,
Chapter 3. Meanings, languages and data

it could have been 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 §6.1.2 under the rubric *Data description 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 those that might be termed ‘secondary’, such as EARLOBE, EYELASH and NOSTRIL (all of which are in the final list), that tend to be named in terms of major body parts such as EAR, EYE and NOSE (none of which are included). On the other hand, body parts should not be allowed to dominate, since they may tend to use the same construction and thus limit the overall structural diversity.

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’).

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 72.5 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 seek to remove 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 distribution (Figure 16). As a result, semantic field cannot form the basis of any comparisons or typological generalizations (unless these are restricted to, say, The body vs. Modern world).

![Figure 16: Distribution of meanings across semantic fields](image)

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 even and therefore more suitable for statistical analysis.

![Figure 17: Distribution of meanings across semantic types](image)
The 100 meanings shown in Table 17 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 the 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. Typological studies are thus invariably based on a subset 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:

- **probability sample** a (stratified) sample that is areally and genetically as unbiased 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 questions. 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 1 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 claimed to be 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). But they argue that their sample is preferable to using just one or two languages,

---

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).
or to relying on intuition. (For a map showing the geographical distribution of WOLD languages, see Figure 18 on page 88).

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 1 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 58 on page 374).

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 given in Miestamo, Bakker & Arppe (2016:272–273) show that a 100-language sample constructed using the DV method with the Ethnologue 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 §6.2.2 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 “controlled opportunism” was adopted. The 41 WOLD languages (see page 52) constituted the foundation of the sample, and to these were added the five languages used for testing the meaning list (§3.1.2). Thereafter, languages were added based on questionnaires received from language experts and, to quote 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 might otherwise have been achieved (for example, within Indo-European). Moreover, the inclusion of multiple languages from a single genus permits comparisons at the micro-typological level (for example, the one that reveals binominals to be rather good indicators of genus, see §6.2.2). There is, as Saulwick writes, “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).

Figure 18: Areal distribution of language sample

The current sample of 99 languages consists of the languages shown in Figure 18 and listed in Table 18 on the next page.¹ (The map includes ISO 693-3 language codes which are legible via 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. For a list of sources, and in order to find the ISO code from the language name, see Appendix B. In the next section I provide a more detailed overview and evaluation of the sample.

¹ Data from the following languages has been received but has not yet been fully integrated into the database: Amarakaeri, Croatian, Trinitario, Tuwari, Wartha Thuntai, Western Farsi.
The sample consists of 99 languages distributed across 37 families and 66 genera. On average there are 1.5 languages per genera, 2.7 languages per family and 1.8 genera per family, but the distribution, as noted above, is very uneven (Table 19). Two families, Indo-European and Afro-Asiatic, account for over a third (34%) of the languages in the sample and 17% of the genera. At the other end of the scale, 23 families are represented by a single language. Of the 66 genera, 18 are represented by more than one language (26).

(26)  Cushitic (5), Germanic (5), Slavic (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)
The typology of binominal lexemes

<table>
<thead>
<tr>
<th>family</th>
<th>languages</th>
<th>genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.  Indo-European</td>
<td>22</td>
<td>7</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>
</tbody>
</table>

Table 19: 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 66 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 20 shows, the sample is unbalanced also in that respect.

<table>
<thead>
<tr>
<th>area</th>
<th>languages</th>
<th>genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.  Africa</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>2.  Eurasia</td>
<td>38</td>
<td>21</td>
</tr>
<tr>
<td>3.  Oceania/SE Asia</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>4.  PNG/Australia</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>5.  North America</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>6.  South America</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>7.  Pidgins/Creoles</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 20: Areal distribution of languages
One 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, PNG/Australia is least well represented, with six languages in all, and thus a balanced subset of the current sample could contain no more than \((6 \times 6 = 36)\) 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 PNG/Australia, North America and South America. Adding data from a further 17 languages would suffice for a balanced sample of 78 languages (13 per area). However, this would still require the elimination of 25 Eurasian languages already in the database.

3.3 Data collection

This section describes the three kinds of sources that were used for the data collection: the WOLD database, questionnaires and dictionaries.

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 52. 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.cld.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. These included the glossing information, but it was buried inside a JSON field and thus not easily accessible to most tools. This 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 an extremely time-consuming and error-prone task. Fortunately it turned out that the data also existed in the form of a

---

1 In the spirit of the “re-doing typology” debate in *Linguistic Typology* 10(1), this section contains a lot of technical detail which is intended to aid the further reuse of the WOLD data and enable replication of the present study.
The typology of binominal lexemes

PostgreSQL dump in another GitHub project at http://github.com/clld/wold-data. This file, once downloaded and uncompressed, could be loaded directly into the DBMS. It was now possible to extract an initial data set from WOLD using an SQL query (27).

(27) 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

This selected every translation equivalent (or “word”) for each of the 100 meanings present in WOLD. For each such word the query extracts 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 (28).

(28) "Hausa";"the nostril";"kàfá-r háncii";"kàfá-r háncii-ı  
    [orifice-GEN nose]"

---

1 My thanks to Robert Forkel for helping me decipher the database.
2 Recall from §3.1.4 that only 82 of the 100 meanings are to be found in WOLD.
This file would not import cleanly into Microsoft Excel, which erroneously interprets a new line that occurs *inside* a quoted string as a record end. The CSV file was therefore modified by replacing such new lines with spaces.\(^1\) Once this had been done, the file could be imported into Excel (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>language</td>
<td>meaning</td>
<td>word</td>
<td>gloss</td>
</tr>
<tr>
<td>2</td>
<td>Hausa</td>
<td>the nostril</td>
<td>káfář háncii</td>
<td>káfá-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>tinzá</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Thai</td>
<td>the nostril</td>
<td>ruucamðuk</td>
<td>ruu-camðuk [hole-nose]</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 *fómford* 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].

---

\(^1\) Regular expressions were used for this purpose. The search string was \(\;:\^[^\"\n]+\) and the replace string \(\n\) (where \(\n\) 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 \(\n\) 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. \(\n\)\n), whereas new lines inside cells are LF only (i.e. \(\n\)).
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 2015). 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 30 colleagues. I owe these linguists an enormous debt. Without them the present project would not have been possible.

---

1 See Typographical and naming conventions on page xxiii for the modified glossing rules.
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.

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 language where the dictionary used an unfamiliar contact language. 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 17 on page 77) 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 clean-up

Before the analysis could proceed some minor clean-up 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 periods instead of hyphens to mark morpheme boundaries (see Typographical and naming conventions on page xxv). Thus, for example, AG.N, used to denote nomen agentis in the Iraqw vocabulary, was changed to AGT (agentive), and 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 (Comrie, Haspelmath & Bickel 2015). For example, the two examples in (30a) had to be amended as shown in (30b), both for the sake of consistency and in order 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 possessum 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 in the database. Despite this, such instances were retained on the grounds 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 (30).

¹ See http://wold.clld.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.
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 369).

3.4 Data annotation

The annotation (or coding) of the data consisted of three parts: distinguishing binominals from other words (§3.4.1), determining the head of each binominal (§3.4.2), and establishing the type of construction (§3.4.3). All three tasks were initially performed on the version of the database consisting of 201 meanings and 50 languages.1 After reducing the meaning list from 201 to 100, the same tasks were performed on data from a further 50 languages.

3.4.1 Identifying binominals

The first task consisted in a rough categorization of each word based on its onomasiological type. 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. bůja 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 21.

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 želazna" and "kolej želazna [course IRON.ADJZ]").2 Columns headed c1 and c2 show the first and second (major) constituent, respectively.

---

1 Recall that the initial annotation of 50 languages provided the basis for reducing the number of meanings from 201 to 100 in a principled manner (cf. §3.1.4).
2 There is also a note (not shown here) indicating that this form is dated. The modern word is kolej.
Table 21: Preliminary classification of structural types

<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>unanalyisable word</td>
</tr>
</tbody>
</table>

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 we consider first thing-roots and thereafter thing-affixes. Distinguishing roots and affixes cross-linguistically is notoriously tricky, but for present purposes the simple definitions found in introductions to morphology such as Lieber (2010) will suffice (31).

(31)   **affix**: A bound morpheme that consists of one or more segments that typically appear before, after, or within, a base morpheme.

   **root**: The part of a word that is left after all affixes have been removed.

**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 clearly denote things (some more prototypically than others):

---

1 For an explanation of onomasiological types, see page 11.
The typology of binominal lexemes

<table>
<thead>
<tr>
<th>E</th>
<th>G</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>word</td>
<td>gloss</td>
<td>meaning</td>
<td>str</td>
<td>c1</td>
<td>c2</td>
<td>pos</td>
</tr>
<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>płatek 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 mlecz.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 zelaz.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.rośl</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

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 clearly denote actions:

*eat, mind/graze, grow*

---

1 The significance of the shaded cells is explained on page 108.
Some of the other glosses are not so 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 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 it 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*,

---

¹ The other criteria are restriction of usage, semantic range, semantic pattern, phonetic shape, morphologic type and criterion of stress.
as is the case in, for example, Welsh *melyn.ŷw* [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 and not a Property (even if, as a thing, it is largely defined in terms of a property). Allowing ourselves to switch 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 2001:54) scale of temporal stability (Figure 21).

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

| noun | adj | adj | verb | verb | verb |

*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.

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, *challwafe* might have been rejected on the internal counterevidence of *wizi.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). 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. 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 trichotomy applied here, but as a locative it can be interpreted as a kind of place, and thus as a thing-morph.

**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. 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 *vrăi.i.tor* [magic.VBLZ.AGT] SORCERER OR WITCH. Constructions like these are treated analogously to Vietnamese *bůa 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. *chorzy* ‘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.ADJZ] ‘milky’ in combination with a thing-root or thing-affix is classified as Type 3, i.e. as a binominal. 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 (Štekaower 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. 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 construction (see §3.4.3 below).

3.4.2 Determining the head

Once binominals had been identified, the English gloss of the primary constituents was add in columns L and M (labelled c1 and c2, respectively). This would prove useful in ensuring consistency in later stages of the analysis. Also, the position of the head was indicated in column T, 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 every binominal (with very few exceptions) exhibits a clear head-modifier structure.

It was a desideratum to know which of the primary constituents was the head, for two reasons. First, the position of the head affects the type of construction. 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

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.
Bourque’s 2014 classification used in Chapter 5, 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.

**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. 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. And finally, only the semantic head can be used to determine the nature of the semantic relation (see Chapter 5).

The semantic head is determined via the ISA condition. The formulaic statement of this condition, given by Scalise and Fábregas in (32), makes the (unwarranted) assumption that the head is the second (right-hand) constituent.

(32) In a compound [ [ ]x [ ]Y ]z, 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
Applying this as a diagnostic to the Polish examples in the shaded cells in Figure 20 (page 100) produces the results in (33):

(33) The head of *zorza polarna* is *zorza* because arctic lights ISA dawn, not a pole\(^1\)
    The head of *platek ucha* is *platek* because an earlobe ISA lobe, not an ear
    The head of *złoty pierścionek* is *pierścionek* because a gold ring ISA ring, not a gold
    ?? The head of *pohudnie* 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 *kamieński 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\(^2\)

Most of these results are uncontroversial, but *pol.u.dnie* [half.IE.day] MIDDAY requires comment. In the database, almost all analysable words meaning MIDDAY have a constituent glossed as ‘day’.\(^3\) When this is the case, the other constituent is always ‘middle’\(^4\) 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 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: 111 ff). 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 as binominals because they contain an actional element.\(^5\) As for constructions such as

---

\(^1\) 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.

\(^2\) 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*.

\(^3\) Or ‘sun’, which often co-lexifies with ‘day’, cf. Urban (2012:703).

\(^4\) Or occasionally – and metaphorically – ‘heart’.

\(^5\) Štekauer’s five-way typology described in §1.2.3 does not account for this type. A revamped eight-way typology (Štekauer 2016), in which only Types 1-3 are remain unchanged, has them as a (new)
Eng. *skin head* and It. *testa rasata* [head shaven] ‘skin head’, 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 *táá guṝ* [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, *guṝ* ‘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.

**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:

Type 5, in which the verbal component (*porta-*) expresses both the base and the determined element of the mark. An alternative approach would be to characterize them as forms in which both constituents of the mark are expressed, but not the onomasiological base.
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).

Six of Bauer’s criteria for establishing the head of a construction are tested for each of Onomasiological Types 1 to 4.¹ The first, and most important, criterion is that of hyponymy. All four types pass this test without any difficulties, “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 (34) with labels that will be more familiar than those used in the original (the onomasiological base (i.e. the head) is emphasised).

(34) 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 the quote given above shows), but he recognizes that the OT4 example might raise objections:

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 timeverb < timenoun.
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 100) would be regarded as the onomasiological base, with the 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 želaz.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.).

The theoretical frameworks in which this research is conducted thus converge on the view that derivational affixes generally constitute the head (or profile determinant) of the constructions in which they occur, and this is reflected in the codes given in Figure 20; as can be seen, all the Polish derivational binominals are right-headed.

3.4.3 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 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.

---

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).
<table>
<thead>
<tr>
<th>id</th>
<th>word</th>
<th>gloss</th>
<th>meaning</th>
<th>str</th>
<th>c1</th>
<th>c2</th>
<th>pos</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>5882</td>
<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>
<td>Mod.LE.Head</td>
</tr>
<tr>
<td>5865</td>
<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>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>5910</td>
<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>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>5917</td>
<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>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>5849</td>
<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>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>5830</td>
<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>
<td>Head Mod.ADJZ</td>
</tr>
<tr>
<td>5883</td>
<td>droga mlecz.na</td>
<td>[road milk.ADJZ]</td>
<td>milky way</td>
<td>NN</td>
<td>road</td>
<td>milk</td>
<td>L</td>
<td>Head Mod.ADJZ</td>
</tr>
<tr>
<td>5897</td>
<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>
<td>Head Mod.ADJZ</td>
</tr>
<tr>
<td>5851</td>
<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>
<td>Base.NMLZ</td>
</tr>
<tr>
<td>5930</td>
<td>wiatr.ak</td>
<td>[wind.NMLZ]</td>
<td>windmill</td>
<td>NN</td>
<td>wind</td>
<td>NMLZ</td>
<td>R</td>
<td>Base.NMLZ</td>
</tr>
<tr>
<td>5896</td>
<td>król.owa</td>
<td>[king.F]</td>
<td>queen</td>
<td>NN</td>
<td>king</td>
<td>F</td>
<td>R</td>
<td>Base.F</td>
</tr>
<tr>
<td>5863</td>
<td>dziewczyn.ka</td>
<td>[girl.DIM]</td>
<td>girl</td>
<td>NN</td>
<td>girl</td>
<td>DIM</td>
<td>R</td>
<td>Base.DIM</td>
</tr>
<tr>
<td>5932</td>
<td>żółt.ko</td>
<td>[yellow.DIM]</td>
<td>yolk</td>
<td>NN</td>
<td>yellow</td>
<td>DIM</td>
<td>R</td>
<td>Base.DIM</td>
</tr>
<tr>
<td>5857</td>
<td>ryb.ak</td>
<td>[fish.AGT]</td>
<td>fisherman</td>
<td>NN</td>
<td>fish</td>
<td>AGT</td>
<td>R</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>5895</td>
<td>garn. arz</td>
<td>[pot.AGT]</td>
<td>potter</td>
<td>NN</td>
<td>pot</td>
<td>AGT</td>
<td>R</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>5894</td>
<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>
<td>Base.ADJZ.DIM</td>
</tr>
</tbody>
</table>

Figure 22: Subset of the Polish binominal data showing constructions
The result of this analysis is illustrated in Figure 43, which shows the binominals from Figure 20 along with their construction. 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.

3.5 Chapter summary

This chapter has described how the list of meanings and language sample were constructed, how the data were collected, and how they was annotated. The latter was covered in some detail in order to highlight certain theoretical and practical issues relating to the task of distinguishing binominals from other words, determining the head and its position, and establishing the construction. The next chapter discusses the classification that was developed on the basis of the annotated data.
Chapter 4. Formal classification

4 Formal 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. Chapter 3 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 classification based as far as possible on purely formal (i.e. structural) features of binominal constructions. In the next chapter I develop a two-level classification based on the nature of the semantic relation \( \mathcal{R} \) between the two major constituents of a binominal.

In §4.1 I review two classifications that serve as points of departure for developing a formal typology: Koptjevskaja-Tamm’s (2002) classification of possessive noun phrases (PNPs) in Europe (§4.1.1), and Croft’s (2003) use of possessive constructions to illustrate a “cross-linguistically valid description of morphosyntactic structures” (§4.1.2). This section also highlights the problems inherent in representing a classification as a hierarchy.

In §4.2 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 eight-way classification is based primarily on the degree of marking and the locus of marking; the more detailed, 16-way classification adds constituent order to these parameters. A detailed description of each of the eight basic types follows in §4.3, 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 §4.4 and explanations are put forward to account for them. §4.5 then exemplifies the kinds of gradient phenomena found in the data and shows how they are handled by the proposed representation.
4.1 Theoretical prerequisites

4.1.1 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 23). 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 hierarchically with six leaf nodes, labelled Juxtaposition, Dependent-marking, Double-marking, Head-marking, Prepositions and Linking pronouns.¹

![Diagram of typology of possessive NPs in Europe](after Koptjevskaja-Tamm 2002:144)

This classification mostly accounts for the data, but as Koptjevskaja-Tamm points out, it is only a “very rough and simplified approximation” to the great structural diversity within European PNPs. This point 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 23 is based on degree of fusion, with an opposition between synthetic and analytic

¹ Koptjevskaja-Tamm also mentions a seventh type (possessive compounding) which is “mainly restricted to Northern Swedish” and is not included in her typology of major structural types.
and with juxtaposition in between. (The latter is usually analytic when it occurs in PNPsis, but it is the paradigm case in the domain of binominals.) 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 2003: 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 three examples presented on page 1:

- Ger. eisen.bahn [iron.way] – belongs in the category of **juxtaposition**
- Fr. chemin de fer [way of iron] – is of the **prepositional** type
- Rus. želez.naja doroga [iron.BADJZ road] – is **dependent-marked**

In addition, two other PNP types are to be found in the list of binominals on page 14 (example 5):

- Tur. demir.yol.u [iron.road.POSS:3SG] – an example of **head-marking**
- Takia ndu.n awa.n [nose.3SG mouth.3SG] – employs **double marking**

However, this set of five types does not cover the full variety found in (5), or in the domain of binominal lexemes in general. Some languages outside Europe (the area investigated by Koptjevskaja-Tamm) have postpositions instead of prepositions, a classic example being the Japanese no construction in kumo no su [spider GEN web]. The category prepositions has therefore to be generalized to **adpositions**. Moreover, the category juxtaposition, when applied to binominals, conceals a number of significantly different types. In addition to the obvious case of **compounds**, we find two others:

- **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’ in §4.5.7);
- **classifier constructions**, as in Murui Huitoto ui.tirai [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.

In order to capture this variety among binominals it is necessary to subdivide the type juxtaposition into three: compounds, derivation and classifier constructions.
A further difference between possessive noun phrases and binominals is that the category (synthetic) dependent-marking covers two distinct types of construction: those with an adjectival suffix, e.g. Rus. želez.naja doroga [iron.ADJZ road], and those with a genitive suffix, e.g. Bezhta kil.os hino [iron.GEN way] RAILWAY. It might be claimed that distinguishing the two as adjectival and genitival introduces a functional element into an otherwise formal typology, 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, for example, Lithuanian gelež.in.kelis [iron.ADJZ.way] RAILWAY and aus.u vaškas [ear.GEN wax] EARWAX: the data therefore require this distinction to be made.

The preceding discussion suggests that a classification of binominals will consist of the eight types listed in Table 22. Five of Koptjevskaja-Tamm’s six PNP types are retained, but with the following modifications: juxtaposition is subdivided into compounding, derivation and classifier (constructions) and dependent-marking is subdivided into genitival and adjectival. In addition, prepositions is broadened to adpositional and head-marking is renamed construct, following Creissels (2017).

<table>
<thead>
<tr>
<th>PNP types</th>
<th>binomial types</th>
</tr>
</thead>
<tbody>
<tr>
<td>juxtaposition</td>
<td>derivation</td>
</tr>
<tr>
<td></td>
<td>compound</td>
</tr>
<tr>
<td></td>
<td>classifier</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>adpositional</td>
</tr>
<tr>
<td>linking pronouns</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Table 22: 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 (2003), who uses the possessive construction as an example in order to illustrate the classification of cross-linguistic structural types.
4.1.2 Croft (2003)

Croft describes three general morphosyntactic encoding strategies, which he calls simple, relational and indexical. He also discusses the “somewhat problematic” case of classifiers and “more grammaticalized strategies” that employ linkers. 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.

Unlike Koptjevskaja-Tamm, Croft makes a distinction between juxtaposition and concatenation, the latter covering both affixation and compounding. The distinction, however, is not easy to maintain, as Koptjevskaja-Tamm (2004) points out:

The border between juxtaposition and compounding 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 boundedness 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,
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*). For these reasons I chose not to maintain a formal distinction between juxtaposition and compounding in my classification.

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 §3.4.1), 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. I therefore propose to reformulate it more succinctly as follows:

> (35) A relational morpheme denotes a semantic relation, whereas an indexical morpheme denotes a participant in such a relation.

Croft’s distinction between relational and indexical strategies is similar (but not identical) to the distinction between flags and person indexes made by Haspelmath (2019):
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 dichotomy 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 2019 §6). In one sense they are both orthogonal to the head/dependent distinction since they are defined in terms of the function of marking rather than its locus. However, applied to binominals, they give almost exactly the same results: Haspelmath’s system by definition, Croft’s by default. Haspelmath’s flags occur on the possessor, i.e. the modifier, and his (person) indexes on the possessor, i.e. the head. Since Croft’s definition makes no mention of the locus of marking, his relators and indexes can in theory occur on either the head or the dependent, 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” (p.c. 2018-12-14). However, these correspondences between indexes and head-marking, on the one hand, and flags/relators and dependent-marking, on the other, do not take account of what Croft calls nonperson indexes. Rather than adopt either Haspelmath’s or Croft’s analysis, I choose to go with the more widely understood head/modifier distinction, since I have no such qualms in using the terms head and modifier in the context of binominals.1 This also means that I do not have to relate to Croft’s concept of linkers. These 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.” The example cited...

---

1 I also believe that neither Croft nor Haspelmath would have a problem with this, since it is the general use of head and dependent that they take issue with. For example, Haspelmath writes: “If head-marking were confined to the notions of ‘head verb’ in a clause and ‘head noun’ in a nominal, as seems to be the case in more recent work by Nichols and Bickel (see §5), then the problems with a general “head” notion would not arise, and head-marking would become more similar to indexing.”
by Croft, the Persian ezāfe, which arose from a relative pronoun, clearly attaches
to the head, as in zanbur.e asal [bee.EZ honey] HONEY BEE (Tehranisa 1987: 56),
and is therefore covered by Koptjevskaja-Tamm’s head-marking type.

Croft describes classifiers as “a class of overt coded dependencies whose status as
indexical or relational is somewhat problematic,” but concludes that the nominal
source of most of them strongly suggests that they are fundamentally indexical in
nature. However, the classifier constructions encountered in the present study and
classified as binominals, such as Murui Huitoto ui.tïraï [eye.CL(hair)] EYELASH,
consist only of a head element (here, the classifier with the meaning HAIR) and a
modifying element (here, EYE) and thus have more in common with compounds
such as Thai khôn.taa [hair.eye] than with indexical (i.e. head-marking) strategies
such as Hausa gáashì.n ídòo [hair.LK eye]. For this reason, and because they are
clearly separate from compounds and derivations, I choose to treat them as a third
type of simple strategy (i.e., one that involves no additional morpheme).

Croft (2003:40) summarizes the different strategies for encoding grammatical
structure and their diachronic relationship in a diagram that represents two basic
grammatical properties for describing grammatical structure, namely:

(i) additional morpheme: none, relational, indexical, linker, (special form);
(ii) degree of fusion of elements: none, concatenation, fusion.¹

These two properties, suitably modified, suggest an alternative way of representing
the eight types of binominals arrived at in Table 22.

4.2 A non-hierarchical alternative

In order to show how the insights offered by Croft might be applied to the typology
of binominals, the contents of Table 22, relating Koptjevskaja-Tamm’s scheme to
the domain of binominals, are reordered as Table 23, and supplemented with data
regarding the number of additional morphs (over and above the two required thing-
morphs), prototypical constructions that exemplify each type, and three-letter mne-
monic labels based on the prototypes. These labels are used extensively in the re-
mainder 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. Fr. chemin de fer), but

¹ Croft adds a third property, (iii) order of elements, to which we return below (see page 125).
also postpositional constructions (e.g. Jap. kumo no su). Similarly, gen labels a synthetic dependent-marking type that allows for any kind of case affix, not just the genitive.¹

So as not to complicate the typology unnecessarily, two or more consecutive morphs attached to either the modifier or the head count as a single morph. For example Bezhta kil.o.s hino [iron.OBL.GEN way] is simplified to kil.os hino [iron.GEN way] and regarded 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). An additional reason is that there are not enough examples of this phenomenon in my data to justify defining separate types to cater for them.

<table>
<thead>
<tr>
<th>additional morphs</th>
<th>binominal type</th>
<th>example</th>
<th>code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>derivational</td>
<td>pavuč.ina [spider.NMLZ] SPIDER WEB</td>
<td>der</td>
</tr>
<tr>
<td></td>
<td>compounding</td>
<td>eisen.bahn [iron.way] RAILWAY</td>
<td>cmp</td>
</tr>
<tr>
<td></td>
<td>classifier</td>
<td>tuu.heju [nose.CL(hole)] NOSTRIL</td>
<td>cls</td>
</tr>
<tr>
<td>1</td>
<td>adjectival</td>
<td>želez.naja daroga [iron.ADJZ.NMLZ] RAILWAY</td>
<td>ñaj</td>
</tr>
<tr>
<td></td>
<td>genitival</td>
<td>kil.os hino [iron.GEN road] RAILWAY</td>
<td>gen</td>
</tr>
<tr>
<td></td>
<td>prepositional</td>
<td>chemin de fer [way PREP iron] RAILWAY</td>
<td>prp</td>
</tr>
<tr>
<td></td>
<td>construct</td>
<td>lala.m.by [road.PER.iron] RAILWAY</td>
<td>con</td>
</tr>
<tr>
<td>2</td>
<td>double</td>
<td>-emo.li sakila.li [nose.POSS aperture.POSS] NOSTRIL</td>
<td>dbl</td>
</tr>
</tbody>
</table>

*Table 23: Reordered binominal types*

It is possible to arrange these eight 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 23) 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 semantic map (Figure 24) which incorporates three different parameters (number of markers, degree of fusion and locus of marking), and allows for the addition of a third dimension that captures the order of components.

¹ For more precise descriptions of each type, see §4.3.
The vertical axis represents the *number of markers* (levels 0, 1 and 2, respectively); it caters for binominals consisting of

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

The types **cmp** (noun-noun compounds), **der** (derivations consisting of a base and a nominalizing affix) and **cls** (noun-classifier constructions) are sited 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 sited on level 1; and **dbl** (constructions in which both the head and the modifier are marked) are sited on level 2.

The horizontal axis represents what Bybee has called the *degree of fusion*. The latter, illustrated in Figure 25, is a continuum that ranges from “the most highly fused means of expression, lexical expression” to “the most loosely joined means of expression, syntactic or periphrastic expression” (1985: 12). 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 on level 0, but not on level 2, which only contains a single type.

![Figure 24: Formal classification of binominal lexemes](image)

**Figure 24: Formal classification of binominal lexemes**

**Figure 25: Degree of fusion (Bybee 1985)**
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 known form of head-marking. The types *gen* and *con* both involve non-transpositional affixes and are therefore sited symmetrically on level 1. The parallels between these two types is underscored by the fact that the term ‘genitive’ is often applied to possessive markers that attach to the head. Dixon (2010b) advocates using the term ‘pertensive’ when the morphological marking is on the possessum in order to avoid confusion. In this study, as noted earlier, I use the term ‘construct’ 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 undiscovered: 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 §4.4.

The use of ovals with gradient fill rather than the boxes with sharp outlines found in more traditional models (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 also enables the representation of “in-between” or gradient phenomena, as will be seen in §4.5.

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 handle situations in which a language exhibits both left-headed and right-headed binominals of one and the same type, such as *Mod Head* and *Head Mod* (both *cmp*) in Vietnamese (36a); and *Mod.ADJZ Head* and *Head Mod.ADJZ* (both *adj*) in Polish (36b).

(36) a. Vietnamese

*hoá xa* [fire vehicle] TRAIN – **Mod Head** (loan from Sinitic)
*
*xe lìa* [vehicle fire] TRAIN – **Head Mod** (native form)

b. Polish

*pchl.i targ* [flea.ADJZ market] FLEA MARKET – **Mod.ADJZ Head**
*koľej želaz.na* [course iron.ADJZ] RAILWAY (arch.) – **Head Mod.ADJZ**
This parameter is orthogonal to those used in Figure 24 and is coded separately in the database. It can be thought of as the third axis of a three-dimensional model, or as an additional plane containing a mirror image of the one depicted in Figure 24. When taken into consideration (as it is in later chapters) the additional parameter transforms the eight-way classification into a 16-way classification, with labels such as **cmpL**, **cmpR**, **adjL**, **adjR**, etc.

4.3 Binominal types

The classification of binominals developed in the preceding section consists of eight types (**cmp**, **der**, **cls**; **prp**, **gen**, **adj**: **con**: **dbl**). These are arranged in a semantic map that captures the degree of marking, the degree of fusion and the locus of marking. 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 §4.4 but first, definitions are provided for the eight attested types, along with examples and more detailed discussion. Intermediate types are alluded to where applicable in this section and discussed more fully in §4.5.

4.3.1 No additional marker: **cmp**, **der** and **cls**

Level 0 of the model contains three types, **cmp**, **der** and **cls**, each of which has exactly two components: the two thing-morphs that are the primary constituents of a binominal.

**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.

The paradigm case of this type is the noun-noun compound in which two nouns are simply juxtaposed or concatenated; the difference, where it is exists, is not taken into consideration in the present study, for the reasons stated in §4.1.2). The **cmp** type is by far the most frequent type in the data set of this study, accounting for over 50% of all the binominals in the database. It occurs in 94 languages and is a

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. In this case, however, the (sub-) category corresponds fairly exactly to the traditional notion of (noun-noun) compound.
significant word-formation type (accounting for at least 10% of binominals) in 79 of these (cf §6.2.1). Furthermore, it accounts for the majority of binominals in 49 languages, and is the only binominal word-formation strategy in the following 14 languages (cf. Figure 51 on page 219):

(37) Caïjia, Ceq Wong, Datooga, Hawaiian, Hmong Daw, Hupdë, Imbabura Quechua, Kam, Norwegian, Seychelles Creole, Srenge, Thai, Vietnamese, Walman.

Examples of compounds (all with the meaning RAILWAY) are given in (38). Less prototypical examples of the cmp type are compounds that contain a linking element; these are discussed in §4.5.4.

(38) 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]
    Kildin Sami rův't čuekas [iron road] Korean chel.kil [iron.road]
    Mandarin Chinese tie3-lu4 [iron road] Mapudungun tren.rüpū [train.way]
    Thai thaay.rótfay [way.train] Vietnamese tròng sáit [road iron]
    Welsh rheil.ffordd [rail.road] Yakut timir suol [iron path]

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 considered to consist of two thing-affixes and thus belongs to this type (see §5.1.2 on neoclassical compounds). Some typical examples of der are listed in (39).¹

(39) a. Polish wiatr.ak [wind.NMLZ] WINDMILL
    b. Mapudungun kiitral.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.íffo [hen.SGLT_M] COCK/ROOSTER

¹ Note that the glosses given here are essentially those of the original contributors and reflect not only functional diversity but also terminological variety across different traditions of linguistics.
Only affixes that contribute some tangible semantic content are considered in scope. The meaning contribution may be very general (THING, 39a) or it may be more specific (INSTRUMENT 39a, LOCATION 39b, AGENT 39c, etc.). Diminutives are deemed to bear the meaning contribution SMALL THING. They can denote a small version of the entity denoted by the base (39d) or something small that is in some way related to the base entity (39e). Gender-denoting affixes are included only when they mark a clear semantic alternation. In (39f–h) QUEEN alternates with KING, MARE with HORSE and NOSTRIL with 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 §3.4.2 the position was advanced that the affix should be regarded as the head of derivational binominals on semantic grounds (see the discussion on page 108 ff). The affix usually has a very general meaning and thus represents a hypernym of the referent. Thus a windmill (39a) is a kind of thing (or instrument), a fireplace is a location (39b), a fisherman is an agent (39c), 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 (39a): 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 arm, cf. Japanese ude.wa [arm.ring].

\[1\] Note that the gloss provides only a rough indication of the meaning contribution 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 (39a).
The prototype of \textbf{cls} (“classifier”) is comprised solely of a nominal base and a noun classifier, and the denotatum of the binominal is different from that of the base.

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 (40) and (41) (examples from Urban 2012: 126–127), which clearly qualify as binominals.

\begin{enumerate}
\item[(40)] Arabela
  \begin{enumerate}
  \item \textit{quitia.aca} [breast/teat.CL(liquid)] MILK
  \item \textit{namiji.aca} [eye.CL(liquid)] TEAR
  \end{enumerate}
\item[(41)] Bora
  \begin{enumerate}
  \item \textit{ííñu.héju} [earth.CL(hole)] CAVE
  \item \textit{túú.heju} [nose.CL(hole)] NOSTRIL
  \end{enumerate}
\end{enumerate}

The classifier morphemes in these examples have exactly the same function as the corresponding head constituents of, say, Thai, \textit{náam.taa} [water.eye] TEAR and \textit{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 \textbf{cmp} type described above. The classifiers also constitute a closed class, which sets them off from the typical nominal constituents of the \textbf{cmp} type. They are much more like thing-affixes, so these binominals could be classified under the \textbf{der} type, but the 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 \textbf{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 and Bora belong to one of seven subtypes called noun classifiers, which are 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 (42).

\begin{enumerate}
\item Aikhenvald identifies seven subtypes altogether: (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.
\end{enumerate}
The typology of binominal lexemes

(42) Minangkabau
  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 cmp, 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.

However, the matter does not end there either. Examples to parallel those from Minangkabau are also found in Atlantic-Congo languages. In (43) 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.

(43) Swahili (Russell 2003)
  a. m.limau / mi.limau [CL3.lemon / CL4.lemon] LEMON TREE(S)
  b. limau / ma.limau [CL5:lemon / CL6.lemon] LEMON FRUIT(S)

And so should the Bandial examples in (44), where the noun class prefixes serve, among much else, to distinguish between animals and their offspring (cf. the Czech diminutive suffix in 39d, above).

(44) 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 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. Consequently, there would be a very substantial cost to admitting words like (44a, b): almost every word in the Swahili and Bandial subvocabularies 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

---

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.
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 (45).

(45) Gawwada


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 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. (40)–(42) would then be categorized as binominals, while (43)–(45) would not. This would have the unfortunate result that (42) and (43), which really are parallel in every way, would be accorded different treatments. It 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 (44) and (45). 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 (44a) and (44b). The basis for making such a distinction is that in (44a) the denotatum of the whole (KID) is different from that of the base (GOAT), whereas in (44b) they are the same (GOAT and GOAT). In (44a) the noun classifier contributes a meaning component that changes the denotatum. In (44b) this is not the case; nor is it in the two examples from Gawwada (45). This analytical choice results in a much smaller harvest of cls binominals in the data set than the second solution outlined above: only 21 out of the total of 3,507 binominals in the database are of this type, and they are distributed across just four languages:1 Äiwoo, Bandial, Murui Huitoto and Swahili, with 3, 3,

---

1 Data from Amarakaeri and Trinitario have been received but not yet integrated into the database.
The typology of binominal lexemes

11 and 4 examples respectively. This is few enough that every one can be presented and discussed here.

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

The three general bound nouns in the Äiwoo examples (46) are borderline thing-morphs since two of them denote numbers and the third denotes a concept that is somewhat property-like. However, since they are classed as (bound) nouns, the words containing them are admitted here as binominals, albeit non-prototypical.

(47) Bandial
   ji.jamen [CL.goat] KID
   ji.piliŋ [CL.horse] FOAL OR COLT
   a.say [CL.sorcery] SORCERER OR WITCH

Given the definition arrived at above, only three words in the Bandial vocabulary qualify as binominals and they share a single construction, CL.Base (47). 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.uŋat [CL(e).beehive] BEEHIVE, sa.mbul [CL(sa).flame] FLAME and mu.fu [CL(mu).tear] TEAR. In the first two examples in (47) a base that would normally belong in the 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 (48).

(48) der Czech kůz.le [goat.DIM]
    cmp Hawaiian kao keiki [goat child]
    con Hausa yá.r àkwíyàa [daughter.LK goat]
    prp Tagalog bata.ng kambing [child.LK goat]

In the third example of (47), 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

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.
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 (49). For example, in *uchawi* [CL.14: 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.

(49) Swahili  
*uchawi* [CL.14: witch] MAGIC  
*fumo* [CL.9: spear] CHIEFTAIN  
*m.zabibu* [CL.3: grape] VINE

The 11 Murui Huitoto binominals listed in (50) differ from the Swahili binominals in (49) 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’.

(50) Murui Huitoto  
*taizĩ.ko.ño* [heel.CL(cover).CL(fem)] ANKLE  
*ui.tĩraï* [eye.CL(hair)] EYELASH  
*jitaĩ.ño* [adolescent.CL(fem)] GIRL  
*jĩfai.ño* [in_law.CL(fem)] MOTHER-IN-LAW (OF A MAN)  
*enaize.ño* [grandson/nephew.CL(fem)] NIECE  
*defo* [nose:CL(cavity)] NOSTRIL  
*moo.rui* [father.CL(day)] SUNDAY  
*riũno.kai* [woman.CL(stem)] THUMB  
*eĩ.kai* [foot.CL(stem)] TOE  
*ra.o* [thing.CL(flex)] VINE  
*onoyĩ.ki* [hand.CL(cluster)] WRIST

Once again, each of the examples in (50) 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 (51) all denote the same concept as the stem to which they are attached, and are therefore not classified as binominals.
4.3.2 One additional marker: \textbf{prp, gen, adj} and \textbf{con}

Level 1 of the model contains four types, three of which (\textbf{prp, gen} and \textbf{adj}) form a constituent with the modifier and the fourth (\textbf{con}) 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 \textbf{prp} (“prepositional”) consists of a head, a modifier and another independent lexeme that forms a constituent with the modifier.

In the most common case (52a–e), the additional lexeme is a preposition (hence the choice of mnemonic for this type), but it may also be a postposition (52f-g) or a particle named according to a language-specific descriptive category, such as a connector (52h), linker (52i), relator (52j), possessive classifier (52k), article (52l) or determiner (52m).

(52)  
\begin{itemize}
  \item a. French \textit{chemin de fer} [road of iron] RAILWAY
  \item b. French \textit{moulin à vent} [mill to wind] WINDMILL
  \item c. French \textit{arc-en-ciel} [bow-in-sky] RAINBOW
  \item d. Lower Sorbian \textit{lapka na woku} [flap on eye] EYELID
  \item e. Welsh \textit{papur lle chwech} [paper for toilet] TOILET PAPER
  \item f. Japanese \textit{kumo no su} [spider POSTP web] SPIDER WEB
  \item g. Hindi \textit{dāmt kā braś} [tooth POSTP brush] TOOTHPUSH
  \item h. Swahili \textit{tundu la pua} [hole CON nose] NOSTRIL
  \item i. Tagalog \textit{daa.ng-bakal} [road.LK-iron] RAILWAY
  \item j. Kupsabiny \textit{loleet nyēē cēēkāsēnē} [bag REL:SG back] BACKPACK
  \item k. Takia \textit{graian sa.n anaŋ} [evening POSS.3SG food] SUPPER
  \item l. Maltese \textit{xatt il-bahar} [shore DEF:sea] SHORE
  \item m. Welsh \textit{asgwrn y cefn} [bone DET back] SPINE
  \item n. Barain \textit{assi ge ŋ kee} [bone REL:3SG:M PREP head] SKULL
\end{itemize}
(52n) 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 instance.

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 Japanese (52f) and Hindi (52g) examples are cases in point. Both of them are glossed by their respective contributors using the abbreviation GEN, and yet both Japanese no and Hindi ka are commonly regarded as postpositions. In this study the Japanese construction has in fact been classified as gen (hence the asterisk), while the Hindi construction has been classified as prp. The issue is discussed in more detail under the rubric Gradience in §4.5.5 and §4.5.6.

The most commonly used adpositions are those whose function includes marking possession, such as the French de (52a) and the Hindi ka (52g), but some languages permit other prepositions to be used as well, as illustrated by the various locatives (52b-d) and the purposive (52e). 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 (52h), the Tagalog linker ng (52i) and the Kupsabiny relational marker nyēē (52j).

**gen**

The prototype for gen (“genitival”) consists of a head and a modifier (both of them independent lexemes), with an additional, non-transpositional affix attached to the modifier. Typically the affix marks the genitive case (53a–c) or possessive function (53d), but other cases occur as well, including dative (53e), locative (53f), lative (53g) and proprietive (53h).

(53)  a. Irish *muileann gaoithe* [mill wind:GEN] WINDMILL  
     b. Ket *maˀm.d.kūb* [breast.GEN.tip] NIPPLE OR TEAT  
     c. Sidamo *sano.te giddo* [nose.GEN inside] NOSTRIL  
     d. Takia *grye.n tatu* [side.3SG bone] RIB  
     e. Gurinji *yawarta.wu marru* [horse.DAT house] STABLE OR STALL  
     f. Manange *toŋko.ri kʰeta* [basement.LOC cattle] STABLE OR STALL  
     g. Western Mari *šand.eš pumaga* [toilet.LAT paper] TOILET PAPER  
     h. Yakut *sa:χar.da:χ trostnik* [sugar.PROP cane] SUGAR CANE  
     i. Bezhta *kil.o.s hino* [iron.OBL.GEN way] RAILWAY  
     j. Tarifit *tisi u.fus* [bottom CON.hand] PALM OF HAND
The typology of binominal lexemes

The Tarifit example (53j) 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 CON belong to the type *gen* (dependent-marked, relational) whereas Semitic words glossed with CON belong to the type *con* (head-marked, indexical), cf. (58c) on page 138.

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.

(54) b. Greek petr.ini γεφίρα [stone.ADIZ bridge] STONE BRIDGE
c. Italian aurora bore.ale [light north.ADIZ] ARCTIC LIGHTS
d. Lithuanian gelež.in.kelis [iron.ADIZ.way] RAILWAY
e. Lower Sorbian nos.owa žeřka [nose.ADIZ hole] NOSTRIL
f. Polish kolej żelaz.na [course iron.ADIZ] RAILWAY (arch.)
g. Polish zlot.y pierścionek [gold.ADIZ ring] GOLD RING
h. Russian želez.naja doroga [iron.ADIZ road] RAILWAY
i. Slovak veter.ný mlyn [wind.ADIZ mill] WINDMILL
j. Welsh buwch laeth.og [cow milk.ADIZ] DAIRY COW
k. Estonian mesi.las.vaha [honey.NMLZ.wax] BEESWAX
l. Kildin Sami mājjt.jes’ lījjhm [milk.ATTR cow] DAIRY COW
m. Hungarian őszak.i fény [north.ADIZ light] ARCTIC LIGHTS
n. Hungarian kép.es.lap [picture.PROP.card] POSTCARD

The database contains 171 instances of this type, the great majority from European languages, either Indo-European or Uralic (54). The five Slavic languages (Czech, Lower Sorbian, Polish, Russian and Slovak) account for 108 of them alone. 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 (54f) or head-final (54g). In Hungarian there

¹ Other terms used are annexed state (‘état d’annexion’, Kossmann 2000) and dependent form (‘avhengig form’, Endresen 1990).
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 (12 in all) are found scattered across the globe in Africa (55a-b), the Caucasus (55c-g), Asia (55h-i), Papua/New Guinea (55j-k) and Central America (55l). The examples from Archi (55c-d) are borderline binominals since the modifying element in both cases is more adverbial than nominal. They are included in the database in order to exemplify the use of relational adjectives, which can also be derived from nouns (Kibrik 1994).

(55)  

a. Kanuri kámú nyiyá.à [woman marriage.adjz] MARRIED WOMAN  
b. Sidamo k’ar.aame k’ale [sharp_blade.adjz wheel] PADDLE WHEEL  
c. Archi aʳ̥ba.tːut iq [on_wednesday.adjz day] WEDNESDAY  
d. Archi ak:onni.ʔ.ːut kummul [in_morning.adjz food] BREAKFAST  
e. Bezhta k.ɐ.ɬ.ʔ.ːut tormoz [hand.OBL.SUP.ATTR brake] HAND BRAKE  
f. Bezhta xidalaʔ.ːo čūr [snot.ATTR scarf] HANDKERCHIEF OR RAG  
g. Bezhta nucodaq t’ot’ [honey:ADJZ fly] BEE  
h. Ket sol.tu tʊq.əl [gold.ADJZ finger.covering] GOLD RING  
i. Yakut tualet.nay kuməγ [toilet.ADJZ paper] TOILET PAPER  
j. Kalamang kokok jago.ten [chicken cock.REL] COCK/ROOSTER ##gen  
k. Kalamang sontum war.ten [person sorcery.REL] SORCERER OR WITCH  
l. Q’eqchi’ k’im.əl kab’l [straw.ADJZ house] THATCH

**con**

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

The term construct is traditionally used in Semitic linguistics but is extended by Creissels (2017) to cover any obligatory marking on a noun that fulfills the role of head in noun-modifier constructions, provided that 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 (56-58). Labels used in traditions other than Semitic include linker, possessive, genitive and pertensive.¹ The latter term, proposed by Dixon (2010b), is restricted to possessive constructions (unlike Creissels’ construct), but does permit cross-referencing.

The type con corresponds in practice to Croft’s (2003:34) indexical strategies, just as prp and gen correspond to his two types of relational strategy. But while the

---

¹ In addition, Haspelmath (2009) proposes the term “anti-genitive”.
latter is subdivided on the basis of the root/affix distinction, in the case of indexical strategies Croft notes “a natural classification of indexical coded dependencies into two types, depending on whether they encode the category of person” which he calls **person indexation** and **nonperson indexation** respectively.

Koptjevskaja-Tamm (2003:645) makes what appears to be a similar distinction, albeit using different terminology, between two subtypes:

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.

However, Koptjevskaja-Tamm’s term ‘indexer’ is broader than Croft’s ‘person indexation’, in that any property of the dependent can cause the marker to vary, whereas Croft’s term only admits of the category of person doing so. There is evidence of both types of distinction in the database of binominals. For example, the form of the marker used in Barain varies between -ji and -(g)eti depending on the gender of the dependent (56a-b), as does that of the Hausa linker -rl-n (56c-d). These would thus be ‘indexers’ but not cases of ‘person indexation’.

(56)  
- b. Barain *nokuno non.ji* [goat child.POSS:3SG:M] KID  
- c. Hausa *kàfá.r hâncii* [orifice.LK nose] NOSTRIL  
- d. Hausa *dóoki.n kârfèe* [horse.LK metal] BICYCLE

On the other hand, the glossing of (57a-d), all of which make reference to the third person, suggests that these are both indexers and cases of person indexation.

(57)  
- a. Kalamang *kanggir pul.un* [eye skin.3POSS] EYELID  
- b. Q'eqchi' *x.na aj xam* [3ERG.place fire] FIREPLACE  
- c. Takia *su mala.n* [breast eye.3SG] NIPPLE OR TEAT  
- d. Yakut *zarax u.ta* [eye water.3SG] TEAR

Finally, the invariant possessive affixes in Kupsabiny (58a), Malagasy (58b) and Hebrew (58c), would be ‘relators’ and examples of ‘nonperson indexation’.

(58)  
- a. Kupsabiny *kariit.aap maata* [car.POSS fire] TRAIN  
- b. Malagasy *lala.m.by* [road.PER.iron] RAILWAY  
- c. Hebrew *mesila.t barzel* [track.CON iron] RAILWAY  
- d. Galibi Carib *manati poti.li* [breast tip.POSS] NIPPLE OR TEAT  
- e. Galibi Carib *upupo kuwai.yi* [head calabash.POSS] SKULL
So too would the Galibi Carib possessive suffix -li and its allomorph -yi (58d-e), whose distributions are phonological and thus not conditioned by features of the head.

4.3.3 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.

(59)  
  b. Takia patu.n kdabog.an [egg.3SG yellow.3SG] YOLK  
  c. Seri i.to i.naail [3:poss.eye 3:POSS.skin] EYELID  
  d. Oroqen dalay.ɲi ɲa:ko.n [sea.GEN bank.3SG:POSS] SHORE  
  e. Central Yupik imarpi.im ceñ.ii [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 bit habûb.āt [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 (59a-c). In some languages the markers appear to cross-reference each other (59b-c), in others the affix on the head cross-references the modifier (59d-e). Somali exhibits two definiteness markers (59f) and Maltese a combination of definite marker and definite preposition (59g). Finally, Akkadian (59h) 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 63 instances of dbl, spread over 11 languages, which is deemed too few for this to be feasible.

4.4 Unattested types

The eight types of binominal discussed in the preceding section were situated on a two-dimensional grid in Figure 24 (page 124). It will be recalled that the three levels on the vertical axis represent the number of components: level 2 contains three types (cmp, der and cls); level 3, four types (prp, gen, adj and con); and level 4, one type (dbl). There is no significance to the horizontal positioning on levels 2 and 4 other than aesthetic, but the positioning on level 3 does have import.
Firstly, level 3 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 25 on page 124). 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 the latter is a non-transpositional affixing strategy with the locus of marking on the dependent, con is a non-transpositional affixing strategy with the locus of marking on the head. Once the eight types are laid out in this manner, two apparent gaps are revealed, both labelled prn and nml in Figure 24. These are the head-marking correlate of prp (to the left) and the head-marking correlate of adj (to the left). In this section I discuss possible explanations for these lacunae.

4.4.1 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 2003: 34). The missing type prn could thus be whatever is the source of con.

According to Crof, 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 -f (singular feminine), which attaches to the head in possessive constructions (60a, 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

---

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.
According to Creissels (2009) this suffix results from the cliticization of a pronoun *nai̱ta* that is co-referent with the head noun in the synonymous construction illustrated by (60b, d).

(60) 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.-router 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** (pace Creissels) is **Head PRON Mod**. The latter construction would be considered an instance of the missing type **mia** if the pronoun forms a constituent with the head (**Head PRON Mod**), but that is not the case. Instead, the pronoun forms a constituent with the modifier (**Head PRON Mod**), which means that (60b) and (60d), 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 (61a). The converse of this process, whereby an element preposed to the head in a head-final construction is reinterpreted as a postposed marker on the modifier, is the case of Germanic constructions involving linking pronouns, discussed by Koptjevskaja-Tamm (2003:670–676) and exemplified by Norwegian (61b).

(61) a. Hausa *kàree {na Daudà} ➔ {kàre.n}* Daudà ‘Dauda’s dog’
   b. Norwegian *Per {sitt hus} ➔ {Per sitt}* hus [Per his house] ‘Per’s house’

Given the strong preference for suffixing rather than prefixing cross-linguistically (Dryer 2013), these processes can be expected to occur fairly regularly when the marker occupies medial position, but it is not inevitable. Koptjevskaja-Tamm (2003:671) cites Givón’s (1979:90–91) example from the English-based creole Krio, in which “the fully grammaticalized head-marked PNP *Jón hin-ós* ‘John his-house’ has developed from *John, his house*” (p. 671).

From the preceding discussion it is clear that constituency must be taken into consideration when looking for examples of the missing type **prn**. There are six logical possibilities (62). The component X might be a pronoun or an article, but it must

---

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.

2 Givón does not identify the language. It may be the West-African Krio (**kr**).
form a constituent with the head. This means that constructions (62c) and (62d) are highly unlikely: they could only occur in a non-configurational language.

\begin{itemize}
  \item[(62)] a. \{X Head\} Mod  
  \item b. \{Head X\} Mod  
  \item c. Head \{Mod\} X  
  \item d. X \{Mod\} Head  
  \item e. Mod \{X Head\}  
  \item f. Mod \{Head X\}
\end{itemize}

Perhaps the most likely candidate is (62e), represented with a form that might be glossed as, say, [nose 3SG hole]. Were this type of binominal to be found, it would correspond to Koptjevskaja-Tamm’s “linking pronoun”, the one type in her PNP classification (see Figure 23 on page 116) that I did not found among the binominal data. While the linking pronoun type of PNP is rare in European languages, its status across the world’s language is unclear and it seems eminently possible that the binominal type prn could exist somewhere. Finding it, however, must remain a topic for further research.

### 4.4.2 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.ADIZ road] RAILWAY. There are two logical possibilities: either the 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.ADIZ, 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.

### 4.5 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 eight major types. It is based primarily on the data collected but include some examples from other sources. Each subsection refers to one of the numbered items in Figure 26.

---

1 The word *eisenbahnisch* has actually been recorded

2 Marie-E Elaine Van Egmond has identified a possible candidate for the nml type in Anindilyakwa, which I will discuss in a later version, but it appears to involve a change of subclass only.
4.5.1 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 on a level 1 in Figure 26, but since they would not be binominals (by definition), 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 2 (der, cmp, cls) and those on a putative level 1 ①. 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 (63a), which was originally a compound but which is no longer identifiable as such, despite the first constituent still being recognisable as ‘nose’. In (63b), 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.

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

4.5.2 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 ②, 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 (64).

(64)  
*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 /mɔn/ 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).

### 4.5.3 Bound nouns cmp ➔ cls

As was pointed out in §4.3.1, the type 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 “six way for nouns to meet nouns” in Äiwoo (Næss 2017). One of the types found in this language is called a “bound noun” (see 46a on page 132), a term suggestive of something intermediate between a noun and a classifier, which would in turn give rise to binominals mid-way between cmp and cls ③. 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 -*mba* in (65). Again, this suggests the possibility of an intermediate type, but further research is required.

(65)  
Harakmbut/Amarakaeri *tare’.mba*’ [manioc.hand] MANIOC LEAF

### 4.5.4 Linking elements gen ➔ cmp

In many languages noun-noun compounds involve linking elements ④. Almost all the examples in the database are from Indo-European languages (66a-d), the only exceptions being from Korean (66e) and occurring in what Yeon & Brown (2011) describe as “compounds in which the two elements are linked together by the ad-
dition of the so-called ‘genitive s’ (p. 31). The latter, which causes tensing (or rein-
forcement) on the following plain consonant, is best regarded as a linking ele-
ment in the modern language. Further afield, Bauer (2001) cites an example from Cambodian (66f) 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 - aa- “which is being increasingly used at present” in Maori (66g).

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 regierungs.chef [government.LE.head] ‘head of govern-
ment’ the linking element - s-, a reflex of the masculine genitive, is here attached to a feminine noun. The Greek linking element - o- (66b) 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 cmp and 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. Whatever their origin, all such binominals are classified as cmp, unless they are clearly (and consistently) isomorphic with a case affix.

(66)
   a. German nase.n.loch [nose.LE.hole] NOSTRIL
   b. Greek sidir.o.dromos [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.o.thaan [vehicle.LK.place] GARAGE
   g. Maori waiata-aa- Ringa [song.LK.hand] ACTION SONG

4.5.5 Adpositions or case affixes? prp ➔ gen

As noted above (§4.4.1), case affixes arise from adpositions that become affixed to the noun. As a result, the status of some binominals as either prp or gen can be hard to determine  webdriver. A classic example is the Japanese no construction which some linguists analyse as a genitive suffix (67a) and others as a postposition (67b). 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 ㅇ [-ny] 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.

(67) 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 (67c), 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 (68a) the possessive morpheme is written, transliterated and referred to as a postposition, in Nepali (68b) 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 ‘binominal fingerprints’ (cf. §6.2.3). This underscores the importance of the gradient approach to classification adopted in §4.2.

(68) a. Hindi  
   मकड़ी का जाला *makṛī kā jālā* [spider POSTP web] SPIDER WEB  
   b. Nepali  
   माकुराको जालो *mākurāko jālo* [spider GEN web] SPIDER WEB

### 4.5.6 Inflection or derivation? **gen** ~ **adj**

It was noted above (footnote 91 on page 140) 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 found in the morphological literature, whereby derivational affixes change the word-class of their base, while inflectional affixes do not, has been shown by to be wrong. 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 6. Defining them in terms of transposition, on the other hand, results in a more clear-cut distinction.

4.5.7 Head replacement adj \(\rightarrow\) der

The type adj belongs to level 3 (§4.3.2) in the classification whereas der belongs on level 2 (§4.3.1); the former has three components whereas the latter has just two. An intermediate for between these two is represented by the Slovak word železnica (69a). The structure of this word parallels that of the Russian železnaja doroga (69b) precisely, except for the presence of the nominalizing suffix -ica instead of a lexical head.

(69) a. Slovak želez.n.ica [iron.adjz.nmlz] RAILWAY
    b. Russian želez.naja doroga [iron.ADIZ 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 7 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.

4.5.8 Morpheme loss dbl \(\rightarrow\) con

The final example of intermediate (gradient) phenomena is that of morpheme loss. Citing data from Hungarian, Kirmandji,1 Arbore and Maltese, Koptjevskaja-Tamm (2003) shows that “the step between double-marking and head-marking [in PNP] is not necessarily big” 8:

Head-marked PNP in Maltese, similarly to head-marked PNP in Kirmandji, have developed from earlier double-marked PNP, partly due to the breakdown of the case

---

1 The precise identity of this language is not clear. It is described as a “an Iranian language spoken on the southeastern periphery of Europe” and seems most likely to be Northern Kurdish (KMR).
The typology of binominal lexemes

system of modern Arabic dialects compared to Classical Arabic, in which the possessor regularly appeared in the genitive case (p. 647).

The same appears to be the case with binominals, and not just between dbl and con, but also between dbl and gen, between prp and gen, and between gen and con on the one hand and cmp on the other. One 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 cmp. Elsewhere in the database there are indications of a similar process at work in Galibi Carib, Tarifit and Swahili.

(70a-c) are dbl, gen and con, respectively. The double-marked pattern (a) may represent an earlier construction from which the others have developed.

(70) Galibi Carib
   a. emo.lı sakila.li [nose.POSS aperture.poss] NOSTRIL
   b. pana.lı weti [ear.POSS dirtiness] EARWAX
   c. manati poti.li [breast tip.POSS] NIPPLE OR TEAT

(71) is one of three words in which the preposition n is given as optional. With the preposition the construction is considered an instance of prp; without it, it is an instance of gen; but both are dependent-marking.

(71) Tarifit tisi (n) ufus [bottom (of) hand:CON] PALM OF HAND

There is also an example in Swahili (72) of a construction in which the associative marker is given as optional. Since it is the only occurrence, it is classified as prp along with other words that exhibit this marker, but it may also indicate gradience.

(72) Swahili gari (la) moshi [car (CON) smoke] TRAIN

4.6 Chapter summary

In this chapter, I described the classification of binominals and gave examples of each type. I highlighted some issues inherent in traditional, hierarchical representations and proposed a semantic map as an alternative. Two gaps revealed by the non-hierarchical representation were accounted for and a number of gradient phenomena were discussed. The next chapter approaches the classification of binominals from a completely different angle, that of semantic relations.
5 Semantic relations

In the previous chapter I developed a formal classification of binominal lexemes. In this chapter 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” (73), where X, Y and Z represent “major lexical categories” (in the case of binominals, these are, of course, nouns or other thing-morphs), and \( \mathfrak{R} \) “represents an implicit relationship between the constituents (a relationship not spelled out by any lexical item)” (p. 107).

(73) \[ X_N \mathfrak{R} Y_N \] Z_N

Jackendoff (2016) provides a nice set of examples (74) to show that the kind of semantic relation can be “hugely varied”, even across compounds that share a common hypernym, such as cake.

(74) 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 cmp) there is no indication at all 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 \( \mathfrak{R} \) 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, §5.1 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 (§5.1.2), including denominal derivations (§5.1.3). This overview is followed in §5.2 and §5.2.2 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 §5.2.3 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 §5.3. This is followed in §5.4 by a short summary. The actual results of the semantic analysis are presented in the next chapter (see §6.3 and §6.4).

5.1 Theoretical prerequisites

The history of research into semantic relations in compounding tends to be rehearsed by every researcher 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.

---

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 Use of the term ‘semantic’ (or ‘conceptual’) ‘relations’ (or ‘associations’) is not restricted to the domain of compounding, or even to the field of linguistics, but is encountered much more broadly in philosophy, psychology, knowledge representation and elsewhere. While relevant to the long-term goals of this research project, such uses are beyond the scope of the present work. Henceforth the term ‘semantic relations’ should be taken as referring only to the domain of binominals.
5.1.1 Background

Interest in semantic relations can be traced at least as far back 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 (1960), Lees (1960; 1970), Marchand (1960; 1969), Brekle (1970), Li (1971), Adams (1973), Downing (1977), Allen (1978), Bauer (1978), Levi (1978), Warren (1978), Ryder (1994) and Jackendoff (2010). A useful way to gain an overview of the various approaches taken in these and other works is the classification of compounding theories suggested by Søgaard (2005), who distinguishes between transformational, slot-filler, reductionist, and pragmatic theories.

**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 7.

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. (Add example.)

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.

---

1 Other early work worthy of mention includes Bergsten (1911) and Carr (1939).
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:

- **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 (headache, land-breeze, side-glance); Time, including duration (nightmare, day-fly); Purpose (beehive); Means (handwriting); Characterizing Feature (sandpaper); Similarity (needle-fish) and Material (gold ring). He then 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, sundial, weathercock and rainbow.

- **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 §5.2.2.

- **Levi (1978)** (Generative Semantics): Levi defines nine “recoverably deletable predicates” (RDPs), of which the first three are reversible: 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) (see also §2.3.1.).

- **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 (girlfriend); Resemblance (clubfoot); and Purpose (ball bat) (cf. Warren’s summary, p. 229).

- **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

---

1 The labels given here are mine. Jespersen employs descriptive phrases.
inverse, CONIN [X is contained in Y] (teapot); SELL [Y sells/delivers X] (fish-wife); 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 (anthill, honey bee); SERVES AS (guard dog); HAVE (AIDS baby, writer’s cramp);3 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’.4

---

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. Note that the number and/or labels of functions differ slightly from those in Jackendoff (2009; 2010). Also, the labels used here are slightly simplified.
3 Regarding this function, Jackendoff writes (p. 30): “This might be further separated into the many senses of ‘have’ in English, some of which are found in (42). I have not attempted a full semantic analysis, which would doubtless benefit from cross-linguistic comparison.”
4 Bauer never entirely abandons this position: “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).
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 (§4.3.1), 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 §7.2.3, 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.

5.1.2 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. The focus here is on their investigation of the semantic relation (what they call the “meaning link” or “meaning relationship”).
first demonstrate (75) and comment on how those six constructions relate to the formal classification developed in the previous chapter (cf. (7) on page 50).

(75) (a) **cmp** 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* is formally parallel to *eisen.bahn* [iron.way] and thus an instance of **cmp**. (Recall that the classification does not distinguish between juxtaposition and concatenation.)

(b) *manual labour* is clearly an instance of **adj**, parallel to Rus. *želez.naja daroga* [iron.ADIZ 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. Despite the traditional term for such constructions, they do not fit the category **cmp** since *hydro-* and -*mancy* are thing-affixes, not thing-roots.\(^1\) Such forms are better regarded as a non-prototypical subtype of **der** and closer to the two morpheme analysis of Slovak *želez.nica* [iron.NMLZ] (page 147) 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 etymology) is established as a prefix\(^2\), *paratroops*

---

\(^1\) Wiktionary lists 349 English words containing the prefix *hydro-* and 98 words containing the suffix -*mancy* (accessed 2018-05-30). (Provide more arguments for the affix analysis. What is a root?)

\(^2\) Wiktionary lists 17 such words, including *paraglider* and *parajump*. Words formed from Ancient Grk. παρά (‘beside; next to, near, from; against, contrary to’) or from Fr. or It. *para-*, (< Lat. *parā* ‘I prepare’) do not count in this context, even though *parachute* itself belongs to the latter type.
constitutes another gradient, affixoid-like phenomenon between \textit{cmp} and \textit{der} (cf. §4.5.2).

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 152), 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 24 provides a complete listing.

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 §4.3 (\textit{cmp}, \textit{adj}, \textit{gen} and \textit{prp}) are covered by their work, which also provides evidence concerning a subtype of \textit{der} (neoclassical compounds).

Notes on Table 24: 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). \textsc{possessor-possessum} is treated by Levi as a subtype of the \textsc{part-whole} (HAVE) relation. See Tarasova (2013:43, 99, App.1) for a naming scheme that uses the numbers 1 and 2 to indicate the direction of the relation (e.g. \textsc{cause1} and \textsc{cause2} for my \textsc{cause} and \textsc{result}); her terms are \textsc{cause1}, \textsc{cause2}, \textsc{possession1}, \textsc{possession2}, \textsc{composition1}, \textsc{composition2}, \textsc{instrument2}, \textsc{essive2}, \textsc{location2}, \textsc{purpose2}, \textsc{source1} and \textsc{topic2}. 
<table>
<thead>
<tr>
<th>Relation</th>
<th>N N</th>
<th>A REL 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>hydromelanos, 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, ophthalmia</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, dinosum</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>thugon, 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, snoblem</td>
</tr>
</tbody>
</table>

Table 24: 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 §4.3) 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 (76)-(78), 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, (76) would be instances of N₁ MAKE N₂ (cf. snowball in Table 24) while (77) and (78) would be instances of N₂ HAVE N₁ (cf. camera phone).

(76)  IRON + TRACK = RAILWAY  
(a) con  Hebrew mesila.t barzel [track.CON iron] RAILWAY  
(b) cmp  German eisen.bahn [iron.track] RAILWAY  

(77)  EGG + YELLOW = YOLK  
(a) dbl  Takia patu.n kdabog.an [egg.3SG yellow.3SG] YOLK  
(b) cmp  Welsh melyn.wy [yellow.egg] YOLK  

(78)  EYE + HAIR = EYELASH  
(a) cls  Murui Huitoto ui.tїraї [eye.CL(hair)] EYELASH  
(b) cmp  Thai khõn.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 (76)-(78) can be found in (79).

(79)  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...
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.

5.1.3 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.\(^1\) 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 is taken severely to task for this by Brdar & Brdar-Szabó (2014). However, the latter’s objections, stripped of *ad hominem* attacks and false attribution of positions which Janda does not hold, 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 (80), (81).

(80) 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’

(81) CONTAINED FOR CONTAINER
   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).
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 (80a-c) and in (81a-c), there can be no doubt that Janda is correct: in (80) it is the PART-WHOLE relation and in (81) 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 (81b) and (81c), 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 24).

Exact parallels for the Russian and Czech words in (80) 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 (82) are just the tip of a very large iceberg.

(82) 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.xiňu* [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 §7.2.3.

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.
5.2 An integrated approach to classification

5.2.1 A low-level classification (Bourque 2014)

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 rather than simply starting from scratch, as I did in (Pepper 2010b).

2. 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 those of Levi, Warren, Downing and Jackendoff, each of which clock in at 12-14 relations. This is important, because it is much more challenging to classify individual binominals correctly and consistently using a very high-level 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.
3. Each relation is accompanied by a good **description** (the presentation extends over 40 pages). This makes it easier for another researcher to understand just what is intended by each one.

4. 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. For example, it makes no sense to use the names of English verbs and prepositions (cf. Levi’s HAVE, BE and FROM) to describe relations in French. This is also reflected in Bourque’s use of NON-HEAD (or modifier) and HEAD instead of N\(_1\) and N\(_2\) (cf. Levi, Jackendoff and many others), or A and B (Hatcher), which imply head-final order.\(^1\)

5. English and French **examples** are provided for each relation. This makes the dissertation accessible to researchers who are not fluent in French.

6. 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. §5.3.1).

7. 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 Aristotelean 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.\(^2\)

8. Finally, Bourque has made his data available on his website, thus facilitating further exploration of his classification.

In sum, these eight aspects of Bourque’s work provide a compelling argument for adopting his classification scheme.

Table 25 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.

---

1 This highlights the importance of using (language-independent) comparative concepts in preference to (language-specific) descriptive categories in cross-linguistic studies (Haspelmath 2010).

<table>
<thead>
<tr>
<th>Label</th>
<th>Type</th>
<th>Template</th>
<th>Linking material</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hyponymy</strong></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><strong>Coordination</strong></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><strong>Similarity</strong></td>
<td></td>
<td>an H that is similar to M</td>
<td>similar to, like</td>
<td>ant lion</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td></td>
<td>an H that serves as M</td>
<td>functions, serves as</td>
<td>buffer state</td>
</tr>
<tr>
<td><strong>Possession</strong></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><strong>Part</strong></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><strong>Location</strong></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</td>
<td></td>
<td>bedroom</td>
</tr>
<tr>
<td></td>
<td></td>
<td>at/near/in</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Composition</strong></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><strong>Source</strong></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><strong>Cause</strong></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><strong>Production</strong></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><strong>Topic</strong></td>
<td></td>
<td>an H about M</td>
<td>about</td>
<td>history conference</td>
</tr>
<tr>
<td><strong>Time</strong></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><strong>Use</strong></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><strong>Purpose and Proper Function</strong></td>
<td></td>
<td>an H intended for M</td>
<td>for</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

*Table 25: Bourque’s semantic relations*
We observe that Bourque’s classification is not fundamentally different from those sketched briefly on pages 152ff and indeed, the same kinds of relations recur again and again in the literature. In his Appendix B, Bourque thoughtfully provides a 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; 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 so far 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. 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.

5.2.2 A high-level classification (Hatcher 1960)

The starting point for Hatcher’s paper is a trenchant critique of Jespersen’s (1942) attempt to classify semantic relations which was sketched briefly on page 152. 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’effets 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 (83), 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

(83)  

a. Subject (sunrise) or Object (sun-worship)  
b. Location in/from/to (headache, land-breeze, side-glance)  
c. Time at/duration (nightmare, day-fly)  
d. Purpose (beehive)  
e. Means (handwriting)  
f. Characterizing Feature (sandpaper)  
g. 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 and Relation, 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 Matzner 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 27. 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 (a broomstick and β castor oil). She then proceeds to reorganize these six relational types:

¹ See the discussion of appositional compounds on pages 176ff. 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 27.
Figure 27: Hatcher’s reworking of Jespersen’s classification

Jespersen’s referential and mixed types

Jespersen’s and Mätzner’s relational types

Hatcher’s four logical-relational types

Key:
- Jespersen’s referential and mixed types
- Jespersen’s and Mätzner’s relational types
- Hatcher’s four logical-relational types

Subject

Object

Place

Time

Purpose

Means

Characterizing Feature

Material

B is a part of A

B obtained from A

α

β

sunshine

sun worship

side glance

seagull

land breeze

life-annuity

spring frost

Ming bronze

life annuity

2-hour discussion, 3-foot pole

‘souls’ of its own activity” (365)

“The sun is that toward which the worship is directed.” (365)

“A = Place to which...

“A = Place in which B is found

“A = Place from which...

“A = Place in which B is found

“A = Time from which...

“A = Time in which B is found

“A is somehow, to some extent, contained, comprehended in B” (363)

“This represents the reverse of our first new category; now we have a pair of opposites” (364)
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” 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”). These two “mutually exclusive categories are symbolized (very roughly) by ⊶ (A is contained in B) and ⌂ (B is contained in A). I will, however, henceforth symbolize these categories as A ⊆ B and A ⊇ B, respectively. 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 ⊆ B and A ⊇ B, this pair exhibit a “dynamic contrast”, which she indicates using the symbols A→B (“A is the source”) and A←B (“A is the destination”).

(a) A ⊆ B “A is contained in B”
(b) A ⊇ B “B is contained in A”
(c) A→B “A is the source of B”
(d) A←B “A is the destination of B”

Having reduced the six relational categories of Jespersen/Mätzner to two pairs of mutually exclusive concepts (84), 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 27, fits clearly into A ⊇ 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→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←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 (⊆ means subset of, a form of containment); æsthetics (A ⊆ B nicely balances A→B); and convenience (⊆ is available in more fonts).
The typology of binominal lexemes

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: \( A \subseteq B \), 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 \( A \rightarrow B \) and \( A \leftarrow B \), respectively. Surely the subject is the “source” of its own activity (in putting *sunshine* under \( A \rightarrow B \), we are merely adding Agent to Agency); and in *sun-worship* (\( A \leftarrow B \)), 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 \( \circ \) and \( \circ \), and the dynamic \( A \rightarrow B \) and \( A \leftarrow B \) (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 (85). I will not pursue her idea here, fascinating though it is, because it is beyond the scope of my study.
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.¹

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 (MAN) “which is an N₂” (frog), and is thus what Hatcher calls appositional compounds. In Coordination (e.g. moules-frites [mussel-chips] “a popular main dish of mussels and fries originating in Belgium”) the two nouns are “ordonnés au

¹ His objection is, alas, quoted uncritically by Eiesland (2016) and used by her to dismiss Hatcher.
mème rang” (on the same level) 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 (even though he confused me for a while for not distinguishing clearly enough between Noailly’s own four types, described in the preceding paragraph, and her use of Hatcher’s four relations to subdivide them). But my debt to him goes far beyond that, to the extent that he deserves a section of his own.

5.2.3 Extending Hatcher’s classification (Arnaud 2016)

Arnaud’s paper on categorizing the modification relations in French relational sub-ordinate *NNn* 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 (§5.3.3).

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 (86).

\[(86)\]
\[
\begin{align*}
\text{a. NON-HEAD (concrete, discrete) is one of the parts of HEAD} \\
& \text{**tailleur pantalon** [suit trouser] ‘trouser suit’}
\end{align*}
\]
\[
\begin{align*}
\text{b. NON-HEAD is the container of HEAD} \\
& \text{**bière bouteille** [beer bottle] ‘bottled beer’}
\end{align*}
\]
c. NON-HEAD is the origin (general) of HEAD
   \textit{arrêt maladie} [stop sickness] ‘sick leave’
d. NON-HEAD is the goal/the purpose/the object (general) of HEAD
   \textit{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 \texttt{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 \texttt{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). This seems again to provide evidence 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 (87). 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.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{pareto_curve}
\caption{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 156, 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.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{jackendoff_model}
\caption{He also examines how Jackendoff’s model (see page 155) can be applied to French compounds.}
\end{figure}
Alternatively, the reader can continue to think in Hatcher’s terms and focus on the English translation rather than the French word or gloss.

(87) a. NON-HEAD (concrete, discrete) is one of the parts of HEAD
tailleur pantalon [suit trouser] ‘trouser suit’
\[\rightarrow A \subset B \quad (\text{trouser} \subset \text{suit})\]
b. NON-HEAD is the container of HEAD
bière bouteille [beer bottle] ‘bottled beer’
\[\rightarrow A \supset B \quad (\text{bottle} \supset \text{beer})\]
c. NON-HEAD is the origin (general) of HEAD
arrêt maladie [stop sickness] ‘sick leave’
\[\rightarrow A \rightarrow B \quad (\text{sickness} \rightarrow \text{leave})\]
d. NON-HEAD is the goal/the purpose/the object (general) of HEAD
carte réponse [card answer] ‘answer (post)card’
\[\rightarrow A \leftarrow B \quad (\text{answer} \leftarrow \text{card})\]

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 (88).

(88) a. ANALOG (denotes analogy in an aspect of the head)
b. BE (denotes a state of the head)
c. NON-HEAD SYMBOLIZES HEAD
d. NON-HEAD SYMBOLIZED BY HEAD

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 (88a), is exemplified, as we have seen, by régime jockey (89a), a diet that is typical of jockeys. But when A typifies B, does it not characterize B? And did not Hatcher subsume Characterizing Feature under A ⊆ B? If so, that is where régime jockey properly belongs; no new high-level category is required in order to accommodate it. In Arnaud’s 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 (89b-d). Let us examine each of them to see if they really do. (It will turn out that one of them does.)

(89) 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 (89b), 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 (89c) 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

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.
this category is mapped to Hatcher’s scheme is not subjective; it has to be rigorous. Now, is it or is it not a Characterizing Feature of the rhesus factor that it was identified on rhesus monkeys? If so, A ⊆ B is once again the appropriate home.

The fourth low-level relation used to justify ANALOG (89d) 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 she restricts herself to 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,

> 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 (p. 4).¹ On the face of it, this definition would seem to exclude Arnaud’s brasse papillon (89d) 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 sub-divide the latter subtype as follows:

¹ The distinction between species-genus and cross-classification is related to two of Jackendoff’s functions, BE and KIND, about which he comments nicely: “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).
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*¹ (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

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 parallel with *brasse papillon* is plain to see. 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 Aristotelean 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

---

¹ A disconcerting number of Hatcher’s examples are military, reflecting the sad state of the post-war world at the time she was writing.
The typology of binominal lexemes

to 13 exemplars, having reassigned three of his four subtypes (89a-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 (§5.1.3) I made the point that the concept underlying both metonymy and the phenomenon discussed by Janda in derivation is Contiguity. I noted (fn. 107 on page 159) 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. 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 (Fagus and Quercus, respectively) under the family Fagaceae within the Plant kingdom.1

The notion of Similarity thus accounts for both species-genus compounds of the type fuel oil and cross-classification compounds of the type 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 7), 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

1 Compounds of the species-genus kind, sometimes called pleonastic, epexegetic or subsumptive compounds, are somewhat peculiar in that the head is essentially redundant. Is tree in 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.”
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 ⊆ B, A ⊇ B) and Source/goal (A→B, A←B). The new relation is symbolized by the mathematical operator for “almost equal or equal to”, A ≊ B. For the limiting case of Identity, we may use A ≡ B; for the subtype Similarity (proper), A ≃ B (“almost equal to”); and for the species-genus subtype involving Taxonomy I suggest the “Up Tack” A⊥B. This relation is reversible (cf. Bourque’s hypernymy relation in Table 25, with the basic example oak tree and the reversed example bear cub, and also the discussion of the latter in §5.3.2); we will therefore also need a genus-species subtype: ATB.

BE

Having now replaced Arnaud’s ANALOG with similarity (A ≊ B), let us briefly apply ourselves to the other difficiliora that led him to define new categories, viz. BE; NON-HEAD SYMB HEAD; and HEAD SYMB NON-HEAD (90)–(92).

(90)  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’

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 might suggest of oak and beech pointing up towards their co-hypernym, tree (right). Add reference to Carr. Cf. Masha’s slides from Tartu.
Arnaud groups six of his low-level relations into the new high-level category BE (90a-f). Like brasse papillon, kirsch fantasie (90a) 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 \cong B. In opéra rock (90b), 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 (90c), 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 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, snowball, snowflake (soap-flakes), snowdrop, tear drop, chocolate drop, butter pat, sugarloaf 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 contrast 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 philosophical 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 \( A \subseteq B \) or \( A \supseteq B \)), and therefore do not justify the creation of Arnaud’s new category, ANALOG.

As for the other subtypes of BE, in (90d) millisecond is surely the extent (in time) of the pulsar and thus belongs in \( A \subseteq B \) (cf. Figure 27 on page 166); in (90e) the delay characterizes the match, hence also \( A \subseteq B \), and the same applies to return in (90f). Consequently I contend that all six subtypes of BE belong elsewhere and that BE can be dispensed with altogether.

**SYMB**

I turn now to Arnaud’s last two innovations, NON-HEAD SYMB HEAD (92) and HEAD SYMB NON-HEAD (91), each of which has just one subtype, with five and 24 exemplars, respectively. I will deal with the second of these first.

(91) **HEAD SYMB NON-HEAD** (one subtype, 5 instances)

NON-HEAD is represented/symbolized by HEAD 

\[ \text{franc or [franc gold]} \] ‘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 \( A \subseteq B \) (cf. Figure 27). 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).\(^1\) 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.

(92) **NON-HEAD SYMB HEAD** (one subtype, 24 instances)

NON-HEAD represents/symbolizes HEAD 

\[ \text{pierre papier [stone paper]} \] ‘shares in real estate’

---

\(^1\) 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). Note: Revise in the light of Arnaud’s comments (p.c.).
The typology of binominal lexemes

The exemplar given for the last new category, *pierre papier* (92), is something of an oddity. We have first to ask, what is the head?\(^1\) 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):

> 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 *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, *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 *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, *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 ⊂ B. This becomes even more clear if we consider a further comment Hatcher makes regarding this category:

---

\(^1\) Recall the ISA condition (page 107).
_means “A is somehow comprehended in B; B somehow comprehends A, involves \textit{A, 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.\footnote{Incidentally, Hatcher has an nice comment regarding the etymology of ‘about’: “Note that OE 	extit{abutan} < 	extit{on butan} meant ‘on the outside of’, i.e. ‘containing’” (footnote 16 on p. 366).} Given that the only exemplar of the only subtype of the new high-level category \textit{NON-HEAD SYMB HEAD} belongs elsewhere, this category, too, can be dispensed with.

\textbf{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 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, \textit{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 §5.3.3), 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 \textit{ANALOG}, in order not to deviate unnecessarily from Aristotle, and to employ the notation \textit{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 (93) and (94), 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 167). I have recast them in terms of two superordinate relations, \textit{CONTAINMENT} and \textit{CAUSATION} (or source/goal).
CONTIGUITY-BASED

Containment
(a) \( M \subset H \) “M is contained in H” (seed orange)
(b) \( M \supset H \) “H is contained in M” (orange seed)

Causation (source/goal)
(c) \( M \rightarrow H \) “M is the source of H” (cane sugar)
(d) \( M \leftarrow H \) “M is the destination of H” (sugar cane)

SIMILARITY-BASED

Similarity
(e) \( M \cong H \) “H is similar or identical to M”
   – \( M \approx H \) “H is similar to M” (butterfly table)
   – \( M \equiv H \) “H is also an M” (fighter-bomber)
   – \( M \perp H \) “M is a subclass of H” (oak tree)
   – \( M \top H \) “H is a subclass of M” (bear cub)

(94)

CONTRAST-BASED

Contrast
(f) \( M \leftrightarrow H \) “C is the opposite of M”
   (Ger. Unkraut [un.useful_plant] ‘weed’)

Figure 28: Summary of Hatcher 2.0

I have included the four subtypes of Similarity (94) and I have also added a third high-level category (95), based on the third of Aristotle’s associative relations, Contrast. This is 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 the binominal lexemes in my data, but it belongs here in order to complete the picture, because it the combination of two thing-roots that denote the opposite does occur, as in Chinese 东西 dōngxī [east.west] THING (Ceccagno & Scalise 2006).¹ 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 coordinative compounds, in addition to non-appositional.

¹ The phenomenon is more common when two property-roots are combined: 高矮 gāoǎi [high.low] HEIGHT, 大小 dàxiǎo [large.small] SIZE.
5.3 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.

Researchers working on semantic relations have a tendency to reinvent the wheel by developing their own systems of classification.¹ 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 particularly concerned with the actual set of relations, but only whether the same ones are found across different types of construction that involve adnominal modification in English. (They chose Levi’s system simply because it is widely known and of a manageable size.)

I feel this is extremely unfortunate, for how is science supposed to advance if we are always going back to square one? What I would like to see is the community of researchers building 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. Apart from Bourque, only Ryder and Arnaud are excused: Bourque because his system was one of relations retained from previous 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 … but only after developing his own low-level 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...

---

¹ I did the same in my study of nominal compounding in Nizaa (Pepper 2010b), so I am as guilty as anyone.
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), and it is easy to spot inconsistencies and errors in other people’s work; we see them 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), as too anglocentric for my French translation equivalents, and likewise Jackendoff’s seemingly eclectic mix of the schematic (CLASSIFY) and the specific (PROTECT (FROM)). I very nearly adopted Warren’s scheme, which appealed to me for its intuitive role-based naming system (e.g. WHOLE-PART vs. PART-WHOLE), but in the end I opted to “let the data speak for themselves” and to let the relations “emerge” during the analysis. In other words, I started from square one. In my defence, I was somewhat 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 opera to pharmaceuticals to Pokemon to WALS. The analysis mainly involved identifying topic types and association types (§1.1.1) for that particular domain; the latter express semantic relations, albeit often of a more specific nature (e.g. COMPOSED-BY) than the rather more schematic relations (e.g. PRODUCED-BY) defined in most classification schemes. So I thought I was an expert and could do better…

This time I am proceeding 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.

---

1 I will have cause to return to Warren’s approach in §7.2.3.
5.3.1 The Bourquifier

The reasons for adopting Bourque’s scheme were detailed above on pages 161ff, 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 POSSSESSION in Figure 29. Each such summary is followed by up to four pages of detailed description. Note the test frames (templates), in both English and French, examples, and use of constituent-order independent terminology, Head and Modifier, in the templates.

<table>
<thead>
<tr>
<th>PRODUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation Type</td>
</tr>
<tr>
<td>Basic</td>
</tr>
<tr>
<td>Reversed</td>
</tr>
</tbody>
</table>

Figure 29: Bourque’s template for POSSSESSION

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 (96).

(96) (a) career girl
    Basic: a girl that possesses (a) career ✓
    Reversed: a girl that (a) career possesses ×

(b) family estate
    Basic: an estate that possesses (a) family ×
    Reversed: an estate that (a) family possesses ✓

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 3,500 binominals, and this increases the risk of errors creeping in. In order to alleviate this problem, I created a little “app” in Excel, which I call the Bourquifier, shown in Figure 30. To use it, one simply types in the binominal, modifier and head 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 devised for my database (POSS and POSS2). The latter is highlighted because it is the one chosen, after examining all 25 in the course of a few seconds, as being the most appropriate. (Of course, this does not happen automatically. The Bourquifier is not that smart. Yet.)
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, if any: the Bourquifier is not able to figure out the count/mass noun status of the head and modifier, and it doesn’t bother to check for words beginning with a vowel. The templates for hyponymy are marked with an asterisk because I felt a need to revise Bourque’s original (cf. Table 25 on page 163), for the reasons given in the next section. The Bourquifier was extremely easy to create, so everyone can have their own: 1 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 (97).

(97) (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 163 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 31.
Of course, it is not always plain sailing, as the quote from Hatcher on page 178 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 be subtypes of one of Hatcher 2.0. 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 (p. 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 31 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

---

\(^{1}\)Henceforth I will use codes (e.g. PROD, PROD2) rather than Bourque’s labels (e.g. PRODUCTION) for reasons of both space and clarity.

\(^{2}\)“Here, surely, [Jackendoff] is napping,” if I may be permitted to quote 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.
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 §5.3.3.

It proved to be surprisingly easy to categorize my 3,500+ binominals using the revised Bourque scheme, but around 100 defied classification. For the most part this was because the motivation underlying the combination of concepts (and hence the semantic relation) could not be ascertained without further information. In some cases this was simply due to the presence of a cranberry morpheme, as in Bezhta hāš 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 108). 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 deleόr tám [day:OF three] WEDNESDAY and Äiwoo dābu mi.polee [day BN:GNL seven]. It could be argued that such words do not belong in the database; they are, at best, 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 label “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 muumuj'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,
upon reflection it became clear that these were covered by Jespersen’s Character-
izing 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
are 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 26).

<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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>shoe.make.r</td>
<td>Ket ē.ted.s [iron.hit.NMLZ]</td>
</tr>
<tr>
<td>Type 2</td>
<td>farm</td>
<td>Navajo ’atsid.ii [he_pounds.NFE]</td>
</tr>
<tr>
<td>Type 3</td>
<td>pott.er</td>
<td>Mandarin Chinese duan4.gong1 [forge.worker]</td>
</tr>
<tr>
<td>Type 4</td>
<td>black.smith</td>
<td>black.smith</td>
</tr>
<tr>
<td>Exocentric</td>
<td></td>
<td>Hawaiian ku’i hao [strike iron]</td>
</tr>
</tbody>
</table>

Table 26: 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 (POT-
TER); 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” (our *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 (98a), 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 (98b). Jackendoff’s functionally based system is much 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.

(98) (a) violinist₁ = PERSON₂α; [OCC (PLAY (α, VIOLIN₁))] 
(b) street₁ sing₃:er₂ = [PERSON₂α; [OCC/CHAR ([SING₃:β(α); [BE (β, IN STREET₁)])]]

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 27 on page 166). 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 §5.3.3.

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

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-arnes/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).
Bourquifier. However, the system could be further improved with a few tweaks. These are discussed in the next section.

5.3.2 Bourque 2.0

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, templates and examples.

Relations

Most of the names used by Bourque denote conceptual relations, with one exception: ‘Hypernymy’ denotes a lexical relation. For consistency I propose to use ‘Taxonomy’ instead. I also propose to rename the ‘Part’, on the grounds that ‘part’ denotes one of the roles of the relation (the other being ‘whole’). The name Part-Whole would do, but there exists a widely established term for this kind of relation, namely Meronymy, which I prefer.

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 it involves. 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. My experience applying Bourque’s system is that the level of granularity he has chosen achieves the right balance for the practical purpose of classification, and I therefore opt not to tinker with it in any major way (but see below regarding containment). 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 that presents will not be insurmountable.

For the same reason I choose not to combine PURP and USE2, as might seem apt. To take Bourque’s example, it is true that a hand brake (a brake that the hand uses = USE2) is also a brake intended for the hand (= PURP). However, the obverse does not work: an animal doctor (a doctor intended for animals = PURP) can hardly be
described as a doctor that animals use (= USE2). Thus combining the two would require a fundamental rephrasing of the templates.

The Hawaiian binominal pahu meli [box honey] BEEHIVE raises another issue. Is a beehive a box that honey is part of (PART2) or that honey is located at/near/in (LOC2)? 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 contains honey. So how does Containment fit into Bourque’s scheme? It is not one of his 25 relations, but Bourque does not ignore the matter. He discusses it at some length in the context of the overlap between PART and LOCATION, using the example of 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 PART or LOCATION. This dual analysis is related to the fact that LOCATION may subsume 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 toolbox or 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 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):

(105)  a. a C without an M is still a C
       b. un C sans M est toujours un C

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 some other relationship (i.e. container-contained). This result is the same for the French boîte à outils (i.e. 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” (beehive) is parallel to 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 toolbox (and thus also “honey box”) as (mere) location. I suggest that it might be better to bite the bullet and add the relation
“container-contained” (i.e. 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 (99), and it would constitute the prototypical (low-level) subtype of Hatcher’s high-level $A \subseteq B$ and $A \supseteq B$. As examples, I will adopt Hatcher’s *orange seed* and *seed orange*.

(99) CONTAINMENT
   (a) CONT: “an H that is contained in M” *orange seed* $A \supseteq B$
   (b) CONT2: “an H that contains M” *seed orange* $A \subseteq B$

**Templates**

In addition to the templates introduced along with the new relation CONTAINMENT, I propose certain other changes, the first of which speaks for itself:

- **COORDINATION**: C (compound) → B (binominal)

Change to the templates for SRC are prompted by the Caijia word for queen (100).

(100) Caijia
   $k^h\eta^\text{55}ja^\text{21}$ [sky.old_woman] 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 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 “a H that originates in 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 exactly the same thing.

A better solution is to reformulate 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, which can be confusing. My proposal is therefore to use the following templates, which also work for Bourque’s examples:

- **SRC**: a H that comes from a M (e.g. sugar cane)
- **SRC2**: a M that H comes from (e.g. cane sugar)
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’ → ‘composed of’
- PRODUCTION (basic and reversed): ‘makes’ → ‘produces’

The final changes are to the templates for the HYPERNYMY relation, which I newly rechristened as TAXONOMY. The original templates are shown in (101).

(101) 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 (102).

(102) KIND (X, Y)

(a) [Y2α; [KIND (X1, α)]], ‘an N2 of kind N1’
   puppy dog, ferryboat, pine tree, gemstone, limestone, girl child
(b) [Y2α; [KIND (α, X1)]], ‘an N2 that is a kind of N1’
   seal pup, bear cub

Notice that (101a) and (102a) are essentially identical (N2 = H, and N1 = M), but that (101b) is the opposite of (102b). 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 (103).

(103) TAXONOMY

(a) TAX: “a M is a kind of H” oak tree
(b) TAX2: “a H is a kind of M” bear cub
The new formulations produce the results shown in Figure 32 and Figure 33. The populated templates make it clear that *oak tree* belongs under TAX (an oak is a kind of tree”) and not under TAX2 (*a tree is a kind of oak), and that *bear cub* belongs under TAX2 (a cub is a kind of bear) and not under TAX (*a bear is a kind of cub).

**Figure 32: The Bourquifier – OAK TREE**

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), Bourque’s HYP – 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 174. 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),

1 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 super-
classes. Some systems of logic allow this, others don’t. For the purists, the alter-
native (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 33).
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, thus demonstrating the power and
utility of a high-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
**CAUS** 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
**CAUS2**, with the paraphrase “a burn that (the) sun causes”. A better example for
**CAUS** is *tear gas*: “a gas that causes tear(s)”, as the Bourquifier shows (Figure 34).

---

1 At least at this level of (folk) taxonomy.
Chapter 5. Semantic relations

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 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 choose it in preference to butterfly table (see page 175) simply because 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 CAUS2. (Real estate is at a premium in both the Bourquifier and Table 28!)

Bourque’s examples for SOURCE, cane sugar (SRC) and sugar cane (SRC2), have the pleasing property of consisting of the same two elements in reverse order. They are what David-Antoine Williams has dubbed “boathouse words”, which were the topic of a Language Log blog posting by Marc Liberman that was inspired by an xkcd cartoon (Figure 35). Like the cartoonist, Randall Munroe, I really 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.

Figure 34: The Bourquifier – SUNBURN

<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>sunburn</td>
<td>sun</td>
<td>burn</td>
<td>CAUS2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation</th>
<th>Code</th>
<th>Basic template</th>
<th>Example</th>
<th>Code</th>
<th>Reversed template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPERNYMY</td>
<td>HYP</td>
<td>a sun is a kind of burn*</td>
<td>oak tree</td>
<td>HYP2</td>
<td>a burn is a kind of sun*</td>
<td>bear cub</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COM</td>
<td>a sunburn is a burn and a sun</td>
<td>boy king</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>a burn that is similar to sun</td>
<td>ant lion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>a burn that serves as a sun</td>
<td>buffer state</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POSS</td>
<td>a burn that possesses a sun</td>
<td>career girl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PART</td>
<td>PART</td>
<td>a burn that is part of a sun</td>
<td>table leg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>a burn located at/near/in a sun</td>
<td>window seat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>a burn made of sun</td>
<td>sugar cube</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>a burn (made) from sun</td>
<td>cane sugar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>a burn that causes sun</td>
<td>tear gas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>a burn that makes sun</td>
<td>honey bee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>a burn that is about sun</td>
<td>history conference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>a burn that occurs at/during sun</td>
<td>summer job</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>a burn that uses sun</td>
<td>steamboat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>a burn intended for sun</td>
<td>animal doctor</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 https://thelifeofwords.uwaterloo.ca/boathouse-words/; http://languagelog.ldc.upenn.edu/nll/?p=39951
The typology of binominal lexemes

<table>
<thead>
<tr>
<th>A THIS THAT HOLDS</th>
<th>CAR</th>
<th>HOUSE</th>
<th>BOAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUCK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOUSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOBILE HOME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOUSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAILER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAILER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I really like the words for "boathouse" and "houseboat" and think we should apply that scheme more consistently.

Figure 35: 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 (CONT2). For CONTAINMENT I therefore adopt Hatcher’s *orange seed* (CONT) and *seed orange* (CONT2). For COMP(osition) Bourque already has *sugar cube*, which is complemented nicely by *cube sugar* for COMP2. For PRODUCTION I go with *song bird* and *bird song* instead of *honey bee* and *beeswax*. *Oil lamp* and *lamp oil* will replace *steamboat* (USE) and *hand brake* (USE2); and *car motor* and *motor car* can do the job of *table leg* (PART) and *wheel chair* (PART2). 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, TIME and CAUSATION.

Summary

The changes to Bourque’s scheme, summarized in Table 27, are very minor, but since they involve extending the classification from 25 to 27 relations, I will henceforth refer to the revised system as Bourque 2.0.

---

1 Actually a houseboat is not really a boat that holds (or contains) a house: it is a boat that serves as a house (i.e. FUNC).

2 Rolf Theil (p.c.) has suggested Norwegian *skadeflom* [damage flood] ‘flood that causes damage’ (CAUS) and *flomskade* ‘flood damage’ (CAUS2), 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 latter four relations.
5.3.3 The Hatcher-Bourque classification

Once Hatcher’s four-way system covering non-appositional compounds had been extended to a five-way system covering all noun-noun compounds, mapping the revised set of 27 relations in Bourque 2.0 was entirely unproblematic, confirming, once again, that Hatcher’s system is essentially flawless. The results are shown in Table 28. The information contained in this table differs from that in Table 25 on page 163, but the two are entirely commensurate, apart from the addition of the new reversible relation CONTAINMENT. However, a number of new columns have been added:

- **B2** contains the database-friendly code for each of Bourque’s relations. These will be used frequently in the next chapter, especially in §6.3.
- **Hatcher 2.0** shows how each of these relations map to the three- and five-way systems (Similarity-Containment-Causation and Hatcher 2.0).
- **H2** provides database-friendly codes for the latter. These too will be used in the next chapter. Note how the linking elements that connect H and M (is, in and to) represent Similarity, Containment and Causation, respectively.
<table>
<thead>
<tr>
<th>Bourque 2.0</th>
<th>B2</th>
<th>Hatcher 2.0</th>
<th>H2</th>
<th>Template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAXONOMY</td>
<td>TAX</td>
<td>M $\sqsubset$ H</td>
<td>MisH</td>
<td>an M is a kind of H</td>
<td>oak tree</td>
</tr>
<tr>
<td></td>
<td>TAX2</td>
<td>M $\sqsubseteq$ 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 $\equiv$ H</td>
<td>MisH</td>
<td>a B is an H and an M</td>
<td>boy king</td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>M $\approx$ H</td>
<td>MisH</td>
<td>an H that is similar to M</td>
<td>kidney bean</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>M $\subset$ H</td>
<td>MinH</td>
<td>an H about M</td>
<td>history book</td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>CONT</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H that is contained in M</td>
<td>orange seed</td>
</tr>
<tr>
<td></td>
<td>CONT2</td>
<td>M $\supseteq$ H</td>
<td>MinH</td>
<td>an H that contains M</td>
<td>seed orange</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POSS</td>
<td>M $\subset$ H</td>
<td>MinH</td>
<td>an H that possesses M</td>
<td>career girl</td>
</tr>
<tr>
<td></td>
<td>POSS2</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H that M possesses</td>
<td>family estate</td>
</tr>
<tr>
<td>PART</td>
<td>PART</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H that is part of M</td>
<td>car motor</td>
</tr>
<tr>
<td></td>
<td>PART2</td>
<td>M $\supseteq$ H</td>
<td>MinH</td>
<td>an H that M is part of</td>
<td>motor car</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H located at/near/in M</td>
<td>house music</td>
</tr>
<tr>
<td></td>
<td>LOC2</td>
<td>M $\supseteq$ H</td>
<td>MinH</td>
<td>an H that M is located at/near/in</td>
<td>music hall</td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H that occurs at/during M</td>
<td>summer job</td>
</tr>
<tr>
<td></td>
<td>TIME2</td>
<td>M $\supseteq$ 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 $\subset$ H</td>
<td>MinH</td>
<td>an H composed of M</td>
<td>sugar cube</td>
</tr>
<tr>
<td></td>
<td>COMP2</td>
<td>M $\supset$ H</td>
<td>HinM</td>
<td>an H that M is composed of</td>
<td>cube sugar</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>M $\rightarrow$ H</td>
<td>MtoH</td>
<td>an H sourced from M</td>
<td>cane sugar</td>
</tr>
<tr>
<td></td>
<td>SRC2</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H that M is sourced from</td>
<td>sugar cane</td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H that causes M</td>
<td>tear gas</td>
</tr>
<tr>
<td></td>
<td>CAUS2</td>
<td>M $\rightarrow$ H</td>
<td>MtoH</td>
<td>an H that M causes</td>
<td>sunburn</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H that produces M</td>
<td>song bird</td>
</tr>
<tr>
<td></td>
<td>PROD2</td>
<td>M $\rightarrow$ H</td>
<td>MtoH</td>
<td>an H that M produces</td>
<td>birdsong</td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>M $\rightarrow$ H</td>
<td>MtoH</td>
<td>an H that uses M</td>
<td>oil lamp</td>
</tr>
<tr>
<td></td>
<td>USE2</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H that M uses</td>
<td>lamp oil</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H that serves as M</td>
<td>buffer state</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>M $\leftarrow$ H</td>
<td>HtoM</td>
<td>an H intended for M</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

Table 28: Hatcher-Bourque classification of semantic relations
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. The Bourquifier has been upgraded to support the new classification (Figure 36).

<table>
<thead>
<tr>
<th>Relation</th>
<th>Code</th>
<th>Basic template</th>
<th>Example</th>
<th>Code</th>
<th>Reversed template</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAXONOMY</td>
<td>TAX</td>
<td>an M is a kind of H</td>
<td>oak tree</td>
<td>TAX2</td>
<td>an H is a kind of M</td>
<td>bear cub</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COOR</td>
<td>a B is an H AND an M</td>
<td>bay king</td>
<td>COOR</td>
<td>a B is an H AND an M</td>
<td>bay king</td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>an H that is similar to an M</td>
<td>kidney bean</td>
<td>SIM</td>
<td>an H that is similar to an M</td>
<td>kidney bean</td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>an H that is about an M</td>
<td>history book</td>
<td>TOP</td>
<td>an H that is about an M</td>
<td>history book</td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>CONT</td>
<td>an H that is contained in an M</td>
<td>orange seed</td>
<td>CONT</td>
<td>an H that contains an M</td>
<td>seed orange</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POS</td>
<td>an H that possesses an M</td>
<td>career girl</td>
<td>POS2</td>
<td>an H that an M possesses</td>
<td>family estate</td>
</tr>
<tr>
<td>MERONEMY</td>
<td>ART</td>
<td>an H that is part of an M</td>
<td>car motor</td>
<td>ART2</td>
<td>an H that an M is part of</td>
<td>car motor</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>an H located at/near/in an M</td>
<td>house music</td>
<td>LOC2</td>
<td>an H that an M is located at/near/in</td>
<td>music hall</td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>an H that occurs at/during an M</td>
<td>summer job</td>
<td>TIME2</td>
<td>an H that an M occurs</td>
<td>golf season</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>an H composed of M</td>
<td>sugar cube</td>
<td>COMP2</td>
<td>an H that an M is composed of</td>
<td>cube sugar</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>an H from an M</td>
<td>cane sugar</td>
<td>SRC2</td>
<td>an H that an M is from</td>
<td>sugar cane</td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>an H that causes an M</td>
<td>tear gas</td>
<td>CAUS2</td>
<td>an H that an M causes</td>
<td>sunburn</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>an H that produces an M</td>
<td>song bird</td>
<td>PROD2</td>
<td>an H that an M produces</td>
<td>birdsong</td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>an H that uses an M</td>
<td>oil lamp</td>
<td>USE2</td>
<td>an H that an M uses</td>
<td>lamp oil</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>an H that serves as an M</td>
<td>buffer state</td>
<td>FUNC2</td>
<td>an H that an M serves as</td>
<td>buffer state</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>an H intended for an M</td>
<td>animal doctor</td>
<td>PURPOSE2</td>
<td>an H intended for an M</td>
<td>animal doctor</td>
</tr>
</tbody>
</table>

Figure 36: The Bourquifier v 2.0

5.4 Chapter summary

In this chapter I started out by providing a brief overview of previous studies on semantic relations in compounding. Then I showed how Bauer and Tarasova’s work, together with that of Janda, provides grounds for optimism that the same kinds of relations applies to binominal lexemes in general. I described in some detail the systems developed by Bourque and Hatcher that I have harnessed in the present work and I upgraded both to version 2.0 (with the help of Arnaud, in the case of Hatcher). The latter system now covers coordinate compounds and both types of determinative compound, appositional and non-appositional. Finally, I unified Hatcher 2.0 and Bourque 2.0 to create the Hatcher-Bourque classification.

In the next chapter, I present the results of my analysis of over 3,500 binominals, in terms of the descriptive annotation described in Chapter 3, the formal typology developed in Chapter 4 and the semantic classification (this chapter).
6 Patterns in the data

This chapter presents statistical analyses of the data. It starts out with a general description based on the annotation described in Chapter 1 (§6.1). This is followed in-depth analyses of formal and semantic patterning. Those presented in §6.2 build on the eight-way formal typology developed in Chapter 4, while those in §6.3 utilize the semantic classification developed in Chapter 5. In §6.4 more advanced statistical methods are applied in order to explore associations between form and meaning, including a lengthy excursus concerning association rules, which have not previously been used in linguistics research. §6.5 provides a brief summary.

6.1 Data description

This section provides a general description of the data based on the pre-theoretical annotation performed during the first phase of the study, i.e. before the formal typology and semantic classification were developed. That annotation, described in §3.4, 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)

Table 29 gives the overall database statistics and Figure 37 shows the distribution by language and meaning.¹

<table>
<thead>
<tr>
<th></th>
<th>Totals</th>
<th>By language (n=99)</th>
<th>By meaning (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words (S)</td>
<td>8,432</td>
<td>mean</td>
<td>85.2</td>
</tr>
<tr>
<td>Analysable words (A)</td>
<td>5,016</td>
<td>mean</td>
<td>50.7</td>
</tr>
<tr>
<td>Binominals (NN)</td>
<td>3,558</td>
<td>mean</td>
<td>35.9</td>
</tr>
</tbody>
</table>

Table 29: Overall database statistics²

¹ Here and elsewhere the use of colours in a plot has no significance if there is no legend.
² 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
In the following sections, these numbers are broken down, first by language, and thereafter by meaning.

6.1.1 Data description by language

Figure 38 gives a breakdown of the numbers in Table 29 by language. The green column shows the number of words in the sub-database for that language (S); the blue column shows the number of analysable words (A), which is always less than or equal to S; and the orange column shows the number of binominals (NN), which is always less than or equal to A. The languages are arranged in ascending order of vocabulary size (S).

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.
Figure 38: Basic data summary (by language)
A number of observations can be made on the basis of this the 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 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 27 (Wawa), at the left-hand edge of Figure 38, 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 tend to appear towards the left-hand end of Figure 38, whereas national languages of large and developed states tend to appear towards the right-hand end. (Exceptions to this tendency are mostly due to the effects of the first two factors.)

1 Viz. Akkadian and Old High German.
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 suitable figure 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 39: 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 98.
For the database as a whole the degree of morphological complexity is roughly **59%**, but the figure varies considerably across languages, from a high of 97% for Mandarin Chinese to a low of 16% for Selice Romani. Figure 39 lists languages that are at the extreme ends of the scale. Mandarin is, of course, an extreme case, often regarded as a language of compound words (Arcodia 2007). The position of Swahili and Murui Huitoto at the high end of the scale is due to the presence of noun classes and noun classifiers in these languages. Also, the high value obtained for Datooga is not really indicative of morphological motivation, since it is mainly due to purely grammatical suffixes (the primary suffix -óó/-ée(n), the singulative suffix -(C)éan, and/or the singular suffix -tal/-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 binominals as a word-formation strategy. 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 84% (Swahili) to a low of 3% (Datooga), with an average value of 41%. Figure 40 shows the ten languages at each extreme of the scale.

The high figure obtained for Swahili is due to the analytical decision, described in §4.3.1, to regard noun class markers as thing-morphs. As a result, any Swahili noun that consists of a thing-root and a noun class marker, is considered to be a binominal. The only non-binominals in the Swahili data are six loanwords (*baiskeli, bangili, glavu, hospitali, reli*, and *treni*), nine deverbal words (mostly denoting professions), and the two weekdays (which contain numerals). The case of Mandarin and the compound-like structure of its lexicon was noted above, and the same trait accounts in large part for the high ranking of Japanese.
At the other end of the scale, both Datooga and Puyuma are rather sparsely represented in the database (with 29 and 54 words, respectively). Those of 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.

6.1.2 Data description by meaning

The same basic data can be described in terms of meanings, as shown in Figure 41. As in Figure 38, the green column shows the total number of words in the database for each meaning (S); the blue column shows the number of these that are analysable (A); and the orange column shows the number of binominals (NN). Meanings are arranged in ascending order of vocabulary size (S). The same 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 33 (PADDLE WHEEL), at the left-hand edge of Figure 41, to 122 (BEE) at the right-hand edge. Figure 42 shows how the size of the meaning vocabularies varies across different semantic fields (see §3.1.5). 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 96 and 114 words respectively (on average, 105). 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, MIDDAY), where the coverage averages 75 words or fewer.
Figure 41: 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 wūpó). 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 88 languages, despite being represented by 114 words in the database.

Figure 42: Vocabulary sizes across semantic fields

**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%, but as before, the figure varies considerably, from a high of 97% for GOLD RING to a low of 25% for BEE. Figure 43 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. The 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.\(^1\)

Otherwise, it is striking that the other eight meanings to the left in Figure 43 (WIND-MILL, 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.

\[\text{Figure 43: Morphological complexity (A/S) by meaning}\]

At the other extreme, to the right in Figure 43, 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 may originally have been an analysable form (e.g. Kanuri bāskùr < Eng. bicycle, Central Yupik nuussnik TOILET < Rus. núžnik ‘latrine’), while others are more-or-less basic level concepts.

\(^1\) The extent to which items denoting concepts like STONE BRIDGE and GOLD RING are lexicalised is sometimes difficult to determine, especially across a large number of languages. The whole 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 it is clear that STONE BRIDGE and GOLD RING are suboptimal in this respect, but their inclusion does not appear to have skewed the results of the study in any major way.
**Binominal frequency**

The degree to which the 100 meanings are expressed by binominals as opposed to other kinds of lexeme ranges from 89% in the case of STONE BRIDGE to 10% in the case of MAGIC, with an average of **44.8%**. Figure 44 shows the ten meanings at each end of the scale; they are identical to those in Figure 43, except that RAILWAY and EARLOBE appear instead of KEYWORD and TOOTHBRUSH. 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 OT1 (e.g. Assamese go.rokh.ia [cow.watch.AGT]) or OT2, (e.g. Malagasy mpi.àndry [AGT.watch]), rather than a binominal.

![Figure 44: Binominal frequency (NN/S) by meaning](image)

Another useful measure of the frequency of binominals by meaning is shown in Figure 45, 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 (93% 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 to an even higher degree.
The typology of binominal lexemes

Figure 45: Binominal frequency (NN/A) by meaning

6.2 Formal patterning

In the preceding section the binominals data were described and explored in terms of very basic properties, viz. total word counts, analysability, binominal frequency and the head/modifier dichotomy. In this section the focus is on structural patterns based on the eight-way formal typology developed in Chapter 4.

Figure 46: Number of binominals by type

6.2.1 Distribution of binominals by type

Overall frequencies

Figure 46 shows the number of binominals by type. The relative frequency of the cmp type (1863 out of 3559, or 52.3%) is striking and confirms the widely-held belief that compounding is the most common method of word formation in the world’s languages. However, 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 47 provides a breakdown by geographical area in terms of percentages. It shows that the areas that are most widely represented in the sample (Eurasia and Africa, with 37 and 24 languages, respectively) actually have lower than average proportions of the cmp type (43% and 33%); this suggests that Figure 46, if anything, underestimates the predominance of this type across the world.

![Figure 47: Areal distribution of binominal types](image)

But is compounding a universal, as many have suggested? On the basis of the present data that question can only be addressed for noun-noun compounds, but since these are supposedly the most widespread kind of compound (cf. Figure 7 on page 40), it is worth considering here. Ideally one should define both the frequency and the degree of productivity required in order to claim that compounds are attested in a given language. The matter of productivity cannot be addressed on the basis of the present data, but that of frequency can.

For the purpose of this discussion we 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 48 shows how widespread the various types of binominal are across the languages in the database. The green column shows the number of languages in which the type is attested (even if only by a single binominal); the blue column shows the number of languages in which the type is common; and the orange column shows the number of languages in which it constitutes over half of the total.
As far as noun-noun compounds (cmp) are concerned, we observe that they are attested in 94 of 99 languages and thus unattested in five (Akkadian, Central Yupik, Kupsabiny, Selice Romani and Slovak). Of course, one cannot claim on the basis of evidence from just 100 meanings that such compounds never occur in the languages in question, but it is perhaps not unreasonable to suggest that any such occurrences are the exceptions that prove the “rule” that these languages do not have noun-noun compounds (just as left-headed English compounds like attorney-general are the exceptions that prove the “rule” that English compounds are right-headed). The numbers also show that noun-noun compounds are uncommon in a further 15 languages, in which they account for fewer than 10% of binominals. In conclusion, on the basis of the present data, it seems that noun-noun compounding, at least, is not a language universal, except in a very literal sense that is probably not very useful.

**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 58 binominals in Norwegian are cmp. Sometimes it is less so: of the 30 binominals in Greek, 12 are gen and 10 are cmp. Applying Dryer’s (1989) relative frequency criterion for “basicness”, whereby a basic form should be at least twice as frequent as any other, 70 out of the 99 languages in the sample can be said to have a basic binominal type. In a few languages there is no one type that occurs more frequently than the others: 17 of the Assamese binominals...
are **cmp** and 17 are **gen**, and of Selice Romani’s eight, four are **gen** and four are **der** (in the following, languages of the latter type are labelled ‘mixed’).

<table>
<thead>
<tr>
<th>type</th>
<th>count</th>
<th>languages (n=99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmp</td>
<td>52</td>
<td>Āiwoo, Baa, Bambara, Bandial, Basque, Cabécar, Caijia, Ceq Wong, Chakali, Datogo, Dutch, English, Finnish, Galibi Carib, Gawwada, German, Gurinji, Hawaiian, Hindi, Hmong Daw, Hungarian, Hupdë, Imbabura Quechua, Indonesian, Japanese, Kalamang, Kam, Ket, Kildin Sami, Korean, Malagasy, Malayalam, Mamara Senoufo, Manange, Mandarin Chinese, Mapudungun, Mbyá Guarani, Navajo, Norwegian, Old High German, Puyuma, Querétaro Otomí, Saramaccan, Seychelles Creole, Srenge, Thai, Vietnamese, Walman, Welsh, Western Mari, Wik-Mungkan, Yaqui</td>
</tr>
<tr>
<td>gen</td>
<td>14</td>
<td>Amharic, Arche, Bezhta, Estonian, Greek, Irish, Kambaata, Kanuri, Latvian, Nepali, Sidamo, Wawa, Wichi, Zinacantán Tzotzil</td>
</tr>
<tr>
<td>con</td>
<td>9</td>
<td>Anindilyakwa, Hausa, Hebrew, Iraaqw, Kupsabiny, Q‘eqchi’, Ticuna, Turkish, Yakut</td>
</tr>
<tr>
<td>prp</td>
<td>8</td>
<td>Barain, French, Italian, Maltese, Romanian, Tagalog, Tarifit, Wolof</td>
</tr>
<tr>
<td>der</td>
<td>5</td>
<td>Central Yupik, Lithuanian, Oroqen, Polish, 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, Swahili</td>
</tr>
<tr>
<td>mixed</td>
<td>2</td>
<td>Assamese, Selice Romani</td>
</tr>
</tbody>
</table>

Table 30: Preferred binominal types by language

Table 30 shows preferences for each language, and Figure 49 displays the same data in the form of a bar chart, in order to enable comparison with the overall distribution of binominal types shown in Figure 46 (see page 214).

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:

\[(104)\quad \text{Overall ranking:} \quad \text{cmp} >> \text{gen} > \framebox{der} > \text{con} > \text{prp} > \text{adj} > \text{dbl} > \text{cls} \]
\[\text{Preference ranking:} \quad \text{cmp} >> \text{gen} > \text{con} > \text{prp} > \framebox{der} > \text{adj} > \text{dbl} > \text{cls} \]

This indicates that while derivational word-formation is rather widespread (recall from Figure 48 that it is the second most common structural type), it is seldom the principal word-formation strategy, at least in the language sample of this study. It is the preferred type in five out of 99 languages (cf. Table 30), and the basic type in just two: Central Yupik and Lithuanian.
The geographical distribution of preferred binominal types is shown in Figure 50. No clear areal tendencies are discernible on the basis of the present data.

6.2.2 Intralingual competition

One striking difference between the languages in the sample is the number of different binominal constructions within each language. Some, like Hmong Daw and Norwegian exhibit only one type; others exhibit two, three, four, or as many as five different competing constructions. Figure 51 illustrates this variety. Note that the counts are based on eight-way classification used throughout this chapter. As a result, facts about languages such as Vietnamese and Polish, which exhibit mixed
head order within one and the same type (cf. page 253) are not captured. Thus, for example, Polish has 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.

Figure 51: Intralingual competition

When a language only exhibits a single type, it is almost invariably cmp: there is just one exception in the present data: Akkadian, which only exhibits the dbl type. This underlines the flexibility of compounding cross-linguistically. However, as 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 52). An explanation for the relatively strong positive correlation between der and adj awaits further investigation. The strongest negative correlations suggest a tendency to avoid cmp when either prp or dbl are available. However, none of these correlations are particularly strong and they will therefore not be pursued here.
6.2.3 Genetic 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 Norwegian and Vietnamese have a simple fingerprint that can be expressed as \([100_{cmp}]\) (cf. NOR and VIE in Figure 51). At the other end of the scale, Polish has a complex fingerprint consisting of a mixture of five different types in varying proportions, viz. \([4_{cmp} \ 7_{pp} \ 5_{gen} \ 34_{adj} \ 50_{der}]\). The question arises whether these fingerprints show any kind of genetic or areal patterning. In other words, can the binominal fingerprint of a language tell us what family it belongs to, or where in the world it is spoken? Here I focus on genetic 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.\(^1\) When such data exist, genera can be compared genetically (within the same language family) or areally (within the same geographical area).

---

\(^1\) This is an example of the kind of question that cannot be addressed using a “balanced” sample.
Figure 53: The Indo-European spectrogram

Figure 53 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/or 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 both Assamese and Nepali. As for Celtic, it is noticeable that Welsh patterns more closely with English than with Irish, a result, no doubt, of both language contact and the loss of the case system in Welsh.

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 an eight-dimensional space (i.e., one for each type of binominal), which is impossible to visualize. In Figure 54 each language is instead represented as a point in a three-dimensional space, the axes of which are the three principal
The typology of binominal lexemes

components, PC1, PC2 and PC3. The six quadrants containing points provide views onto the data as if through the six faces of a cube.¹

![Figure 54: The Indo-European scatter plot](image)

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 observe that the Germanic languages tend to cluster to the left of this quadrant (the partial exception is GOH, Old High German), while Slavic clusters to 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. The difference within Celtic is very clear in R2C1: the Celtic language that clusters

¹ The following biplots provide keys to the three top left quadrants (R2C1, R1C1 and R1C2) of Figure 54, respectively. They are best read by zooming in the electronic version.
with Germanic is Welsh (Cym), while Irish (Gle) is much closer to Greek, Indo-Aryan and Latvian (Lav). R1C1 has Irish close to both Baltic languages, but R1C2 confirms that it most closely resembles Latvian rather than Lithuanian.

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.

### 6.3 Semantic patterning

All the patterns examined in the previous two sections were based on formal, i.e. structural parameters. This section investigates the frequency and distribution of semantic relations. §0 is concerned with the low-level relations of the Bourque 2.0 scheme (represented in the database by the variable stype) and §6.3.2 with the high-level relations of the Hatcher 2.0 scheme (represented by the variable htype). Table 31 gives a summary of the five high-level relations and the 25 low-level relations that map to them. For a more detailed overview, including templates and examples, refer to Table 28 on page 200.

<table>
<thead>
<tr>
<th>htype</th>
<th>stypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MisH</td>
<td>TAX, TAX2, COOR, SIM</td>
</tr>
<tr>
<td>HinM</td>
<td>POSS2, PART, LOC, TIME, COMP2</td>
</tr>
<tr>
<td>MinH</td>
<td>TOP, POSS, PART2, LOC2, TIME2, COMP</td>
</tr>
<tr>
<td>HtoM</td>
<td>SRC2, CAUS, PROD, USE2, FUNC, PURP</td>
</tr>
<tr>
<td>MtoH</td>
<td>SRC, CAUS2, PROD2, USE</td>
</tr>
</tbody>
</table>

*Table 31: Summary of high- and low-level semantic relations*

#### 6.3.1 Low-level semantic relations

The overall frequency of low-level semantic relations in the database is shown in Figure 55 and is summarized in the following scale:

\[
\text{PART} >> \text{PURP} > \text{COOR} > \text{LOC} > \text{POSS2} > \text{COMP} > \text{USE} > \text{TIME} > \text{PROD} > \ldots
\]

The high frequency of PART is largely due to its extreme frequency in binominals that denote body parts (see below). For this reason Figure 55 also shows the values when body parts are excluded. Apart from the reduced frequency of PART, the differences are minimal.
The typology of binominal lexemes

The distribution in terms of the number of languages in which these relations are attested is shown in Figure 56. The frequency scale here is:

\[
\text{PART} > \text{LOC} > \text{POSS}2, \text{COOR}, \text{PURP} > \text{COMP} > \text{LOC}2, \text{PROD}, \text{USE} > \ldots
\]

The same six relations predominate in both distributions, albeit with slightly different rankings. Note that one relation (COMP2) is not attested at all in the database, and that a further five (SRC, TOP, TIME2, TAX2 and CAUS) are rare.

The distribution across meanings (Figure 57) shows a different scale, with TAX as by far the most frequent relation.

\[
\text{TAX} > \text{PART} > \text{COOR} > \text{PURP} > \text{LOC} > \text{USE}2 > \text{USE} > \text{SIM} > \text{POSS}2 > \ldots
\]

This anomaly can be traced to the analytical choice made in §4.3.1 to regard noun class markers as thing-morphs. In a binominal consisting of a root and a noun class marker, such as (105), the semantic relation is almost always taxonomic (“a child is a kind of person”). Consequently, the presence of a single language with many such binominals in the database makes a huge difference. If Swahili is omitted from the calculation, the figure for TAX is 24 rather than 50.

(105) Swahili mtoto [CL1:child] BOY

Finally, Figure 58 shows the number of binominal types that exhibit each kind of relation. Given the small number of types, the frequency scale here is much flatter and therefore of less interest:
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>PART</th>
<th>an H that is part of M</th>
<th>table leg</th>
<th>HinM</th>
<th>M⇒H</th>
</tr>
</thead>
<tbody>
<tr>
<td>PURP</td>
<td>an H intended for M</td>
<td>animal doctor</td>
<td>HtoM</td>
<td>M←H</td>
</tr>
<tr>
<td>COOR</td>
<td>a B is an H and an M</td>
<td>boy king</td>
<td>MisH</td>
<td>M≈H</td>
</tr>
<tr>
<td>LOC</td>
<td>an H that M is located at/near/in</td>
<td>music hall</td>
<td>HinM</td>
<td>M⇒H</td>
</tr>
<tr>
<td>POSS2</td>
<td>an H that M possesses</td>
<td>family estate</td>
<td>HinM</td>
<td>M⇒H</td>
</tr>
<tr>
<td>COMP</td>
<td>an H composed of M</td>
<td>sugar cube</td>
<td>MinH</td>
<td>M⊂H</td>
</tr>
</tbody>
</table>

*Table 32: Most frequent low-level semantic relations*

Figure 59 shows the kinds of semantic relations that occur in different semantic fields. Observe how PART plays the dominant role in *The Body*, whereas PURP does so *Modern World*. Note also the similarity between *Animals* and *Kinship*. 
Figure 59: How low-level relations vary across semantic fields
6.3.2 High-level semantic relations

We turn now from the low-level semantic relations of the 25-way Bourque scheme to the high-level relations of the five-way (revised) Hatcher scheme. The first four figures in the previous section showed how the low-level relations distribute across the database as a whole (both in total and with body parts excluded) and across languages, meanings and formal types. Figure 60 provides similar information regarding 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 advanced statistical analysis.

Figure 61: How high-level relations vary across semantic fields
The typology of binominal lexemes

Figure 61, the high-level equivalent of Figure 59, shows how Hatcher’s relations vary across the ten most populous semantic fields. Whereas the low-level plots highlight the similarity between *Animals* and *Kinship*, the new ones reveal additional commonalities, in particular between *The body*, *The physical world* and *Food and drink*: in all of these Hatcher’s ⊗, or A⇒B (here labelled simply *HinM*) predominates. In these semantic fields there is a tendency for conceptualizations in which (adopting Hatcher’s formulation) the target concept, B, “is somehow, to some extent, contained, comprehended in” the modifying concept, A.

### 6.3.3 Form and semantics

Figure 62 shows how the six most frequent relations (PART, PURP, COOR, LOC, POSS2 and COMP), those that account for 5% or more of the data, distribute across the eight binominal types. The most striking features to note are the very different fingerprints of *der* (which mostly encodes the COOR relation) and *adj* (where PURP and COMP are the most common relations). These contrast strongly with the other six types, in all of which PART predominates. Another striking feature is that *cls* never encodes the PURP, LOC and COMP relations.

![Figure 62: Low-level semantic relations across binominal types](image)

The divergent behaviour of *adj*, *der* and *cls* is also apparent when viewed from the perspective of the five high-level relations (Figure 63). We observe that *cls* and *der* are largely appositional (*HisM*); this is due to the classificatory nature of the former and the tendency for derivational affixes to denote a broad class of referents such as males, females, diminutives, etc. The fact that *adj* (and to some extent *prp*)
usually occur together with some other type (e.g. cmp or gen) explains why they tend to specialize in less prototypical relations (especially HtoM).

Figure 63: High-level semantic relations across binominal types

6.4 Exploratory statistics

Gries (2013) distinguishes two types of empirical study: “hypothesis-generating” and “hypothesis-testing:”

The former means that you are approaching a (typically large) data set with the intentions of detecting structure(s) and developing hypotheses for future studies; your approach to the data is therefore data-driven, or bottom-up [...] The latter … means your approach to the data involves specific hypotheses you want to test (p. 157–158).

As noted in §1.3.2, this study is of the former type: its research questions are very general and do not constitute hypotheses that are amenable to statistical testing. It is therefore appropriate to apply exploratory methods, such as those of data mining, to the binominal data. However, most such methods are designed to be used with numeric data and are therefore not suitable for our data, which are categorical only. In other words, the kind of information associated with each of our binominals is expressed by properties that have a limited number of discrete values: the property language can take one of a discrete set of 99 possible values; the property ftype (binominal types) can take one of eight possible values (cmp, der, cls, etc.); the property stype (semantic relation) can take one of 27 possible values, etc. None of the basic properties in the database can take a numeric value of the kind that most exploratory methods are designed to predict. One method that does however
work with purely categorical data is called association rules; its application to our binominal data is explored in the next section.¹

6.4.1 Mining for associations

The method of association rules, which belongs to the field of machine learning and data mining, is “a popular and well researched method for discovering interesting relations between variables in large databases” (Hahsler, Grün & Hornik 2007). It is often described in the context of a supermarket scenario, in which the aim is to predict which (other) items a shopping basket might contain based on knowledge of which items it already contains. An example rule might therefore be

(106) \{milk, bread\} \Rightarrow \{butter\}

Each rule has an if-then structure: the left-hand side (LHS) expresses a condition and the right-hand side (RHS) a prediction. (106) can therefore be read as if [the basket contains] milk and bread, then [it also contains] butter. Among the measures that can be associated with a rule are support, which tells how much evidence on the rule is based on, and confidence, which expresses the degree of certainty. Let us assume that (106) was inferred from the examination of five shopping baskets, two containing both milk and bread, and one of these also containing butter: support is \(1/5 = 0.2\) (or 20%), since the rule is supported by evidence from one of the five baskets, and confidence is \(1/2 = 0.5\) (or 50%), since half of the baskets containing the item set in the LHS also contain the item set in the RHS. These two measures are the primary parameters by which the behaviour of association rules algorithms can be controlled, as will be seen in the examples below.

The supermarket scenario most often associated with association rule mining might seem a long way from the realm of binominals, but the same method can be used with any large data set that consists of categorical data. In the R package ‘arules’ (Hahsler et al. 2018) the function `apriori()` takes each categorical variable (e.g. `ftype`) with \(n\) levels (in this case, \(n = 8\): `cmp`, `der`, `cls`, etc.) and turns it into \(n\) binary variables of the type `ftype=cmp`, `ftype=der`, etc., each of which is equivalent to the presence (TRUE) or absence (FALSE) of, say, bread or milk. Each binominal is thus the equivalent of a “transaction”, a shopping basket that contains a certain set of items. For example, the shopping basket Fr. *chemin de fer* contains the items `language=French`, `meaning=railway`, `ftype=prp`, `stype=COMP`, etc. Once the data

¹ Thanks to Stefan Gries (p.c.) for pointing me towards this method.
has been transformed in this way, various algorithms (such as the apriori algorithm) can be applied to discover “interesting relations”.

In order to test this method with the binominal data a modified form of the database was constructed in which every variable that could possibly be of interest was thrown in and all non-categorical variables were removed. The resulting data frame had the following structure:

```r
> str(mm)
'data.frame': 3556 obs. of  16 variables:
$ meaning     : Factor w/ 100 levels "ankle","arctic lights",...
$ stype       : Factor w/ 27 levels "TAX","TAX2","COOR",...
$ language    : Factor w/ 99 levels "Aiwoo","Akkadian",...
$ headPos     : Factor w/ 2 levels "L","R"
$ ftype       : Factor w/ 8 levels "cmp","prp","gen",...
$ language2   : Factor w/ 99 levels "Aiwoo","Akkadian",...
$ area        : Factor w/ 7 levels "A","E","O","G",...
$ iso639      : Factor w/ 99 levels "akk","amh","aoi",...
$ glottocode  : Factor w/ 99 levels "akka1240","amha1245",...
$ family      : Factor w/ 37 levels "Afro-Asiatic",...
$ genus       : Factor w/ 66 levels "Adamawa-Ubangi",...
$ htype       : Factor w/ 5 levels "MisH","HinM",...
$ semField    : Factor w/ 16 levels "Agriculture and vegetation",...
$ semType     : Factor w/ 7 levels "person","body part",...
$ area2       : Factor w/ 7 levels "Africa","Eurasia",...
$ ftype2      : Factor w/ 16 levels "adjL","adjR",...
```

Analysing this data set using the default settings (supp=0.1 and conf=0.8) results in a set of 190 rules. However, most of these turn out to be rather uninteresting:

```r
> rules <- apriori(mm) # 642 rules
> inspect(rules)
```

Rules [1] and [3] are trivial because derR is derived from der + R,\(^1\) hence the value of 1.0 (100%) for confidence. Rules [2] and [4] are not exactly trivial, but they reflect the same known fact: that 95% of der binominals are right-headed (cf. Figure 67 on page 250). The first four rules are based on evidence from the same 381 “transactions” (i.e. binominals), which constitute 10.7% of the data (support = 0.1071429).


---

\(^1\) ftype2 represents the 16-way typology in which the 8 basic types combine with head position.
are just different labels for the same entity (O is an abbreviation for Oceania/SE Asia). They are therefore perfectly correlated, which explains why the confidence is (again) 100%. [8] and [9] are non-trivial: they tell of a strong association between cmp and Oceania/SE Asia, but this fact was already known from Figure 47 (see page 215). Finally, [10] and [11] reveal a strong association between the semantic field Modern world and the semantic type ‘advanced’, which comes as no surprise.

From these observations we can conclude that any rules whose confidence is 1.0 are likely to be trivial in one way or another. We therefore extract a subset of rules where the confidence is less than 1.0 (note that the rules get renumbered, so the first five are those that we have already examined as [2], [4], [5], [6] and [11] in the preceding paragraphs):

```r
> rules <- subset(rules, confidence < 1.0) # 51 rules
> inspect(rules)

lhs                     rhs                     support   confidence lift     count
[1]  {ftype=der}             => {ftype2=derR}           0.1071429 0.9501247  8.867830  381
[2]  {ftype=der}             => {headPos=R}               0.1071429 0.9501247  1.486425  381
[3]  {area=O}                => {ftype=cmp}             0.1285152 0.8294011  1.581421  457
[4]  {area2=Oceania/SE Asia} => {ftype=cmp}             0.1285152 0.8294011  1.581421  457
[5]  {semType=advanced}      => {semField=Modern world} 0.1538245 0.9529617  5.647886  547
[6]  {semField=Modern world} => {semType=advanced}      0.1538245 0.9529617  1.581421  457
[7]  {semType=body part}     => {semField=The body}     0.2609674 0.9666667  3.704167  928
[8]  {semField=The body}     => {htype=HinM}            0.2137233 0.8189655  1.887389  760
[9]  {semType=body part}     => {htype=HinM}            0.2215973 0.8208333  1.891694  788
[10] {area=E}                => {headPos=R}             0.4038245 0.8568019  1.340426 1436
[11] {area2=Eurasia}         => {headPos=R}             0.4038245 0.8568019  1.340426 1436
...  
```

There are no longer any trivial rules, which is an improvement, but there is still a lot of redundancy. This can be observed from the fact that [1] and [2] – formerly [2] and [4] – are still providing duplicated information, namely, that 95% of der binominals are right-headed. We conclude that the variable ftype2 is redundant, given that we already have the variables ftype and headPos. Similarly, [3] and [4] provide identical information – that 83% of binominals in Oceania and SE Asia are of type cmp; given area, area2 is redundant (and vice versa). Redundancy of this kind introduces a lot of noise, rendering the results overly verbose and difficult to interpret. It can be reduced by removing variables that are interchangeable with (or can be predicted from) others, as follows:

- **language**: interchangeable with language2, iso639, glottocode; predicts family, genus, area, area2
- **meaning**: predicts semField, semType and semType2
- **ftype**: predicts ftype2 (given headPos)
- **headPos**: predicts ftype2 (given ftype)
- **stype**: predicts htype
The consequence of this analysis is that only those variables shown in bold can be used together without causing redundancy, unless they are replaced with one of the variables with which they are interchangeable or that can be predicted from them. In accordance with this, we update the association rules analysis as follows:

```r
> mm <- nn[,c("language","meaning","ftype","headPos","stype")]
> rules <- apriori(mm) # 1 rule
> inspect(rules)
lhs          rhs  support confidence lift     count
[1] {ftype=der} => {headPos=R} 0.1071429 0.9501247 1.486425 381
```

Unfortunately we have thrown out the baby with the bath water. All the redundant rules have gone, but so too have all the more interesting rules. We are left with the single, rather mundane rule that most derivations are right-headed, a fact already revealed in Figure 67 on page 250. A moment’s reflection tells us that none of the interesting rules that we have lost involved the variables language or meaning, but rather those variables that were excluded on the grounds that they could be predicted from the latter (e.g. area and semType). Since we are engaged in a typological investigation, we can justify ignoring individual languages in favour of areal or genetic groupings. Similarly, we are arguably more interested in groupings of meanings based on semantic field or semantic type than individual meanings. We therefore rerun the analysis using a different set of variables:

```r
> mm <- nn[,c("area","semType","ftype","headPos","htype")]
> rules <- apriori(mm) # 9 rules
> inspect(rules)
[1] {ftype=der} => {headPos=R} 0.1071429 0.9501247 1.486425 381
[2] {area=O} => {ftype=cmp} 0.1285152 0.8294011 1.581421 457
[3] {semType=body part} => {htype=HinM} 0.2215973 0.8208333 1.891694 788
[4] {area=E} => {headPos=R} 0.4038245 0.8568019 1.340426 1436
[5] {area=E,htype=HtoM} => {headPos=R} 0.1077053 0.8218884 1.285805 383
[6] {semType2=body part,ftype=cmp} => {htype=HinM} 0.1276715 0.8269581 1.905809 454
[7] {area=E,htype=HtoM} => {headPos=R} 0.1406074 0.8233306 1.885926 500
[8] {area=E,htype=HinM} => {headPos=R} 0.1470754 0.8517915 1.332587 523
[9] {area=E,ftype=cmp} => {headPos=R} 0.2123172 0.9173755 1.435190 755
```

We now have more rules, all of which are non-trivial, but a new form of redundancy has appeared in the form of supersets. As an example, rule [5] refines rule [4] (its item set is a superset of those in the latter), but it doesn’t buy us any additional confidence (0.8218884 < 0.8568019); it can therefore be discarded. The same thing applies to rule [7] with respect to rule [3]. On the other hand, rule [6] does provide a modicum of extra value with respect to rule [3], since adding ftype=cmp to the LHS results in a (slight) increase in confidence (0.8269581 > 0.8208333). More specific rules that do not increase the confidence of the more general rule can be removed using the function is.redundant():

```r
> rules <- rules[!is.redundant(rules)] # set of 6 rules
> inspect(rules)
[1] {ftype=der} => {headPos=R} 0.1071429 0.9501247 1.486425 381
[2] {area=O} => {ftype=cmp} 0.1285152 0.8294011 1.581421 457
[3] {semType=body part} => {htype=HinM} 0.2215973 0.8208333 1.891694 788
```
The typology of binominal lexemes

We are now left with six rules which are of four major types, viz. area as a predictor of formal type and head position, formal type as a predictor of head position, and semantic type as a predictor of semantic relation (Table 33). [1] and [2] do not tell us anything new. The other rules provide insights of varying degrees of interest that were not revealed by the analyses earlier in this chapter.

<table>
<thead>
<tr>
<th>LHS</th>
<th>RHS</th>
<th>Description</th>
<th>Example</th>
<th>Rule(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>area</td>
<td>ftype</td>
<td>area ➔ formal type</td>
<td>O ➔ cmp</td>
<td>[2]</td>
</tr>
<tr>
<td>area</td>
<td>headPos</td>
<td>area ➔ head position</td>
<td>E ➔ R</td>
<td>[4], [6]</td>
</tr>
<tr>
<td>ftype</td>
<td>headPos</td>
<td>formal type ➔ head position</td>
<td>der ➔ R</td>
<td>[1], [6]</td>
</tr>
<tr>
<td>semType2</td>
<td>htype</td>
<td>semantic type ➔ semantic relation</td>
<td>body part ➔ HinM</td>
<td>[3], [5]</td>
</tr>
</tbody>
</table>

Table 33: Some basic association rules

This is as far as we get using the default settings of apriori(). In order to discover more rules we can reduce the threshold for support. Setting supp=0.01 essentially means we will accept any rule that is based on 10% or more of the data set.

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:

```
> rules <- apriori(mm, parameter=list(supp=0.01)) # 199 rules
> rules <- rules[!is.redundant(rules)] # 117 rules

Rather than inspect the resulting set of 117 non-redundant rules all at once, it is possible to restrict the rules to certain patterns, such as those with a LHS listed in Table 33, for example area:

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:

Once again, rules [1] and [3] express information already known from Figure 47. [2] and [4], on the other hand, suggest that constituent order ought to be investigated more closely (we shall do so in the next chapter). Next, restricting the LHS to ftype reveals strong tendencies for cls and prp to be left-headed and der to be right-headed, for dbl to express HinM relations, and for adj to occur in Eurasia:
The only strong tendency involving semantic type (semType) as predictor is for the denotation of body parts to involve the HinM relation, a fact that does not surprise us once we recall that HinM subsumes the PART relation:

```r
define levels
> semType2s <- paste0("semType2=",levels(mm$semType2))
> inspect(rules)
```

With the high-level semantic type (htype) as a predictor, no rules are generated unless confidence is reduced from 0.8 (the default) to 0.6:

```r
> htypes <- paste0("htype=",levels(mm$htype))
> rules <- apriori(mm,parameter=list(supp=0.01),appearance=list(lhs=htypes)) # 0 rules
> rules <- apriori(mm,parameter=list(supp=0.01,conf=0.6),appearance=list(lhs=htypes)) # 6 rules
```

Once again, most of these rules are less than compelling. Regarding rule [1], we already know that the type cmp accounts for over 50% of all binominals (cf. Figure 46 on page 214) and the prevalence of this type also accounts for rules [3], [5] and [8]. Rules [2], [4] and [6] reflect the general overrepresentation of languages from Eurasia in the data set. This leaves the moderately interesting insight from rule [7] that the majority of binominals exhibiting the relation HinM denote body parts; this, of course, the inverse of rule [3] on page 233.

So far, then, the method of association rules has not produced any startling insights. We could continue to decrease the support and confidence parameters and generate...
even more rules which could then be subjected to further analysis through filtering. However, it seems unlikely that this would produce any really interesting results. Why is that? Well, it could simply be that there are no interesting associations awaiting discovery in our data. But it is also possible that the method is not ideally suited to these data.

Recall that the \texttt{apriori()} function takes each categorical variable (e.g. \texttt{ftype}) with \textit{n} levels and turns it into \textit{n} binary variables of the type \texttt{area=A}, \texttt{ftype=cmp}, etc. Now consider that the data set we have been using consists of a small number of variables – most of which have several levels: \texttt{area} (7), \texttt{ftype} (8), \texttt{headPos} (2), \texttt{semType} (7), \texttt{htype} (5) – yielding a total of (7+7+8+2+5=29) possible “items”, each of which is equivalent to the presence (\texttt{TRUE}) or absence (\texttt{FALSE}) of (say) bread or milk. The resulting situation differs from the shopping basket scenario for which the method of association rules was designed in one crucial respect. It is never the case that the presence of one item in a shopping basket (say, bread) precludes the possibility of another (e.g. milk) \textit{a priori}. With our data, however, the presence of one item (say, \texttt{ftype=cmp}) precludes the possibility of at least one (and in the case of \texttt{ftype=}, as many as seven) other items \textit{a priori} (a binominal of type \texttt{cmp} cannot also be of other structural types). In other words, the configuration of our data means that the number of possible combinations of the 29 different “items” is severely limited, and it seems possible that this might affect the efficacy of the method. Whether that is the case is a question for mathematicians to answer!

Of particular interest is why the formal typology developed for this study does not result in any interesting observations. It is not because we have been using the 8-way classification instead of the 16-way classification (\texttt{ftype2}), as the example below demonstrates. That left-headed \texttt{cls} binominals associate with the \texttt{MisH} relation and are mostly found in Africa is to be expected, given the way Swahili dominates this category; the strong tendency for right-headed \texttt{adj} binominals to occur in Eurasia is a reflection of the general tendency for \texttt{adj} to be found in this area.

\begin{verbatim}
> mm <- nn[,c("area","semType2","ftype2","htype")]
> ftype2s <- paste0("ftype2=",levels(mm$ftype2))
> rules <- apriori(mm,parameter=list(supp=0.01),appearance=list(lhs=ftype2s)) # 3 rules
> inspect(rules)

lhs             rhs          support    confidence lift     count
[1] {ftype2=clsL} => {htype=MisH} 0.01152981 0.8913043  6.729253  41
[2] {ftype2=clsL} => {area=A}     0.01209224 0.9347826  4.547402  43
[3] {ftype2=adjR} => {area=E}     0.03740157 0.9851852  2.090286 133
\end{verbatim}

Thus it makes little difference whether \texttt{ftype} or \texttt{ftype2} is used when mining the data for associations. To understand why, we need to return to the matter of cross-linguistic comparison, now from the perspective of the formal typology.
6.4.2 Formal types and functions

The important distinction between descriptive categories and comparative concepts was introduced §1.2.2. Indeed, this was the very reason for defining the category of binominal lexeme in the first place: definitions of categories such as ‘noun’, ‘adjective’ and ‘compound’ are always language-specific and therefore unsuitable as the basis for cross-linguistic comparison. The category of binominal lexeme, on the other hand, has proven to be relatively easy to apply across languages, and the same applies to the formal typology. But while it is possible to distinguish the same formal types (*cmp*, *prp*, *adj*, etc.) across languages, it should not be assumed that they carry the same functional load. A comparison of Nizaa and Mapudungun is instructive in this respect.

Recall from §1.1.2 that Nizaa exhibits two types of nominal compound, which we can now denote as *cmpL* and *cmpR*, and that according to Pepper (2010b) they involve two disjunct sets of semantic relations. Left-headed compounds involve relations such as location, purpose, activity and appearance that can be broadly 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:

\[
\begin{align*}
\text{cmpL} & : \text{attributive } \Rightarrow \text{MinH, MisH, HtoM} \\
\text{cmpR} & : \text{subordinative } \Rightarrow \text{HinM, MtoH}
\end{align*}
\]

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 it is clear that the functions of *cmpR* (head-final) and *cmpL* (head-initial) are the exact opposite of those in Nizaa: the former encode attributive relations, whereas the latter encode subordinative relations. What this demonstrates is that any association between formal type and semantic relation in any given language is essentially random, at least from a synchronic perspective. As a result, generalizing across languages will result in the neutralization of such associations and no significant results will be obtained.
There are two ways to address this problem. One is to look for associations between formal type and semantic relation within individual languages rather than across languages. The other is to leverage the known facts of formal structure differently, in a way that is valid across different languages. These two approaches are explored in the next two sections.

6.4.3 Motivational grids

The following investigation into the relationship between form and meaning was initially 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 34, operationalises the insight behind Koch’s refinement of Ullmann’s motivational typology that was briefly mentioned earlier (cf. page 207):

Ullmann’s distinction between ‘morphological’ and ‘semantic’ motivation turns out to correspond not to a clear-cut opposition between disjunct motivational devices, but to 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).

The vertical dimension of the grid consists of a list of formal devices that includes affixation, composition and lexical syntagm (all of which are used in binominal word-formation), as well as formal identity, reduplication, alternation, idiom and more. The horizontal dimension constitutes a closed set of what Koch claims to be universal cognitive relations (viz. conceptual identity, contiguity, metaphorical similarity, co-taxonomic similarity, taxonomic superordination, taxonomic subordination and co-taxonomic contrast), all of which are said to ultimately derive from the Aristotelian relations of contiguity, similarity and contrast. Koch’s goal is 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 therefore propose to use the set of eight binominal types in place of his list of formal devices, and the set of five high-level semantic relations in place of his closed set of universal cognitive relations.
<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>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 34: Koch's motivational grid

To illustrate this approach, consider the contingency table generated from the Polish data (Table 35). This is essentially a motivational grid in which each row represents a formal relation and each column a semantic (or cognitive) relation.

```
67 Polish (54x NN, 7x5)
ftype2
htype MisH HinM MinH HtoM MtoH
adjL 0 3 3 8 2
adjR 0 0 3 0 0
cmpl 0 1 0 0 0
cmpr 0 1 0 0 0
derR 16 2 2 3 3
genL 0 2 0 0 1
prpl 0 1 0 3 0

chi-squared p = 4.1193398706098e-05
X-squared  = 61.3420673076923   df = 24
Fisher's exact p = 0.000499750124937531
```

Table 35: Motivational grid for Polish
Polish is represented in the database by 54 binominals (NN), distributed across seven formal types (adjL, adjR, etc.), that express all five semantic relations (MisH, HinM, etc.). This results in a 7×5 “motivational grid” (Table 35). 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 (H₀) 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 of 4.12e-05 with 24 degrees of freedom and an $\chi^2$ value of 61.34, 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 64. The larger the residual (in either positive or negative direction), the stronger the effect of the association.
Figure 64: Residuals for cells in the Polish contingency table

The largest residual is that between adjR and MinH, with a value of 3.83, 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 adjL and HtoM (1.89). The contrast appears to confirm the traditional Polish distinction between (right-headed) classificatory adjectives (108a) and (left-headed) qualifying relational adjectives (108b).

(108) a. złoty pierścionek [gold.ADJZ ring] GOLD RING (COMP = MinH) adjR
    b. pompa wodna [pump water.ADJZ] WATER PUMP (PURP = HtoM) adjL

The next largest residual is that between derR and MisH, with a value of 2.99, which appears to indicate 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 be used to express relations in which the modifier denotes some kind of container (or possessor, whole, place, etc.), in other words: POSS2, PART, 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. For that reason, Table 36 shows the results of applying the same tests to every language in the database. It lists all 29 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 five 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 choice of construction in 24 of the 99 languages. Determining whether this is the case is, however, a topic for further research.
The typology of binominal lexemes

<table>
<thead>
<tr>
<th>binominals</th>
<th>ftypes</th>
<th>htypes</th>
<th>n/IJ</th>
<th>p (χ²)</th>
<th>df</th>
<th>χ²</th>
<th>p (Fisher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anindilyakwa</td>
<td>15</td>
<td>4</td>
<td>5</td>
<td>0.75</td>
<td>12</td>
<td>30.00</td>
<td>0.001</td>
</tr>
<tr>
<td>Barain</td>
<td>21</td>
<td>3</td>
<td>4</td>
<td>1.75</td>
<td>6</td>
<td>11.92</td>
<td>0.007</td>
</tr>
<tr>
<td>*Caijia</td>
<td>29</td>
<td>2</td>
<td>5</td>
<td>2.90</td>
<td>4</td>
<td>18.62</td>
<td>0.016</td>
</tr>
<tr>
<td>Chakali</td>
<td>24</td>
<td>3</td>
<td>5</td>
<td>1.60</td>
<td>8</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>16</td>
<td>45.86</td>
<td>0.002</td>
</tr>
<tr>
<td>*Finnish</td>
<td>55</td>
<td>4</td>
<td>5</td>
<td>2.75</td>
<td>12</td>
<td>26.82</td>
<td>0.054</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>20</td>
<td>5</td>
<td>4</td>
<td>1.00</td>
<td>12</td>
<td>22.86</td>
<td>0.002</td>
</tr>
<tr>
<td>German</td>
<td>72</td>
<td>4</td>
<td>5</td>
<td>3.60</td>
<td>12</td>
<td>33.91</td>
<td>0.000</td>
</tr>
<tr>
<td>Hebrew</td>
<td>43</td>
<td>5</td>
<td>5</td>
<td>1.72</td>
<td>16</td>
<td>48.60</td>
<td>0.000</td>
</tr>
<tr>
<td>Iraqw</td>
<td>21</td>
<td>3</td>
<td>4</td>
<td>1.75</td>
<td>6</td>
<td>17.10</td>
<td>0.006</td>
</tr>
<tr>
<td>Italian</td>
<td>39</td>
<td>5</td>
<td>5</td>
<td>1.56</td>
<td>16</td>
<td>30.71</td>
<td>0.003</td>
</tr>
<tr>
<td>Kalamang</td>
<td>54</td>
<td>6</td>
<td>5</td>
<td>1.80</td>
<td>20</td>
<td>47.08</td>
<td>0.001</td>
</tr>
<tr>
<td>Kanuri</td>
<td>39</td>
<td>5</td>
<td>5</td>
<td>1.56</td>
<td>16</td>
<td>42.85</td>
<td>0.000</td>
</tr>
<tr>
<td>Ket</td>
<td>29</td>
<td>3</td>
<td>5</td>
<td>1.93</td>
<td>8</td>
<td>25.87</td>
<td>0.000</td>
</tr>
<tr>
<td>Latvian</td>
<td>60</td>
<td>4</td>
<td>5</td>
<td>3.00</td>
<td>12</td>
<td>25.46</td>
<td>0.005</td>
</tr>
<tr>
<td>Malagasy</td>
<td>58</td>
<td>4</td>
<td>5</td>
<td>2.90</td>
<td>12</td>
<td>32.73</td>
<td>0.001</td>
</tr>
<tr>
<td>Oroqen</td>
<td>24</td>
<td>5</td>
<td>5</td>
<td>0.96</td>
<td>16</td>
<td>27.92</td>
<td>0.003</td>
</tr>
<tr>
<td>Polish</td>
<td>54</td>
<td>7</td>
<td>5</td>
<td>1.54</td>
<td>24</td>
<td>61.34</td>
<td>0.000</td>
</tr>
<tr>
<td>Russian</td>
<td>43</td>
<td>6</td>
<td>5</td>
<td>1.43</td>
<td>20</td>
<td>61.63</td>
<td>0.000</td>
</tr>
<tr>
<td>*Saramaccan</td>
<td>31</td>
<td>3</td>
<td>5</td>
<td>2.07</td>
<td>8</td>
<td>20.83</td>
<td>0.056</td>
</tr>
<tr>
<td>Seri</td>
<td>15</td>
<td>6</td>
<td>3</td>
<td>0.83</td>
<td>10</td>
<td>30.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Sidamo</td>
<td>29</td>
<td>4</td>
<td>4</td>
<td>1.81</td>
<td>9</td>
<td>24.59</td>
<td>0.001</td>
</tr>
<tr>
<td>Swahili</td>
<td>87</td>
<td>4</td>
<td>5</td>
<td>4.35</td>
<td>12</td>
<td>76.28</td>
<td>0.000</td>
</tr>
<tr>
<td>Turkish</td>
<td>63</td>
<td>3</td>
<td>5</td>
<td>4.20</td>
<td>8</td>
<td>35.63</td>
<td>0.000</td>
</tr>
<tr>
<td>Wichi</td>
<td>35</td>
<td>4</td>
<td>5</td>
<td>1.75</td>
<td>12</td>
<td>27.98</td>
<td>0.003</td>
</tr>
<tr>
<td>Wik-Mungkan</td>
<td>24</td>
<td>3</td>
<td>4</td>
<td>2.00</td>
<td>6</td>
<td>23.43</td>
<td>0.005</td>
</tr>
<tr>
<td>Yakut</td>
<td>39</td>
<td>5</td>
<td>5</td>
<td>1.56</td>
<td>16</td>
<td>45.85</td>
<td>0.000</td>
</tr>
<tr>
<td>Yaqui</td>
<td>17</td>
<td>3</td>
<td>4</td>
<td>1.42</td>
<td>6</td>
<td>27.52</td>
<td>0.005</td>
</tr>
<tr>
<td>*Zinacantán Tzotzil</td>
<td>25</td>
<td>4</td>
<td>5</td>
<td>1.25</td>
<td>12</td>
<td>35.06</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Table 36: Chi-squared and Fisher tests of ftype2 x htype

6.4.4 Degree of marking

A legitimate cross-linguistic comparison of form 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 24 on page 124). Each binominal can therefore be assigned one of three values (0, 1 or 2) for a variable marking, as shown in Table 37.
Chapter 6. Patterns in the data

<table>
<thead>
<tr>
<th>ftype</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>markers</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>

Table 37: Number of markers per binominal type

<table>
<thead>
<tr>
<th>meaning</th>
<th>language</th>
<th>construction</th>
<th>ftype</th>
<th>marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>beehive</td>
<td>Basque</td>
<td>Base.LOC</td>
<td>der</td>
<td>0</td>
</tr>
<tr>
<td>beehive</td>
<td>Wolof</td>
<td>Head CON Mod</td>
<td>prp</td>
<td>1</td>
</tr>
<tr>
<td>beehive</td>
<td>Wawa</td>
<td>Head Mod.POSS.ASS</td>
<td>gen</td>
<td>1</td>
</tr>
<tr>
<td>beehive</td>
<td>Norwegian</td>
<td>Mod.Head</td>
<td>cmp</td>
<td>0</td>
</tr>
<tr>
<td>beehive</td>
<td>Q'eqchi'</td>
<td>3ERG.Head Mod</td>
<td>con</td>
<td>1</td>
</tr>
<tr>
<td>beehive</td>
<td>Akkadian</td>
<td>Head:STC Mod.OBL</td>
<td>dbl</td>
<td>2</td>
</tr>
<tr>
<td>beehive</td>
<td>Swahili</td>
<td>CL.Base</td>
<td>cls</td>
<td>0</td>
</tr>
<tr>
<td>beehive</td>
<td>Czech</td>
<td>Mod.ADJZ Head</td>
<td>adj</td>
<td>1</td>
</tr>
</tbody>
</table>

Weighted averages of the value of marking can now be calculated for each language:

<table>
<thead>
<tr>
<th>language</th>
<th>iso639</th>
<th>glottocode</th>
<th>area</th>
<th>family</th>
<th>genus</th>
<th>marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Äiwoo</td>
<td>nfl</td>
<td>ayiw1239</td>
<td>O</td>
<td>Austronesian</td>
<td>Oceanic</td>
<td>0.3571429</td>
</tr>
<tr>
<td>Akkadian</td>
<td>akk</td>
<td>akka1240</td>
<td>A</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>2.0000000</td>
</tr>
<tr>
<td>Amharic</td>
<td>amh</td>
<td>amha1245</td>
<td>A</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>0.7872340</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>aoi</td>
<td>anini1240</td>
<td>G</td>
<td>Gunwinyguan</td>
<td>Gunwinyguan</td>
<td>0.6250000</td>
</tr>
<tr>
<td>Archi</td>
<td>aqc</td>
<td>archi1244</td>
<td>E</td>
<td>Nakh-Daghestanian</td>
<td>Lezgic</td>
<td>0.7857143</td>
</tr>
<tr>
<td>Assamese</td>
<td>asm</td>
<td>assa1263</td>
<td>E</td>
<td>Indo-European</td>
<td>Indo-Aryan</td>
<td>0.5000000</td>
</tr>
</tbody>
</table>

And the same can be done for each meaning:

<table>
<thead>
<tr>
<th>meaning</th>
<th>semField</th>
<th>semType</th>
<th>semType2</th>
<th>marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>ankle</td>
<td>The body</td>
<td>concrete body part</td>
<td>0.30232558</td>
<td></td>
</tr>
<tr>
<td>arctic lights</td>
<td>The physical world</td>
<td>concrete</td>
<td>natural</td>
<td>0.78787879</td>
</tr>
<tr>
<td>backpack</td>
<td>Modern world</td>
<td>concrete</td>
<td>basic</td>
<td>0.41379310</td>
</tr>
<tr>
<td>bee</td>
<td>Animals</td>
<td>animal</td>
<td>animal</td>
<td>0.08695652</td>
</tr>
<tr>
<td>beehive</td>
<td>Animals</td>
<td>concrete</td>
<td>natural</td>
<td>0.39622642</td>
</tr>
<tr>
<td>beeswax</td>
<td>Animals</td>
<td>concrete</td>
<td>natural</td>
<td>0.38775510</td>
</tr>
</tbody>
</table>

Marking by language

Now it is possible to look for structure in groupings of languages and meanings. In the case of languages, the only relevant grouping is geographical area; the sample of languages is spread too thinly and unevenly across the 36 families and 64 genera for an analysis based on genetic affiliation. There are six geographical areas (we exclude Pidgins and creoles since the grouping is non-areal and only contains two languages). We start by investigating the distribution of the variable marking across languages:
Values range from 0 to 2 with a median of 0.4, and the first 25% of the data fall between 0 and 0.03, so they are clearly not normally distributed. This is confirmed by the plots in Figure 65, and by the Shapiro-Wilks test where \( p < 0.001 \) indicates that the null hypothesis of normally distributed data must be rejected. It is therefore not possible to use parametric methods on this data.

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:

\[
> \text{kruskal.test(marking ~ area, data=1)} \quad \# \quad p < 0.05 \quad \# \text{H0 cannot be rejected}
\]

Kruskal-Wallis rank sum test

data:  marking by area
Kruskal-Wallis chi-squared = 13.127, df = 5, \text{p-value} = 0.02221

The resulting \( p \)-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:

\[
> \text{pairwise.wilcox.test(jitter(marking), area2, p.adjust.method = "BH")}
\]

Pairwise comparisons using \text{wilcoxon rank sum test}

data:  jitter(marking) and area2

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Eurasia</th>
<th>Oceania/SE Asia</th>
<th>PNG/Australia</th>
<th>North America</th>
<th>South America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurasia</td>
<td>0.187</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oceania/SE Asia</td>
<td>0.122</td>
<td>0.091</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PNG/Australia</td>
<td>0.289</td>
<td>0.616</td>
<td>0.504</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North America</td>
<td>0.504</td>
<td>0.989</td>
<td>0.267</td>
<td>0.764</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>South America</td>
<td>0.267</td>
<td>0.504</td>
<td>0.504</td>
<td>0.989</td>
<td>0.602</td>
<td>-</td>
</tr>
</tbody>
</table>
In conclusion, the present data do not provide evidence of significant differences in the amount of marking across the five main geographical areas.

**Marking by meanings**

The situation with respect to the distribution of marking across meanings is quite different:

```r
> summary(marking)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
0.0000  0.1854  0.3431  0.3366  0.4615  0.7879  
> sd(marking)
[1] 0.1815107  
> # test for normality using Shapiro-Wilks  
> shapiro.test(marking)

Shapiro-Wilk normality test
data:  marking
W = 0.978, p-value = 0.09269
```

![Graphs showing the distribution of marking across meanings](image)

**Figure 66: Normal distribution of marking across meanings**

The mean of 0.34 is near the centre of the range from 0 to 0.79, and while the first two plots in Figure 66 show a very slightly bimodal distribution and the QQ-plot has a whiff of an S shape, the Shapiro-Wilks test returns a p-value of 0.09, which means that the null hypothesis cannot be discarded. Moreover, the box plot of
marking 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:

```r
> lm1 <- lm(marking ~ semType, data=m)
> anova(lm1)
Analysis of Variance Table

Response: marking

              Df Sum Sq Mean Sq F value   Pr(>F)
semType        6  0.9636 0.16060   6.499 9.429e-06 ***
Residuals     93  2.2981 0.02471
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # ==> significant differences in degree of marking across sem types
```

The ANOVA shows highly significant differences in the value of marking 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 the Person (the intercept).

```r
> summary(lm1)

Call:
  lm(formula = marking ~ semType, data = m)

Residuals:
         Min        1Q  Median        3Q       Max
-0.35169 -0.10992 -0.00577  0.10117  0.38884

Coefficients:  Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.165710   0.037047   4.473 2.18e-05 ***
semTypeanimal 0.078781   0.070020   1.125 0.263434
semTypelocation 0.073700   0.079470   0.927 0.356120
semTypebody part 0.224720   0.051067   4.400 2.88e-05 ***
semTypenatural 0.276480   0.058578   4.719 8.32e-06 ***
semTypebasic   0.185980   0.051067   3.641 0.000446 ***
semTypeadvanced 0.257489   0.052403   4.914 3.82e-06 ***
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1572 on 93 degrees of freedom
Multiple R-squared:  0.2954,  Adjusted R-squared:  0.25
F-statistic: 6.499 on 6 and 93 DF,  p-value: 9.429e-06
```

An alternative to using `anova()` on a linear model is to use the function `aov()`:

```r
> a1 <- aov(marking ~ semType)
> summary(a1)

            Df Sum Sq Mean Sq F value   Pr(>F)
semType      6 0.9636 0.16060   6.499 9.43e-06 ***
Residuals   93 2.2980 0.02471
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

This confirms that the effects are significant ($p < 0.001$), so we perform post-hoc pairwise comparisons (using the conservative Bonferroni correction) to find out which pairs of groups are significantly different from each other:
Chapter 6. Patterns in the data

> pairwise.t.test(marking, semType, p.adj = "bonf")

Pairwise comparisons using t tests with pooled SD
data: marking and semType

<table>
<thead>
<tr>
<th></th>
<th>person</th>
<th>animal</th>
<th>location</th>
<th>body part</th>
<th>natural</th>
<th>basic</th>
<th>advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>1.00000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>location</td>
<td>1.00000</td>
<td>1.00000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>body part</td>
<td>0.00060</td>
<td>0.78093</td>
<td>1.00000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>natural</td>
<td>0.00017</td>
<td>0.20171</td>
<td>0.36342</td>
<td>1.00000</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>basic</td>
<td>0.00936</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>advanced</td>
<td>8e-05</td>
<td>0.25904</td>
<td>0.48174</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>-</td>
</tr>
</tbody>
</table>

P value adjustment method: bonferroni

These comparisons confirm the previous results, which suggest the following basic scale for degree of marking:

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):

> kruskal.test(marking ~ semType, data=m)

Kruskal-Wallis rank sum test
data: marking by semType
Kruskal-Wallis chi-squared = 29.249, df = 6, p-value = 5.457e-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:

> pairwise.wilcox.test(jitter(marking), semType, p.adjust.method = "BH")

Pairwise comparisons using Wilcoxon rank sum test
data: jitter(marking) and semType

<table>
<thead>
<tr>
<th></th>
<th>person</th>
<th>animal</th>
<th>location</th>
<th>body part</th>
<th>natural</th>
<th>basic</th>
<th>advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>0.3520</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>location</td>
<td>0.2633</td>
<td>0.7061</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>body part</td>
<td>6.4e-05</td>
<td>0.0326</td>
<td>0.0341</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>natural</td>
<td>0.0008</td>
<td>0.0341</td>
<td>0.0258</td>
<td>0.6741</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>basic</td>
<td>0.0087</td>
<td>0.3520</td>
<td>0.2633</td>
<td>0.9893</td>
<td>0.5076</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>advanced</td>
<td>0.0011</td>
<td>0.1720</td>
<td>0.2329</td>
<td>0.5260</td>
<td>0.9893</td>
<td>0.3520</td>
<td>-</td>
</tr>
</tbody>
</table>

P value adjustment method: BH

In addition to confirming previous results, this tests reveals differences between Animal and Location, on the one hand, and Body part and Natural, on the other, albeit at lower significance (p < 0.05). If correct, this suggests a more fine-grained scale of marking, which could be the subject of future research

Person < Animal, Location < Concrete
6.5 Chapter summary

In this chapter I provided a detailed description of the binominal data by language and by meaning, focusing initially on vocabulary size, morphological complexity and binominal frequency. This showed considerable cross-linguistic variation across all three parameters.

I also examined the distribution of binominals by structural type, the preferences shown by individual languages for particular types, and the degree of intralingual competition between different types, as well as genetic fingerprinting based on structural type. Among other things, I confirmed the prevalence of compounding but also disconfirmed its status as a universal, and I showed that the binominal fingerprint can be a good indicator of genus, at least in Indo-European.

The discussion of how low- and high-level semantic relations distribute across the data set revealed the following scale:

\[
\text{PART} \gg \text{PURP} > \text{COOR} > \text{LOC} > \text{POSS2} > \text{COMP} > \text{USE} > \text{TIME} > \text{PROD} > \ldots
\]

It also showed that the kinds of relation varies considerably across semantic fields, PART-WHOLE being particularly important in the domain of The body, and PURPOSE in the domain of the Modern world.

Finally I applied a number of multivariate statistical methods with a view to finding associations in the data. Association rules proved to be inappropriate, partly due to the nature of my data, but the attempt to apply this method led to the realization that formal subtypes of a comparative concept cannot necessarily form the basis of further cross-linguistic comparison. I addressed this issue in two ways. Firstly, I focused on individual languages, using a motivational grid approach. Secondly, I created a simpler typology based on a single property, the number of markers, that produced interesting results in the form of a tentative scale of marking:

\[
\text{Person} < \text{Animal}, \text{Location} < \text{Concrete}
\]

In sum, the data revealed a number of interesting patterns that are worthy of further investigation, but many of them require additional data. Some might be studied using fewer meanings but would require coverage of more languages, others would 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 could not produce immediate results.
7 Typological investigations

The statistical analyses of the preceding chapter provide a potential foundation for typological generalizations of the kind envisaged by Song for the fourth stage of his recipe for doing typology. In this chapter I follow up on some of the more interesting findings. The topics under consideration are word order typology (§7.1), my two-paths (of mental access) hypothesis (§7.2), and the associative relations in general (§7.3). Occasionally I attempt to put forward explanations of the kind Song looked for in the fifth and final stage of his model, but given the exploratory nature of this study this is not possible everywhere. In addition, I once again force the linguist reader out of her comfort zone in order to show how Topic Maps can enrich the discussion.

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 structural 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 2 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. The two principal sources give somewhat different sets of numbers (Table 38), perhaps because different counting methods were employed, but they agree in reporting much higher frequencies (roughly 6:1 and 10:1) for right-headed compounds than the 2:1 ration 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>5.75</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>9.81</td>
</tr>
</tbody>
</table>

*Table 38: Constituent order pace Morbo/Comp*

The binominal data of the present study are, of course, not strictly comparable with those of Bauer and Morbo/Comp since, on the one hand, they only include noun-noun compounds and, on the other hand, also include functional equivalents of the latter. That being said, it is striking that the ratio of right- to left-headed binominals across the database as a whole (64% vs. 36%) is very close to the 2:1 figure obtained by Bauer and myself.

![Figure 67: Constituent order by binominal type](image)

A more appropriate comparison can be made when the numbers are broken down by structural type, as shown in Figure 67. We observe that 68% of cmp binominals are right-headed, while 32% are left-headed, a ratio of slightly over 2:1. Even these
figures are not strictly comparable to those of Bauer and Morbo/Comp, because they do not include compounds of other types (synthetic, verbal, coordinate, etc.). However they do support Bauer as against Morbo/Comp, and they also suggest that the latter’s inflated view of the prevalence of right-headedness may be due to areal and genetic bias. (Recall from page 34 that 21 of the 25 languages in that 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 68) which shows right-headedness to be much more frequent in Eurasia than in other parts of the world. Lest it be thought that the preponderance of right-headed cmps in Eurasia is due to the large number of Indo-European languages in the sample (cf. Table 19, page 90), it turns out that excluding IE languages results in an increase of the percentage of right-headed cmps in Eurasia from 92% to 98%.

The other numbers in Figure 67 should be taken with a grain of salt because of the relatively low proportion of types other than cmp in the database. In particular, the preponderance of left-headed cls binominals can be ascribed to the contribution of a single language, Swahili, which accounts for over 70% of binominals of this type. The values for der demonstrate the overwhelming preponderance of suffixation 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.òvo [AGT.fishtrap] FISHERMAN. The preference for right-headedness in gen and adj constructions, both of which involve dependent-marking, is of the same order of magnitude as for cmp, while there is a slight preference for left-headedness in con and dbl,
which both involve head-marking (the latter in addition to dependent-marking, of course). The figures for prp are as striking as those for der: 94% are left-headed and only 6% 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 prp binominals and these are, of course, all left-headed (as in chemin de fer).

However, that is probably not the whole story. A closer inspection of the data shows that the only language with more than one or two right-headed prp binominals is Hindi. Neither Japanese with its no construction (recall kumo no su) nor Nepali, a close relative of Hindi, are recorded as exhibiting a single binominal of this type. Instead they both exhibit a fair number of the type gen (12 and 6, respectively). This fact leads to the following explanation for the paucity of right-headed prp binominals: Adpositions usually occupy the position between the head and the modifier, so the two basic patterns are left-headed Head [PREP Mod] and right-headed [Mod POSTP] Head. Now, it is much more common for a particle that follows its co-constituent (as POSTP follows Mod) to fuse with the latter than for one that precedes it (as PREP precedes Mod); a postposition develops much more easily into a postposed case affix than does a preposition into a preposed case affix. The reason why Japanese and Nepali do not figure amongst languages with prp binominals, is simply because binominals like Jap. kumo no su and Nep. mākurā ko jālo [spider. GEN web] SPIDER WEB are coded in the database as gen rather than prp. They are examples of one of the gradient phenomena involved in the lexicalisation of binominals (§4.5). An argument could be made for combining prp and gen into a single type (which could be labelled flg on the basis of Haspelmath’s flag/index terminology, §4.1.2), but this would obscure the oppositions found in languages like Polish (§6.2.2). It would also give Irish a binominal fingerprint very similar to that of the Romance languages (§6.2.3).

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 mixed order. (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.)

Figure 69: Distribution of right-headed and left-headed binominals

Generally speaking, in languages that exhibit mixed order 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 *sidir.o.δromos* [iron.LE.road] RAILWAY, while 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 binominal type exhibits both head-initial and head-final order. This is the case with both Vietnamese compounds (109) and Polish adjectival constructions (110). In the case of Vietnamese, this is a 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).

(109) Vietnamese

a. **cmpL** Head Mod *xe liá* [vehicle fire] TRAIN
b. **cmpR** Mod Head *hoá xa* [fire vehicle] TRAIN

(110) Polish

a. **adjL** Head Mod.ADJZ *droga mlecz.na* [road milk.ADJZ] MILKY WAY
b. **adjR** Mod.ADJZ Head *kamień.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 (§3.4.2). As a consequence, when 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 107). Now, in many cases, the result of such an analysis conforms to the patterns otherwise found in the language. For example 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, kiin 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 Est. kesk.päev. In order to maintain analytical consistency, such words must be classed as left-headed. These anomalies are paralleled in the database by three words 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. Nor. midsommer, Eng. midweek, etc.) and occasionally the middle of something else (e.g. Eng. midship). The explanation for these exceptions 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.

A more detailed analysis of the data shows that mixed order is found in 62 of the 99 languages in the sample, but also that mixed order is often somewhat peripheral, as witness the size of the tails (especially the right-hand tail) of the middle section of the plot. In order to investigate what is going on in languages where mixed order is more than just a peripheral phenomenon, 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 and produces a list of 14 languages in which mixed order is a significant feature. These are listed in Table 39, together with the structural types they exhibit (here using the 16-way classification that includes the head position parameter). Structural types are grouped by head position, and those types that are most frequent in the language in question are shown in bold. The Boolean value in the final column flags whether the der type plays a significant role in the comparison.
In the first eight languages the mixed order phenomenon is due to a co-occurrence of right-headed \texttt{der} constructions and one or more left-headed, non-derivational types. In Gawwada and Welsh, the most frequent left-headed construction is \texttt{cmpL}; in Irish and Kanuri it is \texttt{genL}; in the three Romance languages it is \texttt{prpL}; and in Polish it is \texttt{adjL}. We know from Figure 46 (page 214) that the \texttt{der} type is relatively common, from Figure 47 that it is fairly widespread, and from Figure 48 that it is attested in more than half of the languages in the sample. In addition, Figure 67 tells us that 95\% of such binominals are right-headed. That the \texttt{derR} type should be involved in many of the mixed order situations should therefore come as no surprise, so these cases will not be discussed further.

Of more interest, perhaps, are the six cases in which derivation does not play a role. Of these, two (Vietnamese and Mapudungun) show an opposition between two types of \texttt{cmp}; in Greek it is between \texttt{gen} and \texttt{cmp}, as mentioned earlier; in Barain it is between \texttt{prp} on the one hand and \texttt{con} on the other; in Kalamang it is between \texttt{cmp} and \texttt{con}; and in Ticuna it is between \texttt{con} and \texttt{cmp}. These six cases will be investigated individually in the next section.
7.1.3 No basic order

Three of the six languages that exhibit no basic order (where this is not attributable to right-headed derivation) have both left-headed and right-headed compounds, i.e. cmpL and cmpR. Although this phenomenon has not been investigated in depth, it appears to be relatively uncommon cross-linguistically. We have already noted (Figure 48) that compounding is almost (but not quite) ubiquitous, and we have also seen (Figure 67) that left-headed compounds are far from exceptional. In any given language, compounds tend to be either left-headed or right-headed,¹ but there are exceptions. Examples cited in the literature include Mandarin (Packard 2000; Ceccagno & Scalise 2006; cf. Table (21) in Scalise & Fábregas 2010), Welsh (Awbery 2004), Western Farsi (Tehranisa 1987) and Oko (Glottocode okoo1245) (Atoyebi 2010), as well as the languages under consideration here (and, of course, Nizaa, to which I will return in the next section).

One major class of exceptions comprises languages in which the different orders are exhibited by different compound types: this is the case with Mandarin, which has left-headed verbal compounds and right-headed nominal compounds. However, this kind of exception cannot occur in the domain of binominals, since only noun-noun compounds qualify as binominals; thus other explanations must be sought for Vietnamese, Mapudungun and Kalamang.

In the case of Vietnamese, the answer – language contact – has already been noted: native compounds are left-headed, whereas loans from Chinese are right-headed (cf. example 109 on page 253). (The same explanation appears to hold for Welsh and Western Farsi, although this awaits further investigation.) As for Mapudungun, it was mentioned earlier (§6.4.2) 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. 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, but the distinction is the same, as the motivational grid in Table 40 shows. Unfortunately there is too little data from Mapudungun in the current database to provide statistically significant evidence, but the tendency for cmpL to express one kind of relation (HinM) and cmpR to express others (MinH and HtoM) is clear.

¹ Here I ignore exceptions that prove the rule, like Eng. attorney-general (cf. page 218).
Chapter 7. Typological investigations

Table 40: Motivational grid for Mapudungun

<table>
<thead>
<tr>
<th></th>
<th>htype</th>
<th>MisH</th>
<th>HinM</th>
<th>MinH</th>
<th>HtoM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ftype2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cmpL</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>cmpR</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>derR</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>genR</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The situation in Kalamang is somewhat different, as Table 41 shows. While both cmpL and cmpR are present in significant numbers, and the opposition between MisH (cmpL) and HinM (cmpR) corresponds to that between attributive and subordinative relations, the main contrast is between cmpR and conR, two head-final constructions. I therefore defer discussion of this case until later (page 263ff).

Table 41: Motivational grid for Kalamang

<table>
<thead>
<tr>
<th></th>
<th>htype</th>
<th>MisH</th>
<th>HinM</th>
<th>MinH</th>
<th>HtoM</th>
<th>MtoH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ftype2</strong></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>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>cmpL</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>cmpR</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>conR</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>genL</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>prpL</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>chi-squared p = 0.000324078968400871</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-squared = 48.8384646962233    df = 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisher's exact p = 0.000499750124937531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

@@ Should I spend time discussing the last three languages in Table 39? I’m not sure how much can be said about them. Greek and Ticuna show high p-values, so the null hypothesis of independence cannot be rejected, but the value for Barain is significant p-value. Here are the motivational grids:

- **Greek: genL + cmpR**

Table 42: Motivational grid for Greek

<table>
<thead>
<tr>
<th></th>
<th>htype</th>
<th>MisH</th>
<th>HinM</th>
<th>MinH</th>
<th>HtoM</th>
<th>MtoH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ftype2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjR</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>cmpL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>cmpR</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>derR</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>genL</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Fisher's exact p = 0.127936031984008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

@@@ Check Ralli’s work (esp. 2013) for possible explanations.
• Barain: \textbf{prpL} + \textbf{conR}

<table>
<thead>
<tr>
<th>ftype2</th>
<th>cmpL</th>
<th>conR</th>
<th>prpL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

Fisher's exact $p = 0.00699650174912544$

Table 43: Motivational grid for Barain

@ @ Contact Joey Lovestrand. The \textbf{prpL} types look like relative clause constructions. Perhaps they are descriptions rather than names? There is also a lot of minor variation…

• Ticuna: \textbf{conL} + \textbf{cmpR}

<table>
<thead>
<tr>
<th>ftype2</th>
<th>cmpR</th>
<th>conL</th>
<th>dblL</th>
<th>genL</th>
<th>genR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Fisher's exact $p = 0.180409795102449$

Table 44: Motivational grid for Ticuna

@ @ Ticuna could be ignored on the grounds that Fisher’s Exact is not significant.

7.2 The two-paths hypothesis

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 (§7.2.1). Following this I approach the hypothesis from the perspective of salience: in §7.2.2 I look into the relative salience of head and modifier, and in §7.2.3 the salience of elaboration sites. Finally, in §7.2.4 I permit myself an excursus into a more tangential topic relating to the salience of the head and modifier, where I suggest the possibility of a “species-attribute typology”.
7.2.1 Evidence from Nizaa and binominals

The basis for what I originally referred to as the “dual strategy” was the presence of two types of compound in Nizaa, left-headed (cmpL) and right-headed (cmpR). My analysis revealed that these exhibited two disjunct sets of semantic relation (Table 45). 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.

<table>
<thead>
<tr>
<th>Relation type</th>
<th>Total</th>
<th>LH</th>
<th>RH</th>
<th>Relation type</th>
<th>Total</th>
<th>LH</th>
<th>RH</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCATION</td>
<td>26</td>
<td>26</td>
<td>0</td>
<td>PART</td>
<td>45</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>KIN</td>
<td>9</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>ACTIVITY</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>LOCATED</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>APPEARANCE</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>POSSESSION</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>IDEM</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>PRODUCT</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>STATE</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>CONTAINER</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>GENDER</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>RESPONSIBLE</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EVENT</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>opaque</td>
<td>13</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>MATERIAL</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTENTS</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROPER</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMOTION</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOMAIN</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEATURE</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOOD</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 45: Semantic relations in Nizaa
(after Pepper 2010)

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 it generally denotes the role played by one of the participants in the relation. For example, LOCATION is a relation in which the modifier denotes a location (111a), while LOCATED is a relation in which the modifier denotes something which has a location (111b). The non-transparent term IDEM was used for any kind of appositional relation (111c).

(111) Nizaa (Pepper 2010)

1. LOCATION  siŋw nim [bird water] ‘duck’
2. LOCATED   yir nim [eye water] ‘tear’
3. IDEM      siŋw ŋgɔ̀ [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. Perhaps the results are too good to be true,¹ 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’).

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 judgement calls inevitably involve an element of subjectivity”, and that there is reason to believe

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.
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 USE (a windmill is a mill that uses wind), which maps to Hatcher’s MtoH, which according to the mapping given in §6.4.2 is subordinative (Figure 70).

<table>
<thead>
<tr>
<th>Bourifier2</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>Code</td>
</tr>
<tr>
<td>TAXONY</td>
<td>TAX</td>
<td>a wind is a kind of mill</td>
<td>oak tree</td>
<td>TAX2</td>
</tr>
<tr>
<td>COORDINATION</td>
<td>COOR</td>
<td>a windmill is a mill AND a wind</td>
<td>bay king</td>
<td></td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>SIM</td>
<td>a mill that is similar to a wind</td>
<td>ant lion</td>
<td></td>
</tr>
<tr>
<td>TOPIC</td>
<td>TOP</td>
<td>a mill that is about a wind</td>
<td>history book</td>
<td></td>
</tr>
<tr>
<td>CONTAINMENT</td>
<td>CONT</td>
<td>a mill that is contained in a wind</td>
<td>orange seed</td>
<td>CONT2</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>POSS</td>
<td>a mill that possesses a wind</td>
<td>career girl</td>
<td>POSS2</td>
</tr>
<tr>
<td>PART</td>
<td>PART</td>
<td>a mill that is part of a wind</td>
<td>car motor</td>
<td>PART2</td>
</tr>
<tr>
<td>LOCATION</td>
<td>LOC</td>
<td>a mill located at/near/in a wind</td>
<td>house music</td>
<td>LOC2</td>
</tr>
<tr>
<td>TIME</td>
<td>TIME</td>
<td>a mill that occurs at/during a wind</td>
<td>summer job</td>
<td>TIME2</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>COMP</td>
<td>a mill composed of wind</td>
<td>sugar cube</td>
<td>COMP2</td>
</tr>
<tr>
<td>SOURCE</td>
<td>SRC</td>
<td>a mill from a wind</td>
<td>cane</td>
<td>SRC2</td>
</tr>
<tr>
<td>CAUSE</td>
<td>CAUS</td>
<td>a mill that causes a wind</td>
<td>tear gas</td>
<td>CAUS2</td>
</tr>
<tr>
<td>PRODUCTION</td>
<td>PROD</td>
<td>a mill that produces a wind</td>
<td>song bird</td>
<td>PROD2</td>
</tr>
<tr>
<td>USE</td>
<td>USE</td>
<td>a mill that uses a wind</td>
<td>oil lamp</td>
<td>USE2</td>
</tr>
<tr>
<td>FUNCTION</td>
<td>FUNC</td>
<td>a mill that serves as a wind</td>
<td>buffer state</td>
<td></td>
</tr>
<tr>
<td>PURPOSE</td>
<td>PURP</td>
<td>a mill intended for a wind</td>
<td>animal doctor</td>
<td></td>
</tr>
</tbody>
</table>

Figure 70: The Bourifier – WINDMILL

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, cmpL and cmpR); importantly, it does not hinge on the order of constituents. It is the case that the two Nizaa constructions 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 111b), whereas when (sub-)categorisation is employed, the broader concept precedes the modifier (as ‘bird’ precedes ‘water’ in 111a). 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 semantic relation (either at the lower \text{stype} level or the high \text{htype} level) and structural type (which would have to be at the level of \text{ftype2} in order to capture differences between left- and right-headed forms of the same higher level \text{ftype}). 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. It is true that the analysis in §6.3.3 reveals some major differences between structural types (at the level of \text{ftype}) in terms of the kinds of semantic relation (both \text{stype} and \text{htype}) that they express. However, these differences have not been pursued, for the reason given §6.4.2, which is that formal subtypes of a comparative concept are functionally incommensurable across languages. For the same reason, the analysis based on association rules in §6.4.1 could not have provided any evidence, since this, too, was based on an inappropriate lumping together of every language.

Compared to this, the analysis in §6.4.3 using the chi-squared and Fisher’s Exact tests on language-specific motivational grids is more theoretically sound, and the results indicated that there may be a significant association between the kind of relation expressed by a binominal and the choice of “construction in as many as 24 of the 99 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 was expected (on the basis of previous knowledge) to confirm the hypothesis, Mapudungun, did not even make it onto the list of 24 languages in Table 36 on page 242. This is almost certainly because of the sparsity of the data, as witness the motivational grid in Table 40 on page 257.

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 may 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 in any particular case. 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 & McGregory 1996; Aikhenvald & Dixon 2013; Haspelmath 2017b). The kinds of relation found in prototypical possessive constructions (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, it is instructive to look in more depth at a language in which there are substantial competing constructions. Such a language is Kalamang, mentioned previously in §7.1.3 as a language that exhibits no basic order. For convenience, I reproduce the motivational grid for Kalamang as Table 46, here extended with row sums. As we can see 46 of the 55 binominals in the Kalamang data set are distributed almost equally across three structural types: cmpL, cmpR and conR. The question we wish to answer, is what kind of system motivates this distribution.

<table>
<thead>
<tr>
<th>ftype2</th>
<th>adjL</th>
<th>cmpL</th>
<th>cmpR</th>
<th>conR</th>
<th>genL</th>
<th>prpL</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>htype</td>
<td>MisH</td>
<td>HinM</td>
<td>MinH</td>
<td>HtoM</td>
<td>MtoH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjL</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>cmpL</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>cmpR</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>conR</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>genL</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>prpL</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 46: Motivational grid for Kalamang with row sums

The most intriguing feature of the distribution is that two of the types, cmpR and conR, tend to encode the same high-level relation, HinM. I will therefore focus on the data points that underpin this particular statistic. Table 47 lists all 32 binominals that instantiate these two types (grouped by type), together with the corresponding meaning and semantic type, both types of semantic relation and the gloss for the head and modifier: in short, every property for which these binominals have been annotated in the database.

Now, if any of these properties were responsible for the division into two structural types, we would expect there to 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 46 that the high-level relation HinM is found in both types; Table 47 shows that the low-level relation PART is also common to both. 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 assigned to different structural types.
## The typology of binominal lexemes

<table>
<thead>
<tr>
<th>meaning</th>
<th>semType</th>
<th>head</th>
<th>mod</th>
<th>htype</th>
<th>stype</th>
<th>word</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cmpR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANKLE</td>
<td>Body part</td>
<td>##</td>
<td>foot</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kor kasir</td>
</tr>
<tr>
<td>EYELASH</td>
<td>Body part</td>
<td>hair</td>
<td>eye</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kanggir nenen</td>
</tr>
<tr>
<td>RIB</td>
<td>Body part</td>
<td>bone</td>
<td>side</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kir.kang</td>
</tr>
<tr>
<td>SHOULDERBLADE</td>
<td>Body part</td>
<td>bone</td>
<td>shoulder</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>bekiem.kang</td>
</tr>
<tr>
<td>SKULL</td>
<td>Body part</td>
<td>bone</td>
<td>person</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>som.kang~kang</td>
</tr>
<tr>
<td>SKULL</td>
<td>Body part</td>
<td>shell?</td>
<td>head</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>nakal taok</td>
</tr>
<tr>
<td>SPINE</td>
<td>Body part</td>
<td>bone</td>
<td>back</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>suol.kang</td>
</tr>
<tr>
<td>TOE</td>
<td>Body part</td>
<td>##</td>
<td>foot</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kor.parok</td>
</tr>
<tr>
<td>WRIST</td>
<td>Body part</td>
<td>##</td>
<td>hand</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>tan kasir</td>
</tr>
<tr>
<td>TEAR</td>
<td>Body part</td>
<td>drops</td>
<td>water</td>
<td>&quot;MinH&quot;</td>
<td>COMP</td>
<td>per.tam</td>
</tr>
<tr>
<td>FOOTPRINT</td>
<td>Body part</td>
<td>##</td>
<td>foot</td>
<td>&quot;MtoH&quot;</td>
<td>CAUS2</td>
<td>kor.kom</td>
</tr>
<tr>
<td>BEEHIVE</td>
<td>Natural</td>
<td>nest</td>
<td>bee</td>
<td>&quot;HinM&quot;</td>
<td>POSS2</td>
<td>wenawena eun</td>
</tr>
<tr>
<td>FLAME</td>
<td>Natural</td>
<td>embers</td>
<td>fire</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>din paras</td>
</tr>
<tr>
<td>TREE TRUNK</td>
<td>Natural</td>
<td>base/trunk</td>
<td>tree</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>ror ewun</td>
</tr>
<tr>
<td>RAINBOW</td>
<td>Natural</td>
<td>##</td>
<td>rain</td>
<td>&quot;MtoH&quot;</td>
<td>CAUS2</td>
<td>kalis tanggir</td>
</tr>
<tr>
<td>DINNER</td>
<td>Basic</td>
<td>food</td>
<td>evening</td>
<td>&quot;HinM&quot;</td>
<td>TIME</td>
<td>saun muap</td>
</tr>
<tr>
<td>LUNCH</td>
<td>Basic</td>
<td>food</td>
<td>afternoon</td>
<td>&quot;HinM&quot;</td>
<td>TIME</td>
<td>goyuol muap.et</td>
</tr>
<tr>
<td>GLOVE</td>
<td>Basic</td>
<td>cover</td>
<td>hand</td>
<td>&quot;HtoM&quot;</td>
<td>PURP</td>
<td>tan sarong</td>
</tr>
<tr>
<td><strong>conR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOST</td>
<td>Person</td>
<td>old person</td>
<td>house</td>
<td>&quot;MinH&quot;</td>
<td>POSS</td>
<td>kewe tara.un</td>
</tr>
<tr>
<td>FOAL OR COLT</td>
<td>Animal</td>
<td>child</td>
<td>horse</td>
<td>&quot;MisH&quot;</td>
<td>COOR</td>
<td>kuda tumun.un</td>
</tr>
<tr>
<td>LAMB</td>
<td>Animal</td>
<td>child</td>
<td>sheep</td>
<td>&quot;MisH&quot;</td>
<td>COOR</td>
<td>domba tumun.un</td>
</tr>
<tr>
<td>EARLOBE</td>
<td>Body part</td>
<td>under</td>
<td>ear</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kelkam elao.un</td>
</tr>
<tr>
<td>EYEBROW</td>
<td>Body part</td>
<td>skin</td>
<td>eye</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kanggir pul.un</td>
</tr>
<tr>
<td>EYELID</td>
<td>Body part</td>
<td>skin</td>
<td>eye</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>kanggir pul.un</td>
</tr>
<tr>
<td>NIPPLE OR TEAT</td>
<td>Body part</td>
<td>##</td>
<td>breast</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>am bel.un</td>
</tr>
<tr>
<td>NOSTRIL</td>
<td>Body part</td>
<td>hole</td>
<td>nose</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>bustang pos.un</td>
</tr>
<tr>
<td>PALM OF HAND</td>
<td>Body part</td>
<td>under</td>
<td>hand</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>tan el.un</td>
</tr>
<tr>
<td>COLLARBONE</td>
<td>Body part</td>
<td>bone</td>
<td>chest</td>
<td>&quot;HinM&quot;</td>
<td>LOC</td>
<td>aknar kang.un</td>
</tr>
<tr>
<td>EARWAX</td>
<td>Body part</td>
<td>poop</td>
<td>ear</td>
<td>&quot;HinM&quot;</td>
<td>LOC</td>
<td>kolkam kit.un</td>
</tr>
<tr>
<td>DOORPOST</td>
<td>Basic</td>
<td>door</td>
<td>wood</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>anggas ror.un</td>
</tr>
<tr>
<td>FIREPLACE</td>
<td>Basic</td>
<td>place</td>
<td>fire</td>
<td>&quot;MinH&quot;</td>
<td>LOC2</td>
<td>din tompat.un</td>
</tr>
<tr>
<td>LICENSE PLATE</td>
<td>Advanced</td>
<td>plate</td>
<td>car</td>
<td>&quot;HinM&quot;</td>
<td>PART</td>
<td>mobil pelat.un</td>
</tr>
</tbody>
</table>

*Table 47: Competition between cmpR and conR in Kalamang*
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 *(conR) aknar kang.un* COLLARBONE, as well as various *(cmpR)* words denoting the body parts SHOULDERBLADE, SKULL and SPINE. Bones that are part of the human anatomy are surely equally inalienable, whatever particular body part they belong to?

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 (112) of words that both use the PART 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.

(112)  
\[ \begin{align*} 
\text{a. } & \text{cmpR } \text{kanggir nenen [eye hair] EYELASH} \\
\text{b. } & \text{conR } \text{kanggir pul.un [eye skin.3POSS] EYELID} 
\end{align*} \]

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.)

### 7.2.2 Salience of head and modifier

Another way to think about the two-paths hypothesis is to pose the question *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:

- If one of the two concepts being 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.
- Alternatively, if one of the elaboration sites involved is significantly more salient than the other, then the concept it belongs to might constitute the better starting point.

I discuss the first of these here and the second in following section. Both discussions require an understanding of the concept of salience and a way to operationalise it.

---

1 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.
Salience is defined as “the quality of being particularly noticeable or important; prominence”. 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 also assume that salience correlates to a large degree with frequency. With this as our starting point, consider Table 48, 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 27 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. 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 27 might be reduced to 10. The contrast in what we might call “conceptual variation” is even more striking given that the modifiers eyelid, lid and surface only occur once each: of the 65 binominals denoting EYELID, 62 use a word meaning EYE as the modifier. Let us label this pattern ‘Few modifiers, Many heads’ (FM) and denote it symbolically as (1,10); the latter will shortly be used as coordinates in a scatter plot.

<table>
<thead>
<tr>
<th>EYELID (4,27) → (1,10)</th>
<th>TRAIN (10,5) → (10,1)</th>
<th>MARE (8,5) → (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, eyelid, lid, surface</td>
<td>animal skin, back, body hair, brim, cover, cup-like shape, eyelid, flap, half, hut, layer, lid, little cap, meat, rim, roof, root, sack/bag, sheath, shell, skin, skin/hide, small piece, small thing, superior part, top, up</td>
<td>electricity, fire, ground, iron, land, outside, rail, railway, smoke, steam</td>
</tr>
<tr>
<td>4 → 1</td>
<td>27 → 10</td>
<td>5 → 1</td>
</tr>
</tbody>
</table>

Table 48: Conceptual variation in heads and modifiers

---

1 https://en.oxforddictionaries.com/definition/salience
Chapter 7. Typological investigations

The pattern exhibited by the 18 binominals that denote TRAIN is quite different. Here there are five heads and ten modifiers. Two of the five heads (car and cart) are hyponyms of the third (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.inyí [tree.sky] ‘airplane’, Mirjam Möller, p.c.). Altogether these account for 94% of all heads, since long only occurs once;¹ this is tantamount to just one kind of head. Modifiers, by contrast, vary substantially, so the pattern here is one of ‘Many modifiers, Few heads’ (MF), which I denote (10,1).

The meaning MARE exemplifies a third type. At a superficial level there are five heads and eight modifiers, but both heads and modifiers can be reduced to just two basic concepts, HORSE and FEMALE, giving a pattern ‘Few heads, Few modifiers’ (FF), 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 §7.2.4.)

The difference between the very different patterns of conceptual variation exemplified by EYELID, TRAIN and MARE can be illustrated as shown in Figure 71, in which the horizontal axis represents the diversity of modifier concepts and the vertical axis the diversity of head concepts. A meaning like EYELID (FM) belongs in the top left corner of the quadrant; TRAIN (MF) belongs in the bottom right; and MARE (FF) 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.

¹ The analysis of Querétaro Otomi ma.bojá [long.iron] as a long thing made of iron is by analogy with faní.bojá [horse.iron] BICYCLE (a “horse” made of iron).
It seems reasonable to hypothesize that the same two paths of mental access (via a related concept and via a superordinate concept) are inherent in the conceptualizations of EYELID and TRAIN. The concept EYELID has a strong tendency to be conceptualized as something related to the EYE, whereas TRAIN tends to be conceptualized as a VEHICLE of a particular (sub)type. This has repercussions for 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 FM pattern thus corresponds to the reference point path.

With TRAIN, the situation is the reverse: once VEHICLE, the hypernym (head) has been chosen as the starting point for the conceptualization, the only constraint on the choice of modifier is that it should serve to distinguish TRAIN from other types of vehicle. This can be done in multiple ways and, as a consequence, there is room for considerable variety in the choice of modifier, hence electricity, fire, railway, iron, steam, etc. The MF pattern thus corresponds to the hypernym path.

In short, I suggest that EYE and VEHICLE are chosen as starting points for the two conceptualizations because they are salient, basic-level concepts, and that 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 “synsets”. 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 synonym sets, 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 48 this results in the following counts: EYELID (27,4), TRAIN (5,10) and MARE (5,8). 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 to all 100 meanings is shown in Figure 72 and Figure 73.
Figure 72: Scatterplot of heads and modifiers

Figure 72 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=27. We observe a strong tendency for body parts to cluster with EYELID in the top left (FM) 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 boney 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 (10,5), which turns out to be at the higher end of 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 Figure 73. Here the 7 semantic types are separated out and the dots are replaced by hexagons of varying densities. As we just observed, Advanced concepts tend strongly to having 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 FM area with few modifiers and many heads, while most animals, including MARE, are to be found in the lower-left, FF area.

![Figure 73: Heatmap of heads and modifiers by semantic type](image)

All of this would seem to indicate that there is mileage to be gained from studying the variability of head and modifier concepts in more detail, and in relating it to both the concept of salience and the two-paths hypothesis.

7.2.3 Salience of elaboration sites

The prominence, frequency or basicness of heads and modifiers is one aspect of salience, but it is not the only one. 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 (113).

(113) product (*pepper mill, salt mill, flour mill, paper mill*)
    power source (*windmill, water mill*)
    part, use (*saw mill*)

Langacker (1991) defines elaboration site (or e-site) as follows: “In a construction, those facets of one component structure that the other component serves to elaborate.” This can be paraphrased, using the wording in Langacker (2008: 198), as a schematic substructure of one component in a construction which the other component
serves to elaborate, i.e. characterize in finer-grained detail; it thus equates to ‘slot’ in the slot-filler theories described on page 151. 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” (ibid.) Turning to the original 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 the examples above they are elaborated by pepper, wind, saw, etc.

Implicit in the notion of e-site, but not discussed by Langacker, is the existence of what we have been calling 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 pepper; in the case of windmill it is USE; and in the case of saw mill it is PART2. 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 49, which lists the ten heads that participate in the greatest diversity of relations. Thus, house participates in six kinds of relation; head, person, string, thread and tree in five, etc. (Note that no attempt has been made 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 55, might have made it into the Top Ten.)

<table>
<thead>
<tr>
<th>head</th>
<th>semantic relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>house</td>
<td>POSS2 (20), PURP (20), LOC2 (19), LOC (2), COMP (2), PART2 (1)</td>
</tr>
<tr>
<td>head</td>
<td>PART (12), POSS2 (6), LOC2 (2), SRC2 (2), COMP (1)</td>
</tr>
<tr>
<td>person</td>
<td>LOC (4), POSS (1), PROD (1), PURP (1), USE (1)</td>
</tr>
<tr>
<td>place</td>
<td>LOC2 (13), PURP (7), CAUS2 (2), LOC (1), SRC2 (1)</td>
</tr>
<tr>
<td>string</td>
<td>POSS (3), POSS2 (2), PART (1), PART2 (1), PURP (1)</td>
</tr>
<tr>
<td>thread</td>
<td>POSS2 (2), PART (2), USE2 (2), PURP (2), POSS (1)</td>
</tr>
<tr>
<td>tree</td>
<td>PART2 (7), PART (2), LOC (1), LOC2 (1), USE2 (1)</td>
</tr>
<tr>
<td>bone</td>
<td>PART (86), LOC (22), POSS (1), LOC2 (1)</td>
</tr>
<tr>
<td>father</td>
<td>POSS2 (1), PART (1), POSS (1), PROD (1)</td>
</tr>
<tr>
<td>hole</td>
<td>PART (52), LOC2 (3), PURP (2), POSS2 (1)</td>
</tr>
</tbody>
</table>

Table 49: Salience of elaboration sites (by relation)

1 The relations LOC2 and USE are acceptable alternatives. PART2 and LOC2 both map to MinH, whereas USE maps to MtoH. As observed earlier, there is often overlap between different relations.
Interpreting Table 49 is not entirely straightforward since the codes only indicate the kind of relation and not the role played by the head in that relation. For example, while POSS2 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 28 on page 200, where the template for POSS2 tells us that the relation involves “an H that M possesses”. If the head is house, as it is in the first row of the table, then the e-site indicated by the POSS2 relation is the one that would be elaborated by the modifier, i.e. the possessor. Likewise, the template for LOC2 (“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 general, it is the role played by the modifier (the elaborating component) that characterizes the e-site. This suggests that it would be useful to supplement the Hatcher-Bourque classification scheme with information pertaining to the roles played by the head and modifier respectively. I will return to this matter in §7.3.

<table>
<thead>
<tr>
<th>head</th>
<th>elaboration sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSE (6)</td>
<td>Possessor (20), Purpose (20), Located (19), Location (2), Material (2), Part (1)</td>
</tr>
<tr>
<td>HEAD (5)</td>
<td>Whole (12), Possessor (6), Located (2), Result (2), Material (1)</td>
</tr>
<tr>
<td>PERSON (5)</td>
<td>Location (4), Possessum (1), Product (1), Purpose (1), Usee (1)</td>
</tr>
<tr>
<td>PLACE (5)</td>
<td>Located (13), Purpose (7), Cause (2), Location (1), Result (1)</td>
</tr>
<tr>
<td>STRING (5)</td>
<td>Possessum (3), Possessor (2), Part (1), Purpose (1), Whole (1)</td>
</tr>
<tr>
<td>THREAD (5)</td>
<td>Possessor (2), Purpose (2), User (2), Whole (2), Possessum (1)</td>
</tr>
<tr>
<td>TREE (5)</td>
<td>Part (7), Whole (2), Located (1), Location (1), User (1)</td>
</tr>
<tr>
<td>BONE (4)</td>
<td>Whole (86), Location (22), Located (1), Possessum (1)</td>
</tr>
<tr>
<td>FATHER (4)</td>
<td>Possessor (1), Possessum (1), Product (1), Whole (1)</td>
</tr>
<tr>
<td>HOLE (4)</td>
<td>Whole (52), Located (3), Purpose (2), Possessor (1)</td>
</tr>
</tbody>
</table>

*Table 50: Salience of elaboration sites (by role)*

If we reformulate Table 49 using the modifier role instead of the relation (Table 50), the nature of each e-site becomes clear. We see at once that HOUSE has e-sites for the possessor, the things located in it, its purpose and location, the material it is composed of, and its parts. In the actual data, these sites are elaborated by the components shown in Table 51. Given a little familiarity with the meaning list it is easy to see that TOILET and STABLE OR STALL utilize the Purpose e-site (of HOUSE), HOSPITAL and COOKHOUSE utilize the Located e-site, SPIDER WEB and BEEHIVE the Possessor e-site, etc.
Recall that these results are based on the raw, unnormalized data, which is why both ‘animal’ and ‘animals’ occur under Purpose, ‘bee’ and ‘bees’ under Possessor, etc. Nevertheless, they are sufficient to identify e-sites and to provide a measure of the degree to which individual heads are relational (in the sense of ‘relational nouns’); the latter could just be a simple count of the number of e-sites, as in the concept column of Table 50. The salience of individual e-sites, on the other hand, can be measured in two ways. One is the type frequency of the elaborating components, which would give a value of 3 for Possessor (or 2 after normalization) in the case of HOUSE, cf. the values in parentheses in the e-site column of Table 51. The other is the sum token frequencies of the elaborating components, in which case the value for Possessor would be 20 (10+9+1), cf. the values in the elaboration sites column of Table 50, which are the sums of the values in the elaborating component column of Table 51. 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.

### Table 51: Components that elaborate the concept HOUSE

<table>
<thead>
<tr>
<th>e-site</th>
<th>elaborating component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose (12)</td>
<td>excrement (3), urine (3), animal/cattle (2), bovine (2), horse (2), livestock (2), animal (1), animals (1), donkey (1), hand (1), horses (1), pig (1)</td>
</tr>
<tr>
<td>Located (11)</td>
<td>doctor (5), medicine (3), food (2), hospital (2), bath (1), disease (1), fire (1), illnesses (1), kitchen (1), pain (1), stove (1)</td>
</tr>
<tr>
<td>Possessor (3)</td>
<td>spider (10), bee (9), bees (1)</td>
</tr>
<tr>
<td>Location (2)</td>
<td>back (1), side (1)</td>
</tr>
<tr>
<td>Material (2)</td>
<td>leaf (1), straw (1)</td>
</tr>
<tr>
<td>Part (1)</td>
<td>incense (1)</td>
</tr>
</tbody>
</table>

### 7.2.4 Species-framing vs. attribute-framing

In the discussion of the relative salience of head and modifier in §7.2.2 we were mainly concerned with the contrast between the FM pattern exemplified by EYELID and the MF pattern exemplified by TRAIN. In this section I consider the FF 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 73, 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 (114), 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’.
(114)  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 atsinha [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 (114) be analysed as coordinate compounds, which are often said to have two heads (or, alternatively, no head)? 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 conceptualized 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 (115). 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.

(115)  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 (114) and (115) 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:

i) Are languages always consistent in this respect?
ii) 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 (116).

(116) a. **FEMALE X**: MARE, QUEEN, NIECE

   b. **MALE X**: COCK/ROOSTER, WIDOWER

   c. **YOUNG X**: FOAL OR COLT, KID, LAMB

Words that denote one or more of these meanings using the COOR relation\(^1\) are found in 80 of the languages in the sample. Of these, 12 (15%) are consistently species-framed, 54 (68%) are consistently attribute-framed, and 14 (17%) are mixed. Their geographical distribution is shown in Figure 74.

![Figure 74: Species- vs. attribute-framing](image)

All 14 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, in which the meaning LAMB can be expressed as either 羊羔 gāoyáng or 羊羔 yánghāo (where 羊 is glossed as sheep/goat and 羔 as lamb/kid).

Apart from this one exception, Mandarin is consistently species-framed. There is less consistency amongst mixed languages in how they frame meanings of the type “female X”: MARE and QUEEN are framed differently in Baa, 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 konu.mujē [king.woman] or mujē.konu [woman.king]; assuming both of

---

\(^1\) Other relations that are often used to express such meanings are TAX for young animals, POSS\(^2\) for NIECE, and TOP for WIDOWER. Recall also Caijia’s old woman from the sky (SRC), cf. page 195.
these are right-headed, the former would be classed as attribute-framed and the latter as species-framed. However, it is arguable that the system starts to break down when applied to human referents.

In five or the remaining six 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 74 are worthy of further investigation, but that is beyond the scope of the present study.

7.3 A model of associative relations

In Chapter 5 I extended Hatcher’s four-way system of semantic relations to a five-way system that also encompasses appositional and coordinate compounds. I also modified Bourque’s 25-way system (adding the relation CONTAINMENT) and then integrated the two into a single, two-level system. Hatcher herself noted how her four relations paired up into the “static contrast” between A≻B and A≺B, on the one hand, and the “dynamic contrast” between A→B and A←B on the other. In Table 28 on page 200 I cast this as a contrast between two higher-level relations: CONTAINMENT and CAUSATION (or source/goal).1

The Hatcher-Bourque classification is presented in the form of a hierarchy in Figure 75, and connected to the three associative relations of Aristotle, to which reference was made several times in Chapter 5.2 In this section I will extend this model by incorporating other facets of language that appear to be related, including metonymy and polysemy. In §5.1.3 I showed how many of the metonymies that Janda identifies in derivational morphology are based on the same semantic relations found in binominals. In §7.3.1 I extend the discussion in the light of Peirsman & Geeraerts (2006). I then extend the discussion to cognitive relations as discussed by Peter Koch and Andreas Blank (§7.3.2).

1 Here I would like to point out that these correspond to two basic image schemas (container-content and source-path-goal), which Langacker (2008: 32) characterizes as basic, “preconceptual” structures, “schematized patterns of activity abstracted from everyday bodily experience, especially pertaining to vision, space, motion, and force”. This underlines, to my mind, the basicness of these two relations.

2 The dashed lines represent connections that are lost due to the adoption of the hierarchical model.
Finally, in §7.3.3, 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. This will provide further insights and enable us to construct a well-articulated, overall model of associative relations. @@ Ref. Jackendoff’s comment (p. 154).

7.3.1 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 52, 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.

<table>
<thead>
<tr>
<th>Metonymical pattern</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SPATIAL PART &amp; WHOLE</td>
<td>Tony Blair is the Prime Minister of England</td>
</tr>
<tr>
<td>2. TEMPORAL PART &amp; WHOLE</td>
<td>Dutch and German morgen “morning” for “tomorrow”</td>
</tr>
<tr>
<td>3. LOCATION &amp; LOCATED</td>
<td>German Das ganze Haus wurde aus dem Schlaf geschreckt</td>
</tr>
<tr>
<td>4. ANTECEDENT &amp; CONSEQUENT</td>
<td>Greek phobos “flight” for “fear”</td>
</tr>
<tr>
<td>5. SUBEVENT &amp; COMPLEX EVENT</td>
<td>Mother is cooking potatoes</td>
</tr>
<tr>
<td>6. CHARACTERISTIC &amp; ENTITY</td>
<td>French beauté “beauty”</td>
</tr>
<tr>
<td>7. PRODUCER &amp; PRODUCT</td>
<td>I’m reading Shakespeare</td>
</tr>
<tr>
<td>8. CONTROLLER &amp; CONTROLLED</td>
<td>Schwarzkopf defeated Iraq</td>
</tr>
<tr>
<td>9. CONTAINER &amp; CONTAINED</td>
<td>French aimer la bouteille “love the bottle”</td>
</tr>
<tr>
<td>10. MATERIAL &amp; OBJECT</td>
<td>French carton “cardboard” for “cardboard box”</td>
</tr>
<tr>
<td>11. CAUSE &amp; EFFECT</td>
<td>unlock the prisons for “let the prisoners free”</td>
</tr>
<tr>
<td>12. LOCATION &amp; PRODUCT</td>
<td>China</td>
</tr>
<tr>
<td>13. POSSESSOR &amp; POSSESSED</td>
<td>the long straw starts for “the person with the long straw”</td>
</tr>
<tr>
<td>14. ACTION &amp; PARTICIPANT</td>
<td>to author a book</td>
</tr>
<tr>
<td>15. PARTICIPANT &amp; PARTICIPANT</td>
<td>the pen is mightier than the sword for “the writer is mightier than the soldier”</td>
</tr>
<tr>
<td>16. PIECE OF CLOTHING &amp; PERSON</td>
<td>French une vieille perruque “an old wig” for “an old person”</td>
</tr>
<tr>
<td>17. PIECE OF CLOTHING &amp; BODY PART</td>
<td>German Sohle “sole” for “sole”</td>
</tr>
<tr>
<td>18. SINGLE ENTITY &amp; COLLECTION</td>
<td>German Imme “swarm of bees” for “bee”</td>
</tr>
<tr>
<td>19. TIME &amp; ENTITY</td>
<td>French un mardi-gras “a Shrove Tuesday” for “a disguised man”</td>
</tr>
<tr>
<td>20. OBJECT &amp; QUANTITY</td>
<td>French un quart “a quarter” for “a tin of sardines in oil”</td>
</tr>
<tr>
<td>21. CENTRAL FACTOR &amp; INSTITUTION</td>
<td>the press</td>
</tr>
<tr>
<td>22. POTENTIAL &amp; ACTUAL</td>
<td>Can you see him?</td>
</tr>
<tr>
<td>23. HYPONYM &amp; HYPERONY</td>
<td>the pill for “the contraceptive pill”</td>
</tr>
</tbody>
</table>

Table 52: Peirsman & Geeraerts’ metonymical patterns
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 parallel with Bourque’s reversible relations is plain to see. Table 53 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 PROD2) correspond to the directed metonymies PRODUCT FOR PRODUCER and PRODUCER FOR PRODUCT, respectively.

<table>
<thead>
<tr>
<th>PRODUCTION</th>
<th>Type</th>
<th>Code</th>
<th>Template</th>
<th>Example</th>
<th>Modifier role</th>
<th>Head role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>PROD</td>
<td>an H that produces an M</td>
<td>song bird</td>
<td>product</td>
<td>producer</td>
<td></td>
</tr>
<tr>
<td>Reversed</td>
<td>PROD2</td>
<td>an H that an M produces</td>
<td>birdsong</td>
<td>producer</td>
<td>product</td>
<td></td>
</tr>
</tbody>
</table>

Table 53: Bourquifier template for PRODUCTION

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. My claim is that the same kinds of associative relation underpin both metonymy and both kinds of word-formation. Of course, this is partly a matter of terminology, as Janda points out. However, it seems to me that embracing the term ‘associative relation’, which is underutilized in linguistics, has more potential than extending either ‘metonymy’ or ‘semantic relation’, which both have long-established traditions and communities of research interest.

The correspondence between the metonymic pattern PRODUCER & PRODUCT and the semantic relation PRODUCTION is very transparent because the same verb, to produce, is used for both. This applies to six of P&G’s patterns, coded as (T) in Table 54, viz. CONTAINMENT, POSSESSION, LOCATION, COMPOSITION, CAUSE and PRODUCTION: none of these require further explication. Nor do MERONYMY, TIME and TAXONOMY, which are also transparent. These nine are all reversible and thus account for 18 of the 27 relations in the Bourque 2.0 classification.
Lest this evidence be deemed insufficient to support my claim, I will briefly discuss the 14 remaining patterns, for most of which there are correspondences of one kind of another with relations in Bourque’s system. The reason many of these are less transparent than the nine listed above is because the pattern either generalizes (G) or specializes (S) the corresponding relation.

Specialization accounts for eight patterns, four of which specialize the MERO NYMY 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 MERO NYMY covers all four of these patterns.
The remaining four cases in which the metonymical pattern is a specialization of an existing relation are:

- **#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 168 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 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 USE. 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.

Four of P&G’s metonymical patterns generalize one of Bourque’s relations:

- **#4 ANTECEDENT & CONSEQUENT** is a generalization of SOURCE and CAUSE and might find its place as SEQUENCE (with the roles antecedent and consequent) in a less fine-grained classification. The example cited by P&G (Greek *phobos* “flight” for “fear”) is surely an instance of 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 as LEADER & ORGANIZATION or REPRESENTATIVE & GROUP. There is no precise equivalent in Bourque’s system, but extending the latter with LEADERSHIP (leader, led) or REPRESENTATION (representative, organization) would not pose any problem.

- **#12 LOCATION & PRODUCT** is exemplified by *china* (in the sense of household tableware or other objects made of porcelain, which originally came from China).
The typology of binominal lexemes

This, like the old woman from the sky (cf. page 193), is clearly an example of the source relation.\(^1\) In one sense, it generalizes the latter, since location is more general than source; in another, it specializes it, since product is more specific than result.

- #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 USE; it is not even necessary to specialize the role user to wielder.

The two remaining patterns are exceptional, each in its own way.

#6 CHARACTERISTIC & ENTITY would not be found in a system geared towards binominals since a characteristic is a property, not a thing. Having said that, the very name of the pattern brings to mind Jakobsen’s Characterizing Feature, which Hatcher subsumed under her A\(\subseteq\)B, our \(MinH\). In order to incorporate this pattern into Bourque’s system, a new relation, CHARACTERIZATION (characteristic, entity), would be required. This would be a near relative (perhaps a subtype) of TOPIC, for which the Bourquifier example is *history book*.

The only remaining pattern, #22 POTENTIAL & ACTUAL, exemplified by *Can you see him?*, is special because, Panther & Thornburg (1999) state, it is non-referential. Because of this, incorporating it into our classification of semantic relations would require extending the system beyond its present scope, and I therefore 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 equivalents (many of them transparent) in Bourque’s classification. This provides strong evidence to support 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, provides further evidence of this.

---

\(^1\) 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).
7.3.2 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 in (e.g. Blank 1999, 2003; Koch 1999, 2001; Koch & Marzo 2007; Urban 2012). In the following I will employ the first of these since it was introduced earlier in the context of the motivational grid (§6.4.3).

According to Koch and Marzo, all such cognitive relations ultimately go back to the three associative relations of ‘contiguity’, ‘similarity’, and ‘contrast’, 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 55. Two of these, formal identity and syntagmatic contiguity, are based on form rather than meaning, and are ignored by Koch in the works cited. For the semantic axis of his motivational grid, Koch includes seven of the eight relations (shown in bold), noting that the eighth, antiphrastic contrast, is very rare (2001: 1159).

<table>
<thead>
<tr>
<th>Similarity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>metaphorical similarity</td>
<td><em>(foot of mountain)</em></td>
</tr>
<tr>
<td>co-taxonomic similarity</td>
<td><em>(Sp. <em>tigre</em> &gt; <em>jaguar</em> in South America)</em></td>
</tr>
<tr>
<td>taxonomic subordination</td>
<td><em>(hound ‘dog in general’ &gt; ‘hunting dog’)</em></td>
</tr>
<tr>
<td>taxonomic superordination</td>
<td><em>(inverse of preceding, no example)</em></td>
</tr>
<tr>
<td>conceptual identity</td>
<td><em>(tautology “or certain classes of word-formation”)</em></td>
</tr>
<tr>
<td>formal identity</td>
<td><em>(hotel &gt; motel)</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>co-taxonomic contrast</td>
<td><em>(bad &gt; good)</em></td>
</tr>
<tr>
<td><em>(antiphrastic contrast) (“more indirectly opposed concepts”, e.g. Fr. <em>pensionnaire</em> ’guest in a boarding house’ &gt; ‘convict’)</em></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contiguity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>conceptual contiguity</td>
<td><em>(“spatial, temporal, or logical connection between the concepts”)</em></td>
</tr>
<tr>
<td>syntagmatic contiguity</td>
<td>*(“the relation between the parts of complex lexical units”, e.g. <em>motor car</em> ~ <em>car)</em></td>
</tr>
</tbody>
</table>

Table 55: Blank’s ten cognitive relations
Blank is of the opinion that, of the three Aristotelean principles of remembering, “similarity shows by far the greatest diversity” (2003: 46), 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 75). Metaphorical similarity corresponds to the relation SIMILARITY; taxonomic subordination and superordination are the equivalent of TAXONOMY, and co-taxonomic similarity equates to COORDINATION, but conceptual identity does not have any direct correspondence. I have previously treated identity as a limiting case of similarity and subsumed it under SIMILARITY, but I have also suggested (page 177) that it could be regarded as a subtype of the latter, along with Similarity (proper) and Taxonomy. Implementing this idea in our classification would not only harmonise the two systems, it would also provide a very obvious home for reduplication as a word-formation process. The relation of (co-taxonomic) contrast has 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 in connection with Figure 28 (page 182), it does occur, for example in Chinese, and so it should be added in any future revision of Hatcher’s classification.

The upshot of this 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 their one remaining 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 polysemy, word-formation and metonymy. It seems likely that the system can also accommodate semantic roles (cf. Kittilä & Zúñiga 2014), but that is a topic for further research. I would like now to move beyond linguistics to the domain of knowledge representation, where I believe that further insights can be gleaned from Topic Maps.

---

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.”
7.3.3 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 is depicted in Figure 2 (reproduced here as Figure 76 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.

![Diagram of an association](image)

*Figure 76: The anatomy of an association (after Pepper 2010a)*

In Koch and Blank’s terms this is an example of conceptual contiguity (and also the engynomic relation ‘producer/product’). The statement *I am listening to Puccini* when I am in fact listening to *Tosca* would be classified by Peirsman and Geeraerts as a PRODUCER FOR PRODUCT metonymy. There is no binominal lexeme *puccini*
The typology of binominal lexemes

tosca, but the possessive construction Puccini’s Tosca is nicely characterized by Bourque’s PRODUCTION. 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 is generalized by produced-by, i.e PRODUCTION. Corresponding to these three levels are three pairs of roles: work/composer, creation/creator and product/ producer, the names of which could be reflected in three metonymies: WORK FOR COMPOSER, CREATION FOR CREATOR and PRODUCT FOR PRODUCER, all at different levels of granularity. Moreover, it can be seen from Figure 75 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: the one, very abstract and vague relation meaning “there is a connection between” entertained by Bauer in 1979 (cf. page 153).

This example also is also relevant to the naming of semantic relations. Following Bourque, I have mostly used deverbal 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 in the database 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 CONT2, POSS2, COMP2 and PROD2. 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 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 152 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 correspond exactly to the naming conventions used in the study of metonymy (PART FOR WHOLE, WHOLE FOR PART, etc.). In the final version of the Hatcher-Bourque classification, to be appear shortly, I adopt the same convention for directional relations. Adding this to the system will also cater for the need to specify role in order to identify elaboration sites that we experienced in §7.2.3.
### Table 56: Association types in the Italian Opera Topic Map

<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>Florida Tosca has the voice type soprano</td>
</tr>
<tr>
<td>killed-by</td>
<td>victim, perpetrator, cause of death</td>
<td>Tosca killed Scarpia by stabbing</td>
</tr>
<tr>
<td>libretto-by</td>
<td>opera, librettist</td>
<td><em>Tosca</em>’s 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 Florida Tosca</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 Sardou</td>
</tr>
</tbody>
</table>

The preceding discussion centred on a single association type, composed-by, and its correspondence to the various semantic relations, metonymies and cognitive relations discussed in the earlier part of this chapter. That association was taken 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 topic map grew out of the initial assertion that *Tosca* was composed by Puccini: first I added the other Puccini operas, and then other composers and their operas, and at that point I had to add a further association types, in order to capture who wrote the libretti, what they 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 56, along with their corresponding roles and an example of each. It is instructive to examine some of them more closely:
We observe that some association types are isomorphic with Bourque’s relations (located-in == LOCATION, part-of == MERONYMY); subtype-of is equivalent to TAXONOMY; libretto-by and written-by, like composed-by, are specializations of PRODUCTION. Moreover, to judge by their roles, based-on and revision-of are more fine-grained versions of SOURCE. Note also that associations in Topic Maps are not necessarily binary; killed-by is ternary and unfinished is unary.

We see 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 others. For my own part, I believe that the very explicit and carefully articulated model of Topic Maps has helped clarify my understanding of the relations, roles and labels in semantic relations, and made it possible to unify what, on the surface, are 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 standardized 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 one exception, however: we standardized two 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). Standardizing these two types of association results in more interoperability between Topic Maps systems “out of the box”. The question we then faced was whether to standardize more association types in order to gain 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 MERONYMY, 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 assume that these relations should be 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 by definition denote classes of referents and not individual referents.
On the basis of all preceding discussion I propose the integrated model shown in Figure 77. At the top of the hierarchy is Bauer’s (1979) one relation to rule them all, which conveys no information other than that there is some kind of connection. At the next level we find the three Aristotelean relations, similarity, contiguity and contrast. Similarity is divided into four subtypes, one of which, IDENTITY, is new. Given that TAXONOMY is reversible (which is why it alone has two distinct roles), these four cover Blank’s five subtypes of similarity; together with contiguity and contrast they are the seven cognitive relations used by Koch in his motivational grid. Contiguity is subdivided into CONTAINMENT and SEQUENCE (Hatcher’s static vs. dynamic duality of non-appositional compounds); the latter takes the place of CAUSATION (cf. Figure 75, page 277) since, on reflection prompted by Peirsman and Geeraerts’ metonymical pattern #4, this seems to be the best 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.\footnote{Is there not a certain sense in which Puccini was both the cause and source of Tosca?} So, too, are MERONYMY and COMPOSITION, and LOCATION and TEMPORALITY (a term I have co-opted to denote temporal location). The dotted boxes around USE, FUNCTION and PURPOSE indicate uncertainty about how they relate to one another.

Note finally, that there is just one role linked to TOPIC, FUNCTION and PURPOSE. This is not because they are unary relations (like unfinished). Rather it indicates that they are not reversible. Quite why these three, and only these three have that property is a conundrum that awaits further investigation.
7.4 Chapter summary

Any resemblance, apart from the title, between this chapter and Wittgenstein (1953), is entirely coincidental. Except, perhaps, for the slightly speculative nature of the final section. I started out on typological terra firma and focused on word order, consistency and basicness. I then examined the evidence (or lack of it) for the two-paths hypothesis, bringing salience into the discussion and then proposing a typology of species- vs. attribute-framing. Finally, as I think befits an exploratory study, I permitted myself the luxury of broadening the discussion beyond the narrow confines of binominal lexemes in order to synthesis the results of investigating those unstated (or underspecified) semantic relations that first intrigued me about noun-noun compounds. I plead guilty to dragging Topic Maps into the proceedings and apologize for forcing my linguist readers out of their comfort zone, but I think the excuse was both interesting and useful. On the other hand, bringing metonymy into the discussion was inevitable in the light of Janda’s (2011) paper, and I could not avoid Blank and Koch’s cognitive relations once I realized the relevance of the motivational grid approach. While on that topic, I would like to make the point that Koch & Marzo (2007:273) discern three scopes for applying their two-dimensional grid in typology. They are as follows:

(i) motivational profiles of particular languages;
(ii) identification of cross-linguistic motivational tendencies and idiosyncrasies with respect to the language facts (are there more or less formally transparent languages? Are there predominantly metaphorical languages? etc.);
(iii) universal/language type-specific motivational preferences and gaps with respect to combinations in the two-dimensional grid in Table 2.

I believe there is a similar potential for the motivational grids that I have used in this study, with the formal types and semantic relations pertinent to binominals instead of the formal and cognitive relations used by Koch & Marzo. Whereas I had sufficient data to create binominal profiles “fingerprints” for each language in §6.2.3, there was too little to justify creating motivational profiles. However, with more data for each language, it would be perfectly to do so, and then to compare profiles across languages.
8 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 not been able to show that “binominal lexeme” figures in any kind of cross-linguistic generalization, so I have not 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.

8.1 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 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 100 languages, how the data was collected and the initial analysis which revealed a harvest of over 3,500 binominal lexemes.

In Chapters 4 and 5 I developed the kind of classifications that are at the heart of typological research. The first, purely structural classification identified eight types of binominal, each of which can occur as a head-initial or head-final construction.
These were presented in a kind of two-dimensional semantic map in which degree and locus of marking were the primary parameters. Here I also discussed gradient phenomena and apparent gaps in the system. The second classification was devoted to semantic relations and built on the work of Anne Hatcher and Yves Bourque. I refined their high- and low-level systems and then, taking my cue from Pierre Arnaud, integrated these into a single model, which I then applied to my collection of binominal data.

Om Chapter 6 I analysed the resulting data set using a variety of statistical methods. Although no major findings could be reported, probably due to lack of data, there was no shortage of minor findings, many of which are listed below, and a number of interesting observations that formed the basis of the next chapter.

There, in Chapter 7, I carried out investigations into three major topics. I looked first at constituent order, consistency and basic order. I then investigated the two-paths hypothesis and the salience of heads, modifiers and elaboration sites. And finally, I refined and extended the system of semantic relations developed earlier into a general model of associative relations that can be applied in the fields of metonymy studies, lexical semantics and elsewhere as well as word-formation.

8.2 Contribution to science

My primary contribution to science of this work is, I believe, the identification of ‘binomial lexeme’ as a linguistic category and its development as a comparative concept. Most previous work in this field has focused on just a subset 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 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, if accepted as a valid category, constitutes strong evidence that the compartmental view of grammar is artificial and that the constructionist approach accords better with the facts.

Over and above this, I regard the following as being the primary contributions of this study to science:

- the collection and publication of binominal data from 100 languages
- the development of the formal (structural) classification of binominals
- the refinement and unification of Hatcher’s and Bourque’s classifications of semantic relations
- the synthesis of semantic, metonymical and conceptual relations into a holistic system of associative relations

In addition, I believe I can lay claim to the following significant, albeit more minor contributions:

- operationalizing the notion of (relative) salience
- recognizing a typology of species/attribute framing
- identifying the problem of structural types as cross-linguistic comparanda

Furthermore, I have been able to report a number of findings that have a certain significance:
• the prevalence of compounding (**cmp**) as a (binominal) word-formation type, but also the non-universality of noun-noun compounding
• the 2:1 prevalence of right-headedness, both overall and among compounds
• the scale of marking for semantic types (Person < Animal, Location < Concrete) and the patterning of Body part with the latter rather than Person
• the curious case of MIDDAY

8.3 Further research

The present study opens up a plethora of interesting pathways for further research, concerning questions such as the following:

• When there is competition between structural types, what factors influence the choice? Is it the semantic relation (as in Nizaa and Mapudungun), the semantic type (ref. alienability), or something else?
• Does more data and more refined analysis (including use of synsets) confirm the major patterns of salience between head and modifier?
• What is the distribution of species- vs. attribute-framing types in a larger sample of languages?
• Can the scale of marking be confirmed and further refined using more data?
• Can an expanded, genetically and areally balanced sample of languages lead to interesting typological implications?
• Can the approach used here to onomasiological type 3 naming units be applied to other types, e.g. OT1 (NVN constructions, synthetic compounds, etc.)?

8.4 Envoie

In writing this work I have had the honour to stand on the shoulders of giants. My greatest sources of inspiration have been Ron Langacker, Bill Croft and Martin Haspelmath. Without Masja Koptjevskaja-Tamm’s thorough-going 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, much harder. Without Anna Granville Hatcher’s distillation of Jespersen, the ground-breaking 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 to all those who contributed data for this project.
Their names are listed in the Acknowledgements, in Appendix B (*Sources*), and in the Index of names.
References


References


The typology of binominal lexemes


Dixon, R. M. W. 1982. Where have All the Adjectives Gone? And other essays in semantics and syntax. Berlin: de Gruyter.
The typology of binominal lexemes


The typology of binominal lexemes


The typology of binominal lexemes


The typology of binominal lexemes


The typology of binominal lexemes


The typology of binominal lexemes


Pepper, Steve. 2007. As We REALLY May Think: Memex, Topic Maps, and subject-centric computing. Presented at the AToMS 2007, Kyoto.


The typology of binominal lexemes


Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. Linguistic Typology 2(3).


The typology of binominal lexemes


Theil Endresen, Rolf. nd. Field notes on Nizaa: Word lists and accompanying notes.


The typology of binominal lexemes


Appendix A.

Appendices

@@ Insert single level TOC for appendices here
A. Languages

This appendix lists all the languages in the sample,\(^1\) ordered by ISO code\(^2\) 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>AKK</td>
<td>Akkadian</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>AMH</td>
<td>Amharic</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>AMR</td>
<td>Amarakaeri</td>
<td>Harakmbut</td>
<td>Harakmbut</td>
<td>South America</td>
</tr>
<tr>
<td>AOI</td>
<td>Anindilyakwa</td>
<td>Gunwinyguan</td>
<td>Gunwinyguan</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>AQC</td>
<td>Archi</td>
<td>Nakh-Daghestanian</td>
<td>Lezgic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>ARN</td>
<td>Mapudungun</td>
<td>Araucanian</td>
<td>Araucanian</td>
<td>South America</td>
</tr>
<tr>
<td>ASM</td>
<td>Assamese</td>
<td>Indo-European</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>BAM</td>
<td>Bambara</td>
<td>Atlantic-Congo</td>
<td>Mande</td>
<td>Africa</td>
</tr>
<tr>
<td>BOJ</td>
<td>Bandial</td>
<td>Atlantic-Congo</td>
<td>North-Central Atlantic</td>
<td>Africa</td>
</tr>
<tr>
<td>BVA</td>
<td>Barain</td>
<td>Afro-Asiatic</td>
<td>Chadic</td>
<td>Africa</td>
</tr>
<tr>
<td>CAI</td>
<td>Caijia</td>
<td>Sino-Tibetan</td>
<td>Macro-Bai</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>CAR</td>
<td>Galibi Carib</td>
<td>Cariban</td>
<td>Guianan</td>
<td>South America</td>
</tr>
<tr>
<td>CES</td>
<td>Czech</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>CJP</td>
<td>Cabécar</td>
<td>Chibchan</td>
<td>Chibchan</td>
<td>South America</td>
</tr>
<tr>
<td>CLI</td>
<td>Chakali</td>
<td>Atlantic-Congo</td>
<td>Gur</td>
<td>Africa</td>
</tr>
<tr>
<td>CMN</td>
<td>Mandarin Chinese</td>
<td>Sino-Tibetan</td>
<td>Sinitic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>CRS</td>
<td>Seychelles Creole</td>
<td>Pidgins &amp; Creoles</td>
<td>French-based</td>
<td>Pidgins/Creoles</td>
</tr>
<tr>
<td>CGW</td>
<td>Ceq Wong</td>
<td>Austro-Asiatic</td>
<td>Aslian</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>CYM</td>
<td>Welsh</td>
<td>Indo-European</td>
<td>Celtic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>DEU</td>
<td>German</td>
<td>Indo-European</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>DSB</td>
<td>Lower Sorbian</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>ELL</td>
<td>Greek</td>
<td>Indo-European</td>
<td>Greek</td>
<td>Eurasia</td>
</tr>
<tr>
<td>ENG</td>
<td>English</td>
<td>Indo-European</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>EST</td>
<td>Estonian</td>
<td>Uralic</td>
<td>Finnic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>ESU</td>
<td>Central Yupik</td>
<td>Eskimo-Aleut</td>
<td>Yupik</td>
<td>North America</td>
</tr>
<tr>
<td>EUS</td>
<td>Basque</td>
<td>Basque</td>
<td>Basque</td>
<td>Eurasia</td>
</tr>
</tbody>
</table>

\(^1\) As of this draft data from a few languages remain to be added to the database.

\(^2\) 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 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>FIN</td>
<td>Finnish</td>
<td>Uralic</td>
<td>Finnic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>FRA</td>
<td>French</td>
<td>Indo-European</td>
<td>Romance</td>
<td>Eurasia</td>
</tr>
<tr>
<td>GLE</td>
<td>Irish</td>
<td>Indo-European</td>
<td>Celtic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>GNT</td>
<td>Wartha Thuntai</td>
<td>Morehead-Wasur</td>
<td>Morehead-Wasur</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>GOH</td>
<td>Old High German</td>
<td>Indo-European</td>
<td>Germanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>GUE</td>
<td>Gurinji</td>
<td>Pama-Nyungan</td>
<td>Desert Nyungic</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>GUN</td>
<td>Mbyá Guarani</td>
<td>Tupian</td>
<td>Tupi-Guarani</td>
<td>North America</td>
</tr>
<tr>
<td>GWD</td>
<td>Gawwada</td>
<td>Afro-Asiatic</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>HAU</td>
<td>Hausa</td>
<td>Afro-Asiatic</td>
<td>Chadic</td>
<td>Africa</td>
</tr>
<tr>
<td>HAW</td>
<td>Hawaiian</td>
<td>Austronesian</td>
<td>Oceanic</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>HEB</td>
<td>Hebrew</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>HIN</td>
<td>Hindi</td>
<td>Indo-European</td>
<td>Indo-Aryan</td>
<td>Eurasia</td>
</tr>
<tr>
<td>HRV</td>
<td>Croatian</td>
<td>Indo-European</td>
<td>Slavic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>HUN</td>
<td>Hungarian</td>
<td>Uralic</td>
<td>Hungarian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>HUU</td>
<td>Murui Huitoto</td>
<td>Huitotoan</td>
<td>Huitotoan</td>
<td>South America</td>
</tr>
<tr>
<td>IND</td>
<td>Indonesian</td>
<td>Austronesian</td>
<td>Malayo-Sumbawan</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>IRK</td>
<td>Iraqw</td>
<td>Afro-Asiatic</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>ITA</td>
<td>Italian</td>
<td>Indo-European</td>
<td>Romance</td>
<td>Eurasia</td>
</tr>
<tr>
<td>JPN</td>
<td>Japanese</td>
<td>Japonic</td>
<td>Japanese</td>
<td>Eurasia</td>
</tr>
<tr>
<td>JUP</td>
<td>Hupdë</td>
<td>Nadahup</td>
<td>Nadahup</td>
<td>South America</td>
</tr>
<tr>
<td>KAP</td>
<td>Bezhta</td>
<td>Nakh-Daghestanian</td>
<td>Avar-Andic-Tzevic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>KDX</td>
<td>Kam</td>
<td>Atlantic-Congo</td>
<td>Adamawa-Ubangi</td>
<td>Africa</td>
</tr>
<tr>
<td>KEK</td>
<td>Q'eqchi'</td>
<td>Mayan</td>
<td>Core Mayan</td>
<td>North America</td>
</tr>
<tr>
<td>KET</td>
<td>Ket</td>
<td>Yeniseian</td>
<td>Yeniseian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>KGV</td>
<td>Kalamang</td>
<td>West Bomberai</td>
<td>West Bomberai</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>KNC</td>
<td>Kanuri</td>
<td>Nilo-Saharan</td>
<td>Saharan</td>
<td>Africa</td>
</tr>
<tr>
<td>KOR</td>
<td>Korean</td>
<td>Koreanic</td>
<td>Koreanic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>KPZ</td>
<td>Kupsabiny</td>
<td>Nilo-Saharan</td>
<td>Nilotic</td>
<td>Africa</td>
</tr>
<tr>
<td>KTB</td>
<td>Kambaata</td>
<td>Afro-Asiatic</td>
<td>Cushitic</td>
<td>Africa</td>
</tr>
<tr>
<td>KWB</td>
<td>Baa</td>
<td>Atlantic-Congo</td>
<td>Gur</td>
<td>Africa</td>
</tr>
<tr>
<td>LAV</td>
<td>Latvian</td>
<td>Indo-European</td>
<td>Baltic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>LT</td>
<td>Lithuanian</td>
<td>Indo-European</td>
<td>Baltic</td>
<td>Eurasia</td>
</tr>
<tr>
<td>LSR</td>
<td>Srenge</td>
<td>Nuclear Torricelli</td>
<td>Nuclear Torricelli</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>MAL</td>
<td>Malayalam</td>
<td>Dravidian</td>
<td>Dravidian</td>
<td>Eurasia</td>
</tr>
<tr>
<td>MLT</td>
<td>Maltese</td>
<td>Afro-Asiatic</td>
<td>Semitic</td>
<td>Africa</td>
</tr>
<tr>
<td>MRJ</td>
<td>Western Mari</td>
<td>Uralic</td>
<td>Mari</td>
<td>Eurasia</td>
</tr>
<tr>
<td>MWW</td>
<td>Hmong Daw</td>
<td>Hmong-Mien</td>
<td>Hmong-Mien</td>
<td>Oceania/SE Asia</td>
</tr>
<tr>
<td>MYK</td>
<td>Mamara Senoufo</td>
<td>Atlantic-Congo</td>
<td>Kwa</td>
<td>Africa</td>
</tr>
<tr>
<td>MZH</td>
<td>Wichi</td>
<td>Matacoan</td>
<td>Matacoan</td>
<td>South America</td>
</tr>
<tr>
<td>NAV</td>
<td>Navajo</td>
<td>Athabaskan-Eyak-Tlingit</td>
<td>Athabaskan-Eyak-Tlingit</td>
<td>North America</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>NFL</td>
<td>Aíwoo</td>
<td>Austronesian</td>
<td>Oceanic</td>
<td>Oceania/SE Asia</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>ORQ</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>PNG/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>Tarifit</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>SJD</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>TBA</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>TWI</td>
<td>Tuwari</td>
<td>Walio</td>
<td>Walio</td>
<td>PNG/Australia</td>
</tr>
<tr>
<td>TZO</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>PNG/Australia</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>

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 325
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

[@@ Maintained manually in order to include Zotero field information.]
The typology of binominal lexemes

<table>
<thead>
<tr>
<th>Code</th>
<th>Glottocode</th>
<th>Author(s)</th>
</tr>
</thead>
</table>
| Bezhta | KAP | Bernard Comrie and Madzhid Khalilov  
W: (Comrie & Khalilov 2009)  
G: (Kibrik & Testelets 2004) |
| Cabécar | JP | Guillermo González  
G: (Quesada 2007; Verhoeven 2012) |
| Caijia | CAI | Shanshan Lü |
| Ceq Wong | CWG | Nicole Kruspe  
W: (Kruspe 2009) |
| Central Yupik | ESU | D: (Jacobson 2013)  
G: (Jacobson 1995; Miyaoka 2012) |
| Chakali | CLI | Jonathan Brindle  
D: (Brindle 2017) |
| Croatian | HRV | Višnja Cicin-Sain  
G: ##  
D: ## |
| Czech | CES | Pavlina Pešková  
G: ## |
| Datooga | TCC | Alice Mitchell |
| Dutch | NLD | W: (van der Sijs 2009)  
G: (Booij 2002; Don 2009) |
| English | ENG | W: (Grant 2009) |
| Estonian | EST | Reili Argus  
G: ## |
| Finnish | FIN | Matti Miestamo  
G: (Sulkala & Karjalainen 1992; Karlsson 2013) |
| French | FRA | Yves Bourque  
D: (Bourque 2016)  
G: (Bourque 2014) |
| Galibi Carib | CAR | W: (Renault-Lescure 2009)  
G: (Courtz 2008) |
| Gawwada | GW | Mauro Tosco  
W: (Tosco 2009)  
G: (Tosco 2007) |
| German | DEU | Anne Krause  
G: (Whittle et al. 2011) |
| Greek | ELL | Katerina Fragkopoulou  
G: (Holton et al. 2012; Ralli 2013) |
| Gurinji | GUE | W: (McConvell 2009) |
| Hausa | HAU | W: (Awagana & Wolff 2009)  
D: (Newman 1990; Newman 2007)  
| Hawaiian | HA | W: (Jones 2009)  
D: (Pukui & Elbert 1971)  
G: (Elbert & Pukui 1979) |

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.
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Author</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hebrew</td>
<td>HEB</td>
<td>Julian Lysvik</td>
<td>G: (Levi 1976; Glinert 1989; Borer 2009)</td>
</tr>
<tr>
<td>HIndi</td>
<td>HIN</td>
<td>Claus Peter Zoller</td>
<td>D: (McGregor 1993; Verma &amp; Sahai 2003)</td>
</tr>
<tr>
<td>Hmong Daw</td>
<td>MW</td>
<td>Martha Ratliff</td>
<td>W: (Ratliff 2009)</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td></td>
<td>G: (Mottin 1978)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>HUN</td>
<td>Zsofia Schön</td>
<td>G: ##</td>
</tr>
<tr>
<td>Hupé</td>
<td>JUP</td>
<td>Patience Epps</td>
<td>W: (Epps 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Epps 2008)</td>
</tr>
<tr>
<td>Indonesian</td>
<td>IND</td>
<td>W: (Tadmor 2009)</td>
<td>D: (Echols &amp; Shadily 1975)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Sneddon 1996)</td>
</tr>
<tr>
<td>Iraqw</td>
<td>IRK</td>
<td>Maarten Mous</td>
<td>W: (Mous 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D: (Mous, Qorro &amp; Kiessling 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Nordbustad 1988; Mous 1993)</td>
</tr>
<tr>
<td>Irish</td>
<td>GLE</td>
<td>Cormac MacAindir</td>
<td>G: ##</td>
</tr>
<tr>
<td>Italian</td>
<td>ITA</td>
<td>Francesca Masini</td>
<td>G: ##</td>
</tr>
<tr>
<td>Kalamang</td>
<td>KGV</td>
<td>Eline Visser</td>
<td>G: (Visser 2016)</td>
</tr>
<tr>
<td>Kam</td>
<td>KDX</td>
<td>Jakob Lesage</td>
<td></td>
</tr>
<tr>
<td>Kambaata</td>
<td>KTB</td>
<td>Yvonne Treis and Deginet Wotango</td>
<td></td>
</tr>
<tr>
<td>Ket</td>
<td>KET</td>
<td>Edward Vajda and Andrey Nefedov</td>
<td>W: (Vajda &amp; Nefedov 2009)</td>
</tr>
<tr>
<td></td>
<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>
<td>W: (Rießler 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Kert 1971)</td>
</tr>
<tr>
<td>Korean</td>
<td>KOR</td>
<td>Petter Mehlum</td>
<td>G: (Sohn 1994; Yeon &amp; Brown 2011)</td>
</tr>
<tr>
<td>Kupsabiny</td>
<td>KPZ</td>
<td>Kazuhiro Kawachi</td>
<td></td>
</tr>
<tr>
<td>Latvian</td>
<td>LAV</td>
<td>Agnē Navickaitē-Klišauskienē</td>
<td>G: (Praulins 2012; Navickaitē-Klišauskienē 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D: (Turkīna 1982; Velži 1996)</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>LIT</td>
<td>Agnē Navickaitē-Klišauskienē</td>
<td>G: ##</td>
</tr>
</tbody>
</table>
The typology of binominal lexemes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>W:</th>
<th>G:</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>Lower</td>
<td>Sorbian</td>
<td>DSB</td>
<td>Hauke Bartels</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>W: (Bartels 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Schuster-Šewc 1996)</td>
</tr>
<tr>
<td>56</td>
<td>Malagasy</td>
<td>PLT</td>
<td></td>
<td>W: (Adelaar 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Fugier 1999; Ferrand 2010[1903]; Parker 2014[1883])</td>
</tr>
<tr>
<td>57</td>
<td>Malayalam</td>
<td>MAL</td>
<td>Arathi Raghunathan</td>
<td>G: ##</td>
</tr>
<tr>
<td>58</td>
<td>Maltese</td>
<td>MLT</td>
<td>Benjamin Saade</td>
<td>G: (Borg &amp; Azzopardi-Alexander 1997; Fabri 2009)</td>
</tr>
<tr>
<td>59</td>
<td>Mamara Senoufo</td>
<td>MYK</td>
<td>Sekou Coulibaly</td>
<td>G: ##</td>
</tr>
<tr>
<td>60</td>
<td>Manange</td>
<td>NM</td>
<td>W: (Hildebrandt 2009)</td>
<td>G: (Hildebrandt 2004)</td>
</tr>
<tr>
<td>61</td>
<td>Mandarin Chinese</td>
<td>CMN</td>
<td>Shuqiong Wu</td>
<td>W: (Wiebusch 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Yip &amp; Rimmington 2004; Ceccagno &amp; Basciano 2009)</td>
</tr>
<tr>
<td>62</td>
<td>Mapudungun</td>
<td>ARN</td>
<td>Antonio Chihuacura, Belén Villena, Fernando Zúñiga</td>
<td>W: (Golluscio, Fraguas &amp; Mellico 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Smeets 2008; Baker &amp; Fasola 2009; Zúñiga 2014)</td>
</tr>
<tr>
<td>63</td>
<td>Mbyá Guarani</td>
<td>GUN</td>
<td>Bob Dooley</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Murui Huitoto</td>
<td>HUU</td>
<td>Kasia Wojtylak</td>
<td>G: ##</td>
</tr>
<tr>
<td>65</td>
<td>Navajo</td>
<td>NAV</td>
<td>Steve Pepper</td>
<td>D: (Young &amp; Morgan 1980)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Young &amp; Morgan 1980)</td>
</tr>
<tr>
<td>66</td>
<td>Nepali</td>
<td>NPI</td>
<td>Krishna Prasad Parajuli</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Norwegian</td>
<td>NOR</td>
<td>Steve Pepper</td>
<td>G: (Faarlund, Lie &amp; Vannebo 1997)</td>
</tr>
<tr>
<td>68</td>
<td>Old High German</td>
<td>GOH</td>
<td>W: (Schuhmann 2009)</td>
<td>G: ##</td>
</tr>
<tr>
<td>69</td>
<td>Oroqen</td>
<td>ORH</td>
<td>W: (Fengxiang Li &amp; Whaley 2009)</td>
<td>G: (Nedjalkov 1997)</td>
</tr>
<tr>
<td>70</td>
<td>Polish</td>
<td>POL</td>
<td>Pavel Urbanik</td>
<td>G: ##</td>
</tr>
<tr>
<td>71</td>
<td>Puyuma</td>
<td>PYU</td>
<td>Stacy Fang Ching Teng</td>
<td>D: (Cauquelin 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Teng 2008)</td>
</tr>
<tr>
<td>73</td>
<td>Querétaro Otomi</td>
<td>OTE</td>
<td>W: (Bakker &amp; Hekking 2009)</td>
<td>G: ##</td>
</tr>
<tr>
<td>74</td>
<td>Romanian</td>
<td>RON</td>
<td>W: (Schulte 2009)</td>
<td>G: (Mallinson 1986; Grossmann 2012; Halvorsen 2012; Giurgea &amp; Dobrovie-Sorin 2013)</td>
</tr>
<tr>
<td>75</td>
<td>Russian</td>
<td>RUS</td>
<td>Svetlana Sokolova</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Saramaccan</td>
<td>SRM</td>
<td>Jeff Good</td>
<td>W: (Good 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G: (Veenstra 2006; McWhorter &amp; Good 2012)</td>
</tr>
<tr>
<td>Language</td>
<td>Code</td>
<td>Source</td>
<td>W:</td>
<td>G:</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>--------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Selice Romani</td>
<td>RMC</td>
<td>W: (Elšík 2009)</td>
<td>G: (Hancock 1990)</td>
<td></td>
</tr>
<tr>
<td>Seri</td>
<td>SEI</td>
<td>Steve Marlett</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidamo</td>
<td>SID</td>
<td>Kjell-Magne Yri</td>
<td>G: ##</td>
<td></td>
</tr>
<tr>
<td>Slovak</td>
<td>SLK</td>
<td>Pavol Štekauer</td>
<td>G: ##</td>
<td></td>
</tr>
<tr>
<td>Somali</td>
<td>SOM</td>
<td>Nina Hagen Kaldhol</td>
<td>G: (Saeed 1993; 1999)</td>
<td></td>
</tr>
<tr>
<td>Srenge</td>
<td>LSR</td>
<td>Lea Brown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swahili</td>
<td>SWH</td>
<td>W: (Schadeberg 2009)</td>
<td>G: (Van de Velde 2013)</td>
<td></td>
</tr>
<tr>
<td>Tagalog</td>
<td>TGL</td>
<td>Mark Donohue</td>
<td>D: (English 1977; 1986)</td>
<td>G: (Schachter &amp; Otanes 1972)</td>
</tr>
<tr>
<td>Thai</td>
<td>THA</td>
<td>Titima Suthiwan</td>
<td>W: (Suthiwan 2009)</td>
<td>G: (Singnoi 2000; Smyth 2002; Iwaseki &amp; Ingkaphirom 2005)</td>
</tr>
<tr>
<td>Ticuna</td>
<td>TCA</td>
<td>Denis Bertet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trinitario</td>
<td>TRN</td>
<td>Françoise Rose</td>
<td>G: (Rose 2015)</td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>TUR</td>
<td>Ash Gürer</td>
<td>D: (Akdikmen 2006)</td>
<td>G: (Kornfilt 1997; Göksel &amp; Kerslake 2005; Göksel 2009)</td>
</tr>
<tr>
<td>Tuwari</td>
<td>TW</td>
<td>Sylvain Loiseau</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walman</td>
<td>VSN</td>
<td>Lea Brown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wartha Thunthai</td>
<td>GNT</td>
<td>Kyla Quinn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wawa</td>
<td>WW</td>
<td>Marieke Martin</td>
<td>G: ##</td>
<td></td>
</tr>
<tr>
<td>Welsh</td>
<td>CYM</td>
<td>Gwen Awbery</td>
<td>D: (Hawke, Fychan &amp; Roberts 2014)</td>
<td>G: (Thorne 1993; King 2003; Awbery 2004; Awbery 2014)</td>
</tr>
<tr>
<td>Western Mari</td>
<td>MRJ</td>
<td>Polina Pleshak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Farsi</td>
<td>PES</td>
<td>Cormac MacAindir</td>
<td>G: ##</td>
<td></td>
</tr>
</tbody>
</table>
### The typology of binominal lexemes

<table>
<thead>
<tr>
<th></th>
<th>Language</th>
<th>Code</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Wichi</td>
<td>MZH</td>
<td>W: (Vidal &amp; Nercesian 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Nercesian 2014)</td>
</tr>
<tr>
<td>101</td>
<td>Wik-Mungkan</td>
<td>WIM</td>
<td>David Ogarby</td>
</tr>
<tr>
<td>102</td>
<td>Wolof</td>
<td>WOL</td>
<td>Sokhna Bao</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: ###</td>
</tr>
<tr>
<td>103</td>
<td>Yakut</td>
<td>SAH</td>
<td>Innokentij Novgorodov</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D: (Nedjalkov 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Vinokurova 1998; Stachowski &amp; Menz 1998)</td>
</tr>
<tr>
<td>104</td>
<td>Yaqui</td>
<td>YAQ</td>
<td>W: (Fernández 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Fernández &amp; Buitimea Valenzuela 2009)</td>
</tr>
<tr>
<td>105</td>
<td>Zinacantán Tzotzil</td>
<td>TZO</td>
<td>W: (Brown 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G: (Cowan 1969; Haviland 1980)</td>
</tr>
</tbody>
</table>
This appendix lists all 100 meanings in alphabetical order, together with the size of
the subvocabulary, the number of analysable forms and the number of binominals.
These data are plotted in descending order of subvocabulary size in Figure 41 on
page 210. Information regarding semantic fields and semantic types can be found
in Table 17 on page 77.

<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>104</td>
<td>54</td>
<td>43</td>
<td>FIREPLACE</td>
<td>91</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td>ARCTIC LIGHTS</td>
<td>50</td>
<td>42</td>
<td>33</td>
<td>FISHERMAN</td>
<td>94</td>
<td>79</td>
<td>27</td>
</tr>
<tr>
<td>BACKPACK</td>
<td>57</td>
<td>34</td>
<td>28</td>
<td>FISHING LINE</td>
<td>90</td>
<td>57</td>
<td>33</td>
</tr>
<tr>
<td>BEE</td>
<td>120</td>
<td>30</td>
<td>21</td>
<td>FLAME</td>
<td>96</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>BEEHIVE</td>
<td>99</td>
<td>63</td>
<td>52</td>
<td>FLEA MARKET</td>
<td>40</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>BEESWAX</td>
<td>78</td>
<td>53</td>
<td>48</td>
<td>FOAL OR COLT</td>
<td>85</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>BICYCLE</td>
<td>95</td>
<td>30</td>
<td>13</td>
<td>FOOTPRINT</td>
<td>98</td>
<td>58</td>
<td>44</td>
</tr>
<tr>
<td>BICYCLE PUMP</td>
<td>54</td>
<td>48</td>
<td>40</td>
<td>GIRL</td>
<td>108</td>
<td>54</td>
<td>35</td>
</tr>
<tr>
<td>BLACKSMITH</td>
<td>87</td>
<td>49</td>
<td>23</td>
<td>GLOVE</td>
<td>79</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>BOY</td>
<td>109</td>
<td>48</td>
<td>31</td>
<td>GOLD RING</td>
<td>62</td>
<td>60</td>
<td>52</td>
</tr>
<tr>
<td>BRACELET</td>
<td>98</td>
<td>46</td>
<td>27</td>
<td>HAND BRAKE</td>
<td>49</td>
<td>44</td>
<td>38</td>
</tr>
<tr>
<td>BREAKFAST</td>
<td>91</td>
<td>57</td>
<td>30</td>
<td>HANDBAG</td>
<td>72</td>
<td>45</td>
<td>34</td>
</tr>
<tr>
<td>CARPENTER</td>
<td>85</td>
<td>45</td>
<td>24</td>
<td>HANDKERCHIEF OR RAG</td>
<td>108</td>
<td>46</td>
<td>30</td>
</tr>
<tr>
<td>CHIEFTAIN</td>
<td>113</td>
<td>50</td>
<td>25</td>
<td>HERDSMAN</td>
<td>86</td>
<td>55</td>
<td>14</td>
</tr>
<tr>
<td>COCK/ROOSTER</td>
<td>99</td>
<td>40</td>
<td>32</td>
<td>HOSPITAL</td>
<td>94</td>
<td>51</td>
<td>21</td>
</tr>
<tr>
<td>COLLARBONE</td>
<td>79</td>
<td>54</td>
<td>44</td>
<td>HOST</td>
<td>79</td>
<td>47</td>
<td>26</td>
</tr>
<tr>
<td>COOKHOUSE</td>
<td>77</td>
<td>48</td>
<td>17</td>
<td>KEYWORD</td>
<td>44</td>
<td>40</td>
<td>28</td>
</tr>
<tr>
<td>DAIRY COW</td>
<td>54</td>
<td>43</td>
<td>30</td>
<td>KID</td>
<td>88</td>
<td>60</td>
<td>41</td>
</tr>
<tr>
<td>DINNER</td>
<td>79</td>
<td>52</td>
<td>39</td>
<td>LAMB</td>
<td>94</td>
<td>44</td>
<td>31</td>
</tr>
<tr>
<td>DOORPOST</td>
<td>80</td>
<td>47</td>
<td>39</td>
<td>LICENSE PLATE</td>
<td>66</td>
<td>46</td>
<td>33</td>
</tr>
<tr>
<td>EARLOBE</td>
<td>79</td>
<td>66</td>
<td>57</td>
<td>LUNCH</td>
<td>89</td>
<td>50</td>
<td>37</td>
</tr>
<tr>
<td>EARRING</td>
<td>99</td>
<td>61</td>
<td>38</td>
<td>MAGIC</td>
<td>96</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>EARWAX</td>
<td>91</td>
<td>69</td>
<td>64</td>
<td>MAIL BOX</td>
<td>50</td>
<td>43</td>
<td>41</td>
</tr>
<tr>
<td>EYEBROW</td>
<td>103</td>
<td>52</td>
<td>38</td>
<td>MARE</td>
<td>82</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>EYELASH</td>
<td>93</td>
<td>60</td>
<td>50</td>
<td>MARRIED WOMAN</td>
<td>81</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>EYELID</td>
<td>92</td>
<td>70</td>
<td>64</td>
<td>MIDDAY</td>
<td>102</td>
<td>75</td>
<td>46</td>
</tr>
<tr>
<td>FARMER</td>
<td>94</td>
<td>59</td>
<td>30</td>
<td>MILKY WAY</td>
<td>63</td>
<td>54</td>
<td>45</td>
</tr>
</tbody>
</table>
The typology 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>MOTHER-IN-LAW</td>
<td>98</td>
<td>54</td>
<td>29</td>
<td>STONE BRIDGE</td>
<td>55</td>
<td>52</td>
<td>49</td>
</tr>
<tr>
<td>NATIVE COUNTRY</td>
<td>82</td>
<td>66</td>
<td>35</td>
<td>SUGAR CANE</td>
<td>86</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>NEIGHBOUR</td>
<td>91</td>
<td>44</td>
<td>14</td>
<td>SUNDAY</td>
<td>63</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>NIECE</td>
<td>101</td>
<td>58</td>
<td>47</td>
<td>SUPPER</td>
<td>81</td>
<td>52</td>
<td>39</td>
</tr>
<tr>
<td>NIPPLE OR TEAT</td>
<td>103</td>
<td>53</td>
<td>49</td>
<td>TEAR</td>
<td>96</td>
<td>37</td>
<td>31</td>
</tr>
<tr>
<td>NOSTRIL</td>
<td>97</td>
<td>74</td>
<td>70</td>
<td>THATCH</td>
<td>70</td>
<td>36</td>
<td>28</td>
</tr>
<tr>
<td>PADDLE WHEEL</td>
<td>33</td>
<td>29</td>
<td>25</td>
<td>THUMB</td>
<td>100</td>
<td>68</td>
<td>30</td>
</tr>
<tr>
<td>PALM OF HAND</td>
<td>79</td>
<td>59</td>
<td>55</td>
<td>TOE</td>
<td>103</td>
<td>70</td>
<td>60</td>
</tr>
<tr>
<td>POSTCARD</td>
<td>63</td>
<td>37</td>
<td>30</td>
<td>TOILET</td>
<td>108</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>POTTER</td>
<td>80</td>
<td>61</td>
<td>32</td>
<td>TOILET PAPER</td>
<td>52</td>
<td>43</td>
<td>35</td>
</tr>
<tr>
<td>QUEEN</td>
<td>79</td>
<td>46</td>
<td>41</td>
<td>TOOL</td>
<td>94</td>
<td>43</td>
<td>17</td>
</tr>
<tr>
<td>RAILWAY</td>
<td>64</td>
<td>54</td>
<td>52</td>
<td>TOOLBOX</td>
<td>51</td>
<td>46</td>
<td>38</td>
</tr>
<tr>
<td>RAINBOW</td>
<td>101</td>
<td>52</td>
<td>38</td>
<td>TOOTHBRUSH</td>
<td>65</td>
<td>56</td>
<td>37</td>
</tr>
<tr>
<td>RIB</td>
<td>103</td>
<td>39</td>
<td>26</td>
<td>TRAIN</td>
<td>82</td>
<td>29</td>
<td>18</td>
</tr>
<tr>
<td>SHOEMAKER</td>
<td>74</td>
<td>53</td>
<td>20</td>
<td>TREE TRUNK</td>
<td>98</td>
<td>57</td>
<td>50</td>
</tr>
<tr>
<td>SHORE</td>
<td>116</td>
<td>51</td>
<td>38</td>
<td>VEIN OR ARTERY</td>
<td>108</td>
<td>39</td>
<td>26</td>
</tr>
<tr>
<td>SHOULDERBLADE</td>
<td>96</td>
<td>54</td>
<td>42</td>
<td>VINE</td>
<td>81</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>SKULL</td>
<td>102</td>
<td>60</td>
<td>50</td>
<td>WATER PUMP</td>
<td>56</td>
<td>44</td>
<td>40</td>
</tr>
<tr>
<td>SORCERER OR WITCH</td>
<td>114</td>
<td>61</td>
<td>35</td>
<td>WEDNESDAY</td>
<td>61</td>
<td>41</td>
<td>20</td>
</tr>
<tr>
<td>SPECTACLES/GLASSES</td>
<td>89</td>
<td>44</td>
<td>21</td>
<td>WIDOWER</td>
<td>87</td>
<td>49</td>
<td>18</td>
</tr>
<tr>
<td>SPIDER WEB</td>
<td>91</td>
<td>69</td>
<td>60</td>
<td>WINDMILL</td>
<td>45</td>
<td>41</td>
<td>34</td>
</tr>
<tr>
<td>SPINE</td>
<td>102</td>
<td>69</td>
<td>54</td>
<td>WRIST</td>
<td>98</td>
<td>61</td>
<td>51</td>
</tr>
<tr>
<td>STABLE OR STALL</td>
<td>90</td>
<td>29</td>
<td>26</td>
<td>YOLK</td>
<td>85</td>
<td>60</td>
<td>38</td>
</tr>
</tbody>
</table>
This appendix lists every binominal construction found in the data by the language in which it occurs, along with an example and the number of occurrences. Languages are ordered by area, family and genus. The language name is followed by its ISO code and three numbers that indicate the total number of words (W) for that language in the database (in the case of Tarifit, 96), the total number of binominals (18B), and the number of constructions across which they are distributed (4C). Following the arrow ► are the number of instances of each type of binominal, (cmp:1, prp:14 and gen:3). Within each language constructions are listed in descending order of frequency. 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).

(Afro-Asiatic) Berber

**Tarifit (RIF): 96W / 18B / 4C ► prp:14 gen:3 cmp:1**

13 prp Head PREP Mod:CON: *abriḍ n mašina* [road of train:CON] RAILWAY
3 gen Head Mod:CON: *tahr’us’ t umoẓẓun* [ring? ear:CON] EARLOBE
1 cmp Head Mod: *frinu manu* [brake hand] HAND BRAKE
1 prp Head.3SG PREP Mod:CON: *dyāğgi.s n uma* [daughter.3SG of brother:CON] NIECE

(Afro-Asiatic) Chadic

**Barain (BVA): 57W / 21B / 7C ► prp:11 con:7 dbl:2 cmp:1**

7 prp Head REL PREP Mod: *gera ge ŋ gargar* [home REL:3SG:M PREP spider] SPIDER WEB
7 con Mod Head.POSS: *sinja guma.geti* [nose hole.POSS:3SG:F] NOSTRIL
2 prp Head PREP2 Mod: *looli ta lutta* [dirt PREP ear] EARWAX
2 dbl Head.POSS PREP Mod: *bug.eti ŋ mooro* [mouth.POSS:3SG:F PREP river] SHORE
1 cmp Head Mod: *peesī mee* [horse woman] MARE
The typology of binominal lexemes

1 prp Head PREP Mod.POSS: golmo ŋ bal.ti [house PREP 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

Hausa (HAU): 100W / 43B / 4C ► con:40 der:2 cmp:1
40 con Head.LK Mod: kàfà. r hànčii [orifice.LK nose] NOSTRIL
1 der Base.ABST: mài.tàa [sorcerer.ABST] MAGIC
1 der Base.F: sàráu.niyå [king.F] QUEEN
1 cmp Mod Head: ràánàa tsákàa [sun mid] MIDDAY

(Afro-Asiatic) Cushitic

Gawwada (GWD): 74W / 17B / 4C ► cmp:11 der:6
8 cmp Head Mod.ASS: hawdatto sikk.ète [craftsman pots.ASS] POTTER
3 cmp Mod Head: paso pako [lake mouth] SHORE

Iraqw (IRK): 69W / 22B / 5C ► con:15 der:4 cmp:3
15 con Head.CON Mod: foxàr duunga’ [hole:of nose] NOSTRIL
3 cmp Mod Head: afa tlawi [mouth lake] SHORE
2 der Base.AGT: tlee’.usmo [clay.AGT] POTTER
1 der Base.F: wawi’ita.o’o [king.SG:F] QUEEN
1 der Base.M: koonk.amo [hen.SG:M] COCK/ROOSTER

Kambaata (KTB): 100W / 34B / 4C ► gen:31 der:2 cmp:1
29 gen Mod.GEN Head: baabur.i maar.a [train.M:GEN line.M:ACC] RAILWAY
2 der Base.AGT: hoga’.aan.ch.u [plough.AGT.SGLT.M:ACC] FARMER
2 cmp Mod Head: magan.zeeb.uta [God.???.F:ACC] RAINBOW

Sidamo (SID): 98W / 29B / 6C ► gen:20 der:7 cmp:1 adj:1
20 gen Mod.GEN Head: baabur.u doogo [train.GEN road] RAILWAY
3 der Base.F: moot.itte [king.DER_F] QUEEN
3 der Base.SUF: boos.aallo [fireplace.DER] COOKHOUSE
1 der Base.M: lukk.iffo [hen.SGLT_M] COCK/ROOSTER
1 cmp Head Mod: taalo barra [middle day] MIDDAY
1 adj Mod.ADJ Head: k’ar.aame k’ale [sharp_blade.ADJZ wheel] PADDLE WHEEL

Somali (SOM): 95W / 19B / 8C ► dbl:7 gen:5 cmp:4 der:3
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
Appendix D. Constructions (summary)

3  cmp  Head Mod: xuub.caaro [sloughed_snakeskin.spider] SPIDER WEB
2  der  Base.F: bogor.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  cmp  Mod Head: beer.a.ley [farm.PL.owner.PL] FARMER
1  dbl  Mod.DEF Head.POSS: geed.ka gun.tiisa [tree.DEF base.POSS] TREE TRUNK

(Afro-Asiatic) Semitic

Akkadian (AKK): 61W / 19B / 2C ► dbl:19
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

Amharic (AMH): 100W / 47B / 3C ► gen:37 cmp:10
37  gen  GEN.Mod Head: ye.baburi ḥāddī [GEN.train way] RAILWAY
9  cmp  Mod Head: shererīti diri [spider web] SPIDER WEB
1  cmp  Head Mod: ikule k’eni [half day] MIDDAY

Hebrew (HEB): 100W / 43B / 6C ► con:34 der:4 dbl:3 cmp:1 prp:1
34  con  Head:CON Mod: mesil.at barzel [track.STC iron] RAILWAY
4  der  Base.SUF: malk.a [king.F] QUEEN
2  dbl  Head:CON DEF.Mod: sevil he.xalav [path:STC DEF.milk] MILKY WAY
1  cmp  Head Mod: tik gav [bag back] BACKPACK
1  prp  Head PRP.Mod: kešet be.’anan [arc in:DEF:sky] RAINBOW
1  dbl  Head:CON Mod:ADJZ: or.ot arkt.iyim [light.STC arctic:ADJZ] ARCTIC LIGHTS

Maltese (MLT): 96W / 32B / 8C ► prp:25 cmp:4 dbl:2 der:1
20  prp  Head PREP:DEF-Mod: qolla tan-nahal [hive of:DEF-bees] BEEHIVE
3  prp  Head PREP Mod: saba' ta' sieq [finger/toe of foot] TOE
2  prp  Head DEF.Mod: xatt il-bahar [shore DEF-sea] SHORE
2  cmp  Head Mod: gaws.alla [bow.god] RAINBOW
1  der  Base.AGT: skarp.an [shoe.AGT] SHOEMAKER
1  cmp  DEF.Head Mod: il-kelma muftieh [DEF-word key] KEYWORD
1  prp  DEF.Head PREP Mod: It-Triq ta' Sant'Anna [DEF-way of Saint_Anne] MILKY WAY

(Atlantic-Congo) Adamawa-Ubangi

Kam (KDX): 69W / 26B / 2C ► cmp:26
18  cmp  Head Mod: ᴹwé mārī [mouth river] SHORE
The typology of binominal lexemes

8 cmp Head.LE.Mod: à.gùm.i.cmbirèg [NOM.hole.LE.nose] NOSTRIL

(Atlantic-Congo) Bantoid

Swahili (SWH): 104W / 52B / 4C ► prp:39 cmp:9 cls:4
39 prp Head CON Mod: tundu la pua [hole CON nose] NOSTRIL
8 cmp Head Mod: kĩna.upepo [mill.wind] WINDMILL
4 cls CL.Base: fumo [CL9:spear] CHIEFTAIN
1 cmp Mod Head: jua kati [sun middle] MIDDAY

Wawa (WWW): 27W / 13B / 3C ► gen:9 cmp:4
9 gen Head Mod.POSS.ASS: náábò tóni [fufu ear.LOC.ASS] EARWAX
3 cmp Head Mod: mūn nugwéé [child woman] GIRL
1 cmp Mod Head: sāngā tėŋgə [sun middle] MIDDAY

(Atlantic-Congo) Gur

Baa (KWB): 91W / 53B / 6C ► cmp:34 gen:19
34 cmp Head Mod: krà.kiśa [road.train] RAILWAY
11 gen Head Mod.POSS: ná.jivi [mouth.milk:1PL:POSS] NIPPLE OR TEAT

Chakali (CLI): 51W / 25B / 5C ► cmp:20 der:5
17 cmp Mod Head: mu.boa [nose.hole] NOSTRIL
3 cmp Head Mod: bi.nháàŋ [child.woman] GIRL
1 der Base.AGT: nāār.o.koo [shoe.make] SHOEMAKER

(Atlantic-Congo) Kwa

Mamara Senoufo (MYK): 81W / 27B / 2C ► cmp:25 der:2
25 cmp Mod Head: tumɔnɔn.ko.li [iron.road.CL] RAILWAY
2 der Base.AGT: kulu.foɔ [village.AGT] CHIEFTAIN

(Atlantic-Congo) Mande

Bambara (BAM): 81W / 39B / 6C ► cmp:28 gen:10 der:1
25 cmp Mod.Head: tere.n.sira [train.road] RAILWAY
Appendix D. Constructions (summary)

10 gen Mod.SUF.Head: kúlo.la.nége [ear.LOC.iron] EARRING
1 der Base.SUF.DIM: miusó.mani [woman.DIM] GIRL
1 cmp Head Mod: cèñéliyà [man widow] WIDOWER
1 cmp Head.Mod.SUF: bóløtend.kin.ba [finger.head.AUG] THUMB
1 cmp Mod GEN Head: Ala ká mürü [God GEN knife] RAINBOW

(Atlantic-Congo) North-Central Atlantic

Bandial (BQJ): 72W / 16B / 3C ► cmp:9 prp:4 cls:3
9 cmp Head Mod: ka.liba e.mít [CL.knife CL.god/sky] RAINBOW
4 prp Head CON Mod: (sinan) sa këbujom [rice CON morning] BREAKFAST
3 cls CL.Base: ji.jamen [CL.goat] KID

Wolof (WOL): 64W / 17B / 2C ► prp:14 cmp:3
14 prp Head CON Mod: pax.u bakkan [hole.CON nose] NOSTRIL
3 cmp Head Mod: jaaro-nopp [jewelry-ear] EARRING

(Nilo-Saharan) Nilotic

Datooga (TCC): 29W / 1B / 1C ► cmp:1
1 cmp Head.SGLT.SG Mod.SG: fúw.éán.dà bádàyda [thread.SGLT.SG back.SG] SPINE

Kupsabiny (KPZ): 90W / 43B / 5C ► con:34 prp:8 gen:1
34 con Head.POSS Mod: areet.aap kariitaap_maata [road.POSS train] RAILWAY
5 prp Head PREP Mod: tiinkeet nyēpo yoomeet [machine for air] WINDMILL
2 prp Head REL Mod: loleet nyēē cēēk āsēnē [bag REL:SG back] BACKPACK
1 prp Mod REL Head: makeyoonteet wuloo mayayi [egg where yellow] YOLK
1 gen Mod.POSS Head: rwaanteet.aap taraca [stone.POSS bridge] STONE BRIDGE

(Nilo-Saharan) Saharan

Kanuri (KNC): 83W / 39B / 7C ► gen:23 der:12 cmp:3 adj:1
23 gen Head Mod.GEN: súwúlí kónzà.bè [opening nose.GEN] NOSTRIL
6 der Base.AGT: një.má [pot.OWN] POTTER
5 der Base.LOC: kànnù.rám [fire.LOC] COOKHOUSE
2 cmp Mod Head: fâr kùrwúdi [horse female] MARE
1 der Base.NMLZ: kômàwèn.mtí [honey.NMLZ] BEESWAX
1 cmp Head Mod: kúli.kómàwèn [insect.honey] BEE
1 adj Head Mod.ADJZ: kátú nyićjà.à [woman marriage.ADJZ] MARRIED WOMAN
Eurasia

(Altaic) Tungusic

Oroqen (ORH): 60W / 24B / 7C ► der:8 dbl:7 cmp:6 gen:3
7  dbl  Mod.GEN Head.POSS: өңөктөүү дүүн [nose.GEN inside.3SG:POSS] NOSTRIL
5  der  Base.DIM: una:dʒi.kan [young_woman.DIM] GIRL
5  cmp  Mod Head: gək mərm [female horse] MARE
3  gen  Mod.GEN Head: tima:ni bilga [morning.GEN food] BREAKFAST
2  der  Base.REC: amo.roq [stool.REC] TOILET
1  cmp  Head Mod: inəŋ dulin [day middle] MIDDAY

(Altaic) Turkic

Turkish (TUR): 100W / 65B / 5C ► con:47 cmp:10 der:8
47  con  Mod Head.3SG: demir.yol.u [iron.road.3SG] RAILWAY
10  cmp  Mod Head: taş köprü [stone bridge] STONE BRIDGE
6   der  Base.AGT: çömlek.çış [pot.AGT] POTTER
1   der  Base.F: kral.içe [king.F] QUEEN
1   der  Base.SUF: göz.lük [eye.SUF] SPECTACLES/GLASSES

Yakut (SAH): 109W / 39B / 7C ► con:21 cmp:9 der:7 gen:1 adj:1
21  con  Mod Head.3SG: murun χаяyah.a [nose hole.3SG] NOSTRIL
9   cmp  Mod Head: timir suol [iron path] RAILWAY
4   der  Base.AGT: küös.çüt [pot.AGT] POTTER
2   der  Base.NMLZ: χos.poχ [room.NMLZ] COOKHOUSE
1   der  Base.PROP: ap.taχ [magic.PROP] SORCERER OR WITCH
1   adj  Mod.ADJZ Head: tualet.nay kuma:γι [toilet.ADJZ paper] TOILET PAPER
1   gen  Mod.PROP Head: sa:χar.da:χ trostnik [sugar.PROP cane] SUGAR CANE

Basque

Basque (EUS): 102W / 55B / 6C ► cmp:42 der:7 adj:6
41  cmp  Mod Head: tren.bide [train.way] RAILWAY
6   adj  Mod.ADJZ Head: harri.zko zubi [stone.ADJZ bridge] STONE BRIDGE
4   der  Base:AGT: buzti.nari [mud.AGT] POTTER
2   der  Base.LOC: erla.tegi [bee.LOC] BEEHIVE
1   der  Base:SUF: gos.ari [morning.SUF] BREAKFAST
1   cmp  Head Mod: er.puru [finger.head] THUMB
Dravidian

Malayalam (MAL): 100W / 45B / 5C ▶ cmp:44 gen:1

32 cmp Mod.Head: cilanti.vala [spider.web] SPIDER WEB
9 cmp Mod.LE.Head: cevi.k.kāja [ear.LE.wax] EARWAX
2 cmp Mod Head: peñ.kuṭṭi [female.child] GIRL
1 cmp Head.Mod: kamaṇ.kāl [part.foot] ANKLE
1 gen Mod.GEN.Head: āṭṭ.in.kuṭṭi [goat.GEN.child] LAMB

(Indo-European) Baltic

Latvian (LAV): 102W / 59B / 7C ▶ gen:30 cmp:18 der:11

30 gen Mod.GEN Head: zirnek.la tīkls [spider.GEN net] SPIDER WEB
17 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
1 der Base.INS: acenes [eye:INS] SPECTACLES/GLASSES
1 cmp Head Mod: pus.diena [half.day] MIDDAY

Lithuanian (LIT): 102W / 47B / 10C ▶ der:22 gen:10 cmp:8 adj:7

12 der Base.NMLZ: puodž.ius [pot.NMLZ] POTTER
10 gen Mod.GEN Head: aus.y vaškas [ear.GEN wax] EARWAX
7 adj Mod.ADZ.Head: gelež.in.kelis [iron.ADZ.way] RAILWAY
5 cmp Mod.LE.Head: vor.a.tinklis [spider.LE.web] SPIDER WEB
3 der Base.DIM: caille.ín [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

(Indo-European) Celtic

Irish (GLE): 100W / 40B / 6C ▶ gen:27 der:10 cmp:3

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
The typology of binominal lexemes


41 cmp Head Mod: cwyr clust [wax ear] EARWAX
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.o [path milk.ADJZ] MILKY WAY
1 der Base.F: brenin.es [king.F] QUEEN
1 der Base.NMLZ: pen.aeth [chief.NMLZ] CHIEFTAIN
1 der Base.SUF: cwmwd.o [district.SUF] NEIGHBOUR
1 prp Head PREP Mod: papur lle chwech [paper for toilet] TOILET PAPER

(Indo-European) Germanic

Dutch (NLD): 105W / 54B / 7C ► cmp:48 der:3 adj:2 gen:1

41 cmp Mod Head: spoor.weg [track.way] RAILWAY
7 cmp 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

English (ENG): 105W / 44B / 4C ► cmp:39 der:3 adj:2

38 cmp Mod Head: <V></> rail.way RAILWAY
3 der Base.AGT: <V></> pot.AGT POTTER
2 adj Mod.ADJZ Head: <V></> milk.ADJZ way MILKY WAY
1 cmp Head Mod: <V></> middle.day MIDDAY

German (DEU): 131W / 73B / 8C ► cmp:62 der:10 adj:1

49 cmp Mod Head: eisen.bahn [iron.way] RAILWAY
11 cmp Mod.LE.Head: nase.n.loch [nose.LE.hole] NOSTRIL
4 der Base.DIM: maid.chen [maid.DIM] GIRL
3 der Base.AGT: töpf.er [pot.AGT] POTTER
2 der Base.NMLZ: tisch.ler [table.NMLZ] CARPENTER
2 cmp Head Mod: mit.tag [middle.day] MIDDAY
1 der Base.F: konig.in [king.F] QUEEN
1 adj Mod.ADJZ Head: pol.ar.lichter [pole.ADJZ.lights] ARCTIC LIGHTS

Norwegian (NOR): 100W / 57B / 3C ► cmp:57

51 cmp Mod Head: jern.bane [iron.way] RAILWAY
5 cmp Mod.LE.Head: øye.n.bryn [eye.LE.brow] EYEBROW
1 cmp Head Mod: mid.dag [middle.day] MIDDAY
Old High German (GOH): 63W / 23B / 5C ► cmp:18 der:5

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

(Indo-European) Greek

Greek (ELL): 100W / 30B / 7C ► gen:12 cmp:10 der:5 adj:3

9 gen Head Mod:Gen: istrates [web spider.Gen] SPIDER WEB
9 cmp Mod:LE:Head: siðir.o.ðromos [iron.LE:road] RAILWAY
4 der Base:SUf: proi.ino [morning.SUF] BREAKFAST
3 gen Head DEF Mod:Gen: keri tu afti.ou [wax DEF ear.Gen] EARWAX
3 adj Mod:ADJZ Head: uran.io tokso [sky.ADJZ bow] RAINBOW
1 der Base:DIM: katsik.aki [goat:DIM] KID
1 cmp Head Mod: leksi-kliði [word-key] KEYWORD

(Indo-European) Indo-Aryan

Assamese (ASM): 87W / 34B / 3C ► cmp:17 gen:17

17 gen Mod:Gen: Head: nak.ɔr phuta [nose.Gen hole] NOSTRIL
16 cmp Mod Head: ram.dhenu [Lord_Rama.bow] RAINBOW
1 cmp Head Mod: ga.gos [body.tree] TREE TRUNK

Hindi (HIN): 95W / 30B / 4C ► cmp:19 prp:10 der:1

19 cmp Mod Head: karṇa.mal [ear.dirt] EARWAX
7 prp Mod:GEN: Head: makrī kā jālā [spider GEN web] SPIDER WEB
3 prp Mod GEN: Head: ghore kā baccā [horse GEN child] FOAL OR COLT
1 der Base:AGT: lohā.r [iron.AGT] BLACKSMITH

Nepali (NPI): 66W / 11B / 3C ► gen:6 cmp:5

6 gen Mod:GEN: Head: mākurā.ko jālo [spider GEN web] SPIDER WEB
4 cmp Mod:Head: kāne.guji [ear.?] EARWAX
1 cmp Mod:Head: mātri.bhumi [mother.land] NATIVE COUNTRY

Selice Romani (RMC): 88W / 8B / 4C ► gen:4 der:4

4 gen Mod:GEN: Head: kan.en.ger.e khula [ear.OBL:GEN.PL shit.PL] EARWAX
1 der Base:ABST: čohan.ipe [witch/sorcerer.ABST] MAGIC
1 der Base:F: kirá.ckiña [king.F] QUEEN
(Indo-European) Romance

French (FRA): 102W / 49B / 10C ► prp:34 der:9 adj:4 cmp:2

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.ADIZ: voie lactée [way milk:ADIZ] MILKY WAY
  2  prp  Head PREP DEF Mod: lobe de l'oreille [lobe of DEF:ear] EARLOBE
  1  der  Base.F: sorcièr.e [sorcerer.F] SORCERER OR WITCH
  1  prp  Head PREP3 Mod: arc-en-ciel [bow-in-sky] RAINBOW
  1  cmp  Mod Head: omo.plate [shoulder.plate] SHOULDERBLADE

Italian (ITA): 112W / 38B / 8C ► prp:20 der:9 adj:5 cmp:4

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.ADIZ: via latt.ea [way milk.ADIZ] MILKY WAY
  3  der  Base.DIM: capr.etto [goat.NMLZ] KID
  3  cmp  Head Mod: capo.tribù [head.clan] CHIEFTAIN
  1  der  Base.AGT: pesca.tore [fish.er] FISHERMAN
  1  cmp  Mod Head: ferro.via [iron.way] RAILWAY

Romanian (RON): 135W / 41B / 11C ► prp:18 der:14 adj:6 dbl:2 cmp:1

18  prp  Head PREP Mod: pânză de păianjen [cloth of spider] SPIDER WEB
  7  der  Base.AGT: o(a)lă.ar [pot.AGT] POTTER
  6  adj  Head Mod.ADIZ: calea fer.ata [way iron.ADIZ] RAILWAY
  2  der  Base.F: nepo(a)t.ă [grandson/nephew.F] NIECE
  2  dbl  Head.DET Mod.DET.GEN: încheietura mâin.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: brăţ.ară [arm.NMLZ] BRACELET
  1  der  Base.SUF: gălben.us [yellow.SUF] YOLK
  1  cmp  Head Mod: cuvînt cheie [word key] KEYWORD

(Indo-European) Slavic

Czech (CES): 98W / 44B / 7C ► adj:22 der:19 prp:2 cmp:1

21  adj  Mod.ADIZ Head: nos.ní dírka [nose.ADIZ hole] NOSTRIL
Appendix D. Constructions (summary)

Lower Sorbian (DSB): 118W / 43B / 11C ► adj:26 der:11 gen:3 prp:2 cmp:1

Polish (POL): 104W / 54B / 11C ► der:26 adj:19 prp:4 gen:3 cmp:2

Russian (RUS): 103W / 43B / 10C ► adj:20 der:16 gen:4 cmp:2 prp:1
The typology of binominal lexemes

Slovak (SLK): 100W / 57B / 10C ► der:33 adj:21 prp:2 gen:1

1 prp Head PRP Mod: jaščik dlja instrument.ov [box for tools.GEN] TOOLBOX
1 cmp Mod.LE.Head: golen.o.stop [shank.LE.foot] ANKLE

Japanese (JPN): 130W / 81B / 3C ► cmp:67 gen:12 der:2

68 cmp Mod Head: tetsu.dō [iron.road] RAILWAY
12 gen Mod.GEN Head: kumo.no.su [spider.GEN.web] SPIDER WEB
1 cmp Head Mod: mi.ki [body.tree] TREE TRUNK

Korean (KOR): 144W / 63B / 2C ► cmp:59 gen:4

59 cmp Mod Head: chel.kil [iron.road] RAILWAY
4 gen Mod.GEN Head: kho.s.kwumeng [nose.GEN.hole] NOSTRIL

(Nakh-Daghestanian) Avar-Andic-Tsezic

Bezhta (KAP): 93W / 35B / 8C ► gen:26 cmp:4 adj:3 der:2

25 gen Mod.GEN Head: kil.o.s hino [iron.OBL.GEN way] RAILWAY
3 cmp Mod Head: šayt’an.mašina [devil.machine] BICYCLE
2 adj Mod.ATTR Head: xidalaƛ.ko čüryän [snot.ATTR scarf] HANDKERCHIEF OR RAG
1 der Base.NMLZ: ł ‘erec’.madi [(onom).NMLZ] WATER PUMP
1 der Base.PAUC: c’uddo c’emuc’ [red egg] YOLK
1 cmp Head Mod: c’uddo c’emuc’ [red egg] YOLK
1 gen Mod.ABL Head: ƛoba.ƛ’a.s bešiyoli [midday.SUP.ABL meal] LUNCH
1 adj Mod.ADJZ Head: nucodaq t’ot’ [honey:ADJZ fly] BEE
Appendix D. Constructions (summary)

(Nakh-Daghestanian) Lezgic

**Archi (AQC):** 66W / 14B / 3C ► gen:9 cmp:3 adj:2

9. **gen** Mod.GEN Head: *muč.li.n klan* [nose.OBL.GEN hole] NOSTRIL.
3. **cmp** Mod Head: *tenne lo* [female child] GIRL
2. **adj** Mod.ADJZ Head: *ak:onni.t:ut kummul* [in_the_morning.ADJZ food]
   BREAKFAST

Nuclear Torricelli

**Srenge (LSR):** 40W / 20B / 2C ► cmp:20

19. **cmp** Mod Head: *mupə tengə* [nose hole] NOSTRIL.
1. **cmp** Mod Head: REDUP: *ala talmtalmə* [leg/foot branch:RED] TOE

**Walman (VAN):** 48W / 16B / 1C ► cmp:16

16. **cmp** Mod Head: *lō muüi* [hole nose] NOSTRIL

(Uralic) Finnic

**Estonian (EST):** 104W / 65B / 6C ► gen:32 cmp:26 der:6 adj:1

32. **gen** Mod.GEN Head: *ämbliku.vörk* [spider:GEN.web] SPIDER WEB
25. **cmp** Mod Head: *raud.tee* [iron:NOM.way] RAILWAY
4. **der** Base.SUF: *pea.lik* [head.DER] CHIEFTAIN
2. **der** Base.GEN.SUF: *kuninga.nna* [king:GEN.DER] QUEEN
1. **cmp** Head Mod: *kesk.pääev* [middle.day] MIDDAY
1. **adj** Mod.NMLZ Head: *mesi.las.vaha* [honey.NMLZ.wax] BEESWAX

**Finnish (FIN):** 100W / 57B / 6C ► cmp:43 gen:11 der:3

41. **cmp** Mod.Head: *rauta.tie* [iron.road] RAILWAY
11. **gen** Mod.GEN Head: *hämähäki.n.verkko* [spider.GEN.net] SPIDER WEB
2. **der** Base.NMLZ: *isä.ntä* [father.NMLZ] HOST
1. **der** Base.F: *kuninga.tar* [king.F] QUEEN
1. **cmp** Head Mod: *keski.päivä* [middle.day] MIDDAY
1. **cmp** Head.Mod: *keski.viikko* [middle.week] WEDNESDAY

(Uralic) Hungarian

**Hungarian (HUN):** 105W / 57B / 6C ► cmp:42 adj:9 der:6

42. **cmp** Mod Head: *vasút* [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. **adj** Mod.PROP Head: *kép.es.lap* [picture.PROP.card] POSTCARD
2. **der** Base.PROP: *fazék.as* [pot.PROP] POTTER
The typology of binominal lexemes

1  der  Base.F: király.né [king.F] QUEEN

(Uralic) Mari

Western Mari (MRJ): 93W / 43B / 3C ► cmp:42 gen:1
41  cmp  Mod Head: kőrtni kornô [iron road] RAILWAY
1  cmp  Mod.GEN Head: mônô.n sarô [egg.GEN yellow] YOLK
1  gen  Mod.LAT Head: šand.eš pumaga [toilet.LAT paper] TOILET PAPER

(Uralic) Saami

Kildin Sami (SJD): 87W / 35B / 4C ► cmp:29 gen:3 adj:3
26  cmp  Mod Head: rūv't čuekas [iron road] RAILWAY
3  cmp  Mod.Head.DIM: koass.a.all'k.a [goat.DIM.son.DIM] KID
3  adj  Mod.ATTR.Head: ūnc.es’.pierrk [morning.ATTR.meal] BREAKFAST
3  gen  Mod.GEN.Head: oaz.e sājjm [spider.GEN:SG net] SPIDER WEB

Yeniseian

Ket (KET): 70W / 29B / 3C ► cmp:16 gen:12 adj:1
16  cmp  Mod Head: ekŋ.qoˀt [thunder.path] RAINBOW
12  gen  Mod.GEN Head: oln.d qūk [nose.GEN hole] NOSTRIL
1  adj  Mod.ADJZ Head: sol.tu tǝq.ol [gold.ADJZ finger.covering] GOLD RING

Oceania/SE Asia

(Austro-Asiatic) Aslian

Ceq Wong (CWG): 38W / 20B / 1C ► cmp:20
20  cmp  Head Mod: daraŋ mɔ̌h [hole nose] NOSTRIL

(Austro-Asiatic) Vietic

Vietnamese (VIE): 85W / 49B / 2C ► cmp:49
41  cmp  Head Mod: dürông sāt [road iron] RAILWAY
8  cmp  Mod Head: nũ hoàng [female emperor] QUEEN

(Austronesian) Formosan

Puyuma (PYU): 54W / 3B / 3C ► cmp:2 dbl:1
1  dbl  3SG.Head DEF Mod: tu-bira? kana țañiļa [3SG-earlobe DEF:OBL ear] EARLOBE
1  cmp  Mod.PERF.Head: <V>d<in>apal-an</> [foot<PERF>-LOC] FOOTPRINT
1  cmp  TMP.Base.LOC: ka-ľauk-an [TMP-lunch-LOC] MIDDAY
(Austronesian) Greater Barito

**Malagasy (PLT): 89W / 58B / 6C ► cmp:31 con:18 der:8 prp:1**

31 **cmp** Head Mod: *vàva.òrona* [mouth.nose] NOSTRIL
18 **con** Head.PER.Mod: *lala.m.by* [road.PER.iron] RAILWAY
6 **der** AGT.Base: *mpan.dràfitra* [AGT.carpentry] CARPENTER
1 **prp** Head SOC.Mod: *vehivàvà 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

(Austronesian) Greater Central Philippine


19 **prp** Head LK Mod: *daa.ng-bakal* [road.LK-iron] RAILWAY
6 **cmp** 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: *ginto.ng singsing* [gold.LK ring] GOLD RING
1 **der** CIRC.Base.CIRC: *ka.sangkap.an* [CIRC.belongings.CIRC] TOOL
1 **cmp** Mod Head: *kamay preno* [hand brake] HAND BRAKE

(Austronesian) Malayo-Sumbawan

**Indonesian (IND): 108W / 49B / 3C ► cmp:47 der:2**

47 **cmp** Head Mod: *jalan keréta api* [road carriage fire] RAILWAY
1 **der** AGT.Base: *peng.sihir* [AGT.sorcery] SORCERER OR WITCH
1 **der** LOC.Base.CIRC: *per.api.an* [LOC.fire.CIRC] FIREPLACE

(Austronesian) Oceanic

**Äiwoo (NFL): 48W / 14B / 3C ► cmp:6 prp:5 cls:3**

6 **cmp** Head Mod: *nupo.lea* [net.?spider] SPIDER WEB
5 **prp** Head PREP Mod: *talâwu wâ nnu pevaio* [meal of place morning] BREAKFAST
3 **cls** Head BN:Mod: *kio mi.sigiläi* [hen BN:GNL.male] COCK/ROOSTER

**Hawaiian (HAW): 91W / 50B / 2C ► cmp:50**

49 **cmp** Head Mod: *ala.hao* [path.iron] RAILWAY
1 **cmp** Mod Head: *poho lima* [hollow hand] PALM OF HAND

**Takia (TBC): 71W / 24B / 7C ► dbl:11 con:7 cmp:3 prp:2 gen:1**

10 **dbl** Mod.3SG Head.3SG: *ŋdu.n awa.n* [nose.3SG mouth.3SG] NOSTRIL
7 **con** Mod Head.3SG: *su mala.n* [breast eye.3SG] NIPPLE OR TEAT
The typology of binominal lexemes

2 cmp Mod Head: ab mroun [house owner] HOST
2 prp Mod POSS.3SG Head: graia san anay [evening POSS.3SG food] DINNER
1 cmp Head Mod: tamol sos [man Derris_root] WIDOWER
1 gen Mod.3SG Head: grage.n tatu [side.3SG bone] RIB
1 dbl Mod.3SG Head.IPOSS.3SG: pao.n tatu.w.a.n [shoulder.3SG bone.IPOSS.3SG] COLLARBONE

Hmong-Mien

Hmong Daw (MWW): 86W / 46B / 1C ► cmp:46
46 cmp Head Mod: kev tsheb ngaj hlau [way car rail iron] RAILWAY

(Sino-Tibetan) Bodic

Manange (NMM): 61W / 21B / 4C ► cmp:18 con:2 gen:1
18 cmp Mod Head: Ina.kʰuy [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: 1ano.lʌ.tsami [elder_sister.GEN.daughter] NIECE

(Sino-Tibetan) Macro-Bai

Caijia (CAI): 66W / 30B / 2C ► cmp:30
28 cmp Mod Head: ei³⁵ ho³⁵ [iron.road] RAILWAY
2 cmp Head Mod: me²¹.twie³⁵ [horse.sister] MARE

(Sino-Tibetan) Sinitic

Mandarin Chinese (CMN): 137W / 94B / 4C ► cmp:91 der:3
88 cmp Mod Head: tie3.lu4 [iron.road] RAILWAY
3 der Base.SUF: ju1.zi1 [foal.SUF] FOAL OR COLT
2 cmp Mod.Head.SUF: shou3.wan4.zi1 [hand.wrist.SUF] WRIST
1 cmp Mod.Head.SUF: jiao.wan4.zi1 [foot.wrist/joint.SUF]ANKLE

(Tai-Kadai) Kam-Tai

Thai (THA): 103W / 56B / 2C ► cmp:56
51 cmp Head Mod: thaay.rótfay [way.train] RAILWAY
5 cmp Mod Head: kasi.kɔn [agriculture.doer] FARMER
PNG/Australia

Gunwinyguan

Anindilyakwa (AOI): 47W / 16B / 5C ► con:10 der:5 cmp:1
9 con IPOSS.Head Mod: *e.me.dhvrre e.mindha* [CL(neut).IPOSS.hole CL(neut).nose] NOSTRIL
3 der CL.IPOSS.Base: *a.mv.ngrra* [CL(neut).IPOSS.joint] WRIST
2 der CL.IPOSS.Base: *e.nv.ng.a_rrrra* [CL(neut).M.POSS.CL(neut).wind] BICYCLE PUMP
1 cmp Head Mod: *a.mukwa a.ngura* [CL(neut).source CL(neut).fire] FIREPLACE
1 con Mod IPOSS.Head: *yu.kudhukudha yi.nv.ma.dhvdhrra* [CL(masc).chest CL(masc).M.IPOSS.bone] COLLARBONE

Nuclear Torricelli

Srenge (LSR): 40W / 20B / 2C ► cmp:20
19 cmp Mod Head: *mupə tengə* [nose hole] NOSTRIL
1 cmp Mod Head:REDUP: *ala talmtalmə* [leg/foot branch:RED] TOE

Walman (VAN): 48W / 16B / 1C ► cmp:16
16 cmp Mod Head: *lô müi* [hole nose] NOSTRIL

(Pama-Nyungan) Desert Nyungic

Gurinji (GUE): 53W / 8B / 5C ► cmp:4 der:3 gen:1
4 cmp Mod Head: *jitji jarriny* [nose hole] NOSTRIL
1 der Base.INS: *warlu.waji* [fire.INS] FIREPLACE
1 der Base.SUF: *pirnti.yi* [side.SUF] RIB
1 der Base.VBLZ.AGT: *wuyurrurn.karra.aji* [fishing_line.ACT.AGT] FISHERMAN
1 gen Mod.DAT Head: *yawarta.wu marru* [horse.DAT house] STABLE OR STALL

(Pama-Nyungan) Paman

Wik-Mungkan (WIM): 44W / 25B / 3C ► cmp:24 gen:1
22 cmp Mod Head: *kaa’ uuyan* [nose hole] NOSTRIL
2 cmp Head Mod: *puk wuut* [child old_man] BOY
1 gen Head Mod.DAT: *yuk mee’.akana* [tree/thing eye.DAT] SPECTACLES/GLASSES

West Bomberai

Kalamang (KGV): 90W / 56B / 6C ► cmp:33 con:14 gen:6 adj:2 prp:1
19 cmp Mod Head: *os.ket* [sand.above?] SHORE
The typology of binominal lexemes

14 **cmp** Head Mod: *jaring pueselet* [web spider] SPIDER WEB
14 **con** Mod Head.3POSS: *bustang pos.un* [nose hole.3POSS] NOSTRIL
6 **gen** Head Mod.POSS: *sapu tan.kin* [broom hand.POSS] HANDKERCHIEF OR RAG
2 **adj** Head Mod.REL: *sontum war.ten* [person sorcery.ADJZ] SORCERER OR WITCH
1 **prp** Head DEM Mod: *los (wame) yar* [bridge PROX stone] STONE BRIDGE

North America

Athabaskan-Eyak-Tlingit

**Navajo (NAV): 77W / 27B / 5C ► cmp:16 con:9 prp:1 gen:1**

15 **cmp** Mod.Head: *tó.bāqgh* [water.edge] SHORE
9 **con** Mod 3SG.Head: *bēésh bi.tiin* [iron 3SG.path] RAILWAY
1 **gen** 3SG.Mod.Head: *'a.ké.ts'iin* [3SG.foot.bone] ANKLE
1 **prp** Mod INS Head: *tsé bee na'ni'á* [stone with_it bridge] STONE BRIDGE
1 **cmp** Mod.LEEHead: *nák'ee.sh.to'* [eye_area.LIG.water] TEAR

(Eskimo-Aleut) Yupik

**Central Yupik (ESU): 70W / 21B / 15C ► der:18 dbl:3**

3 **der** Base.TAQ2: *ataku.taq* [evening.TAQ2] DINNER
3 **dbl** Mod.REL Head.POSS: *imarpi.im ceñ.ii* [sea.REL shore.POSS:ABS] SHORE
2 **der** Base.AQ3: *arnar.aq* [woman.AQ3] GIRL
2 **der** Base.LLEQ1: *keni.lleq* [fire.LLEQ1] FIREPLACE
1 **der** Base.AR(AQ): *cuk.ar(aq)* [person.AR(AQ)] TOE
1 **der** Base.CENGAG: *quka.cengaq* [waist.CENGAG] BEE
1 **der** Base.CUUN: *anuqe.ssuun* [wind.CUUN] WINDMILL
1 **der** Base.ILITAQ: *tayarner.ilitaq* [wrist.ILITAQ] BRACELET
1 **der** Base.IRIN: *pingay.irin* [three.IRIN] WEDNESDAY
1 **der** Base.LEK: *emu.lek* [mother's milk/breast.LEK] NIPPLE OR TEAT
1 **der** Base.QLIQ: *ela.qliq* [outside?.QLIQ] NEIGHBOUR
1 **der** Base.QUQ: *epul.quq* [shaft.QUQ] TREE TRUNK
1 **der** Base.UAQ: *pacig.uaq* [gills.UAQ] NOSTRIL
1 **der** Base.YAGAQ: *qusngi.yagaq* [reindeer/sheep.YAGAQ] LAMB
1 **der** Base.YAQ: *aglur.yaq* [ridgepole.YAQ] RAINBOW

(Mayan) Core Mayan

**Q'eqchi’ (KEK): 101W / 40B / 14C ► con:22 cmp:10 der:5 dbl:2 adj:1**

20 **con** 3ERG.Head Mod: *x.k’ot xik* [3ERG.excrement ear] EARWAX
8 **cmp** Head Mod: *sa’ u’aj* [belly nose] NOSTRIL
1 **con** 3ERG.Head M Mod: *x.kem aj am* [3ERG.web M spider] SPIDER WEB
Appendix D. Constructions (summary)

1  

dbl 3ERG.Head Mod.DER: x.na’ ixaq’il [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

1  

der Base.DER: jolom’il [head.DER] CHIEFTAIN

1  

der Base.INS.NMLZ: k’uub’.leb’.aal [hearthstone.INS.NMLZ] COOKHOUSE

1  

der Base.NMLZ: k’oteb’.aal [shit.NMLZ] TOILET

1  

dbl Head.UNPOSS 3ERG.Mod: b’aq.el x.b’eenTel [bone.UNPOSS 3ERG.upper_arm] SHOULDERBLADE

1  

cmp Head.UNPOSS Mod: b’aq.el jolom [bone.UNPOSS head] SKULL

1  

cmp Mod Head: asuukr utz’ajl [sugar ?] SUGAR CANE

1  

adj Mod.ADJZ Head: k’im.al kab’l [straw.ADJZ house] THATCH

Zinacantán Tzotzil (TZO): 64W / 28B / 4C ► gen:15 cmp:13

14  

gen Head Mod.IPOSS: niʔ chu’.ul [nose breast.IPOSS] NIPPLE OR TEAT

11  

cmp Head Mod: sano.te giddo [road mucous] NOSTRIL

2  

cmp Mod Head: shokan na [side house] NEIGHBOUR

1  

gen Mod Head.IPOSS: ʔual bak.el [chaplet/rosary bone.IPOSS] COLLARBONE

(Otomanguean) Otopamean

Querétaro Otomi (OTQ): 134W / 33B / 3C ► cmp:32 prp:1

31  

cmp Head Mod: oku.xiñu [hole.nose] NOSTRIL

1  

prp Head PL Mod: hai ya mboxita [land DEF:PL ancestor] NATIVE COUNTRY

1  

cmp Mod Head: do.xumo [stone.bowl] SKULL

Seri

Seri (SEI): 61W / 16B / 8C ► dbl:6 cmp:5 con:3 gen:2

5  

dbl POSS.Mod POSS.Head: Ø.yanopj i.t [3:POSS.fist 3:POSS.base] WRIST

4  

cmp Head Mod: quihehe cmaam [chief_person female] QUEEN

2  

con Mod POSS.Head: pnaal ii.me [honey_bee 3:POSS.abode] BEEHIVE

1  

dbl ABS.Mod POSS.Head: ha.mt i.teen [ABS.breast 3:POSS.opening/mouth] NIPPLE OR TEAT

1  

cmp Mod Head: xepe poosj [sea line] FISHING LINE

1  

con Mod POSS.Head: xepe i.teel [sea 3:POSS.edge] SHORE

1  

gen POSS.Mod Head: i.to.cams [3:POSS.eye.??] EYEBROW

1  

Tupian) Tupi-Guarani

Mbyá Guarani (GUN): 61W / 33B / 6C ► cmp:23 gen:4 con:3 der:3

20 cmp Mod Head: yy rembe [water edge] SHORE
4 gen 3.Mod Head: h.uvixa kunha [3.leader woman] QUEEN
3 cmp Mod Head.DIM: kavara ra'y.'i [goat son.DIM] KID
3 con Mod Head.PST: axi'y kä.gue [shoulder bone.PST] COLLARBONE
2 der Base.DIM: kunha.'i [woman.DIM] GIRL
1 der 3.Mod ABL.Head: h.exa re.gua [3.eye ABL.NMLZ] SPECTACLES/GLASSES

Uto-Aztecan) Southern Uto-Aztecan

Yaqui (YAQ): 92W / 20B / 5C ► cmp:18 der:2

15 cmp Mod Head: yeka wojo'oria [nose hole] NOSTRIL
2 cmp DIM Base: ili jamut [DIM woman] GIRL
1 der Base.LOC: sis'i'iwoo.chi [iron.LOC] TOOL
1 der Base.POSS.LIG.NMLZ: jo'a.ak.a.me [home.POSS.LE.NMLZ] HOST
1 cmp Mod Head.APPL: mam betala.riam [hand plain.APPL] PALM OF HAND

South America

Araucanian

Mapudungun (ARN): 81W / 25B / 5C ► cmp:21 der:3 gen:1

14 cmp Head Mod: wechoz yu [hole nose] NOSTRIL
7 cmp Mod Head: tren.rüpü [train.way] RAILWAY
2 der Base.NMLZ: ruka.fe [house.NMLZ] CARPENTER
1 der Base.LOC: kütral.we [fire.LOC] FIREPLACE
1 gen Mod.ESS.NONF Head: fiita.nge.n zomo [husband.ESS.NONF woman] MARRIED WOMAN

(Cariban) Guianan

Galibi Carib (CAR): 65W / 20B / 7C ► cmp:8 con:6 der:4 gen:1 dbl:1

8 cmp Mod Head: moyowai etaweini [spider web] SPIDER WEB
6 con Mod Head.POSS: manati poti.li [breast tip.POSS] NIPPLE OR TEAT
2 der Base.DIM.DEV: kapala.me.npo [sheep.DIM.DEV] LAMB
1 der Base.DEV: kamisa.npo [piece_of_tissue.DEV] HANDKERCHIEF OR RAG
1 der Base.DIM: kapilita.menpo [goat.DIM] KID
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
Chibchan

**Cabécar (CJP): 81W / 31B / 2C ► cmp:24 con:7**

24 **cmp** Mod Head: **kóbáká ńgalq** [train road] RAILWAY
7 **con** Mod Head.SPEC: **kuká ńejl** [ear excrement:SPEC] EARWAX

Huitotoan

**Murui Huitoto (HUU): 48W / 18B / 4C ► cls:11 cmp:4 con:3**

10 **cls** Base.CL: **defo** [nose:CL(cavity)] NOSTRIL
4 **cmp** Mod Head: **iyé fuue** [river mouth/edge] SHORE
3 **con** Mod ANAPH.Head: **jefi i.goī** [ear:CL(cavity) ANAPH.CL( leather)] EARLOBE
1 **cls** Base.CL: **taizjikoño** [heel.CL(cover).CL(FEM)] ANKLE

Matacoan

**Wichi (MZH): 82W / 36B / 8C ► gen:14 cmp:11 der:9 con:2**

14 **gen** IPOSS.Mod Head: **to.ch’ute’ lhits’i** [IPOSS.ear wax] EARWAX
11 **cmp** Mod Head: **tewuik lhip** [river part] SHORE
4 **der** IPOSS.Base.LOC: **to.nhes.pe’** [IPOSS.nose.LOC] NOSTRIL
2 **der** Base.AGT: **sapatu.wu** [shoe.AGT] SHOEMAKER
2 **der** Base.LOC: **y’amekw hi** [excrement.LOC] TOILET
1 **der** IPOSS.Base.AGT: **to.lhokwe.wu** [IPOSS.container.AGT] POTTER
1 **con** Mod CL.GEN.Head: **chuhut ka.niyhoy** [spider CL:GEN.rope:PL] SPIDER WEB
1 **con** Mod Head.LOC: **hep lhetek.ch’u** [hut head.LOC] THATCH

Nadahup

**Hudpë (JUP): 61W / 28B / 2C ► cmp:28**

27 **cmp** Mod Head: **tøj mšy** [nose hole/house] NOSTRIL
1 **cmp** Head Mod: **tșp.g’ẹt** [house.leaf] THATCH

(Quechuan) Quechua II

**Imbabura Quechua (QVI): 71W / 18B / 2C ► cmp:18**

17 **cmp** Mod Head: **singa utuju** [nose hole] NOSTRIL
1 **cmp** Head Mod: **chajpi puncha** [half day] MIDDAY

Ticuna-Yuri

**Ticuna (TCA): 73W / 20B / 6C ► con:10 cmp:5 gen:4 dbl:1**

10 **con** POSS.Head.Mod: **ná.má.Ιrǎũ** [POSS.hole.nose] NOSTRIL
5 **cmp** Mod.Head: **pawu.chiđũ** [spider.web] SPIDER WEB
2 **gen** POSS.Mod.Head: **ná.chǐnũ.pěrũ** [POSS.outer_ear.extremity] EARLOBE
The typology of binominal lexemes

1 gen Head.Mod.GEN: ānērù ọrò.ārũ [ring gold.GEN] GOLD RING
1 gen Head.Mod.PURP: pē’.me’e.rũ[ũ] [cover.hand/finger.PURP] GLOVE
1 dbl POSS.Head.Mod.GEN: ná.āpāch’i kōtù.ārũ [POSS.roof palm_species.GEN]
   THATCH

Pidgins/Creoles

(Pidgins & Creoles) English-based

Saramaccan (SRM): 65W / 32B / 3C ► cmp:29 der:2 prp:1
29 cmp Mod Head: talán fútu [train foot] RAILWAY
2 der Base.AGT: koósu.ma [skirt.AGT] GIRL
1 prp Head OF Mod: fínga u fútu [finger of foot] TOE

(Pidgins & Creoles) French-based

Seychelles Creole (CRS): 107W / 21B / 1C ► cmp:21
21 cmp Head Mod: trou nennen [hole nose] NOSTRIL
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.

A handful of languages have yet to be added to the database, so this listing is not final.

_Africa_

(Afro-Asiatic) Berber

Tariffit (75): 96W / 16B / 4C

Head Mod (1)

frinu manu [brake hand] HAND BRAKE

Head Mod:CON (3)

taher ‘as ‘ut amez‘en [ring? ear:CON] EARLOBE
tisif ufas [bottom hand:CON] PALM OF HAND
tarbow ufas [palm hand:CON] PALM OF HAND

Head Prep Mod:CON (13)

aqurab n wata ‘ur [bag of back:CON] BACKPACK
upumpa n baakli [pump of bike:CON] BICYCLE PUMP
taher ‘as ‘ut amez‘en [ring? ear:CON] EARLOBE
afurra n tit ‘i [back of eye:CON] EYELID
tadant n wáy [ring of gold:CON] GOLD RING
azyn n nháa [middle of day:CON] MIDDAY
tisif n ufas [bottom of hand:CON] PALM OF HAND
tarbow n ufas [palm of hand:CON] PALM OF HAND
abrid n masina [road of train:CON] RAILWAY
tasfí wózz ‘á [bridge of rain:CON] RAINBOW
künd ‘it n wáy ‘á [bridge of stone:CON] STONE BRIDGE
fiyed ‘n bit sama [paper of toilet:CON] TOILET PAPER

carmel na díyctt [brush of teeth:CON] TOOTHBRUSH

diyyi.‘s n uma [daughter.3SG of brother:CON] NIECE

(Afro-Asiatic) Chadic

Barain (9): 57W / 21B / 7C

Head Mod (1)

peesi mee [horse woman] MARE

Head Prep Mod:POS (1)

golmo y bal.ni [house Prep back.POSS:3SG:F] TOILET

Head Prep Mod (2)

moosso ta peye [cow Prep milk] DAIRY COW
lool ti luta [dirt Prep ear] EARWAX

Head REL Mod:POS (1)

mee de mer.geti [women REL:SG:F husband.POSS:3SG:F] MARRIED WOMAN

Head REL Prep Mod (7)

tii de y bonte [eat:INF REL:SG:F Prep morning] BREAKFAST
tii de y bodo [eat:INF REL:SG:F Prep night] DINNER
tii de n buuso [eat:INF REL:SG:F Prep noon] LUNCH

assí ge n kee [bone REL:3SG:M Prep head] SKULL
gera ge n gargar [home REL:3SG:M Prep spider] SPIDER

A handful of languages have yet to be added to the database, so this listing is not final.

Hausa (33): 100W / 43B / 4C

Base.ABST (1)

matáa [sorcerer.ABST] MAGIC

Base.F (1)

sàrippi.‘i.yi [king.F] QUEEN

Head.LK Mod (40)

káká.‘n záníi [wash.LK bee] BEESWAX
dóóki.‘n kárfe [horse.LK metal] BICYCLE
báá.‘n hánni [thing.LK hand] BRACELET
káši.‘n káǒáíádú [bone.LK shoulder] COLLARBONE
báńcí.‘n dááťe [meal.LK night] DINNER
káá.‘n kóló [wood.LK door] DOORPOST
káá.‘n júi [skin.LK ear] EARLOBE
yá.‘n kúnne [children.LK ear] EARRING
yáá.‘x kúnne [dirty.LK ear] EARWAX
yááshi.‘n jidó [hair.LK eye] EYELASH
yáá.‘r jidó [skin.LK eye] EYELID
árá.‘n wááti [tongue.LK fire] FLAME
bíkí.‘n hánni [brake.LK hand] HAND BRAKE
jáká.‘n hánni [bag.LK hand] HANDBAG
dí.‘n akwíyá [son.LK goat] KID
yá.‘r akwíyá [daughter.LK goat] KID
yá.‘x túnká [daughter.LK ewe] LAMB
dí.‘n rángó [son.LK sheep] LAMB
káá.‘x mótú [number.LK car] LICENSE PLATE
káńcí.‘n rándá [meal.LK day] LUNCH
ákwáí.‘n wáíkkí [box.LK mail] MAIL BOX
yá.‘r wáá [daughter.LK elder_brother] NIECE
báákí.‘n máumá [mouth.LK breast] NIPPLE OR TEAT
káá.‘r hánni [orifice.LK nose] NOSTRIL
yáá.‘n hánni [palm.LK hand] PALM OF HAND
The typology of binominal lexemes

axjan.it [nephew,F] NIECE
malt.a [king,F] QUEEN

Head Mod (1)
tik qav [bag back] BACKPACK

Head PRP Mod (1)
kelet ha'anan [arc in:DEF,sky] RAINBOW

Head CON DEF Mod (2)
tenux ha.żon [lobe:STC DEF,ear] EARLOBE
sevil be.xalax [path:STC DEF, milk] MILKY WAY

Head CON Mod (34)
kaver.et dkap [roof:STC straw] STRAW

The typology of binominal lexemes

Maltese (58): 96W / 32B / 8C

Base.AGT (1)
skarpan [shoe,AGT] SHOEMAKER

DEF Head PREP:DEF Mod (2)
l-għaqlar tal-widn-ejn [DEF-nasty stuff of:DEF,ear-DUAL] EARWAX

l-isfar tal-bajda [DEF-yellow of:DEF,egg] YOLK

DEF Head Mod (1)
il-kelma mpstt [DEF-word key] KEYWORD

DEF. Head PREP:DEF Mod (1)
l-Triq ta’ Sant’Anna [DEF-way of Saint_Anne] MILKY WAY

Head DEF Mod (2)
tebqet il-għajn [half:DEF,eye] EYELID
xatt il-bahar [shore:DEF,sea] SHORE

Head Mod (2)
nofs.inhar [half:day] MIDDAY
qaws.alla [bow, god] RAINBOW

Head PREP Mod (3)
kap ta’ tribu [head of tribe] CHIEFTAIN
ikla ta’ nofs.inhar [meal of midday] LUNCH

saba ta’ sieq [finger/toe of foot] TOE

Head PREP:DEF Mod (20)
qolla tan-nahal [hive of:DEF,bees] BEEHIVE
sema tan-nahal [wax of:DEF,bees] BEESWAX
pomma tar-rota [pump of:DEF,bicycle] BICYCLE PUMP
ghadm-a tal-għonq [bone:SGLT of:DEF,-collar] COLLARBONE
bagra tal-halib [cow of:DEF,milk] DAIRY COW
kaxxa tal-bieb [post of:DEF,door] DOORPOST
lenza tas-sajaż [cord of:DEF,fishing] FISHING LINE
riża tas-siq [read of:DEF,footprint] FOOTPRINT
curkt tad-deheb [ring of:DEF,gold] GOLD RING
raghay tal-marluš [shepherd of:DEF,herd] HERDSMAN
kaxxa tal-xitrir [box of:DEF,letters] MAIL BOX
art tat-twelid [land of:DEF,birth] NATIVE COUNTRY
pala tal-id [shovel of:DEF,hand] PALM OF HAND
gewna tal-ispallu [wing of:DEF,shoulder] SHOULDERBLADE
point tal-gebel [bridge of:DEF,stone] STONE BRIDGE
kaxxa tal-għodda [box of:DEF,tool] TOOLBOX
skapilja tas-sijni [brush of:DEF,teeth] TOOTHBRUSH
zokk tas-siġra [root of:DEF,tree] TREE TRUNK
pomma tal-ilma [pump of:DEF,water] WATER PUMP
mitna tar-rir [mill of:DEF,wind] WINDMILL

(Atlantic-Congo) Adamawa-Ubangi

Kam (45): 69W / 26B / 2C

Head Mod (18)
pi kār [thing ear] EARRING
gūn.ni [round?,eye] EYEBROW
sūr.ni [hair,eye] EYELASH
lū.ni [skin,eye] EYELID
alnūn.t [tongue fire] FLAME
a.lī [NOM,child/boy,female/woman/wife] GIRL
sît.t [room medicine] HOSPITAL
ni gbw [person guest] HOST
i.lī [child/boy/DM,sheep] LAMB
iŋlā [horse,female/wife/woman] MARE
a.gwā.ŋgwāŋ [NOM,friend,house] NEIGHBOUR
a.dîr.ñ [NOM, stomach,hand] PALM OF HAND
ngwā māt [mouth river] SHORE
a.ŋgwā.ŋwâ [NOM,house,spider] SPIDER WEB
ibnîn.âr [bridge,stone] STONE BRIDGE
a.mîn [NOM,dey,eye] TEAR
a.lî [NOM,head,room] THATCH
kpo.bî [body,tree] TREE TRUNK

Head.IE Mod (8)
a.ângî.î [NOM,eye,LE,leg] ANKLE
lâm.i.ķâr [??,LE,ear] EAR
a.giamenti [NOM, hole,LE,nose] NOSTRIL
awân.i.ghânb [chief,LE,female/woman/wife] QUEEN
kây.i.îgwîn [box,LE,back] SPINE
a.fîr.t.îkîn [sugar,LE,??] SUGAR CANE
a.giamenti [NOM, hole,LE,shit] TOILET
a.ângî.î [NOM,eye,LE,hand] WRIST

(Atlantic-Congo) Bantoid

Swahili (86): 104W / 52B / 4C

CL Base (4)
mto [CL1:child] BOY
fumo [CL9:spear] CHIEFTAIN
a.chawi [CL14,witch] MAGIC
m.zabibu [CL3,grape] VINE

Head CON Mod (39)
kiviko cha mgua [wrist CON foot] ANKLE
kifundo cha mgua [joints CON foot] ANKLE
miatl yaka kashinji ya dundu [flames CON north:LOC CON world] ARCTIC LIGHTS
pampu ya batiskeli [pump CON bicycle] BICYCLE PUMP
The typology of binominal lexemes

pe.lor [sheep.F:YNG] LAMB
Base.M (2)
zm.bal [fowl.M:ADLT] COCK/ROOSTER
bumbled [goat.M:YNG] KID
Head Mod (3)
bimbaal [child.man] BOY
bimbaali [child.woman] GIRL
nebi.kangkawal [finger.??] THUMB
Mod Head (17)
nda.jog [leg.joint?] ANKLE
nda.sii [leg.eye] ANKLE
sii.koyo [eye.???] EYEBROW
sii.pop [eye.hair] EYELASH
sii.sii [eye.animal_skin] EYELID
nda.pii [leg.flat] FOOTPRINT
il.noo [breast.mouth] NIPPLE OR TEAT
mza.xu [nose.hole] NOSTRIL
nepa.tatry [hand.stomach] PALM OF HAND
kug.dabii [trib_edge коллектив] RIB
mog.nda [sea_big_river.mouth] SHORE
gant.albaanha [back.middle] SPINE
sii.nu [eye.wine] TEAR
nda.bii [leg.seed] FOOT
lu.laau [funeral.man] WIDOW
ne.fug [arm.joint?] WRIST

(Atlantic-Congo) Kwa

Mamara Senoufo (61): 81W / 27B / 2C

Base.AGT (2)
kulo.fxo [village_AGT] CHIEFTAIN
fungon.fxo [witchcraft.AGT] SORCERER OR WITCH

Mod Head (25)
pileki.ru [night.meal.CL] DINNER
niki.tumux0.pi [ear.iron.CL] EAR EARRING
niki.furi [ear.excreament.CL] EARWAX
pu.furi [eye.hair.CL] EYELASH
pu.xel.xe [eye.skin.CL] EYELID
na.yolo.yi [fire.???] FLAME
kemji.kwo [arm.mark] FOOTPRINT
kemji.wisi [hand.brake] HAND BRAKE
dke0xo3.0o [doctor.house] HOSPITAL
jomene [speech.core.CL] KEYWORD
campa.ru [day.meal.CL] LUNCH
fu.xo0.yi [horse.female.CL] MARE
gata.c [marriage.woman] MARRIED WOMAN
minaw.ye [nose.hole.CL] NOSTRIL
sama.c [king.woman] QUEEN

Bandial (8): 72W / 16B / 3C

CL.Base (3)
ji.pili [CL.horse] FOAL OR COLT
ji.jamen [CL.goat] KID
a.suy [CL.sorcery] SORCERER OR WITCH

Head CON Mod (4)
(sinan) sa këbuom [rice CON morning] BREAKFAST
(sinan) sa këbuom [rice CON evening] DINNER
(sinan) sa këbuom [rice CON afternoon] LUNCH
(sinan) sa këbuom [rice CON evening] SUPPER

Head Mod (9)
ji.ño e.pili [CL.young CL.horse] FOAL OR COLT
ba.et ka.ñen [CL.bag CL.hand] HANDBAG
ji.ño e.jamen [CL.young CL.goat] KID
ji.ño e.jamen e.mandiy [CL.young CL.goat AG.mandingue] LAMB

(Atlantic-Congo) Mande

Bambara (7): 81W / 39B / 6C

Base.SUF.DIM (1)
misi.san [woman.DIM] GIRL.
Head Mod (1)
co.filyats [man widow] WIDOW

Head.SUF.DIM (1)
bëlojyenden.kin.bà [finger.head.AUG] THUMB

Mod GEN Head (1)
Ala kà màru [God GEN knife] RAINBOW

Mod.Head (25)
ñôm.kò [foot.back] ANKLE
klàl mòplàn [back pack] BACKPACK
du.dén [honey.son] BEE
ngi.so [iron.horse] BICYCLE
kùlò.bwo [ear.excreament] EARWAX
nì.sìi [eye.hair] EYEBROW
nì.sì [eye.hair] EYELASH
nì.nolo [eye.skin] EYELID
íasu.mu [fire.light] FLAME
so.dén [horse.son] FOAL OR COLT
bìlù.nàn [hand.mark] FOOTPRINT
dìgàye.sò [doctor.house] HOSPITAL
bà.dén [goat.kid] KID
saga.dén [sheep.kid] LAMB
mobili.plakì [ear-plate] LICENSE PLATE
sù.mùso [horse.woman] MARE
tile.camaci [num.middle] MIDDAY
beni.dén [mother_s.parents.kid] NIECE
nù.wò [nose.hole] NOSTRIL
tìkì.kiína [hand.interior] PALM OF HAND
teren.sira [train.road] RAILWAY
sìbàgu.mùso [sorcerer.woman] SORCERER OR WITCH
sìkàsì.kàla [sugar.stem] SUGAR CANE
nì.yì [eye.wine] TEAR
bòlo.kà [hand.neck] WRIST

(Atlantic-Congo) North-Central Atlantic
Appendix E. Binominal data set

Wolof (97): 64W / 17B / 2C

Head CON Mod (14)

fajur. u bëj-gànnúar [dawn, CON north] ARCTIC LIGHTS
ndambu, yambu ["pagnet"]. CON bee] BEEHIVE
pomp. u welo [pump, CON bicycle] BICYCLE PUMP
xale bu gòir [child CON male] BOY
njitian xee [chief, CON clan] CHIEFTAIN
mbñatur, m-bët [but, CON CL-eye] EYELID
gone bu njéëen [girl CON child] GIRL
laxaxu, u loxo [hit, CON hand] HAND BRAKE
fakwa njéëen [horse CON girl] MARE
goro bu njéëen [## CON girl] MOTHER-IN-LAW (OF A MAN)
dëkk. u casoaan [country, CON origin] COUNTRY
jaarbat bu njéëen [nephew CON girl] NIECE
paxu. bu barkan [hole, CON nose] NOSTRIL
geet. u fas [enclosure, CON horse] STABLE OR STALL

Head Mod (3)

hukk dgg-gannaav [bag back] BACKPACK
jaaroroopp [jewelry-ear] EARRING
jauro wuras [ring gold] GOLD RING

(Nilo-Saharan) Nilotic

Datoga (89): 29W / 18 / 1C

Head SGLT.SG Mod.SG (1)

five, an, da bàdyà. du [thread, SGLT.SG back] SPINE

Kupsabiny (51): 90W / 43B / 5C

Head PREP Mod (5)

purekit njéëpo yewuat [brake for hand] HAND BRAKE
soontuktu njéëpo pëdëroòndòò [box for mail] MAIL BOX
ankuvereet njéëpo mukwàny [leaves for toilet] TOILET PAPER
soontuktu njéëpo karneeunt [box for metals] TOOLBOX
tinkêtee njéëpo yoomeet [machine for air] WINDMILL

Head REL Mod (2)

looleet njéëpo cëëkkàśënë [bag REL: SG back] BACKPACK
mòron njéëpo musuk [man REL: SG part] WIDOWER

Head.POSS Mod (34)

koowñèet. aap gwoonyoo [bone, POSS anke] ANKLE
taarinkeek. aap attic [lamps, POSS Arctic] ARCTIC LIGHTS
muyat. aap sakumuk [oil, POSS bees] BEESWAX
paampuat. aap pooskìilit [pump, POSS bicycle] BICYCLE PUMP
âmik. aap koorròn [food, POSS morning] BREAKFAST
koowñèet. aap kàntu [bone, POSS neck] COLLARBONE
fakwa aap sikéo [cow, POSS dairy] DARY COW
âmik. aap kàèëlnyu [food, POSS evening] DINNER
kiit. aap yiitti [thing, POSS ear] EARRING
muyat. aap yiitti [oil, POSS ear] EAR WAX
puutëck. aap koonta [hairs, POSS eye] EYEBROW
puut oxyontëit. aap koonta [hair, POSS eye] EYELASH
peent. aap koonta [meat, POSS eye] EYELID
vapeenimintëet. aap purpurèok òììñkèe [fisherman, POSS fish] FISHERMAN
lëëkñèët-aap faraassit [child, POSS horse] FOAL OR COLT
peharyataat. aap keerenter [sign, POSS leg] FOOTPRINT
kamooolooyeet. aap kooënti [ring, POSS gold] GOLD RING
taampàariit. aap yeuret [rag, POSS hand] HANDKERCHIEF OR RAG

Root POSS.MAG back [sine, POSS back] SPINE
kòótt. aap skiriyeet [house, POSS donkey] STABLE OR STALL
âmik. aap komoomoo [food, POSS night] SUPPER
mooreneet. aap karenkaa [finger, POSS leg] TOE
pureasit. aap keelac [brush, POSS teeth] TOOTHBRUSH
karit. aap maata [car, POSS fire] TRAIN
paampuat. aap peeks [pump, POSS water] WATER PUMP

Mod REL Head (1)

makeyoonreet wulo mayasi [egg where yellow] YOLK

MOD.POSS Head (1)

raawaatteet. aap taraa [stone, POSS bridge] STONE BRIDGE

(Nilo-Saharan) Saharan

Kanuri (49): 83W / 39B / 7C

Base.AGT (6)

kàgàl. mà [anvil, OWN] BLACKSMITH
bàrère. mà [farming, OWN] FARMER
bònë. mà [fish, OWN] FISHERMAN
në. mà [pot, OWN] POTTER
kùlù. mà [shoemaking, OWN] SHOEMAKER
kùdù. mà [witchcraft, OWN] SORCERER OR WITCH

Base.LOC (5)

mùsù. ràm [hand, LOC] BRACELET
kàmmù. ràm [fire, LOC] COOKHOUSE
sèmmù. ràm [ear, LOC] EARRING
litì. rì [doctor, LOC] HOSPITAL
kùlùwù. ràm [spinal, LOC] SPINE

Base.NMLZ (1)

kùmàgûn. mú [honey, NMLZ] HONEY

Head Mod (1)

kàlì. kùmàgûn [insect, honey] BEE

Head Mod.ADJZ (1)

kùmà nyàà [woman marriage, ADJZ] MARRIED	

Head Mod.GEN (23)

shìm shì. bè [eye, FOOT, GEN] ANKLE
fààtì. kàllìkàmmûgûn. bè [house, bee, GEN] BEEHIVE
bòrì. kàjìfì [meal, afternoon, GEN] DINNER
kàsdì. cìnnà. bè [tree, door, GEN] DOORPOST
fààtì. sëmmù. [root, GEN] EARBONE
kùdàwù. sàmù. [diet, GEN] EARWAX
kùndàwì. shìm. [hair, eye, GEN] EYEBROW
fààtì. shìm. [root, GEN] EYELID
blàmà. kàmmùa. [tongue, fire, GEN] FLAME
lììtà. mùòtà. [number, GEN] LICENSE PLATE
bòrì. kàsàììù. [meal, noon, GEN] LUNCH
sàwwììlì. kùnçà. bè [opening, nose, GEN] NOSTRIL
shìllà. shììtì. [bone, side, GEN] RIB
čì. nì. [mouth, water, GEN] SHORE
bàwò. kàllà. [cover, head, GEN] SKULL
jè. tàtuù. òìì [rope, spider, GEN] SPIDER WEB
bòrì. bàndù. [meal, afternoon, GEN] SUPPER
ngàndòò. sì. bè [finger, foot, GEN] TOE
kàrè. cìlà. [things, work, GEN] TOOL
mù. mà. cìlà. [vehicle, ground, GEN] TRAIN
zàrì. bà. [root, blood, GEN] VEIN OR ARTERY
dàwà. mûskò. [neck, hand, GEN] WRIST
chìì. ngëñòëlì. [red egg, GEN] YOLK

Mod Head (2)

for kùrmà [horse female] MARE
kùmû. dàwà [sun middle] MIDDAY

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 363
The typology of binominal lexemes

Eurasia

(Altaic) Tungusic

Oroqen (69): 60W / 24B / 7C

Base.AGT (2)

targan.tjo [farm_land.AGT] FARMER
adu.tjo [domestic_animal.AGT] HERDSMAN

Base.DIM (5)

tarğa.kan [vehicle.DIM] BICYCLE
f:`.kan [ear.DIM] EARRING
unc.dy.kan [young_woman.DIM] GIRL
im.lam [goat.DIM] KID
konir.kan [sheep.DIM] LAMB

Base.REC (1)
am.rok [stool.REC] TOILET

Head Mod (1)

inoj.dulin [day middle] MIDDAY

Mod Head (5)
amana:n kakara [father chicken] COCK/ROOSTER
adga.boys [master person] HOST
inoj.dulin bilqa [midday food] LUNCH
gök.merim [female horse] MARÉ
diy.o ana.diy [nephew girl] NIECE

Mod.Gen.Head (3)
tma.pi bilqa [morning.GEN food] BREAKFAST
unoken.pi firoko [fishhook.GEN thread] FISHING LINE
fiро.pi bilqa [evening.GEN food] SUPPER

Mod.Gen.Head.POSS (7)
dyg.dudu.pi duc.n [bee.GEN house.3SG.POSS] BEEHIVE
ye:fa.pi wun [eye.GEN up.3SG.POSS] EYELID
umpık.pi d5.n [noose.GEN inside.3SG.POSS] NOSTRL
dalay.pi ko:yn [sea.GEN bank.3SG.POSS] SHORE
dil.pi grama:n [head.GEN bone.3SG.POSS] SKULL
ata:kz.pi addim [spider.GEN net.3SG.POSS] SPIDER WEB
umokta.pi fiyarnim [egg.GEN yellow.3SG.POSS] YOLK

(Altaic) Turkic

Turkish (92): 100W / 65B / 5C

Base.AGT (6)
demir.ci [iron.AGT] BLACKSMITH
cift.ci [yoke.AGT] FARMER
balıkci [fish.AGT] FISHERMAN
cümleci [pot.AGT] POTTER
kündura.cr [shoe.AGT] SHOEMAKER
büyük.ci [spell.AGT] SORCERER OR WITCH

Base.F (1)
kral.ice [king.F] QUEEN

Base.SUF (1)
göz.lük [eye.SUF] SPECTACLES/GLASSES

Mod Head (10)
erkek.cocuk [male child] BOY
esas.yemek [base food] DINNER
alton.üzdük [gold ring] GOLD RING
anahtar.sözük [key word] KEYWORD
kaynak valide [brother-in-law,mother] MOTHER-IN-LAW (OF A MAN)
ana.vatan [mother.country] NATIVE COUNTRY
kız.şeyen [girl nephew/niece] NIECE
taş.köprü [stone bridge] STONE BRIDGE
bas.parmak [head.finger] THUMB
dal.erkek [widow/widower man] WIDOWER

Mod.Head.3SG (47)
ayak.bileği [foot.wrist.3SG] ANKLE
kuzey.fecri [north.dawn.3SG] ARCTIC LIGHTS

Yakut (79): 109W / 39B / 7C

Base.AGT (4)
silgi.hi [horse.AGT] HERDSMAN
inaq.cit [cow.AGT] HERDSMAN
köüs.cit [pot.AGT] POTTER
sappiki.hi [boots.AGT] SHOEMAKER

Base.NMLZ (2)
bak.tlik [bas.NMLZ] CHIEFTAIN
gos.pog [room.NMLZ] COOKHOUSE

Base.PROP (1)
ap.tar [magic.PROP] SORCERER OR WITCH

Mod Head (9)
timir.ar [iron horse] BICYCLE
uol.oyo [boy child] BOY
bilğii irr:ino [Flea market] FLEA MARKET
kis.oyo [girl child] GIRL
turba.üülük [finger mitten] GLOVE
iye.kilb [mother wife’ s relatives] MOTHER-IN-LAW (OF A MAN)
goton.ira:ya:yi [lady tsar/king] QUEEN

timir.suo [iron path] RAILWAY
tu:z.ker.be [stone bridge] STONE BRIDGE
Appendix E. Binominal data set

Basque

Basque (25): 102W / 55B / 6C

Base.LOC (2)
- eirla.tegi [bee.LOC] BEEHIVE
- be.koki [eye.LOC] EYEBROW

Base:AGT (4)
- errement.ari [tool.AGT] BLACKSMITH
- ostal.ari [host.ALGT] HOST
- buztin.ari [maid.AGT] POTTER
- zapat.ari [shoe.AGT] SHOEMAKER

Base:SUF (1)
- gos.ari [morning.SUF] BREAKFAST

Head Mod (1)
- er.puru [finger.head] THUMB

Mod Head (41)
- ipar-argi [north-light] ARCTIC LIGHT
- bizar koro [back purse] BACKPACK
- erle argizari [bee wax] BEESWAX
- su.aroz [fire.carpenter] BLACKSMITH
- oil.arru [chicken.male] COCKROOSTER
- lepu.uctz [neck.ring] COLLARBONE
- suk.alde [fire zone] COOKHOUSE
- ene behi [milk cow] DARY COW
- ate zango [door stilt] DOORPOST
- belarrt-lobulu [ear-lobe] EARLOBE
- bet.ile [eye.hair] EYELASH
- bet.azal [eye skin] EYELID
- su.puzter [fire.corner] FIREPLACE
- zaldi.kume [horse.child] FOAL OR COLT
- esku.lurra [hand.fur] GLOVE
- gako-hit [key-word] KEYWORD
- ant.xume [goat.child] KID
- ar.kume [sheep.child] LAMB
- post.onzzi [post-container] MAIL BOX
- egun.erdiz [day.hall] MIDDAY
- esne.bide [milk.way] MILK WAY
- titi.buru [teat.head] NIPPLE OR TEAT
- sudar.zulo [nose.hole] NOSTRIL
- arrarau-gurpil [paddle-wheel] PADDLE WHEEL
- esku.barru [hand.inside] PALM OF HAND

Basque

Basque (21)
- djuke: bil uot.a [Yukaghir fire.3SG] ARCTIC LIGHT
- iniria.nya.ta [bee est.3SG] BEEHIVE
- timir u:h a [iron master.3SG] BLACKSMITH
- mas u:h a [wood master.3SG] CARPENTER
- an.sinya:a [door check.3SG] DOORPOST
- kulga:y emingey.e [ear earlobe.3SG] EARLOBE
- kulga:y goyo.too.to [ear druff.3SG] EARWAX
- atag soslu: [leg path.3SG] FOOTPRINT
- ili: tournah.a [hand brake.3SG] HAND BRAKE
- puočta ja:hi.g [mail box.3SG] MAIL BOX
- kün orto.to [day middle.3SG] MIDDAY
- jalla:n s:i.g.e [sky seam.3SG] MILKY WAY
- marun yayaya:u [nose hole.3SG] NOSTRIL
- oyoyso uyuyee:u [side bone.3SG] RIB
- sarin yaptayay.ay [shoulder flat.3SG] SHOULDERBLADE
- tōbō uyō. [head bone.3SG] SKULL
- o:yuy sitim.e [spider thread.3SG] SPIDER WEB
- gara: u:ta [eye water.3SG] TEAR
- u: nassoh.a [water pump.3SG] WATER PUMP
- simi:t uoyah. [egg colostrum.3SG] YOLK
- simi:t arayah. [egg yellow.3SG] YOLK

Malayalam (57): 100W / 45B / 5C

Head Mod (1)
- kanan.kāl [part foot] ANKLE

Mod Head (2)
- peŋ.kutti [female.child] GIRL
- pen.kutira [female.horse] MARE

Mod:GEN.Head (1)
- ātt.in.kutti [goat.GEN.child] LAMB

Mod Head (32)
- i०.ίncu [honey.fly] BEE
- i०.ίncicca मेलूको [bee wax] BEESWAX
- saikkil [stone.ADJZ bridge] STONE BRIDGE
- prabhāta bhākṣaṇa [fire vehicle] BREAKFAST
- kalatt.ello [neck.bone] COLLARBONE
- kān.ṭī [eye.lash] EYELASH
- kān.ṭīlā [eye.lid] EYELID
- ā.ṛvā [fire flame] FLAME
- kāl.ṭīlā [head] FOOTPRINT
- svarṇaḥ mūrtirā [gold ring] GOLD RING
- kai.ṭī [hand bag] HANDBAG
- ucca.bhākṣaṇa [midday food] LUNCH
- kṣīra.patāhā [mile way] MILKY WAY
- amūgi.yamā [aunt.mother] MOTHER-IN-LAW (OF A MAN)
- uṭa.cakram [paddle-wheel] PADDLE WHEEL
- māja.vilā [rain bow] RAINBOW
- vāry.ella [foot mark] KID
- tōl.palaka [shoulder blade] SHOULDERBLADE
- tala.yotti [head.shell] SKULL
- cilanti.valā [spider.web] SPIDER WEB
- kutira.pantā [horse.row] STABLE OR STALL
- kalu pālā [stone bridge] STONE BRIDGE
- nāyaq.ālā [sun.day] SUNDAY
- kān.ṭīrō [eye.water] TEAR
- tital.virāl [mother.finger] THUMB
- kāl.virāl [leg.finger] TOE
- mūtra.purā [urine house] TOILET
- ī.वाट [fire.vehicle] TRAIN
- ī.मार [mother.tree] TREE TRUNK
- muntiri.valī [grape string] VINE
- nir.purā [water pump] WATER PUMP
- būdhā.ālīca [Mercury.day] WEDNESDAY

Dravidian

Mod:ADJZ Head (6)
- urre. zuko erasun [gold.ADJZ ring] GOLD RING
- esku. zoko balaza [hand.ADJZ brake] HAND BRAKE
- esku. ko polṣa [hand.ADJZ bag] HANDBAG
- harri. zoko zubi [stone.ADJZ bridge] STONE BRIDGE
- lasto- zko teiulato [straw.ADJZ roof] THATCH

Komune. ko.paper [toilet.ADJZ paper] TOILET PAPER

peŋ. kutti [female child] GIRL
pen. kutira [female horse] MARE
Mod:GEN. Head (1)
ātt.in. kutti [goat. GEN. child] LAMB

Mod Head (32)
i.०.िंcu [honey. fly] BEE
i.०.िंcicca मेलूको [bee wax] BEESWAX
saikkil [stone. A DJZ bridge] STONE BRIDGE
prabhāta bhākṣaṇa [fire vehicle] BREAKFAST
kalatt. ello [neck. bone] COLLARBONE
kān.ṭī [eye. lash] EYELASH
kān.ṭīlā [eye. lid] EYELID
ā.ṛvā [fire flame] FLAME
kāl.ṭīlā [head] FOOTPRINT
svarṇaḥ mūrtirā [gold ring] GOLD RING
kai.ṭī [hand bag] HANDBAG
ucca. bhākṣaṇa [midday food] LUNCH
kṣīra. patāhā [mile way] MILKY WAY
amūgi. yamā [aunt. mother] MOTHER-IN-LAW (OF A MAN)
uṭa. cakram [paddle-wheel] PADDLE WHEEL
māja. vilā [rain bow] RAINBOW
vāry. ella [foot mark] KID
tōl. palaka [shoulder blade] SHOULDERBLADE
tala. yotti [head. shell] SKULL
cilanti. valā [spider. web] SPIDER WEB
kutira. pantā [horse. row] STABLE OR STALL
kalu pālā [stone bridge] STONE BRIDGE
nāyaq. ālā [sun. day] SUNDAY
kān. ṭīrō [eye. water] TEAR
tital. virāl [mother. finger] THUMB
kāl. virāl [leg. finger] TOE
mūtra. purā [urine. house] TOILET
ī. vaṭ [fire. vehicle] TRAIN
ī. mār [mother. tree] TREE TRUNK
muntiri. valī [grape string] VINE
nir. purā [water pump] WATER PUMP
būdhā. ālīca [Mercury. day] WEDNESDAY

Mod I. E. Head (9)
i.०.िंcicca क.कृष्ण [bee. L. nest] BEEHIVE
The typology of binominal lexemes

celn.č.š.ūdž [ear.LE.wax] EARWAX
katira.ka.tu [horse.LE.child] FOAL OR COLT
tav.v.u [hand.LE.cover] GLOVE
mėla.ku [breast.LE.eye] NIPPLE OR TEAT
ku.p.pat [hand.LE.leaf] PALM OF HAND
nat.t.ell [mid.LE.bone] SPINE
ąyndau.p.petri [tool.LE.box] TOOLBOX
ku.t.eu [hand.LE.stem] WRIST

(Indo-European) Baltic

Latvian (54): 102W / 59B / 7C

Base.AGT (4)
gold.nieks [table.AGT] CARPENTER
svej.nieks [fish.AGT] FISHERMAN
poz.nieks [pot.AGT] POTTER
kuņ.snieks [shoe.AGT] SHOEMAKER

Base.DIM (4)
post.ī [post.DIM] ANKLE
vakar.inas [evening.DIM] DINNER
kaz.l.īns [goat.DIM] KID
lāp.st. [spade.DIM] SHOULDERBLADE

Base.F (2)
meit.ene [daughter.F] GIRL
karal.ene [king.F] QUEEN

Base.INS (1)
acenes [eye.INS] SPECTACLES/GLASSES

Head Mod (1)
pus.desna [half.day] MIDDAY

Mod Head (17)
ziemen.īlāzma [north.glow] ARCTIC LIGHTS
meģur.soma [back.bag] BACKPACK
atslēg.kauls [key.bone] COLLARBONE
lauk.saimnieks [field.owner] FARMER
makšķer.aktu [fishing_rod.line] FISHING LINE
sīkum.tirus [trifle.market] FLEA MARKET
ka.bat.lāztoj [pocket.kerchief] HANDKERCHIEF OR RAG
atslēg.vārds [key.word] KEYWORD
past.kastīte [mail.box] MAIL BOX
past.karte [post.card] POSTCARD
dzelz.čēls [iron.way] RAILWAY
meģur.kauļs [back.bone] SPINE
akmen.titi [stone.bridge] STONE BRIDGE
cukar.niedre [sugar.cane] SUGAR CANE
vin.ošājas [wine.berry_bush] VINE
ūdens/sūkis [water_pump] WATER PUMP
vēj.dzirnavs [wind.mill] WINDMILL

Mod.GEN Head (30)
biš.u krops [bee.GEN hive] HIVE
biš.u vasks [bee.GEN wax] BEESWAX
velosīpēd.sūkis [bicycle.GEN pump] BICYCLE PUMP
rok.as.sprāde [hand.GEN.buckle] BRACELET
atslēg.as kauls [key.bone] COLLARBONE
lauk.āru tuvāte [field.GEN kitchen] COOKHOUSE
durvju.s.tendere [door.GEN.jamb] DOORPOST
auš.lipina [ear.GEN lobe] EARLOBE
auš.u.sērs [ear.GEN wax] EARWAX
pēl.as nospecums [foot.GEN imprint] FOOTPRINT
zelt.a.gredzen [gold.GEN ring] GOLD RING
rok.as.bremce [hand.GEN.brake] HAND BRAKE
rok.as.soma [hand.GEN.bag] HANDBAG
nam.a.tēvs [house.GEN.father] HOST
numur.a.zime [number.GEN sign] LICENSE PLATE
pien.a.cēla [milk.GEN way] MILK WAY
vīr.a māte [husband.GEN.mother] MOTHER-IN-LAW (OF A MAN)
brāl.is.mēta [brother.GEN.daughter] NIECE
krūts.s.gals [breast.GEN.end] NIPPLE OR TEAT
galv.as.kauss [head.GEN.bone] SKULL
zirnek.a.tiks [spider.GEN net] SPIDER WEB
jūnt.a klājums [roof.GEN coating] THATCH
kāj.as.pirksts [foot.GEN finger] TOE
tualete.s.papirs [toilet.GEN paper] TOILET PAPER
darb.a.riks [work.GEN.instrument] TOOL
instrument.u.kaste [instrument.GEN box] TOOLBOX
gol.su [tooth.GEN brush] TOOTHBRUSH
kok.a.stumbrs [tree.GEN trunk] TREE TRUNK
plastik.as.locītava [palm.GEN joint] WRIST
ol.as.dzeltenums [egg.GEN yellow] YOLK

Lithuanian (55): 102W / 47B / 10C

Base.AGT (2)
seim.ininkas [family.AGT] HOST
bart.ininkas [spells.AGT] SORCERER OR WITCH

Base.COLL (1)
kaim.ynas [village.COLL] NEIGHBOUR

Base.DIM (3)
kumel.ikas [mare.DIM] FOAL OR COLT
ož.ikas [goat.DIM] KID
spen.elis [nipple.DIM] NIPPLE OR TEAT

Base.F (4)
vakar.inen [evening.F] DINNER
uosv.inen [father_in_law.F] MOTHER-IN-LAW (OF A MAN)
karal.inen [king.F] QUEEN
vakar.inen [evening.F] SUPPER

Base.NMLZ (12)
kupr.inė [hump.NMLZ] BACKPACK
kal.vis [hammer.NMLZ] BLACKSMITH
bern.ikas [lad.NMLZ] BOY
ųk.ininkas [farm.AGT] FARMER
svej.ys [fish.NMLZ] FISHERMAN
pirkt.ine [finger.NMLZ] GLOVE
rank.ine [hand.NMLZ] HANDBAG
nos.ine [nose.NMLZ] HANDKERCHIEF OR RAG
ligon.ine [patient.NMLZ] HOSPITAL
duker.ėčia [daughter.NMLZ] NIECE
puodžius [pot.NMLZ] POTTER
ak.iniu [eye.NMLZ] SPECTACLES/GLASSES

Head Mod (1)
vidur.dienis [middle.day] MIDDAY

Mod Head (2)
būtas.turgis [feet.F] FEET

Mod.ADJZ.Head (7)
piev.ikav [milk ADZ cow] DAIRY COW
rank.inis stubdys [hand.ADZ brake] HAND BRAKE
gint.oi sāls [birth.ADZ country] NATIVE COUNTRY
gelėz.inikelis [iron.ADZ.way] RAILWAY
mūr.inis.tilas [masonry.ADZ bridge] STONE BRIDGE
šiaud.inis.stogas [straw.ADZ roof] THATCH
tualete.s.papīrs [toilet.ADJZ paper] TOILET PAPER

Mod.GEN Head (10)
dviar.cio.pompa [bicycle.GEN pump] BICYCLE PUMP
auks.ies.spen.ēlis [ear.GEN nipple.DIM] EARLOBE
auks.u.vaksts [ear.GEN wax] EARWAX
auks.o.ziedas [gold.GEN ring] GOLD RING
pašto.dėžutė [post.GEN box] MAIL BOX
puški.ī.sūdas [hirds.GEN way] MILK WAY
koj.as.pīrītas [leg.GEN finger] TOE
ivrankių.dėžė [tool.GEN box] TOOLBOX
medžio.kamienas [tree.GEN trunk] TREE TRUNK
vand.ens.sūdylas [water.GEN pump] WATER PUMP

Mod.I.E.Head (5)
rak.t.ikalis [key.LE.bone] COLLARBONE
rak.t.azodis [key.LE.word] KEYWORD
Appendix E. Binominal data set

vaiv.o.rykitě [Vaiva.LE.cane] RAINBOW
vor.a.tinklis [spider.LE.web] SPIDER WEB
cukr.a.nendré [sugar.LE.reed] SUGAR CANE

(Indo-European) Celtic

Irish (28): 100W / 40B / 6C
Base.AGT (3)
feirm.air [farm.AGT] FARMER
iase.aire [fish.AGT] FISHERMAN
gréas.as [embroidery.AGT] SHOEMAKER

Base.DIM (5)
tinte.dn [fires.DIM] FIREPLACE
caille.in [veil.DIM] GIRL
miot.òg [mitt.DIM] GLOVE
slim.dn [slate/tile.DIM] SHOULDERBLADE
òrd.òg [sledge/hammer.DIM] THUMB

Base.NMLZ (1)
draì.acht [magician,NMLZ] MAGIC

Base.SUF (1)
roth.ar [wheel.SUF] BICYCLE

Head Mod GEN (28)
mdla droma [back:GEN] BACKPACK
céir bheach [wax bees:GEN] BEESWAX
saor adhmait [craftsman wood:GEN] CARPENTER
bò bàinne [cow milk:GEN] DAIRY COW
bun na cluaise [bottom DET ear:GEN] EARLOBE
fàinne cluaise [ring ear:GEN] EARRING
céir cluaise [wax ear:GEN] EARWAX
cap.in sìle [cap.DIM eye:GEN] EYELID
lorg coise [trace foot:GEN] FOOTPRINT
fàinne àir [ring gold:GEN] GOLD RING
mdla [láimhe] [bag hand:GEN] HANDBAG
bosca poist [box post:GEN] MAIL BOX
mein lae [middle day:GEN] MIDDAY
bealach na bò finne [way DET cow:GEN fair:GEN] MILKY WAY
mòthaìr céile [mother companion/spouse:GEN] MOTHER-IN-LAW (OF A MAN)
tìr dìchais [land heritage:GEN] NATIVE COUNTRY
poll sòirne [hole/Pool nose:GEN] NOSTRIL
roth luàin [wheel trowel/blade:GEN] PADDLE WHEEL
cárta poist [card post:GEN] POSTCARD
bogha bàistí [base rain:GEN] RAINBOW
droichead lia [bridge stone:GEN] STONE BRIDGE
de Dòmhnagho [day Sunday:GEN] SUNDAY
pàipéar leithris [paper toilet:GEN] TOILET PAPER
bosca urús [box tools:GEN] TOOLBOX
scuab fiaca [brush teeth:GEN] TOOTHBRUSH
de Cèadaoin [day Wednesday:GEN] WEDNESDAY
mulleàm gasaith [mill wind:GEN] WINDMILL

Mod Head (3)
ainbhir.clar [number.board] LICENSE PLATE
tarm.òid [iron.road] RAILWAY
cloth.gean [stone head] SKULL

Welsh (18): 106W / 62B / 9C
Base.AGT (4)
ffarm.wr [farm.AGT] FARMER
pysgot.wr [fishes.AGT] FISHERMAN
gweset.wr [guest.AGT] HOST
crochann.ydd [cauldron.AGT] POTTER

Base.F (1)
brenin.es [king.F] QUEEN

Base.NMLZ (1)
pen.aeth [chief,NMLZ] CHIEFTAIN

Base.SUF (1)
cwmwd.o [district.SUF] NEIGHBOUR

Head DET Mod (5)
goleumi’r gogoledd [light DET north] ARCTIC LIGHTS
pont yr ysgrwydd [bridge DET shoulder] COLLARBONE
tor (y) llaw [palm DET hand] PALM OF HAND
bwa Drindod [bow DET trinity] RAINBOW
asgwyn y cefn [bone DET back] SPINE

Head Mod (41)
bag ysgrwydd [bag back] BACKPACK
gwenyn mél [bees honey] BEE
ccwch gwynyn [hive bees] BEEHIVE
cyfr gwenyn [wax bees] BEESWAX
pmpw heic [pump bike] BICYCLE PUMP
saer coed [carpenter wood] CARPENTER
post drws [post door] DOORPOST
lled clur [lappet ear] EARLOBE
tws clur [jewel ear] EARING
cwy clur [wax ear] EARWAX
blew amran [hairs eyelid] EYELASH
le tân [place fire] FIREPLACE
lein bysgota [line fishing] FISHING LINE
fiair petheuach [market bits_and_pieces] FLEA MARKET
ól troed [mark foot] FOOTPRINT
modrwy aur [ring gold] GOLD RING
brèc llaw [brake hand] HAND BRAKE
bag llaw [bag hand] HANDBAG
cadach posied [ragnocket] HANDKERCHIEF OR RAG
wyn gafr [kid goat] KID
phl astr [plate number] LICENSE PLATE
boes ilythrau [box letters] MAIL BOX
canol dydd [middle day] MIDDAY
cær arianrhod [fort Arianrhod] MILKY WAY
cær gywodon [fort Gwydion] MILKY WAY
olwyn bado [wheel paddle] PADDLE WHEEL
cersyn post [card post] POSTCARD
pont garreg [bridge stone] STONE BRIDGE
cansen siwgr [cane sugar] SUGAR CANE
dydd sul [day sun] SUNDAY
to gwellt [roof straw] THATCH
bys bawd [finger thumb] THUMB
bys troed [finger foot] TOE
blwch offer [box tools] TOOLBOX
brws dannedd [brush teeth] TOOTHBRUSH
bòn.cyll [base trunk] TREE TRUNK
pmpw díf [pump water] WATER PUMP
dydd mercher [day Mercury] WEDNESDAY
govr gwedd [man widow] WIDOWER
melin wint [mill wind] WINDMILL
melyn.sì [yellow.egg] YOLK

Head Mod.ADJZ (2)
bwch laeth.o [cow milk:ADJZ] DARY COW
lwybr llæth.o [path milk:ADJZ] MILKY WAY

Head PREP Mod (1)
papur lle chwech [paper for toilet] TOILET PAPER

Mod Head (6)
clùst.lws [ear.jewel] EARRING
ysbyd.ò [hospital.house] HOSPITAL
allwedd.gair [key.word] KEYWORD
mam.gwlad [mother.country] NATIVE COUNTRY
rheil.llforad [rail.road] RAILWAY
gwin.gwydd [wine.trees] VINE

(Indo-European) Germanic

Dutch (65): 105W / 54B / 7C
Base.DIM (1)
geit.je [goat.DIM] KID
The typology of binominal lexemes

Mod Head (38)
back.pack BACKPACK
cycle.pump BICYCLE PUMP
collar.bone COLLARBONE
door.post DOORPOST
ear.lobe EARLOBE
ear.ring EARRING
ewe.ear EAR.WAX
eye.brow EYEBROW
eye.lid EYELID
fire.place FIREPLACE
flea.market FLEA MARKET
foot.print FOOTPRINT
gold.ring GOLD RING
hand.brake HAND BRAKE
hand.bag HANDBAG
hand.kerchief HANDBERCHIEF OR RAG
herd.POSS.man HERDSMAN
key.word KEYWORD
number.plate LICENSE PLATE
mail.box MAIL BOX
home.land NATIVE COUNTRY
nose.hole NOSTRIL
paddle.wheel PADDLE WHEEL
post.card POSTCARD
rail.way RAILWAY
rain.bow RAINBOW
shoulder.blade SHOULDERBLADE
stone.bridge STONE BRIDGE
sugar.cane SUGAR CANE
sun.day SUNDAY
toilet.paper TOILET PAPER
tool.box TOOLBOX
tooth.brush TOOTHBRUSH
tree.trunk TREE TRUNK
water.pump WATER PUMP
woden.POSS.day WEDNESDAY
wind.mill WINDMILL

Mod.Head AdjZ Head (2)
north.ADZ lights ARCTIC LIGHTS
milk.ADZ way MILK WAY

German (19): 131W / 73B / 8C

Mod.Head AdjZ Head (2)
north.ADZ lights ARCTIC LIGHTS
milk.ADZ way MILK WAY

Base.F (3)
fisch.er [fish.AGT] FISHERMAN
töpf.er [pot.AGT] POTTER
zauber.er [magic.AGT] SORCERER OR WITCH

Base.DIM (4)
knöch.el [bone.DIM] ANKLE
mäd.chen [maid.DIM] GIRL
zick.lein [goat.DIM] KID
gieß.lein [goat.DIM] KID

Base.F (1)
könig.in [King,F] QUEEN

Base.NMLZ (2)
tisch.ler [table,NMLZ] CARPENTER
schrein.er [cupboard,NMLZ] CARPENTER

Head.Mod (2)
mitt.ag [middle.day] MIDDAY
mitt.woch [middle.week] WEDNESDAY

Mod.Head AdjZ Head (1)
pol.ar.lichter [pole.ADZ.light] ARCTIC LIGHTS

Mod.Head (49)
nord.lichter [north.lights] ARCTIC LIGHTS
rack.sack [back.sack] BACKPACK
luft.pumpe [air.pump] BICYCLE PUMP
fahren.pumpe [bicycle.pump] BICYCLE PUMP
Appendix E. Binomial data set

arm.band [arm.rib] BRACELET
zimmer.mann [timber.man] CARPENTER
schlüssel.bein [key.bone] COLLARBONE
milch.kuh [milch.cow] DAIRY COW
abend.brot [evening.bread] DINNER
abend.essen [evening.food/meal] DINNER
tür.pfosten [door.post/jamb] DOORPOST
ohr.lüpp.chen [ear.lobe.DIM] EARLOBE
ohr.ring [ear.ring] EARRING
land.wirt [land.host/keeper] FARMER
floh.markt [flea.market] FLEA MARKET
hand.schuhe [hand.shoe] GLOVE
gold.ring [gold.ring] GOLD RING
hand.bremse [hand.brake] HAND BRAKE
hand.tasche [hand.bag] HANDBAG
schlüssel.wort [key.word] KEYWORD
mittag.essen [noon.food] LUNCH
brief.kasten [letter.box] MAIL BOX
milch.strasse [milch.street] MILK STREET
vater.land [father.land] NATION COUNTRY
heimat.land [home.land] NATION COUNTRY
brust.war [breast.wart] NIPPLE OR TEAT
schaufler.rad [blade/paddle.wheel] PADDLE WHEEL
hand.fläche [hand.surface] PALM OF HAND
post.karte [post.card] POSTCARD
eisen.bahn [iron.way] RAILWAY
regen.bogen [rain.bow] RAINBOW
schulter.blatt [shoulder.blade] SHOULDERBLADE
wirbel.säule [vertebra.pillar] SPINE
rück.grat [back.ridge] SPINE
stein.brücke [stone.bridge] STONE BRIDGE
zucker.rohr [sugar.cane] SUGAR CANE
sonn.tag [sun.day] SUNDAY
abend.essen [evening.food] SUPPER
abend.brot [evening.bread] SUPPER
stroh.dach [straw.roof] THATCH
klo.papier [toilet.paper] TOILET PAPER
werkzeug [work.stuff/gear] TOOL
zahn.bürste [toothbrush] TOOTHBRUSH
baum.stamm [tree.stem/trunk] TREE TRUNK
wein.rebe [wine.tendril/vine] VINE
wasser.pumpe [water.pump] WATER PUMP
wind.mühle [wind.mill] WINDMILL
hand.gelenk [hand.joint] WRIST
ei.ëlb [egg.yellow] Yolk

Mod.L.E.Head (11)
biene.n.korb [bee.L.basket] BEEHIVE
biene.n.stock [bee.L.stick] BEEHIVE
biene.n.wachs [bee.L.wax] BEESWAX
ohr.en.schmalz [ear.L.land] EARWAX
auge.n.braue [eye.L.brow] EYEBROW
auge.n.lid [eye.L.lid] EYELID
tasche.n.tuch [pocket.L.cloth] HANDKERCHIEF OR RAG
nummer.n.schild [number.L.plate] LICENSE PLATE
nase.n.loch [nose.L.hole] NOSTRIL
spinne.n.netz [spider.L.net/web] SPIDER WEB
toilette.n.papier [toilet.L.paper] TOILET PAPER

Norwegian (67): 100W / 57B / 3C

Head Mod (1)
mid.dag [middle.day] MIDDAY

Mod.Head (51)
nord.lys [north.light] ARTIC LIGHTS
rygg.sekk [back.sack] BACKPACK
bi.kube [bee.cube] BEEHIVE
bi.voks [bee.wax] BEESWAX
sykkel.pumpe [bicycle.pump] BICYCLE PUMP
arm.band [arm.band] BRACELET
krauge.bein [collar.bone] COLLARBONE
eld.hus [fire.house] COOKHOUSE
dør.stolpe [door.post] DOORPOST
dør.flipp [door.??] EARLOBE
dør.dobb [door.button] EARRING
dør.voks [door.wax] EARWAX
eye.vipp [eye.??] EYELASH
eye.lok [eye.lid] EYELID
fd.sted [fireplace] FIREPLACE
loppe.marked [flea.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
omme.tk [pocket.kerchief] HANDKERCHIEF OR RAG
nøkkel.ord [key.word] KEYWORD
bil.skill [car.sign] LICENSE PLATE
brev.kasse [letter.box] MAIL BOX
melke.vei.en [milk.way.DEF] MILK WAY
hjem.land [home.country] NATION COUNTRY
bryst.vorte [breast.wart] NIPPLE OR TEAT
nose.bor [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
skulder.blad [shoulder.blade] SHOULDERBLADE
hode.skalle [head.skel] SKULL
troll.mann [troll.man] SORcerer OR WITCH
rygg.rad [back.row] SPINE
stein.ho [stone.bridge] STONE BRIDGE
sukker.rør [sugar.tube] SUGAR CANE
søn.dag [sun.day] SUNDAY
tolett.papir [toilet.paper] TOILET PAPER
verktry.ksasse [toolbox] TOOLBOX
tann.børste [toothbrush] TOOTHBRUSH
tre.stamme [tree.trunk] TREE TRUNK
wind.mühle [wind.mill] WINDMILL
hand.teidd [hand.joint] WRIST

Mod.L.E.(Head) (5)
melk.e.ku [milch.L.cow] DAIRY COW
eye.n.bryn [eye.L.brow] EYEBROW
kvelds.mar [evening.LE.food] SUPPER
ons.dag [Ondin.LE.day] WEDNESDAY
egg.e.pломme [egg.L.yolk] Yolk

Old High German (29): 63W / 23B / 5C

Base.AGT (3)
hirt.i [herd.AGT] HERDSMAN
hafan.âri [pot.AGT] POTTER
zoubar.âri [magic.AGT] SORcerer OR WITCH

Base.DIM (1)
zickn [goat.DIM] KID

Base.F (1)
kung.in [king.SUF] QUEEN

Head Mod (2)
mitt(i).tag [middle.day] MIDDAY
mitta.wehha [middle.week] WEDNESDAY

Mod.Head (16)
bikar [bee.container] BEEHIVE
zimbar.man [timber.man] CARPENTER
turi.studal [door.post] DOORPOST
dr.lappa [ear.cloth] EARLOBE
dr.ring [ear.ring] EARRING
dr.smeoro [ear.wax] EARWAX

Draft. Please do not cite. Comments to: pepper.steve@gmail.com.
The typology of binominal lexemes

(Indo-European) Greek

Greek (21): 100W / 30B / 7C
Base.DIM (1)
katís.kí [gōt.DIM] kid
Base.SUF (4)
proi.ino [morning.SUF] BREAKFAST
velfar.íb [eyelid.SUF] EYELID
mesimer.íano [noon.SUF] LUNCH
vráo.ino [night.SUF] SUPPER
Head DEF Mod.GEN (3)
kolona tis portas [post DEF door.GEN] DOORPOST
lovos tu afitou [boe DEF ear.GEN] EARLOBE
keri tu afitou [wax DEF ear.GEN] EARWAX
Head Mod (1)
leksi-klíi [word-key] KEYWORD
Head Mod.Gen (9)
ayelada yalaíkos [cow dairy.GEN] DAIRY COW
spasímo heriou [hand break.GEN] HAND BRAKE
tsanta heriós [bag hand.GEN] HANDBAG
pinákida kiklóforía [sign traffic.GEN] LICENSE PLATE
istos aráxnis [web spider.GEN] SPIDER WEB
charti výsa [paper hygiene.GEN] TOILET PAPER
kati erýalo.n [box tool.GEN] TOOLBOX
kormos xentu [trunk.tree.GEN] TREE TRUNK
adília ner.u [pump water.GEN] WATER PUMP
Mod.AdjZ Head (3)
vorío seleús [north.ADJZ lights] ARCTIC LIGHTS
urán.io tokso [sky.ADJZ bow] RAINBOW
petri.inī yerína [stone.ADJZ bridge] STONE BRIDGE
Mod.LE.Head (9)
klid.o.kokalo [collar.LE.bone] COLLARBONE
ik.o.despotis [house.LE.lord] HOST
yrmatom.o.kivotio [letters.LE.box] MAIL BOX
sílot.o.áromos [iron.LE.road] RAILWAY
om.o.plát [shoulder.LE.blade] SHOULDERBLADE
zachar.ó.kalamo [sugar.LE.cane] SUGAR CANE
achir.o.skei [straw.LE.cover] THATCH
odóliō.vurtso [tooth.LE.brush] TOOTHBRUSH
anem.o.milos [wind.LE.mill] WINDMILL

(Indo-European) Indo-Aryan

Assamese (6): 87W / 34B / 3C
Head Mod (1)
gau.gza [body.tree] TREE TRUNK
Mod Head (16)
mou.makhí [honey.bee] BEE
mou.sak [bee.hive] BEEHIVE
kath.mistr [wood.carpenter] CARPENTER
kan.phali [ear.flower] EARRING
sor bòzar [chief market] FLEA MARKET
hat.mozá [hand.sock] GLOVE
sikitā.loy [treatment.place] HOSPITAL
xucık əxbɔ [index word] KEYWORD
əmaya.ponsɔ [license.plate/certificate] LICENSE PLATE

daś sīhi [postal letter] POSTCARD
ram.dheña [Lord_Rama bow] RAINBOW
kami.ḥar [chest.bone] RIB
xil.xako [stone.bridge] STONE BRIDGE
soka.puri [eye.water] TEAR
muntra.loy [urine.place] TOILET
budh.bar [Mercury.day] WEDNESDAY

Mod.Gen Head (17)
saikel.ə hava [cycle.GEN gas] BICYCLE PUMP
ratipuə r ahar [morning.GEN food] BREAKFAST
ratə r ahar [night.GEN food] DINNER
duwar.ə khuta [door.GEN post] DOORPOST
kərə lott [ear.GEN lobe] EARLOBE
kan.ə məl [ear.GEN dirt] EARWAX
sokə r nom [eye.GEN hair] EYELASH
sokə r pus [eye.GEN lid] EYELID
ghora.ə ponali [horse.GEN call] FOAL OR COLT
bhori.ə sap [foot.GEN print] FOOTPRINT
son.ə anguli [gold.GEN ring] GOLD RING
sihti.ə baks [letter.GEN box] MAIL BOX
nį.ə dɛx [self.GEN country] NATIVE COUNTRY
nuk.ə phuta [nose.GEN hole] NOSTRIL
hat.ə tolava [hand.GEN backside] PALM OF HAND
hat.ə burha.anguli [hand.GEN old.finger] THUMB
bhori.ə burha.anguli [leg.GEN old.finger] TOE

Hindi (36): 95W / 30B / 4C
Base.AGT (1)
lohā.ṛ [iron.AGT] BLACKSMITH

Mod.Gen Head (3)
rāt kā khānā [evening.GEN meal] DINNER
ghore kā baccū [horse.GEN child] FOAL OR COLT
sāc kā kāgaz [purification.GEN paper] TOILET PAPER

Mod Head (19)
madhama.ṛ [honey.wax] BEESWAX
karaṇa.phîl [ear.flower] EARRING
karaṇa.ṇal [ear.dirt] EARWAX
kabārī bāzār [scrap_merchant market] FLEA MARKET
amugī pēr [license plate] LICENSE PLATE
jādi-tōnā [magic-charm] MAGIC
dāk-baksā [post-box] MAIL BOX
ākā gangā [sky.Ganges] MILKY WAY
mār.bhūmi [mother.earth] NATIVE COUNTRY
bhām.jī [sister.daughter] NIECE
bhat.jī [brother.daughter] NIECE
post.kari (post.card) POSTCARD
indra.ḥanus [Indra.bow] RAINBOW
ghar.sūl [horse.hall] STABLE OR STALL
śau.ḍay [purification.area] TOILET
yantra.pēṛ [instrument.box] TOOLBOX
rel.gāṛ [rail.cart] TRAIN
jal pum [water pump] WATER PUMP
pavan cakkū [wind mill] WINDMILL

Mod.Gen Head (7)
pair kā cinta(ə) [foot.GEN sign] FOOTPRINT
sone kī angāhā [gold.GEN ring] GOLD RING
dōpahār kā khānā [noon.GEN food] LUNCH
mukhē kā jālā [spider.GEN web] SPIDER WEB
rāt kā khānā [night.GEN food] SUPPER
pāmō kā angāhā [foot.GEN finger] TOE
dāntē kā brāsi [tooth.GEN brush] TOOTHBRUSH

Nepali (68): 66W / 11B / 3C
Mod Head (1)
mātri.bhūmi [mother.land] NATIVE COUNTRY

Mod.Gen Head (6)
kukhā.ṛ ko bhāle [hen.GEN cock] COCK/ROOSTER
sūn.ṛ kāǔhi [gold.GEN ring] GOLD RING
dubh.ṛ kā mūnto [milk.GEN head] NIPPLE OR TEAT
Appendix E. Binominal data set

mákara.ko jâlo [spider.GEN web] SPIDER WEB
hât.ko budhi âlû [hand.GEN old Finger] THUMB
khûtâ.ko budhi âlû [leg.GEN old finger] TOE

Mod.Head (4)
biswa.karmà [universe act] BLACKSMITH
kâne.gûjî [ear.?] EARWAX
âkâ.bhânu [eye brow] EYEBROW
âkâs.gânjâ [sky.Ganges] MILKY WAY

Selice Romani (76): 88W / 8B / 4C
Base.ABST (1)
êohan.ipe [witch/sorcerer.ABST] MAGIC

Base.DIM (2)
chav.oro [Gypsy .child.DIM] BOY
châ.jôri [Gypsy .child.DIM] GIRL

Base.F (1)
kirâ.chiâ [king.F] QUEEN

Mod.GEN Head (4)
kan.en.gêr.e khûlu [ear.OBL.GEN.PL shit.PL] EARWAX
pral.es.ker.i chây [brother.OBL.GEN.SG.F daughter] NIECE
phê.nâ.ker.i chây [sister.OBL.GEN.SG.F daughter] NIECE
çûc.en.gêr.o sêmo [breast.OBL.GEN.SG.M grain] NIPPLE OR TEAT

(Indo-European Romance)

French (27): 102W / 49B / 10C
Base.DIM (4)
brac.elet [arm.DIM] BRACELET
cherv.eau [fork.DIM] KID
mamel.on [chest.DIM] NIPPLE OR TEAT
poing.êr [fist.DIM] WRIST

Base.F (1)
sorcér.e [sorcerer.F] SORCERER OR WITCH

Base.NMLZ (4)
ferm.ier [farm.NMLZ] FARMER
pèch.eur [fish.NMLZ] FISHERMAN
pot.iêr [pot.NMLZ] POTTER
écu.rie [horseman.NMLZ] STABLE OR STALL

Head Mod (1)
mot-Êle [word-key] KEYWORD

Head Mod.ADJZ (4)
aurore boreâlé.e [light north] ADJZ ARCTIC LIGHTS
voie lactée [way milk] ADJZ MILKY WAY
carte post.ale [card post] ADJZ POSTCARD
colonne vertebrae.e [column vertebra] ADJZ SPINE

Head Prep DEF Mod (2)
lohe de l’oreille [lobe of DEFear] EARLOBE
paume de la main [pal of hand] PALM OF HAND

Head Prep Mod (17)
ruche d’abeille.s [hive of:bee:PL] HIVE
claire d’abeille [wax of:bee] BEESWAX
chef de clan [chief of clan] CHIEFTAIN
cuisine de chantier [kitchen of worksite] COOKHOUSE
jambage de porte [jamb of door] DOORPOST
boucle d’oreille [ring of:ear] EARRING
cire d’oreille [wax of:ear] EARWAX
ligne de pêche [line of fishing] FISHING LINE
bague en or [ring of:gold] GOLD RING
lard de porc [lard of:hog] HAM
lard de pêche [lard of fish] FISHING LINE
fendre à main [break to hand] HAND BREAK
rayon d’un arbre [trunk of:tree] TREE TRUNK
jaune d’œuf [yellow of:egg] YOLK

Head PREP2 Mod (14)
sac à dos [bag to back] BACKPACK
pompe à vélo [pump to bicycle] BICYCLE PUMP
vache à lait [cow to milk] DAIRY COW
ligne à pêche [line to fish] FISHING LINE
marché aux pâtes [market to:DEP.PL fleas] FLEA MARKET
frein à main [brake to hand] HAND BRAKE
sac à main [bag to hand] HANDBAG
boule aux lettres.s [box to:DEP.PL.letter.PL] MAIL BOX
boule à aubes.s [wheel to paddle PL] PADDLE WHEEL
canne à sucre [cane to sugar] SUGAR CANE
boule à outils.s [box to tool.PL] TOOLBOX
mousse à dent.s [brush to teeth.PL] TOOTHBRUSH
pompe à eau [pump to water] WATER PUMP
moulin à vent [mill to wind] WINDMILL

Head PREP3 Mod (1)
arc-en-ciel [bow-in-sky] RAINBOW

Mod Head (3)
omo.plate [shoulder plate] SHOULDERBLADE

Italian (41): 112W / 38B / 8C
Base.AGT (1)
pesca.tore [fish.er] FISHERMAN

Base.DIM (3)
camin.etto [fireplace/chimney.DIM] FIREPLACE
bors.etta [bag.DIM] HANDBAG
cap.erro [coat.NMLZ] KID

Base.NMLZ (5)
ovecchi.âno [ear.NMLZ] EARRING
ceram.ista [ceramics.NMLZ] POTTER
streg.âne [witch.NMLZ] SORCERER OR WITCH
occhi.âli [eye.NMLZ] SPECTACLES/GLASSES
spazzol.ino [brush.NMLZ] TOOTHBRUSH

Head Mod (3)
capo.tribù [head.clan] CHIEFTAIN
parola chiave [word key] KEYWORD
mezzo.giorno [half day] MIDDAY

Head Mod.ADJZ (5)
aurore boreale [light north] ADJZ ARCTIC LIGHTS
via lattea [way milk] ADJZ MILKY WAY
spina dorsale [bone back] ADJZ SPINE
carta igien.ica [paper hygiene] ADJZ TOILET PAPER
pompa idraulica [pump water] ADJZ WATER PUMP

Head Prep Mod (13)
cerva.d’api [wax of:bees:PL] BEESWAX
cucina da letto [kitchen from camp] COOKHOUSE
mucca da latte [cow from milk] DAIRY COW
anello d’oro [ring of:gold] GOLD RING
freno a mano [brake to hand] HAND BRAKE
padrone di casa [owner of house] HOST
paese d’origine [country of origin] NATIVE COUNTRY
ruota a pale [wheel to blade PL] PADDLE WHEEL
ponte di pietra [bridge of stone] STONE BRIDGE
canna da zucchero [cane from sugar] SUGAR CANE
petto di gallina [roof of:farm female] THATCH
mulino a vento [mill to wind] WINDMILL
rosso d’uovo [red of:egg] YOLK

Head PREP DET Mod (7)
pompa della bicicletta [pump of:DEP bicycle] BICYCLE PUMP
mercato delle pulci [market to:DEP.PL fleas] FLEA MARKET
cassetta delle lettere [box of:DEP letter:PL] MAIL BOX
dito del piede [finger of:DEP foot] TOE
cassetta degli attrezzi [box of:DEP tools] TOOLBOX
tronco dell’albero [trunk of:DEP tree] TREE TRUNK
pompa dell’acqua [pump of:DEP water] WATER PUMP

Mod Head (1)
ferro.via [iron.way] RAILWAY

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com.
The typology of binominal lexemes

Romanian (77): 135W / 41B / 11C

Base.ABST (1)
capitan.ie [captain.NMLZ] CHIEFTAIN

Base.AGT (7)
fier.ar [iron.AGT] BLACKSMITH
potcoaieăr.r [boof.AGT] BLACKSMITH
uș(a).or [door.SUF] DOORPOST
țar.ăr [earth.AGT] FARMER
pășt.or [pasture.AGT] HERDSMAN
oaf(a).ar [pot.AGT] POTTER
cizmă.ar [boot.AGT] SHOEMAKER

Base.AGT.ABST (1)
bucată.ar.ie [piece_of_food.AGT.NMLZ] COOKHOUSE

Base.DIM (1)
mână.us.ă [hand.SUF.F] GLOVE

Base.F (2)
soacăr.ă [father_in_law.F] MOTHER-IN-LAW (OF A MAN)
nepo(a)t.ă [grandson/nephew.F] NIECE

Base.NMLZ (1)
braț.ară [arm.NMLZ] BRACELET

Base.SUF (1)
gălben.uș [yellow.SUF] YELLOW

Base.SUF.F (1)
wětš.ník

Base.SUF.F (2)
ruka.jca [hand.SUF] HAND

Base.SUF.F (3)
man.ță [hand.SUF] HAND

Base.SUF.F (4)
rh.ș.ă [key.AGT] DOOR

Base.SUF.F (5)
ruč.ní dírka [hand.AGT] HAND

Base.SUF.F (6)
dveř.ní sloupek [door.AGT] DOOR

Head Mod (1)
cuvint cheie [word key] KEYWORD

Head Mod.AdJZ (6)
lumină polară [light polar] ARCTIC LIGHTS
cutie post.ălă [box post.ADJZ] BOX

calea lactee [way milk:ADJZ] MILKWAY

carte poștală [card postal] POSTCARD

calea fer.ată [way iron.ADJZ] RAILWAY

coloană vertebrală [column vertebra] SPINE

Head PREP Mod (18)
pompă de bicicletă [pump on bicycle] BICYCLE PUMP
vaća de lapte [cow of milk] DAIRY COW
sfărăș de undiță [thread of fishing_rod] FISHING LINE
urmă de picior [trace of foot] FOOTPRINT
inel de auri [ring of gold] GOLD RING
frână de mâna [brake of hand] HAND BRAKE
guantă de mâna [glove of hand] GLOVE
roată cu zbarări [wheel with paddle] PADDLE WHEEL
până de păsăjen [cloth of spider] SPIDER WEB
pov de pietra [bridge of stone] STONE BRIDGE
trestie de zahăr [sugar cane] SUGAR CANE
acoperiș de paie/stuf [roof of straw/ reed] REED
hârtie igien.ica [paper hygiene.ADJZ] TOILET PAPER
trusa de scute [kit of tools] TOOLBOX
periță de dânti [brush of tooth] TOOTHBRUSH
viță de vie [creeper_plant of vine/vineyard] VINE
pompă de apa [pump of water] WATER PUMP
moran de vant [mill of wind] WINDMILL

Head.DEF Mod.DET.GEN (2)
lopațica umăr.u.lui [shoulder.DEF hand.DEF.GEN] SHOULDERBLADE
incheietura mâin.ii [joint:DEF hand.DEF.GEN] WRIST

(Indo-European) Slavic

Czech (12): 98W / 44B / 7C

Base.DIM (4)
chlap.ec [man.DIM] BOY
kabel.ka [pulse.DIM] HANDBAG
kiz.le [groat.DIM] KID
lopat.ka [shovel.DEF] SHOULDERBLADE

Base.F (2)
čaroděj.nice [wizard.F] SORCERER OR WITCH
střed.a [middle.F] WEDNESDAY

Lower Sorbian (20): 118W / 43B / 11C

Base.AdJZ.NMLZ (1)
lezela.nica [iron.ADJZ.NMLZ] RAILWAY

Base.ATG (3)
yrb.ăr [fish.AGT] FISHERMAN
gósć.ina [word.AGT] KEYWORD

Base.F (1)
krall.owsa [king.F] QUEEN

Base.LOC (1)
wognil.ś.ko [fire.LOC] FIREPLACE

Base.NMLZ (4)
naklesbat.nik [on.back.NMLZ] BACKPACK
doj.ka [milk.NMLZ] DAIRY COW

cukor.ina [sugar.NMLZ] SUGAR CANE

Base.SUF (1)
ruka.jca [hand.SUF] GLOVE

Head Mod (1)
pol.ță [half.day] MIDDAY

Head Mod.GEN (1)
kašćik rěd.a [box tool.GEN] TOOLBOX
Appendix E. Binominal data set

tabu z wuchu [wax PREP ear] EARWAX
lapka na woku [flap on eye] EYELID

Head PREP Mod (2)
tabu z wuchu [wax PREP ear] EARWAX

Head Mod.ADJZ (16)
porządka polar.na [dawn pole] ADIZ ARCTIC LIGHTS
wosk pszczeński [wax bee] ADIZ BEESWAX
pompa rower.owa [pump bicycle] ADIZ BICYCLE PUMP
kuchnia pol.owa [kitchen field] ADIZ COOKHOUSE
krowa mlecz.na [cow milk] ADIZ DAIRY COW
żylka wręca.ka [line fishing rod] ADIZ FISHING LINE
hamulce ręcz.na [brake hand] ADIZ HAND BRAKE
słońce oświetl.owe [sun light] ADIZ HEADLAMP

Mod.Gen Head (2)
listowa kačišć [mail ADIZ box] BOX
ptakowa droga [bird ADIZ road] MILK WAY
psychod na maś [future ADIZ mother] MOTHER-IN-LAW OF A MAN
woschina na krw. [ancestor ADIZ country] NATIONAL COUNTRY
nosiowa sërka [nose ADIZ hole] NOSE
postowa kórka [post ADIZ card] POSTCARD
kamienica most [stone ADIZ bridge] BRIDGE
slomiany kőy [straw ADIZ roof] THATCH
toletowa papierowa [toilet ADIZ paper] TOILET PAPER
zubna szczotka [toothbrush] TOOTHBRUSH

Polish (72): 104W / 54B / 11C

Base.ADJZ.DIM (1)
poczówka [post ADIZ DIM] POSTCARD

Base.AGT (8)
chłopak [farmer ADIZ] BOY
stolarz [table ADIZ] CARPENTER
rolnik [farmland ADIZ] FARMER
rybak [fish ADIZ] FISHERMAN
gospodarz [inn ADIZ] HOST
garncarz [pot ADIZ] POTTER
czarnosnűrińczik [black prince ADIZ] SORCERER OR WITCH
wodnica [wet ADIZ] WINDMILL

Base.DIM (8)
kostka [bone ADIZ] ANKLE
kolczyk [thorn ADIZ] EARLOBE
komin.ek [chimney ADIZ] CHIMNEY
rękawiczka [sleeve ADIZ] GLOVE
torebka [bag ADIZ] HANDBAG
chusteczka [scarf ADIZ] SCARF
laptopka [shovel ADIZ] SHOULDERBLADE
żółta kờ [yellow ADIZ] YOLK

Base.F (6)
meżatka [husband ADIZ] MARRIED WOMAN
teściowa [father in law ADIZ] MOTHER-IN-LAW OF A MAN
bratanica [nephew ADIZ] NIECE
siostroweczka [nephew ADIZ] NIECE
królowa [king ADIZ] QUEEN
pałecznica [spider ADIZ] SPIDER

Base.NMLZ (3)
plecak [back NMLZ] BACKPACK
woskowa wina [wax LE NMLZ] EARWAX
wiatrak [wind INS] WINDMILL

Russian (78): 103W / 43B / 10C

Base.AGT (1)
sapożnik [boot ADIZ] SHOEMAKER

Base.DIM (7)
ključica [key ADIZ] CLOTH BAG
zhereb enok [colt ADIZ] FOAL OR COLT
devuška [maiden ADIZ] GIRL
platok [piece of cloth ADIZ] SCARF
kurczak [goat ADIZ] GOAT
kamienny most [stone ADIZ bridge] BRIDGE

Base.F (3)
šećka [father in law ADIZ] MOTHER-IN-LAW OF A MAN
pelenjakica [nephew ADIZ] NIECE
koroljuka [king ADIZ] QUEEN

Base.M (1)
vdovets [widow ADIZ] WIDOWER

Base.RLT (2)
hleb [bread ADIZ] FOOD
plachta [sheet ADIZ] EMBROIDERY

Head Mod (1)
polden [half day] MIDDAY

Head Mod Gen (4)
mośka uz [lobe ear ADIZ] EARLOBE
palec nogi [finger toe ADIZ] TOE
nabor instrumentov [kit tools ADIZ] TOOLBOX
stolnova [trunk tree ADIZ] TREE TRUNK

Head PRP Mod (1)
jačički [diya ADIZ] TOOLBOX

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com.
The typology of binominal lexemes

Mod.ADJZ Head (20)

Mod.L.E.Head (1)
golen.ost.stop [shank.LE.foot] ANKLE

Slovak (83): 100W / 57B / 10C

Base.ADJZ.NLMZ (1)
želez.n.ico [iron.ADJZ.NMLZ] RAILWAY

Base.AGT (5)

Base.DIM (4)

Base.F (1)
kapr.ovn [king.F] QUEEN

Base.LOC (2)
ohn.isko [fire.LOC] FIREPLACE nemoc.nica [disease.LOC] HOSPITAL

Base.NLMZ (1)
večer.a [evening,NMLZ] DINNER ná.ús.nica [on.ear.NMLZ] EAR RING

Base.SUF (18)

Head Mod.GEN (1)
kmeh.strum.ú [trunk.tree.GEN] TREE TRUNK

Head PREP Mod (2)
pumpa na bicykel [pump for bicycle] BICYCLE PUMP skrinka na nástroje [box for tools] TOOLBOX

Mod.ADJZ Head (21)

(Japonic) Japanese

Japanese (42): 130W / 81B / 3C

Head Mod (1)
mi.ki [body.tree] TREE TRUNK

Mod Head (68)

374 Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 2019-01-01
piima.lehm [milk:GEN.cow] DAIRY COW
óhú.súök [evening:GEN.meal] DINNER
ukse.piit [door:GEN.post] DOORPOST
körra.niiv [ear:GEN.lobe] EARLOBE
körra.röngas [ear:GEN.ring] EARRING
körra.vahu [ear:GEN.wax] EARWAX
silma.laug [eye:GEN.lid] EYELID
kütte.kölle [Heating:GEN.hearth] FIREPLACE
önge.núör [hook:GEN.cord] FISHING LINE
kirbutarga [flea:GEN.market] FLEA MARKET
sörme.jälg [finger:GEN.print] FOOTPRINT
käe.kort [hand:GEN.flag] HANDBAG
vömr.sönú [key:GEN.word] KEYWORD
kitse.tall [goat:GEN.kid] KID
numbr.i.märk [number:GEN.mark] LICENSE PLATE
linnu.tee [bird:GEN.wax] MILK WAY
tiina.mau [father:GEN.land] NATIVE COUNTRY
venn.tútor [brother:GEN.daughter] NIECE
peo.pesa [palm:GEN.next] PALM OF HAND
änblíku.vörk [spider:GEN.web] SPIDER WEB
kivi.síld [stone:GEN.bridge] STONE BRIDGE
subhra.roog [sugar:GEN.cane] SUGAR CANE
óthu.súök [evening:GEN.flag] SUPPER
tööriista.kast [tool:GEN.box] TOOLBOX
hamba.hari [tooth:GEN.brush] TOOTHBRUSH
puu.triivi [tree:GEN.trunk] TREE TRUNK
vee.pump [water:GEN.pump] WATER PUMP
tuule.veski [wind:GEN.mill] WINDMILL
Mod.NML.Z Head (1)
mesi.las.vaha [honey:GEN.wax] BEESWAX

Finnish (26): 100W / 57B / 6C

Base.F (1)
kuninga.tar [king.F] QUEEN
Base.NML.Z (2)
käsi.ne [hand:GEN.wax] GLOVE
isik.níät [father:GEN.LZ] HOST

Head Mod (1)
kesi.páivi [middle.day] MIDDAY

Head.Mod (1)
kesi.viikko [middle.week] WEDNESDAY

Mod.Gen Head (11)
revo.n.tulet [fox:GEN.fire:PL] ARCTIC LIGHTS
pyörä.n.pumpu [bicycle:GEN.pump] BICYCLE PUMP
ov.n.pielit [door:GEN.post] DOORPOST
korva.n.ipukka [ear:GEN.tip] EARLOBE
jala.n.jälki [foot:GEN.trace] FOOTPRINT
linnu.n.rata [bird:GEN.track] MILK WAY
velje.n.tytär [brother:GEN.daughter] NIECE
sate.n.kuari [rain:GEN.bow] RAINBOW
hämähäkki.n.verkko [spider:GEN.net] SPIDER WEB
puu.n.runko [tree:GEN.trunk] TREE TRUNK
muna.keltoinen [egg:GEN.yolk] YOLK

dörra.reppu [back:bag] BACKPACK
mehiläis.pesä [bee.net] BEEHIVE
mehiläis.vaha [bee.wax] BEESWAX
ranne.rengas [wrist:ring] BRACELET
puu.seppä [tree:smith] CARPENTER
solis.lau [??,bone] COLLARBONE
kenttä.keitto [field.kitchen] COOKHOUSE
korva.rengas [ear:GEN.ring] EARRING
korva.vahka [ear:GEN.wax] EARWAX
kulma.karva [brow:hair] EYEBROW
silmä.riipi [eye.lash] EYELASH
silmä.luomi [eye:lid] EYELID
kirppu.tori [flea:GEN.market] FLEA MARKET
kulta.sormus [gold:ring] GOLD RING
käsi.jarru [hand.brace] HAND BRAKE
käsi.laukku [hand:bag] HANDBAG
nääl.liina [nose:kerchief] HANDKERCHIEF OR RAG
avain.sana [key:word] KEYWORD
posti.lauttiko [post:box] MAIL BOX
koti.ma [home.land] NATIVE COUNTRY
niipi.rattas [blade.wheel] PADDLE WHEEL
posti.kortti [post:card] POSTCARD
ranta.tie [iron.road] RAILWAY
kylki.luu [side:bone] RIB
lapa.luu [shoulder:bone] SHOULDERBLADE
pää.kallo [head:skull] SKULL
silmä.lasi [eye:glass.PL] SPECTACLES/GLASSES
selkä.runka [back:stem] SPINE
hevos.tall [horse:stalk] STABLE OR STALL
kivi.silta [stone:bridge] STONE BRIDGE
sokeri.ruoko [sugar.cane] SUGAR CANE
iltu.pala [evening:piece] SUPPER
olk.katto [straw:roof] THATCH
vesa.paperi [toilet:paper] TOILET PAPER
työ.kalu [work:thing] TOOL
työkulu.pakki [tool:case] TOOLBOX
hammas.harja [tooth:brush] TOOTHBRUSH
veri.suoni [blood:vein] VEIN OR ARTERY
viini.köynnös [wine:creep] VINE
vest.pumppu [water:water] WATER PUMP
tuuli.mylly [wind:mill] WINDMILL

(Uralic) Hungarian

Hungarian (37): 105W / 57B / 6C

Base.F (1)
király.ne [king.F] QUEEN
Base.NML.Z (3)
ország.os [table:GEN.NML.Z] CARPENTER
balакс [fish:GEN.NML.Z] FISHERMAN
cipő.esz [shoe:GEN.NML.Z] SHOEMAKER

Base.PROP (2)
csorda.s [herd.PROP] HERDSMAN
fazék.as [pot:PROP] POTTER

Mod.Head (42)
mezsz.hat [bee.carriage] BEEHIVE
mezsz.vízsz [bee.wax] BEESWAX
bicikli.pumpa [bicycle:pump] BICYCLE PUMP
törzs.főnök [tribe.chief] CHIEFTAIN
kulcs.csont [key:bone] COLLARBONE
ajtó.keret [door:frame] DOORPOST
fül.címa [ear:pin] EARLOBE
fül.sós [ear:fat] EARWAX
szem.öldök [eye.??] EYEBROW
szem.pilla [eye.??] EYELASH
szem.hőf [eye:shell] EYELID
üz.hey [fire:place] FIREPLACE
bolha.piac [flea:market] FLEA MARKET
idő.bat [leg:trace] FOOTPRINT
arany.sárga [gold:ring] GOLD RING
zseb.kendő [pocket:shawl] HANDKERCHIEF OR RAG
kórház [disease:house] HOSPITAL
kulcs.szo [key:word] KEYWORD
kocske.gida [goat:kid] KID
rendszám.tabla [license:gen.plate] LICENSE PLATE
posta.lída [post:box] MAIL BOX
filet.útt [milk:way] MILK WAY
szőlő.polg [parent:earth] NATIVE COUNTRY
unok.kid [grandchild:younger_sister] NIECE
mell.bimbo [breast:sprout] NIPPLE OR TEAT
orr.lyuk [nose:hole] NOSTIL
király.nő [king:woman] QUEEN
vas.útt [iron:road] RAILWAY

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 377
The typology of binominal lexemes

Western Mari (59): 93W / 43B / 3C

Mod Head (41)

kid.šo [hand.?] BRACELET
där kākč [morning meal] BREAKFAST
keld.e.iu [seine.bone] COLLARBONE
dāš kākč [evening meal] DINNER
amasa kōsāk [door post] DOORPOST
pōčč mēčas [ear end] EARLOBE
dāš,i kāngā [ear.coin] EARRING
pāčč kā [ear wax] EARWAX
sēngča.xal [eye.brow] EYEBROW
sēngča.pšn [eye.hair] EYELASH
sēngča.komđst [eye.cover] EYELID
tōl vacak [firePLACE] FIREPLACE
āŋgor sōrdō [river thread] FISHING LINE
tōl salim [fire flame] FLAME
sōmčnāt sārgāi [gold ring] GOLD RING
kid tormoc [hand brake] HAND BRAKE
kēšč īq [goat baby animal] KID
kehrčvāl kākč [midday meal] LUNCH
kē morbō kānā [bird goose road] MILKY WAY
ner raž [nose.?] NOSTRIL
kid lapa [hand pad] PALM OF HAND
cōrsok māstur [pot master] POTTER
kēnti kōrnā [iron road] RAILWAY
tōrēš liu [side.bone] RIB
kem şqroqš [boot shoemaker] SHOEMAKER
pūšč savala [shoulder.spoon] SHOULDERBLADE
vui.karka [head.scoop] SKULL
āŋgōremšk vad [spider.cauldon] SPIDER WEB
up.śno [back.core] SPINE
inni saraj [horse barn] STABLE OR STALL
dā kāvēr [stone bridge] STONE BRIDGE
saxar trošnū [sugar cane] SUGAR CANE
dāš kākč [evening meal] SUPPER
sēngča.vd [eye.water] TEAR
jai.lavn′a [foot.finger] TOE
šāštāx māsād koroplā [make.NMLZ thing box] TOOLBOX
puščang ṭing [tree trunk] TREE TRUNK
vōr,šā [blood.?] VEIN OR ARTERY
vinogrud padšaŋgā [grape vine] VINE
vōr,kuš [blood.?],day WEDNESDAY
kid.piūn [hand.?] WRIST

Mod.GEN Head (1)
māščn savā [egg.GEN yellow] YOLK

Mod.LAT Head (1)
šand ek pulmaga [toilet.LAT paper] TOILET PAPER

(Uralic) Saami

Kildin Sami (82): 87W / 35B / 4C

Mod Head (26)
stēl'kk vuss [back sack] BACKPACK
veazvuk-kiess [wasp-nest] BEEHIVE
veazvuk-țārō [wasp-tar] BEESWAX
vuell'-kax's [shoulder-bone] COLLARBONE
piejv-værr [day-soup] DINNER
piędli-kiedl'k [ear.stone] EARLOBE
cāll'm-rūmta [eye-brim] EYELID
kōll's nārmas [gold ring] GOLD RING
kād vuss [hand bag] HANDBAG
noso-rripp ex [nose.kerchief] HANDBKERCHIEF OR RAG
piejv-pierr [day-meal] LUNCH
poččuāk [post box] MAIL BOX
nīŋgless jēh'ēś [female horse] MARE
piejv-kessk [day-middle] MIDDAY
tāss't tōl [star fire] MILKY WAY
rū̂v't čuekas [iron road] RAILWAY
fěr mes' jākk's [thunder.bow] RAINBOW
jēhr t-tāxz' [flank.bone] RIB
vuejv-kālī [head-cap] SKULL
kiedl'k mōst [stone bridge] STONE BRIDGE
saxar kōrbā [sugar cane] SUGAR CANE
jāľl'k-eiph [foot.toe] TOE
vērr-siān [blood-thread] VEIN OR ARTERY
vinn-'mür [vine.tree] VINE
liess-k-kăll' 'es [widow-old man] WIDOWER
kēd-lāhp [handsole] WRIST

Mod Head.DIM (3)
jēb'ēś all'.k'a [horse.son.DIM] FOAL OR Colt
howas.a.all'.k'a [goat.DIM.son] DIM
lām'pes'.all'.k [sheep.DIM] LAMB

Mod.ATTR Head (3)
jōc.es' pierr [morning.ATTR.meal] BREAKFAST
mājši.jeś' fijjhm [milk.ATTR cow] DAIRY COW
jēk'ex [goat] [evening.ATTR.meal] SUPPER

Mod.GEN.Head (3)
tōl,nūxčes' [fire.GEN:SG tongue] FLAME
mašina nūm'ér [car.GEN number] LICENSE PLATE
sox.e sājim [spider.GEN:SG net] SPIDER WEB

Yeniseian

Ket (47): 70W / 29B / 3C

Mod Head (16)
hìk.dài [male.child] BOY
ili.tāb [arm.ring] BRACELET
okde.kāb [ear.point.end] EARLOBE
bul.sej [leg.place] FOOTPRINT
bul.quāk [leg.way] FOOTPRINT
qīn.dài [woman.child] GIRL
hāp.kōn [female.horse] MARE
Alba kāj [Alba hunting_trail] MILKY WAY
qīn.am [grandmother.mother] MOTHER-IN-LAW (OF A MAN)
qay.qiım [kwan.woman] QUEEN
ebq'[qo' [throat.path] RAINBOW
dō'y'ol [brain.covering] SKULL
se'n.qu'is [reindeer.PL:tent] STABLE OR STALL
ho'q.d.oks baž [excrement.POSS.stick place] TOILET
Appendix E. Binominal data set

Oceania/SE Asia

(Austro-Asiatic) Asian

Ceq Wong (17): 38W / 20B / 1C
Head Mod (20)
səraŋ wæŋ [nest bee] BEEHIVE
səraŋ wæŋ [nest bee] BEESWAX
jœg galoŋ [bone base_of_cassava_press] COLLABORATE
jŋâlaŋ pintuŋ’ [wood doorway] DOORPOST
haleŋ hiːn‡ay [leaf ear] EARLOBE
ʒuːc hntaŋ [excreta ear] EARWAX
sək mɛr [hair eye] EYELASH
hoŋ mɛt [skin eye] EYELID
cɪnɛn mɛt [ring gold] GOLD RING
hyâ? spîtal [spinal hospital] HOSPITAL
pɔwɔ naɔ'/ [middle day] MIDDAY
darâŋ mɔʔ [hole nose] NOSTRIL
tapak cas [palm hand] PALM OF HAND
bâray talon [shadow snake] RAINBOW
joŋŋy bahoŋ? [bone shoulder] SHOULDERBLADE
joŋŋy kyem [bone brain] SKULL
sɔm cɔmbiŋ [nest spider] SPIDER WEB
joŋŋy hŋkŋy [bone back] SPINE
tum mɛt [water eye] TEAR
wɔŋ cən [offspring toe] TOE

(Austro-Asiatic) Vietic

Vietnamese (95): 85W / 49B / 2C
Head Mod (41)
cố chân [neck leg] ANKLE
sò ong [nest bee] BEEHIVE
sòp ong [wax bee] BEESWAX
thơ rèn [skilled_laborer forge] BLACKSMITH
con trai [child boy] BOY
vòŋ tạig [circle hand] BRACELET
thơ mêc [skilled_laborer wood] CARPENTER
gà tròng [chicken male] COCK/ROOSTER
xuông dòn [bone lever] COLLABORATE
nha bếp [house stove] COOKHOUSE
giɔŋg ciɔ [scaffolding door] DOORPOST
trãi tạig [fruit ear] EARLOBE
hoà tạig [flower ear] EARRING
lông mœ [body hair eyebrow] EYELASH
ngοŋ li(da) [peak fire] FLAME
dâu châm [mark foot] FOOTPRINT
con gái [child female] GIRL
bào tạig [bag hand] GLOVE
khịn tạig [small_cloth hand] HANDKERCHIEF OR RAG
cha nhà [master house] HOST

(Austronesian) Formosan

Puyuma (73): 54W / 3B / 3C
3SG.Head DEF Mod (1)
tu-bira? kana tontiŋa [3SG-carlrobe DEF:OBL car] EARLOBE
Mod.PERF.Head (1)
din>apul-an [foot<PERF>LOC] FOOTPRINT
TMP.Base.LOC (1)
kà-laik-an [TMP-lunch-LOC] MIDDAY

(Austronesian) Greater Barito

Malagasy (71): 89W / 58B / 6C
AGT.Base (6)
mpa:mára[AGT. carpentry] CARPENTER
mpa:jono [AGT.set] FISHERMAN
mpa:hãko [AGT.game/produce_of_hunting] FISHERMAN
mpa:ntana [AGT.set] FISHERMAN
mpa:mo[v] AGT.witchcraft] SORcerer OR witch

Head Mod (31)
řiœv, tolehy [mother, honey] BEE
mpa:nø̄vy [moulder/maker iron] BLACKSMITH
žãsa, lûhî [child, male] BOY
sakô̄fо marãina [meal morning] BREAKFAST
akô̄hî lûhî [chicken male] COCK/ROOSTER
mâŋa, máso [? eye] EYEBROW
volo, máso [hair eye] EYELASH
hôdîtra, máso [skin eye] EYELID
lôfà, lôfà [tongue/blade, fire] FLAME
žãna, sos-volume [child/young horse] FOAL OR COLT
žãsa, vîvy [child, female] GIRL
žãnak’, õsy [offspring, goat] KID
žãna,kôndry [offspring, sheep] LAMB
sakô̄fо antoandro [meal noon] LUNCH
sâvoạ vîvy [horse female] MARE
rañfaza,vîvy [parent_in_law,female] MOTHER-IN-LAW (OF A MAN)

sûl.âp [blood, rope] VEIN OR ARTERY
la'r a'd [wrist, bone] WRIST

Hang só [board number] LICENSE PLATE
nguà cûi [horse female] MARE
buôl trẹa [time_of_day late_day] MIDDAY
mẹ vọr [mother wife] MOTHER-IN-LAW (OF A MAN)
hãŋg wam [row hamlet] NEIGHBOUR
châu gãi [the_sibling’ s_or_child’ s child female] NIECE
lố mût [hole nose] NOSTRIL
thor góm [skilled_laborer pottery] POTTER
đông sâi [road iron] RAILWAY
đông xe liéa [road train] RAILWAY
xuông sâm [bone rib] RIB
xuông vai [bone shoulder] SHOULDERBLADE
dâi ôc [brain head] SKULL
phà thi [talisman water] SORcerer OR WITCH
mùng mën [web spider] SPIDER WEB
xuông sông [bone life] SPINE
ngôn chà [finger foot] TOE
xe làa [vehicle fire] TRAIN
góc cãy [base tree] TREE TRUNK
cô y kho [tree grape] VINE
cô tay [neck hand] WRIST

Mod Head (8)
côc quâng [pole light] ARCTIC LIGHTS
nông dân [agriculture person] FARMER
mu thàt [ghost art] MAGIC
tôi quà [ancestor nation] NATIVE COUNTRY
hau thiệp [post card] POSTCARD
níh hòang [female emperor] QUEEN
chu nhát [lord day] SUNDAY
hoà xa [vehicle fire] TRAIN

Mod Head (20)

Mod Head (12)
more d bôk [sea GEN fire] ARCTIC LIGHTS
dôŋ da qoʔ? [dog GEN trail] MILKY WAY
ma/m.d kib [breast, GEN tip] NIPPLE OR TEAT
olm. d qãk [nose, GEN hole] NOSTRIL
dsə d baiŋ [river, GEN land] SHORE
elim.d unay [spider, GEN net] SPIDER WEB
dês.d til [eye, GEN water] TEAR

Mod Head (31)

Mod Head (8)
The typology of binominal lexemes
The typology of binominal lexemes

**Mod Head (28)**

jiang4.mai2.a-1 [ocean.horse.DIM] BICYCLE
e1v3.la2.kua3-1 [iron.craftsman] BLACKSMITH
tsx3.pan3-1 [son.child] BOY
s3v3.la2.kua3-1 [tree.craftsman] CARPENTER
n1k³.ku³.zi1-1 [ear.sh.1] EARWAX
n1h³.m3-1 [eye.hair] EYEBROW
n1p³.m3-1-1 [eyelid.hair] EYELASH
n1h³.p3-1 [eye.skin] EYELID
so1.ti1-1 [hand.cover] GLOVE
k3h3.so1.nio²¹-1 [gold.hand.ring] GOLD RING
ts³.ku³-1 [lunch.moment] MIDDAY
mi1.ts³-1 [breast/head] NIPPLE OR TEAT
pi³.k³.k³-1 [nose.hole] NOSE
so³.ti1-1 [hand.heart] PALM OF HAND
k³y³.ja²¹-1 [sky.old.woman] QUEEN
e1³.h³-1 [iron.road] RAILWAY
s1³.p³-1-1 [water.side] SHORE
fi1.tan²¹-1 [eye.mirror] SPECTACLES/GLASSES
mi³.ts³-1 [horse.pon] STABLE OR STALL
tsa³.k³-1 [stone.bridge] STONE BRIDGE
k³n³.ti³.k³-1-1 [star.period] SUNDAY
ni³.s³-1 [eye.water] TEAR
k³³.ts³-1-1 [foot.finger/toe] TOE
tse³.ja³-1 [fees.pool] TOILET
ja³.su¹-1-1 [tooth.brush] TOOTHBRUSH
so³.ts³-1 [tree.root] TREE TRUNK
ta³.pi³-1.ka²¹.s³-1 [grape.tree] VINE
so³.ti³-1 [hand.section/joint] WRIST

**Sino-Tibetan**

**Bodic**

**Manange (66): 61W / 21B / 4C**

Head.LOC.Mod (1)

**Mod Head (18)**

3pale koko [leg/foot ankle] ANKLE
4k³.e2napray [honey.flying.insect] BEE
2naka 4palle [chicken.rooster] COCK/ROOSTER
2muniste 3siki (evening.food) DINNER
1yjemps to [ear.tail] EARLOBE
1jima ko¹veg [ear.earring/jewel?] EARRING
2mi 2mits'e [eye.?] EYEBROW
1ja 2fa [hand.cover] GLOVE
4nê.kay [medicine.house] HOSPITAL
4k³.ju 1sitsa [sheep.young.animal] LAMB
2xx 3m³.o [horse.female] MARE
3gjoko 2p¹ug [breast.egg] NIPPLE OR TEAT
1nx.k³y³u [nose.hole] NOSE
4p³lí 3tsay [spider.nest/bed] SPIDER WEB
1m.ku [eye.tear/water] TEAR
1ja 1p³u [hand.thumb] THUMB
3pale 3t³ [foot/leg finger] TOE
2ji3sung 4ru¹ [tree.base/stump] TREE TRUNK

**Mod Head.DET (1)**

2mare titi.ko [door.side.DET] DOORPOST

**Mod.GEN.Head (1)**

1uno.-ts³ami [elder.sister.GEN.daughter] NIECE

**Sino-Tibetan**

**Macro-Bai**

**Calija (10): 66W / 30B / 2C**

Head Mod (2)

k³³.p³-1 [chicken.male] COCK/ROOSTER
me³.ts³-1 [horse.sister] MARE

**Mandarin Chinese (15): 137W / 94B / 4C**

Base.SUF (3)

zh4.zi1 [post.pillar/column.SUF] DOORPOST
la2.zi1 [stone.SUF] FIREPLACE
ji1.zi1 [fossil.SUF] FOAL OR COLT

**Mod Head (88)**

bei4.buo1 [back.bag] BACKPACK
mi³.feng1 [honey.wax] BEE
feng1.woli1 [bee.wasp.comb/nest] BEELINE
feng1.chao2 [bee.wasp.nest] BEEHIVE
mi³.lu³ [honey.wax] BEESWAX
feng1.lu³ [bee.wasp.wax] BEESWAX
qi4.gong2 [gas.pump] BICYCLE PUMP
duan4.gong1 [forge.work] BLACKSMITH
nan3.bai2 [male.child] BOY
shou3.zhu2 [hand.brace] BRACELET
za3.ang1 [early.morning.meal/food] BREAKFAST
za3.fan4 [early.food] BREAKFAST
mu4.ji1m⁴ [wood.craftsman] CARPENTER
qi1.2.hung3 [chief.of.a.trIBE.seNIor/leader] CHEFTAIN
zi2.2.hung3 [clan.seNIor/leader] CHEFTAIN
gong1.ji1 [male.animal.hen] COCK/ROOSTER
xiōng1.ji1 [male.animal.chicken] COCK/ROOSTER
na1�.nu2 [milk.cow] DAIRY COW
wan3.fan4 [evening.meal] DINNER
er5.huo2 [ear.ring] EARRING
er3.sh1 [ear.excret/dung] EARWAX
yan3.mei2 [eye.brow] EYEBROW
me3.muo [eyebrow.body_hair] EYEBROW
ji2.muo2 [eyelash.body_hair] EYELASH
yan3.pu2 [eye.skin] EYELID
nong2.min2 [farmer/agriculture.people/representatives.of.certain.professions] FARMER

Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 2019-01-01
The typology of binominal lexemes

PNG/Australia

Gunwinyguan

Anindilyakwa (3): 47W / 16B / 5C

CI.POSS.Base (3)

n.env.m.alivka [35G:M_CI.POSS.foot] FOOTPRINT
a.nv.m.ningga [CI(neut).CI.POSS.joint] WRIST
yi.nv.m.a.dhangkwa [CI(masc).M_CI.POSS.flesh] YOLK

CI.POSS.Base (2)

e.nv.m.a'avrr [CI(neut).M_CI.POSS.CL(neut).wind] BICYCLE PUMP

e.nv.m.ensha [CI(neut).M_CI.POSS.CL(VEG).eye] SPECTACLES/GLASSES

Head Mod (1)
a.mukwa a.mguru [CI(neut).source CI(neut).fire] FIREPLACE

IPOSS.Head Mod (9)

ma.ma.yama m.ensha [CL(VEG).IPOSS.body hai

CI(VEG).eye] EYEBROW

ma.m.mukwa m.ensha [CL(VEG).IPOSS.body hai

CI(VEG).eye] EYELASH

ma.m.akuila m.ensha [CL(VEG).IPOSS.skin CL(VEG).eye] EYELID

a.m.ungena m.menbina [CL(neut).IPOSS.round

CL(neut).breast] NIPPLE OR TEAT
e.m.dhvrra e.mindha [CI(neut).IPOSS.hole CL(neut).nose] NOSTRIL

a.m.mengengawa a.aryrka [CL(neut).IPOSS.life/pulse

CI(neut).hand] PALM OF HAND

a.m.dhivrra a.rvngka [CL(neut).IPOSS.bone

CI(neut).head] SKULL

a.m.mamuna a.aryrka [CL(neut).IPOSS.round

CI(neut).hand] THUMB

a.m.mamuna a.likha [CL(neut).IPOSS.round CL(neut).foot] TOE

Mod IPOSS.Head (1)
yu.dhuvhilhu y.i.nv.m.a.dhurva [CL(masc).chest

CL(masc).M_CI.POSS.bone] COLLARBONE

Nuclear Torricelli

Srenge (56): 40W / 20B / 2C

Mod Head (19)
ala yolto [leg/foot eye] ANKLE

bibi m.wusona [bee testicles] BEEHIVE

ape ake [stomach mouth] BREAKFAST

karkar bana [chicken male._non_human] COCK/ROOSTER

ape ake [stomach mouth] DINNER

yagoldi pi [ear excreta] EARWAX

yolto wol [eye hair/fur/feather] EYEBROW

yolto i [eye ??] EYELID

ala mleks [leg/foot scar/scab] FOOTPRINT

srenge suku [child water] GIRL

biliwiz [spirit?skin] HANDKERCHIEF OR RAG

ape ake [stomach mouth] LUNCH

srenge suku [child water] MARKED WOMAN

mapa tinga [nose hole] NOSTRIL

wanda biningi [head bone] SKULL

bili wiro [spirit initiated_man] SORCERER OR WITCH

ape ake [stomach mouth] SUPPER

nungu yotim [hand thumb/big toe] THUMB

nungu awars [hand knot] WRIST

Mod Head:REDUP (1)
ala talmvam [leg/foot branch;RED] TOE

Walman (94): 48W / 16B / 1C

Mod Head (16)
kayal chkuel [foot eye] ANKLE

oputo yuukel [thing food] DINNER

trai tai [fruit ear] EARLOBE

NA NA EARWAX

NA NA EYEBROW

long mi [body_hair eyebrow] EYELASH

NA NA EYELID

bita an trea [meal eat noon] LUNCH

me vo [mother wife] MOTHER-IN-LAW [OF A MAN]

l o mi [hole nose] NOSTRIL

ngama oto [sun flat leaf] RAINBOW

du o ci [head brain] SKULL

oputo yuukel [thing food] SUPPER

motu chuto [finger/digit female] THUMB

ngon choa [finger foot] TOE

ci tyu [neck hand] WRIST

(Pama-Nyungan) Desert Nyungic

Gurinji (30): 53W / 8B / 5C

Base.INS (1)
warlu.waji [fire.INS] FIREPLACE

Base.SUF (1)
pirmiti.yi [side.SUF] RIB

Base.VBLZ.AGT (1)
waayurru.karra.aji [fishing_line.ACT.AGT] FISHERMAN

Mod Head (4)
jiti jarrini [nose hole] NOSTRIL

wartarn majai [hand stomach] PALM OF HAND

wartarn ngamanyi [hand mother] THUMB

jama ngamanyi [foot mother] TOE

Mod.DAT Head (1)
yawarta.wu marru [horse.DAT house] STABLE OR STALL

(Pama-Nyungan) Paman

Wik-Mungkan (96): 44W / 25B / 3C

Head Mod (2)
puk wut [child old_man] BOY

aat puul [place fathers_father] NATIVE COUNTRY

Head.DAT (1)
yuk mek.akanu [tree/thing eye.DAT] SPECTACLES/GLASSES

Mod Head (22)
tha' kont [foot prawn] ANKLE

man kaanch [neck bone] COLLARBONE

kon mak [ear stickiness] EARWAX

mee puu [eye palm] EYEBROW

mee' pench [eye body_hair] EYELASH

thun nganith [fire light] FLAME

tha' pem [foot track] FOOTPRINT

puk wanch [child woman] GIRL

piny-kench [fathers_sibling-mother_in-law] MOTHER-IN-LAW

[OF A MAN]

thuut me' [breast eye] NIPPLE OR TEAT

kad' uyaan [nose hole] NOSTRIL

ma' un [hand front] PALM OF HAND

ma' puu [hand palm] PALM OF HAND

ngang ek [heart shell] SHOULDERBLADE

kuchek kaanch [head bone] SKULL

ma' mung [hand small_of_back] SORCERER OR WITCH

pikpung kaanch [back bone] SPINE

mee' kam [eye juice] TEAR

tha' pu [foot child] TOE

yuk think [tree lower_back] TREE TRUNK

ma' ochungan [hand round_shell] WRIST

ma' kongk [hand nuts] WRIST
West Bomberai

Kalamang (48): 90W / 56B / 6C

Head DEM Mod (1)
los (wame) yar [bridge PROX stone] STONE BRIDGE

Head Mod (14)
pompa sepedu [pump bike] BICYCLE PUMP
tumun canam [child male] BOY
tukan ror [worker wood] CARPENTER
ganup palu [head tribe] CHIEFTAIN
kakok canam [chicken male] COCK/ROOSTER
sontum amdir [person garden] FARMER
tumun pas [child female] GIRL
tanggaranggar marau [ring gold] GOLD RING
muap yuon:nawari rin [food midday] LUNCH
lajarang pas [horse female] MARE
ketan an pas [inlaw.1SG:POSS female] MOTHER-IN-LAW (OF A MAN)
leit pas [king female] QUEEN
jiang paselet [web spider] SPIDER WEB
pompa air [pump water] WATER PUMP

Head Mod.POSS (6)
muap gosaa.un.kin [food evening.POSS] DINNER
nika war.kin [line fish.POSS] FISHING LINE
tas tan.kin [bag HAND.POSS] HANDBAG
sapu tan.kin [broom hand.POSS] HANDBRUSH OR RAG
konuk sarat.kin [box LETTER.POSS] MAIL BOX
muap gosaa.un.kin [food evening.POSS] SUPPER

Head Mod.REL (2)
kakok jago.to [chicken cock.ADJ3] COCK/ROOSTER
sontum war.ten [person sorcery.ADJ3] SOURCERER OR WITCH

Mod Head (19)
kor kasi [foot ?] ANKLE
wenawena eur [beet next] BEEHIVE
saun muap [evening food] DINNER
gangir neren [eye hair] EYELASH
din paras [fire embers] FLAME
kor.kom [foot?] FOOTPRINT
tan sarong [hand cover] GLOVE
goyool muap.et [afternoon food?] LUNCH
kalis tungiri [rain?] RAINBOW
kir.kang [side bone] RIB
os.ket [sand above?] SHORE
bekiem.kang [shoulder bone] SHOULDERS
nakal taok [head shell?] SKULL
som.kang-kang [person bone-RED] SKULL
suol.kang [back bone] SPINE
per. tam [water drops] TEAR
kor.parok [foot?] TOE
tor eruon [tree base/trunk] TREE TRUNK
tan kasir [hand?] WRIST

Mod Head.3POSS (14)
akhair kang.un [chest bone.3POSS] COLLARBONE
anggas ror.un [door wood.3POSS] DOORPOST
kelkam elau.un [ear under.3POSS] EARLOBE
kolmak kit.un [ear poop.3POSS] EARWAX
kanggir pal.un [eye skin.3POSS] EYEBROW
kanggir pal.un [eye skin.3POSS] EYELID
din topmat.un [fire place.3POSS] FIREPLACE
kuda tumun.un [horse child.3POSS] FOAL OR COLT
keve tara.un [house old_person.3POSS] HOUSE
dombu tumun.un [sheep child.3POSS] LAMB
mobol petel.un [car plate.3POSS] LICENSE PLATE
am bel.un [breast ?] ANKLE
bustang pos.un [nose hole.3POSS] NOSETRIL
tan el.un [hand under.3POSS] PALM OF HAND

North America

Athabaskan-Eyk-Tlingit

Navajo (63): 77W / 27B / 5C
3SG.Mod.Head (1)
’akèle.ts’iin [3SG:foot bone] ANKLE

Mod 3SG.Head (9)
tsís’ñii bi.ghan [bee 3SG:home] BEEHIVE
ntsís’ñii bi.jee [bee 3SG:resin] BRAINWAX
naal’ohóohi bi.č’a [chicken 3SG:male:NFE] COCK/ROOSTER
č’óó’_bee-hadléeshé bi.č’a [fishing_rod 3SG:rope] FISHING LINE
béeso bi.zís [money 3SG:bag] HANDBAG
béesó bi.ttim [iron 3SG:path] RAILWAY
na’asht’é’i bi.č’a [spider 3SG:string] SPIDER WEB
li’í bi.ghan [horse 3SG:home] STABLE OR STALL
tsn bi.tsín [tree 3SG:handle] TREE TRUNK

Mod INS Head (1)
tsé bee na’i’í [stone with_it bridge] STONE BRIDGE

Mod.Head (15)
’ajééh.ts’i [ear.wax] EARWAX
’andí.ts’in [eye.bone] EYEBROW
’andí.dz [eye?] EYELASH
’antí.ez [eye.sack/bag] EYELID
’kō.’k’eh [fire.place] FIREPLACE
kō.’k’eh [foot place] FOOTPRINT
’alá.jish [hand bag] GLOVE
li’í.ts’a [horse female] MARE
’abe.’úlahad [head tip:NFE] NIPPLE OR TEAT
’alá.tl’áh [hand.bottom] PALM OF HAND
’alá.t’áh [hand.bottom] PALM OF HAND
’álá.ts’é [rib bone] RIB
’ino.bqá [water.edge] SHORE
’ogqís.tsín [??:bone] SHOULDERBLADE
tsi.ts’in [head/hair bone] SKULL
’alá.tsín [hand bone] WRIST

Mod.I.E.Head (1)
ñak’ee.sh.to [eye_area.LIG:water] TEAR

( Eskimo-Aleut) Yupik

Central Yupik (24): 70W / 21B / 15C

Base.AQ3 (2)
áulír.zay [arm.AQ3] BRACELET
arnar.zay [woman.AQ3] GIRL

Base.AR(AQ) (1)
cuk.ar[to] [person.AR(AQ)] TOE

Base.CENG(AQ) (1)
quaka.cenga [waist.CENG(AQ)] BEE

Base.CUUN (1)
anuq.ssuun [wind.CUUN] WINDMILL

Base.IILITAQ (1)
tayah.zer [wrist.IILITAQ] BRACELET

Base.IRIN (1)
pingay.ir [three.IRIN] WEDNESDAY

Base.LEK (1)
emuk.ez [mother’s_milk/breast.LEK] NIPPLE OR TEAT

Base.LLEQ(1)
cinó.ileq [foots.string.LLEQ] ANKLE
kenu.ile [fire.LLEQ] FIREPLACE

Base.QLIQ (1)
ela.qil [outside?.QLIQ] NEIGHBOR

Base.QUQ (1)
epul.enuq [shaft.QUQ] TREE TRUNK

2019-01-01  Draft. Please do not cite. Comments to: pepper.steve@gmail.com. 385
The typology of binominal lexemes

Base.TAQ2 (3)
ataku.taq [evening.TAQ2] DINER
erneq.taq [later today.TAQ2] LUNCH
ataku.taq [evening.TAQ2] SUPPER

Base.UAQ (1)
pacig.uaq [gills.UAQ] NOSTRIL

Base.YAGAQ (1)
qungi.yagq [reindeer/sheep.YAGAQ] LAMB

Base.YAQ (1)
aglar.yaq [ridgepole.YAQ] RAINBOW

Mod.REL. Head.Poss (3)
erne.em quk.allr.a [day.REL.middle.PST.Poss:ABS] MID-DAY
tulukaru.um tanglar.allr.i [raven.REL] SNOWSHOE.PST.Poss:ABS MILKY WAY
imarpim.imeqii [sea.REL.shore.Poss:ABS] SHORE

(Mayan) Core Mayan

Q’teq’chi’ (46): 101W / 40B / 14C

3ERG.Head.M Mod (1)
x.kem aj am [3ERG.web.m spidér] SPIDER WEB

3ERG.Head Mod (20)
x.mapaq 33 [3ERG.palm.foot] ANKLE
x.ch’e’el kab’3 [3ERG.tree.ERG>bee] BEEHIVE
r.rochob kab’3 [3ERG.house.bee] BEEHIVE
x.ma’aj xam [3ERG.place fire] COOKHOUSE
x.k’ot xik [3ERG scrimmage ear] EARWAX
r.ismal u 33 [3ERG.hair.surface] EYELASH
r.is u 33 [3ERG.back.surface] EYELID
x.m’a’q jik [3ERG.place fire] FIREPLACE
r.al kawawu 33 [3ERG.offspring_of_horse female] POAL OR COLD
x.ta uq’3 [3ERG.covering hand] GLOVE
r.al yak [3ERG.offspring goat] KID
x.ma’kawawu 33 [3ERG.mother horse] MARE
r.alal ikam [3ERG.offspring_of_mother_uncle] NIECE
x.xab in as [F.daughter_of_mother_elder_brother] NIECE
r.al ch’u’ma’a’3 [3ERG.son.mature.mother] NIECE
x.xub in 33 [F.daughter_of_mother_younger_brother] NIECE
r.al uq’t [3ERG.nose_breast nipple or teat] FOAL OR COLD
x.xa’q jik [3ERG.place cow] STABLE OR STALL
x.y.xa’u 33 [3ERG.liquid.surface] TEAR
r.al uj qo [3ERG.offspring nose foot] TOE

3ERG.Head.DER. DER (1)
x.xa’i cax.il [3ERG.mother.woman.ERG>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)
ar.kah’om [AGT cornfield.AGT] FARMER

Base.ANTIP.NMLZ (1)
awas.in.el [secret.ANTIP.NMLZ] SORCERER OR WITCH

Base.DER (1)
jolom.ii [head.ERG>DER] CHIEFTAIN

Base.INS.NMLZ (1)
k’ub’al.taal [heartstone.INS.NMLZ] COOKHOUSE

Base.NMLZ (1)
k’otob’al [shit.NMLZ] TOILET

Head Mod (8)
kux oq [joint leg] ANKLE
b’uq’uq’ ch’ich’ [corn cob iron] BICYCLE
tz’ox kaxlan [male chicken] COCK/ROOSTER
sa’u’t [belly nose] NOSTRIL
sa’uq’ [belly hand] PALM OF HAND
tarjeet postal [card postal] POSTCARD

na’uq’m [mother hand] THUMB
kux uq’m [neck hand] WRIST

Head.UNPOSS 3ERG.Mod (1)
b’uq.el x.e’en tel [bone.UNPOSS 3ERG.upper_arm] SHOULDERBLADE

Head.UNPOSS Mod (1)
b’uq.el jolom [bone.UNPOSS head] SKULL

Mod Head (1)
assukr utz’aj’ [sugar ?] SUGAR CANE

Mod.ADTZ Head (1)
k’im.al kab’l [straw.ADTZ house] THATCH

Zinacantón Tzotzil (93): 64W / 28B / 4C

Head Mod (11)
ton pom [stone incense] BEE
chanul pom [bug incense] BEE
mishik’ pom [belly button incense] BEE
kuchhipa pom [pipe incense] BEE
nu pu pom [house incense] BEEHIVE
\’at’k’ok’ [penis fire] FLAME
\’ol k’ak’al [middle sun/day] MIDDAY
sano.te gido [road mucous] NOSTRIL
me’d’k’inbaal [mother drizzle] RAINBOW
hol’\’anima [head soul] SKULL
nen sar [mirror eye] SPECTACLES/GLASSES

Head Mod.IPOSS (14)
nuk’\’ok.kal [neck foot.IPOSS] ANKLE
tak’in chixin’il [metal ear.IPOSS] EARRING
shik’ sat’il [wing/fin.eye.IPOSS] EYELASH
\’av\’ok.al [jog foot.IPOSS] FIREPLACE
na’k’ob.al [house hand.IPOSS] GLOVE
ni\’\’ch’us’al [nose breast.IPOSS] NIPPLE OR TEAT
\’at’k’ob.al [inside hand.IPOSS] PALM OF HAND
pachomal\’k’ob.al [basket hand.IPOSS] PALM OF HAND
ch’il.te’t’il [wooden_frame.tree.IPOSS] RIB
ch’ich pat’il [thorn/spin/quill back.IPOSS] SPINE
me’d’k’ob.al [mother hand.IPOSS] THUMB
nuk’\’ok’al [nose foot.IPOSS] TOE
be ch’ich’el.il [road blood.IPOSS] VEIN OR ARTERY
nuk’\’k’ob’al [neck hand.IPOSS] WRIST

Head Mod (2)
shokan na [side house] NEIGHBOUR
me’d’umal vinik [poverty/miser/sad man] WIDOWER

Mod Head.IPOSS (1)
\’u2al bak.el [chapellet/roary bone.IPOSS] COLLARBONE

(Ohmanguean) Otopamean

Quer 仏atar Otomi (70): 134W / 33B / 3C

Head Mod (31)
sera kolmena [wax bee] BEESEWAX
fani.bąjä [horse.iron] BICYCLE
mede.oni [cock.chicken] COCK/ROOSTER
ndo’yo tiki [bone incense] COLLARBONE
\’mai goxt’i [post/pole door] DOORPOST
ndi.gu [beginning/base ear] EARLOBE
xi.da [body.hair.eye] EYELASH
xi.da [body_hair.eye] EYELID
nt’o’ye [sheath.hand] GLOVE
tsa.fini [female.horse] MARE
made.mpa [middle day] MIDDAY
do’ba [stone milk] NIPPLE OR TEAT
ok.yx.xi [hole.nose] NOSTRIL
\’be’hù́ hài [woman king] QUEEN
hù̂ri.dehe [bank/shore/edge.water] SHORE
nengi.dehe [bank/shore.water] SHORE
Appendix E. Binominal data set

(80): 61W / 16B / 8C
ABS.Mod.POSS.Head (1)
ha.m.t lteen [ABS.breast 3:POSS.opening/mouth] NIPPLE OR TEAT

Head Mod (4)
xiquixii ccam [child male] BOY
tootar ccam [male chicken] COCK/ROOSTER
cacy cmaam [female horse] MARE
qihehe cmaam [chief_person female] QUEEN

Mod Head (1)
xepe poosoj [sea line] FISHING LINE

Mod.POSS.Head (2)
pnalla i.me [honey_bee 3:POSS.abode] BEEHIVE
hehe i [plant 3:POSS.base] TREE/STUMP

Mod.POSS.Head (1)
xepe i.teel [sea 3:POSS.edge] SHORE

POSS.Mod.Head (1)
to i/cms [3:POSS.eye.??] EYEBROW

POSS.Mod.POSS.Head (5)
to i.pcz [3:POSS.eye.PL 3:POSS.??] EYELASH
to i.lmaa1 [3:POSS.eye 3:POSS.skin] EYELID
to i.pac [3:POSS.eye 3:POSS.back] EYELID
to i.l.auc [3:POSS.head 3:POSS.bone] SKULL

POSS:Head (1)
a.queect cmaam [3:POSS.parent_in_law female] MOTHER-IN-LAW (OF A MAN)

(Tupian) Tupi-Guarani

Mbyo Guaran 푀 (31): 61W / 33B / 6C

3.Mod ABL.Head (1)
h.exa.r.gua [3.eye ABL.NMLZ] SPECTACLES/GLASSES

3.Mod Head (4)
h.wixaa kunha [3.leader woman] QUEEN
h.exay [3.eye.water] TEAR
h.uguy rape [3.blood path] VEIN OR ARTERY
i.po.apy [3.hand.extremity] WRIST

Base.DIM (2)
ava.i [male.DIM] BOY
kunha.i [woman.DIM] GIRL

Mod Head (20)
po.ap' [hand.piece] ANKLE
ei.ru [honey.father] BEE
ei rapa' [honey sphere] BEEHIVE
i.raity [honey.next] BEESWAX
ara aro [large.bird male] COCK/ROOSTER
oke yi [door support] DOORPOST
nami.x [ear.tie] EARRING
exa.py [eye.extremity] EYEBROW
exa pepi [eye supérieur_part] EYELID
tata ypy [fire beginning] FIREPLACE
pinda x [hook cord] FISHING LINE
kavaju kunha [horse woman] MARE
kuaray mbyte [sun middle] MIDDAY
nip'i rape [tapir path] MILK WAY
kâ ypy [breast beginning] NIPPLE OR TEAT
po mbyte [hand middle] PALM OF HAND
yy rembe [water edge] SHORE
nhandu kya [spider hammock] SPIDER WEB
pyo apl [foot.nail] TOE

Mod Head.DIM (3)
kavaju ra'y.i [horse son.DIM] FOAL OR COLT
kavara ra'y.i [goat son.DIM] KID
ovexa ra'y.i [sheep son.DIM] LAMB

Mod Head.PST (3)
axâ kã.gue [shoulder bone] PASTE/SLAB
axâ kã.gue [shoulder bone] PASTE/SLAB

Mod Head.DIM (15)
mama jo'ara [bee house] BEEHIVE
pueta mayoa [door edge] DOORPOST
naka bwiita [ear secretion] EARWAX
mam.bosam [hands.bags] GLOVE
kaba' jamut [horse female] MARE
bwiita toosa [hand next] NATIVE COUNTRY
pipim.koba [breast.head] NIPPLE OR TEAT
yeku wojo'oria [nose hole] NOSE
bawe mayoa [sea edge] SHORE
koba ota [head bone] SKULL
joo o'ota [back bone] SPINE
mam.pusam [hand.finger] THUMB
bwe' a yokia pesam [big feet.finger] TOE
ayho we' [blood thread] VEIN OR ARTERY
kaba jaapsi [egg heart] EGG

Mod Head.APPL. (1)
mam betula.ram [hand plain.APPL] PALM OF HAND

South America

Araucanian

Mapudungun (5): 81W / 25B / 5C

Base.LOC (1)
kütral.w [fire.LOC] FIREPLACE
The typology of binominal lexemes

Base.NMLZ (2)
ruka.fe [house.NMLZ] CARPENTER
challwa.fe [fish.NMLZ] FISHERMAN

Head Mod (14)
we,che wentr [new.person man] BOY
llaal'ik pili [sea ear] EARLOBE
feri pilan [dirt ear] EARWAX
gen ruka [owner house] HOST
zono kawella [female horse] MARE
rangtantsi [middle day] MIDDAY
longko moyoy [head breast] NIPPLE OR TEAT
wechoc.yu [hole nose] NOSTRIL
pulay kawí [palm hand] PALM OF HAND
fütra changíll kawí [big finger/toe hand] THUMB
chängill ganyung [finger/toe foot] TOE
chill moltiñ [branched blood] VEIN OR ARTERY
troy kawí [joint hand] WRIST
choz karam [yellow egg] YOLK

Mod Head (7)
milia.winal.kawí [gold.round hand] GOLD RING
kawí,pirina [hand.brake] HAND BRAKE
malle hove [father's brother daughter] NIECE
tren.ripü [train.way] RAILWAY
kura.kaykuy [stone.bridge] STONE BRIDGE
ufas foki [grape plant] VINE
jantu wentr [widow man] WIDOWER

Mod.ESS.NONF Head (1)
füta.nge.n zomo [husband.ESS.NONF woman] MARRIED WOMAN

(Cariban) Guianan

Galibi Carib (11): 65W / 20B / 7C

Base.DEV (1)
kamisa.npo [piece_of_tissue.DEV] HANDKERCHIEF OR RAG

Base.DIM (1)
kapilita.mepo [goat.DIM] KID

Base.DIM.DEV (2)
kavale.mepo [horse.DIM.DEV] FOAL OR COLT
kapala.mepo [sheep.DIM.DEV] LAMB

Head Mod (8)
wano wett [biexcrement] BEEESWAX
amekun,un[ir] [un]=U+0268+ [wrist.base] BRACELET
eru1u.piupi [eye.skin] EYELID
wato aponi [fire place] FIREPLACE
ime nofi [son grandmother] MOTHER-IN-LAW OF A MAN
moyowai etawei [spider web] SPIDER WEB
aina yami [hand father] THUMB
weve winti [tree base] TREE TRUNK

Mod.Head.POSS (6)
puna sepele [ear lobe.POSS] EARLOBE
manutti poi [breast tip.POSS] NIPPLE OR TEAT
aina lo.li [hand flat.POSS] PALM OF HAND
upuow kawiyi [head calabash.POSS] SKULL
pupu sikkili [foot little.end.POSS] TOE
i’mo kami.li [egg yellow.POSS] YOLK

Mod.POSS Head (1)
puna.li wett [ear.POSS dirtiness] EARWAX

Mod.POSS Head.POSS (1)
emo.li sakila.li [nose.POSS aperture.POSS] NOSTRIL

Chibchan

Cab 退 car (13): 81W / 31B / 2C

Mod Head (24)
klá tali [foot spur] ANKLE
bukala ju [bee house] BEEHIVE
bulu ñu [bee excrement] BEEESWAX
óshkoro jàyiri [chicken male] COCK/ROOSTER
wá ká [face hair] EYEBROW
wóbla ká [eye hair] EYELASH
wá kiu [eye skin] EYELID
zhííjí wá [smoke road] FIREPLACE
bikólo kíchà [hook liana] FISHING LINE
yókó wá [fire fruit/ball] FLAME
kalwá yaba [horse child] FOAL OR COLT
yaba alákká [child woman] GIRL
due yaba [goat child] KID
ovejá yaba [sheep child] LAMB
kalwá alá [horse female] MARE
kúvá míg [sun half] MIDDAY
juýl yak [man mother_in law] MOTHER-IN-LAW OF A MAN
tsí bata [teat tip] NIPPLE OR TEAT
jula kí [hand underside] PALM OF HAND
kóhká wúlg [train road] RAILWAY
wá riá [eye liquid] TEAR
julásí mikó [finger grandmother] THUMB
tehélí sakú [knife mass] TOOL
pi kíchú [blood liana] VEIN OR ARTERY

Mod.Head.SPEC (7)
kúkó ohí [ear excrement:SPEC] EARWAX
bák chichí [shoulder bone:SPEC] SHOULDERBLADE
tókí chichí [head bone:SPEC] SKULL
kalwá ju.i [horse house:SPEC] STABLE OR STALL
hú ju.i [excrement house:SPEC] TOILET
kal wáckúchí [tree trunk:SPEC] TREE TRUNK
ava káli [grape tree:SPEC] VINE

Huitotoan

Murui Huitoto (38): 48W / 18B / 4C

Base.CL (10)
ui tiraí [eye.CL(hair) ] EYELASH
jitái.ho [adolescent.CL(FEM) ] GIRL
jiíí.ho [in_law.FEM] MOTHER-IN-LAW OF A MAN
enaze.ho [grandson/nephew.CL(FEM) ] NIECE
defo [nose.CL(cavity) ] NOSTRIL
moorui [father.CL(DAY)] SUNDAY
rúho.kái [woman.CL(stem) ] THUMB
eí.kái [foot.CL(stem) ] TOE
ra.o [thing.CL(flex)] VINE
onoví.kí [hand.CL(cluster)] WRIST

Base.CL.CL (1)
taizí.ko.ho [heel.CL(covr).CL(FEM)] ANKLE

Mod ANAPH.Head (3)
jefo i.goí [ear.CL(cavity) ANAPH.CL(leather)] EARLOBE
onoví i.ko [hand ANAPH.CL(covr)] GLOVE
jiíí.goí korai [head.CL(ovl) ANAPH.CL(skl)] SKULL

Mod Head (4)
irai fuue [fire mouth] FIREPLACE
rafue náma [celebration owner.CL(masc)] HOST
onoví jereí [hand inside] PALM OF HAND
iye.fuue [river mouth/edge] SHORE

Matacoan

Wichi (62): 82W / 36B / 8C

Base.AGT (2)
tshoyowu [animals.AGT] HERDSMAN
sapatwu [shoe.AGT] SHOEMAKER

Base.LOC (2)
kana.li [needle.LOC] SUGAR CANE
y’anekw.hi [excrement.LOC] TOILET
The typology of binominal lexemes

koósu.ma [skirt.AGT] GIRL

Head OF Mod (1)

fínga u fútu [finger of foot] TOE

Mod Head (29)

höni.wósu [bee.house] BEEHIVE
höni.kák [bee.excrement] BEESWAX
wómi.ganja [man.chicken] COCK/ROOSTER
dóó.pídi [door.tree] DOORPOST
jész.bíka [ear.mouth] EARLOBE
wójo.kóóó [eye.shell] EYELID
füji.kumian [fire.place] FIREPLACE
füji.tónjó [fire.tongue] FLAME
úsí.míí [horse.child] FOAL OR COLT
góótu andél [gold ring] GOLD RING
koósu.pi [cloth.piece] HANDKERCHIEF OR RAG
háti.wósu [pain.house] HOSPITAL
sikápu.míí [sheep.child] LAMB
sónu.háti [sun.heart] MIDDAY
delwí.písi [dairy.path] MILKY WAY
bóóbi.bíka [breast.mouth] THUMB
nújí.bóóku [nose.hole] NOSE
mújé.könu [woman.king] QUEEN
könu.mújé [king.woman] QUEEN
talân.fútu [train.foot] RAILWAY
télba.bíka [land.mouth] SHORE
hiéli.bóóku [head.bone] SKULL
andísí.wósu [spider.house] SPIDER WEB
bóóba.múndí [back.middle] SPINE
mboóti.wósu [animal.house] STABLE OR STALL
ndéití.táfa [night.table] SUPPER
wójo.wáta [eye.water] TEAR
mamá.fínga [mother.finger] THUMB
ránla bósí [tooth brush] TOOTHBRUSH

(Pidgins & Creoles) French-based

Seychelles Creole (16): 107W / 21B / 1C

Head Mod (21)

banan lalimyer artík [PL light arctic] ARCTIC LIGHTS
mous démyel [fly honey] BEE
kaka zorey [feces ear] EARWAX
lapo lizye [skin eye] EYELID
met lakour [master home] HOST
plak nýmero [plate number] LICENSE PLATE
bout tete [end breast] NIPPLE OR TEAT
trou nýmen [hole nose] NOSE
plak lánm [plate hand] PALM OF HAND
kart postal [card postal] POSTCARD
bor-lammer [edge-sea] SHORE
plak zepol [plate shoulder] SHOULDERBLADE
bonnonm dibwa [man woods] SORCERER OR WITCH
lakaz bib [house spider] SPIDER WEB
lezo leren [bone kidney] SPINE
kolonn vertebral [string spinal] SPINE
twa fey [roof straw] THATCH
ledwa lipye [finger foot] TOE
pouso.lipye [thumb foot] TOE
pye rzen [tree grape] VINE
zonn dzez [yellow egg] YOLK
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 4, with values such as \texttt{cmp}, \texttt{prp}, \texttt{gen}, etc. The two semantic types, \texttt{semType} and \texttt{semTypeH}, represent my revision of the system proposed by Hatcher and the original, respectively (see page 76).

```r
> str(nn)
'data.frame': 3556 obs. of 24 variables:
$ meaning : Factor w/ 100 levels "ankle","arctic lights",...: 1 1 1 1 1 1 1 1 ... $ stype : Factor w/ 27 levels "TAX","TAX2","COOR",...: 6 6 6 6 6 6 6 6 6 ... $ language: Factor w/ 99 levels "Äiwoo","Akkadian",...: 38 63 13 82 88 36 62 ... $ word : chr "mata kaki" "'akéts'iin" "klat' ta'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 1 2 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",...: 1 3 1 2 5 1 8 1 1 ... $ language2: Factor w/ 99 levels "Äiwoo","Akkadian",...: 38 63 13 82 88 36 62 ... $ area : Factor w/ 7 levels "A","E","O","G",...: 3 5 6 1 2 6 6 ... $ iso639 : Factor w/ 99 levels "akk","amh","aoi",...: 39 63 13 86 92 43 38 ... $ glottocode : Factor w/ 99 levels "akka1240","amha1245",...: 39 63 13 86 92 43 38 ... $ family : Factor w/ 37 levels "Afro-Asiatic",...: 7 4 10 5 2 21 15 7 16 18 ... $ genus : Factor w/ 66 levels "Adamawa-Ubangi",...: 39 4 13 7 62 43 30 47 ... $ latitude: chr "-7.33458" "36.2106" "9.67031" "-8.25605" ... $ longitude: chr "109.716" "-110.082" "-83.4102" "37.624" ... $ area2 : Factor w/ 7 levels "Africa","Eurasia",...: 3 5 6 1 2 6 6 ... $ htype : Factor w/ 5 levels "MisH","HinM",...: 2 2 2 2 2 ... $ semField : Factor w/ 16 levels "Agriculture and vegetation",...: 12 12 12 12 ... $ semType : Factor w/ 7 levels "person","animal",...: 4 4 4 4 4 4 4 ... $ semTypeH : Factor w/ 6 levels "person","animal",...: 3 3 3 3 3 3 ... $ ftype2 : Factor w/ 16 levels "adjL","adjR",...: 5 14 6 15 8 6 4 5 6 ...
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` contains abbreviated language names used for reasons of space in some figures in Chapter 6 (e.g. “Seych. Creole”).

```r
> str(l)
'data.frame': 99 obs. of 11 variables:
$ language : Factor w/ 99 levels "Äiwoo","Akkadian",..: 1 2 3 4 5 6 7 8 9 10 ...
$ language2: Factor w/ 99 levels "Äiwoo","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 1 1 1 1 1 ...
$ iso639 : Factor w/ 99 levels "akk","amh","aoi",..: 64 1 2 3 4 5 6 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",..: 47 54 54 27 37 32 40 ...
$ latitude : chr  "-10.2302" "33.1000" "11.7082" "-13.9999" ...
$ longitude : chr  "166.21" "44.1000" "39.5435" "136.6414" ...
$ 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 ...
```

The data frame `m` contains information relating to the 100 meanings. See above for the difference between `semType` and `semTypeH`.

```r
> str(m)
'data.frame': 100 obs. of 4 variables:
$ 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 1 1 ...
$ semTypeH: Factor w/ 7 levels "person","animal",..: 3 3 3 2 3 3 3 3 1 1 ...
```

The data frame `s` contains information relating to the two classifications of semantic relations (cf. Table 28 on page 200), with “B” and “H” standing for Bourque and Hatcher, respectively. Somewhat inconsistently, `stype` contains the codes for the Bourque 2.0 classification and `htype` those for the Hatcher 2.0 classification. The variable `atype` maps these two systems to the three classic associative relations identified by Aristotle.

```r
> str(s)
'data.frame': 27 obs. of 7 variables:
$ 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 3 3 2 3 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 3 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" ...
```

The typology of binominal lexemes

**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` contains abbreviated language names used for reasons of space in some figures in Chapter 6 (e.g. “Seych. Creole”).

```r
> str(l)
'data.frame': 99 obs. of 11 variables:
$ language : Factor w/ 99 levels "Äiwoo","Akkadian",..: 1 2 3 4 5 6 7 8 9 10 ...
$ language2: Factor w/ 99 levels "Äiwoo","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 1 1 1 1 1 ...
$ iso639 : Factor w/ 99 levels "akk","amh","aoi",..: 64 1 2 3 4 5 6 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",..: 47 54 54 27 37 32 40 ...
$ latitude : chr  "-10.2302" "33.1000" "11.7082" "-13.9999" ...
$ longitude : chr  "166.21" "44.1000" "39.5435" "136.6414" ...
$ 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)
'data.frame': 100 obs. of 4 variables:
$ 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 1 1 ...
$ semTypeH: Factor w/ 7 levels "person","animal",..: 3 3 3 2 3 3 3 3 1 1 ...
```

**semantic relation (s)**

The data frame `s` contains information relating to the two classifications of semantic relations (cf. Table 28 on page 200), with “B” and “H” standing for Bourque and Hatcher, respectively. Somewhat inconsistently, `stype` contains the codes for the Bourque 2.0 classification and `htype` those for the Hatcher 2.0 classification. The variable `atype` maps these two systems to the three classic associative relations identified by Aristotle.

```r
> str(s)
'data.frame': 27 obs. of 7 variables:
$ 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 3 3 2 3 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 3 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)

<table>
<thead>
<tr>
<th>Contact</th>
<th><a href="mailto:pepper.steve@gmail.com">pepper.steve@gmail.com</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Slide show</td>
<td>SLE 2016 presentation</td>
</tr>
<tr>
<td>Description</td>
<td>Project description</td>
</tr>
</tbody>
</table>

**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>FRENCH</td>
<td></td>
</tr>
<tr>
<td>railway</td>
<td>eisenbahn</td>
<td>eisen.bahn [iron.way]</td>
<td>GERMAN</td>
<td></td>
</tr>
<tr>
<td>railway</td>
<td>železnaja doroga</td>
<td>želez.naja doroga [iron.ADJZ road]</td>
<td>RUSSIAN</td>
<td></td>
</tr>
<tr>
<td>railway</td>
<td>reli</td>
<td>-</td>
<td>SWAHLI</td>
<td></td>
</tr>
<tr>
<td>railway</td>
<td>železnica</td>
<td>želez.n.ica [iron.ADJZ.NMLZ]</td>
<td>SLOVAK</td>
<td></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>лодыжка</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>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>colmen</td>
<td>ruche d’abeilles</td>
<td>beehive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>влос</td>
<td>oër</td>
<td>cire</td>
<td>beeswax</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>велот</td>
<td>bicic</td>
<td>vélo</td>
<td>bicycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>насос</td>
<td>bombom</td>
<td>pom</td>
<td>bicycle pump</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>кузнечик</td>
<td>herr</td>
<td>forg</td>
<td>blacksmith</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>мальч</td>
<td>chic</td>
<td>garç</td>
<td>boy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ключица</td>
<td>clavícula</td>
<td>clavicule</td>
<td>collarbone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>вострик</td>
<td>desa</td>
<td>petit</td>
<td>breakfast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>плот</td>
<td>carpl</td>
<td>char</td>
<td>carpenter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>вожкт</td>
<td>caciq</td>
<td>chef</td>
<td>chieftain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>петух</td>
<td>gall</td>
<td>coq</td>
<td>cock/rooster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ключ</td>
<td>clav</td>
<td>clavicule</td>
<td>collarbone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>камбуз</td>
<td>cocina</td>
<td>cuisine</td>
<td>cookhouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>дойная корова</td>
<td>vaca lechera</td>
<td>vache à lait</td>
<td>dairy cow</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>дверной косяк</td>
<td>jamba de puerta</td>
<td>jambage de porte</td>
<td>doorpost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ушная сера</td>
<td>cera de oído</td>
<td>cire d’oreille</td>
<td>earwax</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>бровь</td>
<td>ceja</td>
<td>sourcil</td>
<td>eyebrow</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ресница</td>
<td>pest</td>
<td>cil</td>
<td>eyelash</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>веко</td>
<td>párp</td>
<td>paup</td>
<td>eyelid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>фермер</td>
<td>agric</td>
<td>ferm</td>
<td>farmer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>камин</td>
<td>chimenea</td>
<td>cheminée</td>
<td>fireplace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>рыбак</td>
<td>pesc</td>
<td>pêche</td>
<td>fisherman</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>леска</td>
<td>seda</td>
<td>ligne</td>
<td>fishing line</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>пламя</td>
<td>llama</td>
<td>flam</td>
<td>flame</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 "δεκα.πέντε.α.σιλαβος" ‘fifteen-syllable metre’. The second “linking element” field was also used to record the POSS:3SG suffix of the Turkish izafet construction, as in "καλδηρής μηχανετής" [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>NO (NOR)</td>
</tr>
<tr>
<td>compound</td>
<td>orthographic form (Latin script)</td>
<td>arbeidsdag</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>arbeiđ</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>s</td>
</tr>
<tr>
<td>2nd const (C₂)</td>
<td>root form</td>
<td>dag</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>work+day(12 hours)=working-day</td>
</tr>
</tbody>
</table>

Table 57: Morbo/Comp database structure
The typology of binominal lexemes

Štekauer et al’s language sample

Cf. §2.2.2 on page 43. 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-Khoekhoe)</td>
<td>Japanese JPN (Japanese)</td>
</tr>
<tr>
<td>Datooga TCC (Niloctic)</td>
<td>Georgian KAT (Kartvelian)</td>
</tr>
<tr>
<td>Luo LUO (Niloctic)</td>
<td>Estonian EST (Finnic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>North America (N) 8</th>
<th>Eurasia (E) 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slave SCS (Athapaskan-Eyak-Tlingit)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Kalaallisut KAL (Eskimo-Aleut)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Zinacantán Tzotzil TZO (Core Mayan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Clallam CLM (Salishan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Lakota LKT (Siouan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>TotonacTKU (Totonacan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Pipil PPL (Southern Uto-Aztecan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Kwakiutl KWK (Wakashan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>South America (S) 4</th>
<th>Eurasia (E) 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maipure MAIP1246 (Arawakan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Jaqaru JQR (Aymara)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Wichí MZH (Matacoan)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
<tr>
<td>Movima MZP (Movima)</td>
<td>Hawaiian HAW (Austronesian)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southeast Asia &amp; Oceania (O) 9</th>
<th>Eurasia (E) 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vietnamese VIE (Vieitic)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Anejom ATY (Eastern Malayo-Polynesian)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Kumak NEE (Eastern Malayo-Polynesian)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Maori MRI (Eastern Malayo-Polynesian)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Indonesian IND (Malayo-Sumbawan)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Iloko ILO (Northern Luzon)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Karao KYJ (Northern Luzon)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Tibetan BOD (Bodic)</td>
<td>Korean KOR (Korean)</td>
</tr>
<tr>
<td>Mandarin Chinese CMN (Sinitic)</td>
<td>Korean KOR (Korean)</td>
</tr>
</tbody>
</table>

| South America | | Eurasia (E) 17 |
|---------------|----------------|
| Maipure MAIP1246 (Arawakan) | Hawaiian HAW (Austronesian) |
| Jaqaru JQR (Aymara) | Hawaiian HAW (Austronesian) |
| Wichí MZH (Matacoan) | Hawaiian HAW (Austronesian) |
| Movima MZP (Movima) | Hawaiian HAW (Austronesian) |

Table 58: Language sample (Štekauer, Valera & Körtvélyessy 2012)
Alphabetical list of the initial 201 meanings

Cf. §3.1.3 and §3.1.4 on page 69ff.

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 59: The original 201 meanings (alphabetical order)

<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 of binominal lexemes

<table>
<thead>
<tr>
<th>Merchant</th>
<th>15</th>
<th>Servant</th>
<th>11</th>
<th>Waterfall</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midday</td>
<td>39</td>
<td>Shoemaker</td>
<td>23</td>
<td>Weapons</td>
<td>12</td>
</tr>
<tr>
<td>Milky way</td>
<td>46</td>
<td>Shore</td>
<td>30</td>
<td>Wedding</td>
<td>11</td>
</tr>
<tr>
<td>Mistake</td>
<td>2</td>
<td>Shoulderblade</td>
<td>33</td>
<td>Wednesday</td>
<td>30</td>
</tr>
<tr>
<td>Mother-in-law</td>
<td>29</td>
<td>Sibling</td>
<td>10</td>
<td>West</td>
<td>2</td>
</tr>
<tr>
<td>Murder</td>
<td>4</td>
<td>Skull</td>
<td>45</td>
<td>Whetstone</td>
<td>9</td>
</tr>
<tr>
<td>Native country</td>
<td>46</td>
<td>Sorcerer or witch</td>
<td>35</td>
<td>Whirlpool</td>
<td>15</td>
</tr>
<tr>
<td>Necklace</td>
<td>17</td>
<td>Spectacles/glasses</td>
<td>31</td>
<td>Widow</td>
<td>13</td>
</tr>
<tr>
<td>Neighbour</td>
<td>18</td>
<td>Spider web</td>
<td>63</td>
<td>Widower</td>
<td>21</td>
</tr>
<tr>
<td>Niece</td>
<td>51</td>
<td>Spine</td>
<td>52</td>
<td>Windmill</td>
<td>32</td>
</tr>
<tr>
<td>Nipple or teat</td>
<td>50</td>
<td>Spring or well</td>
<td>11</td>
<td>Womb</td>
<td>22</td>
</tr>
<tr>
<td>Nostril</td>
<td>77</td>
<td>Stable or stall</td>
<td>29</td>
<td>Wrist</td>
<td>43</td>
</tr>
<tr>
<td>Nurse</td>
<td>6</td>
<td>Stepdaughter</td>
<td>13</td>
<td>Yolk</td>
<td>10</td>
</tr>
<tr>
<td>Nut</td>
<td>11</td>
<td>Stepparent</td>
<td>16</td>
<td>Young man</td>
<td>6</td>
</tr>
<tr>
<td>Old woman</td>
<td>5</td>
<td>Stepmother</td>
<td>15</td>
<td>Young woman</td>
<td>8</td>
</tr>
<tr>
<td>Older sister</td>
<td>6</td>
<td>Stepson</td>
<td>6</td>
<td>Younger brother</td>
<td>10</td>
</tr>
<tr>
<td>Paddle wheel</td>
<td>26</td>
<td>Stone bridge</td>
<td>40</td>
<td>Younger sister</td>
<td>14</td>
</tr>
<tr>
<td>Palm of hand</td>
<td>68</td>
<td>Stranger</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td>2</td>
<td>Sugar cane</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perjury</td>
<td>6</td>
<td>Sunday</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pestle</td>
<td>12</td>
<td>Supper</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pity</td>
<td>2</td>
<td>Swelling</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plaintiff</td>
<td>3</td>
<td>Tailor</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postage stamp</td>
<td>24</td>
<td>Teacher</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcard</td>
<td>55</td>
<td>Tear</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter</td>
<td>51</td>
<td>Thatch</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Praise</td>
<td>0</td>
<td>Thief</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostitute</td>
<td>11</td>
<td>Thumb</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pupil</td>
<td>9</td>
<td>Toe</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarrel</td>
<td>0</td>
<td>Toilet</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Queen</td>
<td>53</td>
<td>Toilet paper</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Railway</td>
<td>39</td>
<td>Tool</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainbow</td>
<td>33</td>
<td>Toolbox</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rape</td>
<td>2</td>
<td>Toothbrush</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Razor</td>
<td>8</td>
<td>Train</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remains</td>
<td>0</td>
<td>Tree trunk</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rib</td>
<td>26</td>
<td>Twins</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roof</td>
<td>24</td>
<td>Vein or artery</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>3</td>
<td>Victory</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screwdriver</td>
<td>4</td>
<td>Vine</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sculptor</td>
<td>21</td>
<td>Water pump</td>
<td>29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
List of 201 meanings (binominality order)

Cf. §3.1.3 and §3.1.4 on page 69ff.

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 60: 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>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nostril</td>
<td>77</td>
<td>32</td>
<td>footprint</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>earlobe</td>
<td>71</td>
<td>33</td>
<td>stone bridge</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>palm of hand</td>
<td>68</td>
<td>34</td>
<td>supper</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>arctic lights</td>
<td>65</td>
<td>35</td>
<td>midday</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>earwax</td>
<td>65</td>
<td>36</td>
<td>railway</td>
<td>39</td>
</tr>
<tr>
<td>6</td>
<td>spider web</td>
<td>63</td>
<td>37</td>
<td>thatch</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>license plate</td>
<td>60</td>
<td>38</td>
<td>mare</td>
<td>39</td>
</tr>
<tr>
<td>8</td>
<td>beeswax</td>
<td>58</td>
<td>39</td>
<td>fireplace</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>eyelid</td>
<td>56</td>
<td>40</td>
<td>earring</td>
<td>37</td>
</tr>
<tr>
<td>10</td>
<td>Sunday</td>
<td>56</td>
<td>41</td>
<td>ankle</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>postcard</td>
<td>55</td>
<td>42</td>
<td>farmer</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>glove</td>
<td>54</td>
<td>43</td>
<td>toothbrush</td>
<td>36</td>
</tr>
<tr>
<td>13</td>
<td>queen</td>
<td>53</td>
<td>44</td>
<td>vine</td>
<td>36</td>
</tr>
<tr>
<td>14</td>
<td>spine</td>
<td>52</td>
<td>45</td>
<td>mail box</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>toe</td>
<td>52</td>
<td>46</td>
<td>sorcerer or witch</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>potter</td>
<td>51</td>
<td>47</td>
<td>bracelet</td>
<td>34</td>
</tr>
<tr>
<td>17</td>
<td>niece</td>
<td>51</td>
<td>48</td>
<td>bicycle pump</td>
<td>33</td>
</tr>
<tr>
<td>18</td>
<td>nipple or teat</td>
<td>50</td>
<td>49</td>
<td>cock/rooster</td>
<td>33</td>
</tr>
<tr>
<td>19</td>
<td>collarbone</td>
<td>50</td>
<td>50</td>
<td>rainbow</td>
<td>33</td>
</tr>
<tr>
<td>20</td>
<td>milky way</td>
<td>46</td>
<td>51</td>
<td>toilet paper</td>
<td>33</td>
</tr>
<tr>
<td>21</td>
<td>beehive</td>
<td>46</td>
<td>52</td>
<td>toolbox</td>
<td>33</td>
</tr>
<tr>
<td>22</td>
<td>doorpost</td>
<td>46</td>
<td>53</td>
<td>fisherman</td>
<td>33</td>
</tr>
<tr>
<td>23</td>
<td>gold ring</td>
<td>46</td>
<td>54</td>
<td>fishing line</td>
<td>33</td>
</tr>
<tr>
<td>24</td>
<td>native country</td>
<td>46</td>
<td>55</td>
<td>shoulder blade</td>
<td>33</td>
</tr>
<tr>
<td>25</td>
<td>skull</td>
<td>45</td>
<td>56</td>
<td>thumb</td>
<td>32</td>
</tr>
<tr>
<td>26</td>
<td>wrist</td>
<td>43</td>
<td>57</td>
<td>windmill</td>
<td>32</td>
</tr>
<tr>
<td>27</td>
<td>dinner</td>
<td>42</td>
<td>58</td>
<td>girl</td>
<td>32</td>
</tr>
<tr>
<td>28</td>
<td>eyelash</td>
<td>42</td>
<td>59</td>
<td>train</td>
<td>31</td>
</tr>
<tr>
<td>29</td>
<td>tree trunk</td>
<td>42</td>
<td>60</td>
<td>carpenter</td>
<td>31</td>
</tr>
<tr>
<td>30</td>
<td>kid</td>
<td>41</td>
<td>61</td>
<td>spectacles/glasses</td>
<td>31</td>
</tr>
<tr>
<td>31</td>
<td>host</td>
<td>40</td>
<td>62</td>
<td>tear</td>
<td>31</td>
</tr>
</tbody>
</table>

2019-01-01   Draft. Please do not cite. Comments to: pepper.steve@gmail.com.
The typology of binominal lexemes

| 94 | sugar cane | 22 | 130 | prostitute | 11 | 166 | screwdriver | 4 |
| 95 | tool | 21 | 131 | spring or well | 11 | 167 | teacher | 4 |
| 96 | widower | 21 | 132 | servant | 11 | 168 | bad luck | 4 |
| 97 | hospital | 21 | 133 | nut | 11 | 169 | murder | 4 |
| 98 | chieftain | 21 | 134 | wedding | 11 | 170 | deceit | 4 |
| 99 | bicycle | 21 | 135 | married man | 10 | 171 | east | 4 |
| 100 | sculptor | 21 | 136 | sibling | 10 | 172 | disease | 4 |
| 101 | bee | 21 | 137 | younger brother | 10 | 173 | school | 3 |
| 102 | lipstick | 20 | 138 | yolk | 10 | 174 | grief | 3 |
| 103 | married woman | 20 | 139 | beggar | 9 | 175 | dawn | 3 |
| 104 | bow tie | 20 | 140 | airplane | 9 | 176 | freeman | 3 |
| 105 | magic | 19 | 141 | descendants | 9 | 177 | plaintiff | 3 |
| 106 | firewood | 19 | 142 | pupil | 9 | 178 | election | 2 |
| 107 | calf | 19 | 143 | intention | 9 | 179 | rape | 2 |
| 108 | neighbour | 18 | 144 | arson | 9 | 180 | earthquake | 2 |
| 109 | necklace | 17 | 145 | whetstone | 9 | 181 | parents | 2 |
| 110 | toilet | 17 | 146 | young woman | 8 | 182 | drink | 2 |
| 111 | bark | 17 | 147 | hummingbird | 8 | 183 | twins | 2 |
| 112 | stepfather | 16 | 148 | ancestors | 8 | 184 | mistake | 2 |
| 113 | capital city | 16 | 149 | razor | 8 | 185 | west | 2 |
| 114 | meeting house | 15 | 150 | waterfall | 7 | 186 | bruise | 2 |
| 115 | whirlpool | 15 | 151 | tailor | 7 | 187 | stranger | 2 |
| 116 | cave | 15 | 152 | judgment | 7 | 188 | pity | 2 |
| 117 | merchant | 15 | 153 | defendant | 7 | 189 | beginning | 2 |
| 118 | stepmother | 15 | 154 | young man | 6 | 190 | idea | 2 |
| 119 | armpit | 14 | 155 | stepson | 6 | 191 | food | 2 |
| 120 | younger sister | 14 | 156 | envy | 6 | 192 | anger | 2 |
| 121 | widow | 13 | 157 | nurse | 6 | 193 | darkness | 0 |
| 122 | stepdaughter | 13 | 158 | older sister | 6 | 194 | defeat | 0 |
| 123 | captive | 13 | 159 | perjury | 6 | 195 | divorce | 0 |
| 124 | guard | 12 | 160 | end | 5 | 196 | itch | 0 |
| 125 | pestle | 12 | 161 | old woman | 5 | 197 | praise | 0 |
| 126 | afternoon | 12 | 162 | meal | 5 | 198 | quarrel | 0 |
| 127 | weapons | 12 | 163 | thief | 5 | 199 | remains | 0 |
| 128 | birth certificate | 11 | 164 | crowd | 5 | 200 | swelling | 0 |
| 129 | baby | 11 | 165 | anxiety | 4 | 201 | victory | 0 |
Lost constructions (84 meaning sample)

Cf. §3.1.4 on page 69ff.

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. Those marked with a dagger (†) did not make it into the final data set based on 100 meanings.

*Table 61: Constructions lost with a sample of 84 meanings*

<table>
<thead>
<tr>
<th>Language</th>
<th>Role</th>
<th>Example</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bezhta</td>
<td>† Base.SUP.ATTR</td>
<td>māhā.ƛ’ā.kō [door_frame.SUP.ATTR]</td>
<td>SERVANT</td>
</tr>
<tr>
<td></td>
<td>Head Mod</td>
<td>c ’uddo c’emuc’ [red egg]</td>
<td>YOLK</td>
</tr>
<tr>
<td></td>
<td>Mod.ADJZ Head</td>
<td>nucodaq t’ot’ [honey:ADJZ fly]</td>
<td>BEE</td>
</tr>
<tr>
<td>Dutch</td>
<td>Base.M</td>
<td>weduwn.aar [widow.M]</td>
<td>WIDOWER</td>
</tr>
<tr>
<td>Hausa</td>
<td>Base.ABST</td>
<td>mûndâafû.ncii [unfaith.ABST]</td>
<td>DECEIT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mâi.táa [sorcerer.ABST]</td>
<td>MAGIC</td>
</tr>
<tr>
<td>Hindi</td>
<td>Base.AGT</td>
<td>lohā.r [iron.AGT]</td>
<td>BLACKSMITH</td>
</tr>
<tr>
<td></td>
<td>† Mod.ADJZ Head</td>
<td>havā.ī jahāz [air.ADJZ ship]</td>
<td>AIRPLANE</td>
</tr>
<tr>
<td>Indonesian</td>
<td>† Mod Head</td>
<td>satpam &lt; satuan pengamanan</td>
<td>[security unit] GUARD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mata air [water eye]</td>
<td>SPRING OR WELL</td>
</tr>
<tr>
<td>Irish</td>
<td>Base.NMLZ</td>
<td>draí.acht [magician.NMLZ]</td>
<td>MAGIC</td>
</tr>
<tr>
<td></td>
<td>† Head Mod</td>
<td>tráth.nóna [occasion.noon]</td>
<td>AFTERNOON</td>
</tr>
<tr>
<td>Kanuri</td>
<td>Head Mod.ADJZ</td>
<td>kwôngâ nyiyâ.à [man marriage.ADJZ]</td>
<td>MAN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kâmû nyiyâ.à [woman marriage.ADJZ]</td>
<td>MARRIED WOMAN</td>
</tr>
<tr>
<td>Ket</td>
<td>† Mod.PRIV Head</td>
<td>saan.an.ke’d [squirrels.PRIV.person]</td>
<td>DEFENDANT</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>† Base.ABST</td>
<td>žmogžud.ystê [murderer.ABST]</td>
<td>MURDER</td>
</tr>
<tr>
<td></td>
<td>Base.COLL</td>
<td>kaim.ynas [village.COLL]</td>
<td>NEIGHBOUR</td>
</tr>
</tbody>
</table>
The typology of binominal lexemes

**Lower Sorbian**

- † Base.ABST: lut.osć [sorrow.ABST] PITY
- † Base.DIM: góle.tko [child.DIM] BABY
- † Base.INST: šrub.owak [screw.INST] SCREWDRIVER
- † Base.VBLZ.AGT: wik.owa.ître [market.VBLZ.AGT] MERCHANT

**Malagasy**

- Head SOC.Mod: lehilâhy manaN.vàdy [man with.spouse] MARRIED MAN
- vehivâvy manaN.vàdy [woman with.spouse] MARRIED WOMAN
- NMLZ.Base: fi.lòha [NMLZ.head] CHIEFTAIN
- NMLZ.Base.CIRC: fanambakàna < faN.ambâka.ana [NMLZ.deceit.CIRC] DECEIT
- fifidiànana < fi.fîdy.(an)ana [NMLZ.choice.CIRC] ELECTION
- fialònana < fi.àlona.ana [NMLZ.jealousy.CIRC] ENVY OR JEALOUSY
- fikasâna < fi.kàsa.ana [NMLZ.jealousy.CIRC] INTENTION
- famosaviana < faN.mosàvy.ana [NMLZ.witchcraft.CIRC] MAGIC
- fiadiâna < fi.àdy.ana [NMLZ.fight.CIRC] WEAPONS

**Mapudungun**

- Mod.VBLZ.AGT Head: kure.nge.n wentru [wife.ESS.NON:F3 man] MARRIED MAN
- fîta.nge.n zomo [husband.ESS.NON:F3 woman] MARRIED WOMAN
- Mod.VBLZ.LOC Head: wingka lawen.tu.we ruka [medicine.VBLZ.LOC house] HOSPITAL

**Old High German**

- † Base.DIM: kindi.lîn [child.DIM] BABY

**Oroqen**

- Base.REC: amo.rok [stool.REC] TOILET
- † Base.VBLZ.AGT: mayma.la.ri [business.VBLZ.AGT] MERCHANT

**Otomí**

- † Head Mod LIM: jwûdâ bâtsi.tho [brother child.LIM] YOUNGER BROTHER
- nju bâtsi.tho [sister child.LIM] YOUNGER SISTER

**Polish**

- † Base.AUG: maco.cha [mother.AUG] STEPMOTHER
### Appendix H. Additional tables

#### Q'eqchi’

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT Base</td>
<td>aj limoox [AGT alm]</td>
<td>BEGGAR</td>
</tr>
<tr>
<td></td>
<td>aj tz’alam [AGT prison]</td>
<td>CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td>DER</td>
<td>jolom.il [head.DER]</td>
<td>CHIEFTAIN</td>
</tr>
<tr>
<td>INST.NMLZ</td>
<td>k’uub’leb’aal [hearthstone.INST.NMLZ]</td>
<td>COOKHOUSE</td>
</tr>
<tr>
<td></td>
<td>ch’ut.leb’aal [group.INST.NMLZ]</td>
<td>MEETING HOUSE</td>
</tr>
<tr>
<td>PL.NMLZ</td>
<td>k’ot.eb’aal [shit.NMLZ]</td>
<td>TOILET</td>
</tr>
<tr>
<td>SUF</td>
<td>sum.la.jik [mate.DER.SUF]</td>
<td>WEDDING</td>
</tr>
<tr>
<td>opaque</td>
<td>b’aqlaq ch’iich’ [corncob iron]</td>
<td>BICYCLE</td>
</tr>
</tbody>
</table>

#### Romanian

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABST</td>
<td>capitan.ie [captain.NMLZ]</td>
<td>CHIEFTAIN</td>
</tr>
<tr>
<td></td>
<td>cunun.ie [crown.SUF]</td>
<td>WEDDING</td>
</tr>
<tr>
<td>AGT.ABST</td>
<td>bucătă.ar.ie [piece_of_food.AGT.NMLZ]</td>
<td>COOKHOUSE</td>
</tr>
<tr>
<td>VBLZ.AGT.ABST</td>
<td>vrăjitor.ie [magic.VBLZ.AGT.ABST]</td>
<td>MAGIC</td>
</tr>
</tbody>
</table>

#### Selice Romani

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABST</td>
<td>čohan.ipe [witch/sorcerer.ABST]</td>
<td>MAGIC</td>
</tr>
<tr>
<td>Mod.ADJZ Head</td>
<td>phivli džuvli [widow.ADJZ woman]</td>
<td>WIDOW</td>
</tr>
<tr>
<td></td>
<td>özvědni džuvli [widow.ADJZ woman]</td>
<td>WIDOW</td>
</tr>
</tbody>
</table>

#### Seychelles Creole

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base.AGT</td>
<td>prizon.nyen [prison.AGT]</td>
<td>CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td></td>
<td>bazar.dye [market.AGT]</td>
<td>MERCHANT</td>
</tr>
</tbody>
</table>

#### Takia

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Mod</td>
<td>tamol sos [man Derris_root]</td>
<td>WIDOWER</td>
</tr>
<tr>
<td>Head Mod LOC</td>
<td>nanuk swa te [child breast at]</td>
<td>BABY</td>
</tr>
<tr>
<td>Mod.3SG COM PFV</td>
<td>tamol iwo.n da ya [man spouse.3SG COM PFV]</td>
<td>MARRIED MAN</td>
</tr>
<tr>
<td></td>
<td>pein iwo.n da ya [woman spouse.3SG COM PFV]</td>
<td>MARRIED WOMAN</td>
</tr>
</tbody>
</table>

#### Welsh

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMLZ</td>
<td>pen.aeth [chief.NMLZ]</td>
<td>CHIEFTAIN</td>
</tr>
<tr>
<td>SUF</td>
<td>carchar.or [prison.SUF]</td>
<td>CAPTIVE OR PRISONER</td>
</tr>
<tr>
<td></td>
<td>cwmwd.og [district.SUF]</td>
<td>NEIGHBOUR</td>
</tr>
</tbody>
</table>

#### Wichí

<table>
<thead>
<tr>
<th>category</th>
<th>word</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT</td>
<td>tshotoy.wu [animals.AGT]</td>
<td>HERDSMAN</td>
</tr>
<tr>
<td></td>
<td>tiena.wu [market.AGT]</td>
<td>MERCHANT</td>
</tr>
<tr>
<td></td>
<td>sapatu.wu [shoe.AGT]</td>
<td>SHOEMAKER</td>
</tr>
<tr>
<td>LOC</td>
<td>kanu.hi [needle.LOC]</td>
<td>SUGAR CANE</td>
</tr>
<tr>
<td></td>
<td>y’amekw.hi [excrement.LOC]</td>
<td>TOILET</td>
</tr>
</tbody>
</table>
The typology of binominal lexemes

**Yakut**

Base.NMLZ  
*bah.ilik* [bas,NMLZ] CHIEFTAIN  
*χοσ.ποχ* [room.NMLZ] COOKHOUSE  

† Base.VR.NMLZ  
*uot.taː.:hin* [fire.VR.NMLZ] ARSON  
*sanaː.:ryaː.:hin* [thought.VR.NMLZ] GRIEF  

† Base.VR.REFL.NMLZ  
*saya.laː.niː:* [time.VR.REFL.NMLZ] BEGINNING  
*iːs.te.niː:* [handicraft.VR.REFL.NMLZ] TAILOR

**Yaqui**

Base.LOC  
*sisi`iwoo.chi* [iron.LOC] TOOL  

† Base.NMLZ  
*ko`oko.a* [pain.NMLZ] DISEASE  

† Mod.PL Head  
*waim asoa* [sister/brother.PL daughter] STEPDAUGHTER  
*waim marai* [sister/brother.PL daughter] STEPDAUGHTER  
*waim achai* [sister/brother.PL father] STEPFATHER  
*waim maala* [sister/brother.PL mother] STEPMOTHER

**Zinacantán Tzotzil**

Mod Head  
*shokan na* [side house] NEIGHBOUR  
*hmulavi ?antz* [sinner woman] PROSTITUTE  
*meʔanal ?antz* [poverty/misery/grief woman] WIDOW  
*meʔanal vinik* [poverty/misery/grief man] WIDOWER  

opaque  
*k`op ?oʔon* [word/argument heart] ANXIETY  
*kachimpa pom* [pipe incense] BEE  
*mishik` pom* [belly_button incense] BEE  
*ton pom* [stone incense] BEE
Index of names

Abraham, Roy Clive, 328
Adams, Valerie, 151, 164, 192
Adelaar, Alexander, 330
Ahmad, Mustapha, 328
Aikhenvald, Alexandra Y., 21, 41, 42, 44, 45, 59, 129, 131
Akdikmen, Resuhi, 67, 331
Allan, Keith, 129
Allen, Margaret Reece, 151
Alsdorf-Bollée, Annegret, 331
Arcodia, Giorgio Francesco, 208
Arends, Jacques, 331
Aristotle, 159, 176, 181, 276, 283, 392
Arnaud, Pierre J.L., xxi, 18, 21, 29, 32, 59, 66, 81, 87, 154, 164, 169, 170, 171, 172, 173, 175, 178, 181, 183, 190, 201, 292
Aronoff, Mark, 9
Artiagoitia, Xabier, 327
Atkinson, Bill, 4
Atoyebi, Joseph Dele, 256
Awagana, Ari, 328
Awbery, Gwenllian Mair, xxi, 67, 256, 331
Baker, Mark C., 330
Bakker, Dik, 56, 84, 85, 86, 87, 330
Bartels, Hauke, xxi, 330
Bauer, Laurie, xxi, 8, 14, 15, 18, 21, 22, 23, 24, 25, 26, 27, 28, 29, 34, 36, 37, 40, 47, 49, 50, 51, 59, 86, 108, 109, 110, 119, 145, 149, 151, 153, 154, 156, 158, 169, 183, 201, 249, 250, 251, 286, 292, 293
Belzėja, Z., 329
Bergsten, Nils, 151
Berners Lee, Tim, 4
Bisetto, Antonietta, 7, 21, 33, 34, 35, 36, 38, 260, 292
Blank, Andreas, 159, 186, 276, 283, 284, 290
Bloomfield, Leonard, 34, 58
Booij, Geert, 9, 34, 144, 293, 328
Borer, Hagit, 329
Borg, Albert, 330
Bourque, Yves Stephen, xxi, xxii, 18, 21, 66, 150, 152, 154, 161, 162, 164, 183, 184, 185, 187, 188, 191, 192, 194, 196, 197, 198, 199, 200, 201, 223, 227, 272, 276, 279, 284, 286, 292, 328, 392
Brdar, Mario, 159
Breckle, Herbert E., 151
Brindle, Jonathan, xxi, 328
Brown, Cecil H., xxii, 38, 144, 329, 331, 332
Bush, Vannevar, 4, 5
Butnariu, Cristina, 154
Bybee, Joan, 124
Carr, Charles Telford, 151, 153, 154, 177
Cauquelin, Josiane, 330
Ceccagno, Antonella, 33, 182, 256, 330
Cetnarowska, Bożena, 253
Chumakina, Marina, 38, 327
Cole, Peter, 329
Comrie, Bernard, xxi, 53, 86, 94, 97, 328, 335
Corne, Chris, 331
Courtz, Henk, 328
Cowan, Marion M., 332
Coyos, Jean-Baptiste, 327
Creissels, Denis, 118, 137, 141
Croft, William, xxi, 1, 12, 16, 17, 25, 117, 118, 119, 120, 121, 122, 138, 140, 293, 294
Cyffer, Norbert, 329
Dinh, Dien, 28, 331
The typology of binominal lexemes

Dixon, R. M. W., 1, 16, 102, 125, 137
Dokulil, Miloš, 10, 12
Don, Jan, 328
Downing, Pamela, 151, 161, 187
Dryer, Matthew S., 8, 34, 56, 57, 84, 85, 86, 87, 141, 323
Dumestre, Gérard, 327
Eachus, Francis, 330
Egmond, John M., 329
Egmond, Marie-Elaine van, xxii, 142, 327
Eiesland, Eli-Anne, 150, 154, 169, 189
Elbert, Samuel H., 328
Elšík, Viktor, 331
Engelbart, Doug, 4
Epps, Patience, xxii, 329
Evan, H. Meurig, 67
Faarlund, Jan Terje, 330
Fabri, Ray, 330
Fallbaum, Christiane, 268
Fengxiang Li, 151, 330
Fernández, Zarina Estrada, 332
Ferrand, Gabriel, 330
Foolen, Ad, 13
Fortescue, Michael D., 29
Fugier, Huguette, 330
Georg, Stefan, 329
Girju, Roxana, 154
Giurgea, Ion, 330
Givón, Talmy, 102, 141
Glinert, Lewis, 329
Gök, Aslı, 331
Goldberg, Adele E., 293
Golluscio, Lucía, 330
Good, Jeff, xxi, 330
Grant, Anthony, 328
Greenberg, Joseph H., 86, 249
Grimm, Jacob, 151
Grosssmann, Maria, 330
Guevara, Emilianó, 7, 15, 21, 32, 33, 34, 37, 38, 39, 40, 149, 250, 292, 395
Hacken, Pius ten, 150, 151
Hahsler, Michael, 230
Halvorsen, Arne, 330
Hammarström, Harald, xxv, 84
Hancock, Ian F., 331
Harmelin, Bryan, 237
Hasegawa, Yoko, 68, 329
Haseplamath, Martin, iii, xxi, 8, 12, 15, 21, 38, 44, 51, 52, 53, 57, 59, 84, 86, 94, 97, 120, 121, 137, 143, 146, 162, 206, 293, 294, 335
Haviland, John, 332
Heath, Jeffrey, 28
Hildebrandt, Kristine A., 330
Hinds, John, 68, 329
Holton, David, 328
Hutchison, John P., 329
Iwakoshi, Shocchi, 331
Jackendoff, Ray, 16, 149, 151, 153, 154, 161, 162, 164, 187, 190, 194, 292
Jacobson, Steven A., xxiii, 160, 328
Jaggar, Philip J., 328
Jakobson, Roman, 159, 283
Janda, Laura A., xxi, 18, 21, 159, 160, 176, 183, 201, 276, 277, 279, 294
Jensen, Bård Uri, xxi
Jespersen, Otto, 47, 152, 153, 154, 164, 165, 167, 168, 187, 294
Johnston, Michael, 151
Jones, ʻōiwi Parker, 328
Kachru, Yamuna, 329
Kageyama, Taro, 68, 329
Kakati, Banikanta, 327
Karlsson, Fred, 328
Keenan, Edward L., 86
Kert, Georgij M., 329
Kibrik, Aleksandr E., 16, 137, 327, 328
Kiefer, Ferenc, 137
Kihn, Garet, 331
Kirtchuk-Halevi, Pablo I., 329
Kiskilä, Seppö, 284
Koch, Peter, 9, 55, 56, 159, 238, 266, 276, 283, 285, 289, 290
Koptjevskaja-Tamm, Maria, xxi, 17, 21, 50, 115, 116, 117, 118, 119, 121, 122, 125, 138, 141, 142, 147, 294
Kornfilt, Jaklin, 67, 331
Körtvélyessy, Lívia, 7, 10, 21, 43, 47, 87, 147
Kossmann, Maarten, xxii, 136, 331
Kővecses, Zoltan, 277
Kruspe, Nicole, xxii, 328, 331
Langacker, Ronald W., xxii, 1, 6, 12, 16, 37, 110, 111, 260, 270, 271, 276, 293, 294
Lees, Robert B., 151
Leslau, Wolf, 327
Levi, Judith N., 14, 21, 47, 59, 151, 152, 153, 156, 161, 162, 184, 329
Lieber, Rochelle, 7, 8, 24, 36, 99, 109, 151, 260
Lødrup, Helge, xxii
Löhr, Doris, 329
Lovestrand, Joseph, xxii, 258, 327
Lukas, Johannes, 329
Maksunova, Zoya, 329
Mallinson, Graham, 330
Marchand, Hans, 34, 101, 108, 109, 151
Masini, Francesca, xxii, 51, 292, 293, 329
Matthews, P.H., 144
Mätzner, Eduard, 47, 151, 165, 167
McConvell, Patrick, 328
McGregor, Ronald Stuart, 329
McWhorter, John, 330
Michaelis, Susanne, 331
Miestamo, Matti, xxii, 85, 87, 328
Mihatsch, Wiltrud, 13
Miyaoka, Osahito, 328
Moldovan, Dan, 154
Mottin, Jean, 329
Mous, Maarten, 329
 Munroe, Randall, 197
Næss, Åshild, xxii, 144, 327
 Nagóenko, Alicja, 128
Nakov, Preslav, 154
 Navickaitė-Klišauskienė, Agnė, xxii, 329
Nedjalkov, Igor, 330, 332
 Nelson, Ted, 4
 Nercesian, Verónica, 332
 Newman, Paul, 141, 328
 Newman, Roxana Ma, 328
 Nguyêñ, Dinh-Hoà, 28, 253, 331
 Nichols, Johanna, 57, 85, 86, 121, 123
 Noailly, Michèle, 169, 170, 172
 Nordbustad, Frøydis, 329
 Ó Séaghdha, Diarmuid, 154
 Olsen, Susan, 37
 Packard, Jerome L., 256
 Panther, Klaus-Uwe, 282
 Parker, G. W., 96, 330
 Peirsman, Yves, 159, 276, 277, 285, 289
 Perkins, Revere, 85
 Plank, Frans, 291
 Prauliņš, Dace, 329
 Pukui, Mary Kawena, 328
 Pustejovsky, James, 151
 Putten, Marijn van, 331
 Quesada, Juan Diego, 328
 Rainer, Franz, 14, 21, 47, 49, 59, 158
 Ralli, Angela, 145, 328
 Rattiff, Martha, xxii, 329
 Renault-Lescure, Odile, 328
 Rendón, Jorge A. Gómez, 329
 Richardson, James, 96
 Rießler, Michael, xxii, 329
 Rijkhoff, Jan, 56, 85, 86
 Rose, Françoise, xxii, 331
 Ross, Malcolm, xxii, 331
 Russell, Joan, 130
 Ryder, Mary Ellen, 150, 151, 152, 153, 183
 Sadock, Jerrold M., 29
 Saeed, John, 331
 Saltarelli, Mario, 327
 Saulwick, Adam, 88
 Saussure, Ferdinand de, 207
 Scalise, Sergio, 7, 21, 32, 33, 34, 35, 36, 37, 38, 39, 40, 106, 107, 125, 149, 182, 250, 256, 260, 292, 395
 Schaaik, Gerjan van, 67
 Schachter, Paul, 41, 331
 Schadeberg, Thilo, 331
 Schäfer, Martin, 154, 162
 Schmidt, Christopher K., 329
 Schuhmann, Roland, 330
 Schulte, Kim, 330
 Schuster-Śewc, Heinz, 330
 Shadbolt, Nigel, 5

2019-01-01 Draft. Please do not cite. Comments to: pepper.steve@gmail.com 407
<table>
<thead>
<tr>
<th>Language</th>
<th>Notes</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Äiwoo (nfl, Austronesian, Oceanic)</td>
<td>xxv, 89, 103, 131, 132, 144, 188, 217, 231, 243, 325, 327, 349, 391, 392</td>
<td></td>
</tr>
<tr>
<td>Akkadian (akk, Afro-Asiatic, Semitic)</td>
<td>89, 139, 206, 216, 217, 219, 231, 243, 323, 327, 337, 359, 391, 392</td>
<td></td>
</tr>
<tr>
<td>Amarakaeri (amr, Harakmbut)</td>
<td>131, 144, 327</td>
<td></td>
</tr>
<tr>
<td>Amharic (amh, Afro-Asiatic, Semitic)</td>
<td>89, 216, 217, 243, 274, 323, 327, 337, 359, 396</td>
<td></td>
</tr>
<tr>
<td>Anindilyakwa (aoi, Gunwinyguan)</td>
<td>89, 142, 217, 242, 243, 323, 327, 351, 384</td>
<td></td>
</tr>
<tr>
<td>Arabela (arl, Zaparoan)</td>
<td>129, 130</td>
<td></td>
</tr>
<tr>
<td>Arbore (arv, Afro-Asiatic, Cushitic)</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>Archi (aqc, Nakh-Daghestanian, Lezgie)</td>
<td>52, 53, 62, 72, 89, 111, 137, 217, 243, 323, 327, 347, 376</td>
<td></td>
</tr>
<tr>
<td>Assamese (asm, Indo-European, Indo-Aryan)</td>
<td>89, 213, 216, 217, 221, 243, 323, 327, 343, 370</td>
<td></td>
</tr>
<tr>
<td>Baa (kwb, Atlantic-Congo, Gur)</td>
<td>89, 127, 217, 267, 275, 324, 327, 338, 361</td>
<td></td>
</tr>
<tr>
<td>Bambara (bam, Atlantic-Congo, Mande)</td>
<td>30, 89, 127, 217, 323, 327, 338, 362</td>
<td></td>
</tr>
<tr>
<td>Bandial (bqj, Atlantic-Congo, North-Central Atlantic)</td>
<td>89, 130, 131, 132, 133, 217, 323, 327, 339, 362</td>
<td></td>
</tr>
<tr>
<td>Barain (bva, Afro-Asiatic, Chadic)</td>
<td>89, 134, 138, 217, 242, 255, 257, 258, 275, 323, 327, 355, 357</td>
<td></td>
</tr>
<tr>
<td>Basque (eus, Basque)</td>
<td>22, 30, 33, 89, 90, 217, 243, 323, 327, 340, 365</td>
<td></td>
</tr>
<tr>
<td>Bezhta (kap, Nakh-Daghestanian, Avar-Andic-Tsepic)</td>
<td>14, 48, 52, 55, 73, 74, 89, 118, 123, 135, 137, 155, 188, 217, 324, 328, 346, 375, 401</td>
<td></td>
</tr>
<tr>
<td>Bora (boa, Boran)</td>
<td>11, 48, 129, 130</td>
<td></td>
</tr>
<tr>
<td>Cabécar (cip, Chibchan)</td>
<td>89, 217, 323, 328, 355</td>
<td></td>
</tr>
<tr>
<td>Caijia (caij1234), Sino-Tibetan, Macro-Bai)</td>
<td>xxv, 89, 127, 193, 217, 242, 323, 328, 350, 382</td>
<td></td>
</tr>
<tr>
<td>Cantonese (yue, Sino-Tibetan, Sinitic)</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>Central Yupik (esu, Eskimo-Aleut, Yupik)</td>
<td>xxiii, 88, 89, 96, 139, 160, 212, 216, 217, 323, 328, 352, 385</td>
<td></td>
</tr>
<tr>
<td>Ceq Wong (cwq, Austro-Asiatic, Aslian)</td>
<td>52, 72, 89, 127, 206, 217, 323, 328, 348, 379</td>
<td></td>
</tr>
<tr>
<td>Chakali (cli, Atlantic-Congo, Gur)</td>
<td>89, 217, 242, 323, 328, 338, 361</td>
<td></td>
</tr>
<tr>
<td>Cora (crn, Uto-Aztecan)</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Croatian (hrv, Indo-European, Slavic)</td>
<td>33, 39, 324, 328</td>
<td></td>
</tr>
<tr>
<td>Datooga (tcc, Nilo-Saharan, Nilotic)</td>
<td>89, 127, 206, 208, 209, 217, 325, 328, 339, 363, 396</td>
<td></td>
</tr>
<tr>
<td>Dutch (nld, Indo-European, Germanic)</td>
<td>33, 39, 52, 72, 73, 74, 89, 93, 94, 144, 217, 242, 254, 278, 325, 328, 342, 367, 396, 401</td>
<td></td>
</tr>
<tr>
<td>English (eng, Indo-European, Germanic)</td>
<td>xxv, 8, 23, 24, 30, 31, 33, 38, 39, 41, 42, 47, 49, 50, 52, 55, 61, 65, 69, 72, 81, 89, 93, 94, 95, 96, 97, 99, 103, 105, 107, 108, 110, 119,</td>
<td></td>
</tr>
</tbody>
</table>
The typology of binominal lexemes


Estonian (est, Uralic, Finnic), 41, 89, 136, 217, 323, 328, 347, 376, 396

Fijian (fij, Austronesian, Oceanic), 56

Finnish (fin, Finn, Uralic, Finnic), 24, 254

French (fra, Indo-European, Romance), 1, 8, 13, 14, 23, 24, 30, 31, 33, 39, 42, 45, 65, 66, 67, 70, 73, 75, 81, 89, 94, 117, 122, 127, 128, 134, 135, 154, 155, 161, 162, 169, 170, 171, 179, 180, 184, 185, 192, 197, 212, 217, 230, 252, 255, 278, 281, 283, 323, 324, 328, 344, 356, 371, 390, 396

Galibi Carib (car, Cariban, Guianan), 48, 52, 72, 73, 89, 138, 139, 148, 217, 242, 323, 328, 354, 388

Gawwada (gwd, Afro-Asiatic, Cushitic), 52, 71, 72, 89, 127, 128, 131, 217, 255, 324, 328, 336, 358

German (deu, Indo-European, Germanic), xxv, 1, 8, 14, 23, 30, 33, 39, 41, 49, 51, 70, 71, 72, 73, 89, 117, 128, 143, 145, 151, 154, 158, 160, 182, 217, 242, 254, 278, 323, 328, 342, 368, 396

Greek (ell, Indo-European, Greek), 33, 89, 136, 145, 165, 216, 217, 221, 223, 253, 255, 257, 278, 281, 323, 328, 343, 370, 396

Gurinji (gue, Pama-Nyungan, Desert Nyungic), 52, 71, 72, 89, 135, 217, 324, 328, 351, 384

Hausa (hau, Afro-Asiatic, Chadic), 52, 56, 72, 89, 92, 93, 122, 127, 132, 138, 140, 141, 160, 217, 324, 328, 336, 357, 396, 401

Hawaiian (haw, Austronesian, Oceanic), 52, 72, 89, 93, 127, 132, 189, 192, 217, 274, 324, 328, 349, 381


Hindi (hin, Indo-European, Indo-Aryan), 50, 70, 74, 76, 89, 134, 135, 146, 217, 221, 252, 324, 329, 343, 370, 396, 401

Hixkaryána (hix, Cariban), 28, 250

Hmong Daw (mww, Hmong-Mien), 52, 71, 72, 89, 127, 217, 218, 324, 329, 350, 381

Hungarian (hun, Uralic, Hungarian), 30, 33, 49, 70, 72, 89, 127, 136, 147, 217, 324, 329, 347, 377, 396

Hupdë (jup, Nadahup), 52, 72, 89, 127, 217, 324, 329, 355

Imbabura Quechua (qvi, Quechuan, Quechua II), 72, 89, 127, 217, 274, 325, 329, 355, 389

Indonesian (ind, Austronesian, Malayo-Sumbawan), 52, 70, 72, 89, 93, 217, 254, 324, 329, 349, 380, 396, 401

Iraqw (irk, Afro-Asiatic, Cushitic), 52, 71, 72, 89, 93, 96, 188, 216, 217, 242, 324, 329, 336, 358

Irish (gle, Indo-European, Celtic), 70, 72, 74, 76, 89, 127, 135, 148, 217, 221, 223, 252, 254, 255, 324, 329, 341, 367, 401


Japanese (jpn, Japonic, Japanese), xxv, 24, 33, 39, 46, 48, 52, 66, 68, 72, 89, 93, 117, 120, 123, 127, 128, 134, 135, 145, 146, 188, 208, 217, 252, 324, 329, 346, 374, 396

Kalamang (kgv, West Bomberai), 89, 104, 137, 138, 217, 242, 255, 256, 257, 263, 265, 276, 324, 329, 351, 385

Kambata (ktb, Afro-Asiatic, Cushitic), 89, 103, 139, 217, 324, 329, 336, 358
Index of languages

Kanuri (knc, Nilo-Saharan, Saharan), 22, 28, 52, 71, 72, 75, 89, 93, 137, 212, 217, 242, 249, 255, 324, 329, 339, 363, 401
Kanuri (knc, Saharan, Kanuric), 22, 28, 52, 71, 72, 75, 89, 93, 137, 212, 217, 242, 249, 255, 324, 329, 339, 363, 401
Ket (ket, Yeniseian), 52, 71, 72, 73, 89, 135, 137, 189, 217, 242, 275, 324, 329, 348, 378, 396, 401
Khmer (khm, Austroasiatic, Khmeric), 22, 24, 30, 81
Khoekhoe (naq, Khoe), 56, 396
Kildin Sami (sjd, Uralic, Saami), 14, 52, 89, 127, 136, 217, 276, 325, 329, 348, 378
Korean (kor, Koreanic), 33, 39, 89, 127, 144, 145, 206, 217, 324, 329, 346, 375
Kupsabin (kpz, Nilo-Saharan, Nilotic), 89, 134, 135, 138, 216, 217, 324, 329, 339, 363
Latvian (lav, Indo-European, Baltic), 89, 217, 223, 242, 324, 329, 341, 366
Lenakel (tnl, Austronesian, Oceanic), 56
Lithuanian (lit, Indo-European, Baltic), 70, 73, 89, 118, 136, 145, 217, 223, 324, 329, 341, 366, 401
Malagasy (plt, Austronesian, Malayo-Polynesian), 89, 217, 223, 242, 324, 329, 341, 366
Lenakel (tnl, Austronesian, Oceanic), 56
Lithuanian (lit, Indo-European, Baltic), 70, 73, 89, 118, 136, 145, 217, 223, 324, 329, 341, 366, 401
Malagasy (plt, Austronesian, Malayo-Polynesian), 89, 217, 223, 242, 324, 329, 341, 366
Malayalam (mal, Dravidian), 89, 217, 223, 242, 324, 329, 341, 365, 396
Maltese (mlt, Afro-Asiatic, Semitic), 89, 134, 139, 146, 147, 217, 324, 330, 337, 360
Manange (mmn, Sino-Tibetan, Bodic), 52, 72, 89, 135, 217, 325, 330, 350, 382
Mandarin Chinese (cmn, Sino-Tibetan, Sinitic), 34, 52, 72, 89, 107, 127, 189, 208, 211, 217, 275, 323, 330, 350, 382
Mapudungun (arn, Arhuaco, Arawakan), 52, 72, 75, 89, 103, 127, 128, 217, 237, 255, 256, 257, 262, 294, 323, 330, 354, 387, 402
Marra (mec, Mangarrayi-Maran), 22, 28, 250
Mbyá Guaraní (gun, Tupian, Tupi-Guarani), 89, 217, 324, 330, 354
Minangkabau (min, Austronesian, Malayic), 130
Murui Huitoto (huu, Huitotoan), 14, 89, 117, 122, 131, 133, 134, 158, 206, 208, 217, 324, 330, 355, 388
Navajo (nav, Athabaskan-Eyak-Tlingit), 88, 89, 96, 160, 189, 217, 324, 330, 352, 385
Nepali (npi, Indo-European, Indo-Aryan), 89, 146, 217, 221, 252, 325, 330, 343, 370
Nizaa (sgi, Atlantic-Congo, Mambiloid), 5, 6, 7, 37, 66, 67, 68, 108, 154, 183, 237, 256, 259, 260, 261, 263, 294
Northern Kurdish (kmr, Indo-European, Iranian), 147
Norwegian (nor, Indo-European, Germanic), 8, 33, 39, 89, 120, 127, 141, 154, 189, 198, 216, 217, 218, 220, 243, 254, 265, 279, 325, 330, 342, 369
Oko (okoo1245, Atlantic-Congo, Oko-Eni-Obayan), 256
Old High German (goh, Indo-European, Germanic), 52, 73, 89, 206, 217, 222, 324, 330, 343, 369, 402
Oroqen (orph, Altaic, Tungusic), 52, 72, 89, 139, 217, 242, 275, 325, 330, 340, 364, 402
The typology of binominal lexemes

Portuguese (por, Indo-European, Romance), 33, 41, 42, 396
Puyuma (pyu, Austronesian, Formosan), 88, 89, 209, 217, 325, 330, 348, 379
Q’eqchi’ (kek, Mayan, Core Mayan), 52, 72, 89, 137, 138, 217, 243, 324, 330, 352, 386, 403
Querétaro Otomi (ote, Otomanguean, Otopamean), 52, 89, 160, 211, 217, 267, 325, 330, 353
Rama (rma, Chibchan), 56, 343, 370
Romanian (ron, Indo-European, Romance), 52, 89, 93, 104, 217, 252, 255, 325, 330, 344, 372, 396, 403
Russian (rus, Indo-European, Slavic), xxv, 1, 8, 14, 32, 33, 39, 42, 89, 94, 117, 118, 136, 142, 145, 146, 147, 155, 159, 160, 212, 217, 242, 325, 330, 345, 373, 396
San Mateo del Mar Huave (huv, Huavean), 56
Saramaccan (srm, Pidgins & Creoles, English-based), 52, 72, 89, 91, 217, 242, 275, 325, 330, 356, 389
Selice Romani (rnc, Indo-European, Indo-Aryan), 52, 89, 208, 216, 217, 221, 325, 331, 343, 371, 403
Seri (sei, Seri), 89, 90, 139, 217, 242, 325, 331, 353, 387
Seychelles Creole (crs, Pidgins & Creoles, French-based), 52, 72, 74, 76, 89, 91, 127, 211, 217, 323, 331, 356, 390, 403
Sidamo (sid, Afro-Asiatic, Cushitic), 89, 127, 135, 137, 217, 242, 325, 331, 336, 358
Slovak (slk, Indo-European, Slavic), 10, 11, 14, 48, 89, 136, 147, 155, 216, 217, 325, 331, 346, 374, 396
Somali (som, Afro-Asiatic, Cushitic), 89, 139, 217, 325, 331, 336, 359
Spanish (spa, Indo-European, Romance), 13, 108, 283
Srenge (lsr, Nuclear Torricelli), 89, 127, 206, 217, 324, 331, 347, 351, 376, 384
Swahili (swh, Atlantic-Congo, Bantoid), 48, 52, 56, 70, 71, 72, 89, 93, 130, 131, 133, 134, 135, 148, 208, 217, 224, 236, 242, 243, 251, 325, 331, 338, 360, 396
Tagalog (tgl, Austronesian, Greater Central Philippine), 41, 89, 132, 134, 135, 217, 274, 275, 325, 331, 349, 380
Takia (tbc, Austronesian, Oceanic), 12, 14, 52, 71, 72, 73, 75, 89, 96, 97, 108, 117, 134, 135, 138, 139, 158, 188, 217, 325, 331, 349, 381, 403
Taklema (tkm, extinct isolate), 22, 24
Tarifit (rif, Afro-Asiatic, Berber), xxv, 48, 52, 63, 73, 89, 135, 136, 148, 217, 325, 331, 335, 357
Thai (tha, Tai-Kadai, Kam-Tai), 22, 52, 72, 89, 93, 95, 122, 127, 129, 158, 217, 325, 331, 350, 383
Ticuna (tea, Ticuna-Yuri), 89, 90, 217, 255, 257, 258, 325, 331, 355, 389
Toaripi (tqo, Eleman), 56
Trinitario (trn, Arawakan), 131, 160, 325, 331
Turkish (tur, Altaic, Turkic), 14, 30, 32, 33, 66, 67, 89, 94, 217, 242, 275, 325, 331, 340, 364, 395
Tuwari (tww, Walio), 325, 331
Tz’utujil (tzj, Mayan, Core Mayan), 250
Vietnamese (vie, Austro-Asiatic, Vietic), 14, 22, 28, 52, 72, 89, 97, 98, 104, 125, 127, 217, 218, 220, 253, 255, 256, 325, 331, 348, 379, 396
Walman (van, Nuclear Torricelli), 89, 127, 217, 325, 331, 347, 351, 376, 384
Wartha Thuntai (gnt, Morehead-Wasur), 324, 331
Western Farsi (pes, Indo-European, Iranian), 95, 122, 256, 325, 331
Western Mari (mrj, Uralic, Mari), 89, 135, 217, 324, 331, 348, 378
Wichi (mzh, Matacoan), 52, 71, 89, 217, 242, 274, 324, 332, 355, 388, 396
Wik-Mungkan (wim, Pama-Nyungan, Paman), 89, 217, 242, 325, 332, 351, 384
Wolof (wol, Atlantic-Congo, North-Central Atlantic), 89, 217, 243, 325, 332, 339, 363

Yakut (sah, Altaic, Turkic), 52, 71, 72, 89, 127, 135, 137, 138, 217, 242, 254, 325, 332, 340, 364, 404
Yaqui (yaq, Uto-Aztecan, Southern Uto-Aztecan), 52, 72, 74, 75, 89, 188, 217, 242, 325, 332, 354, 387, 404
Yimas (yee, Lower Sepic Ramu), 22, 24
Zinacantán Tzotzil (tzo, Mayan, Core Mayan), 52, 72, 73, 74, 89, 211, 217, 242, 325, 332, 353, 396, 404
Index of subjects

action-affix, 12, 104, 105
action-morph, 99
action-root, 12, 48, 70, 98, 103, 105, 108, 209
association type, 2, 3, 184, 287, 288
associative relation, 4, 18, 159, 160, 176, 182, 249, 276, 277, 279, 283, 284, 285, 292, 293, 392
binominal fingerprint, 146, 220, 223, 248, 252
binominal quantifier constructions, 13, 281
binominal type
cls, 123, 124, 126, 129, 130, 131, 139, 143, 144, 149, 158, 160, 217, 228, 229, 230, 234, 236, 243, 251, 338, 339, 349, 355
The typology of binominal lexemes

classifier, 11, 12, 13, 14, 48, 49, 51, 117, 118, 122, 123, 124, 129, 130, 131, 133, 134, 144, 176
CLDF (Cross-Linguistic Data Formats), 91
CLLD (Cross-Linguistic Linked Data), 285
Cognitive Grammar, 1, 6, 13, 16, 110, 150, 151, 152, 260
colour terms, 12
compound types
aggregative, 23
attributive, 7, 35, 36, 40, 256
avyayibhava, 23
coordinative, 9, 23, 35, 36, 38, 40, 42, 98, 99, 105, 107, 170, 176, 182, 183, 201, 241, 251, 260, 274, 276
copulative, 23, 45, 175
determinative, 7, 10, 23, 29, 164, 165, 168, 170, 183, 201
dvandva, 23
endocentric, 35, 39, 40, 42, 49, 156
exocentric, 23, 30, 35, 39, 40, 42, 45, 107, 189
possessive, 116
prepositional, 8, 13, 14, 180
root, 8, 14, 24, 40
subordinate, 7, 9, 35, 36, 38, 40, 170, 237, 256, 257, 260, 261, 266
synthetic, 10, 11, 12, 13, 14, 23, 24, 42, 43, 53, 116, 118, 123, 251, 294
tatpuruṣa, 23, 24
verbal-nexus, 23, 35
contiguity, 6, 57, 58, 59, 159, 160, 176, 182, 238, 239, 277, 283, 284, 285, 289
CSV (comma-separated values), 63, 91, 92, 93
elaboration site, 110, 151, 258, 265, 270, 271, 272, 273, 286, 292
expressive binominal NPs, 13
genetic grouping, 85, 233, 396
geographical area, 25, 26, 46, 57, 84, 86, 87, 91, 215, 220, 243, 245, 251
glotocode, xxv, 231, 232, 243, 328, 391, 392
Glottolog, xxv, 22, 87, 323, 396
ICM (Idealized Cognitive Model), 277
intralingual competition, 220, 248
Leipzig Glossing Rules, xxv, 53, 94, 97, 335
Lucca, 2, 287
Madame Butterfly, 2
metaphor, 36, 175, 176
metonymy, 21, 36, 59, 159, 160, 176, 276, 277, 278, 279, 281, 282, 284, 285, 286, 288, 290, 292, 294
Morbo/Comp project, 21, 32, 34, 37, 40, 59, 61, 62, 87, 250, 251, 395
occurrence, 2, 81, 97, 148, 216, 285, 335, 397, 399
Onomasiological Types, 10, 11, 12, 50, 65, 70, 99, 104, 109, 189, 213, 294
Principal Components Analysis (PCA), 221
property-affix, 12, 104, 105
property-root, 12, 83, 105, 182
Puccini, 2, 285, 287, 289
Radical Construction Grammar, 1, 16
RDF (Resource Description Framework), 5, 91, 285
reference point ability, 6, 37, 260, 262
relational adjective, 1, 14, 47, 49, 102, 124, 137, 158, 241
role type, 4
semantic category, 44, 45, 47, 53, 62, 63, 65, 80, 97
semantic relation, 6, 7, 8, 10, 11, 23, 24, 29, 34, 37, 47, 48, 49, 56, 57, 58, 59,
Index of subjects


SGML (Standard Generalized Markup Language), 54, 184

similarity, 9, 57, 58, 59, 152, 159, 163, 165, 176, 177, 181, 182, 196, 199, 200, 201, 225, 228, 238, 239, 280, 283, 284, 289

subordination, 35, 37, 238, 239, 260, 283, 284

synset, 268, 271, 294

TAO of Topic Maps, 2, 285


thing-morph, 12, 13, 62, 98, 99, 104, 111, 122, 131, 132, 149, 208, 224, 242


topic, 2, 4, 5, 17, 18, 23, 59, 249, 285, 292

topic map, 2, 5, 184, 287, 288

Topic Maps, 2, 4, 5, 7, 18, 54, 184, 249, 277, 284, 285, 287, 288, 290

topic type, 2, 3, 184

Tosca, 2, 285, 287, 289

type binominals, 13

WALS (World Atlas of Language Structures), 84, 85, 87, 285, 396

WOLD (World Loanword Database), 1, 51, 52, 53, 54, 55, 61, 62, 63, 65, 66, 68, 69, 70, 76, 80, 82, 86, 87, 91, 92, 93, 94, 95, 96, 97, 111, 143, 206, 212, 285, 327

word order, 18, 25, 26, 27, 29, 46, 180, 249, 252, 290

WordNet, 268
Steve Pepper
The typology 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  Draft 2019-01-01