BINOMINAL WORD-FORMATION IN THE WORLD’S LANGUAGES
Outline

• Introduction
  – Background, comparative concept, theoretical framework, research questions

• Previous research
  – Compounds, prefiguring “binominals”, lexical typology

• Data and methodology
  – Method, meaning sample, data sources, language sample

• Preliminary findings
  – […]

• Analysis
  – Formal, semantic and statistical analyses, typological classification
Background
The comparative concept
Research questions
Theoretical framework

INTRODUCTION
Nominal compounding

• Previous cross-linguistic studies of compounding
  – Bauer (2001) – well-balanced 36 language sample
    • few generalizations to report
  – Guevara & Scalise (2009) – 80,000 compounds, 21 languages
    • scales of preference only
  – Štekauer, Valera & Körtvélyessy (2012) – 70 language sample
    • scope includes derivation
    • main concern: presence or absence of different types of compound

• Surprisingly few insights…
  – Possible reasons
    • Scope too broad
    • Scope too narrow (!)
Narrowing the scope

• Object of study defined too broadly
  – Attempt to cover every type of compound
    • nominal, verbal, adjectival, …
    • noun-noun, noun-adjective, …
    • root compounds, synthetic compounds, …
    • determinative compounds, coordinative compounds, …
  – Complicates the typology, obscures interesting tendencies
    • e.g. Mandarin’s LH verbal compounds and RH nominal compounds

• Solution
  – Narrow the focus to N+N compounds
    • or more precisely, determinative N+N compounds
    • Wälchli (2005) has already done a good job with co-compounds
Broadening the scope

• How to define **N+N compound**?
  – *eisen.bahn* [iron.path]
  – *sidir.o.dromos* [iron.LE.road] (?)
  – ? *chemin de fer* [way PREP iron]
  – ?? *želez.naja doroga* [iron.ADJZ road]

• Comparative concept
  – Comparative concept must be language-independent
  – Cannot be defined in purely formal terms (Haspelmath 2010)

• Conclusion
  – Employ a functional definition

**entity**: something that has separate and distinct existence and objective or conceptual reality (Merriam-Webster)
Binominal naming constructions

• Constructional schemas for naming units that consist principally (and prototypically) of two “thing-roots”, i.e. roots that denote physical objects (Haspelmath 2012)

• Instances of such schemas are called binominal naming units, or binominals for short
  
  • Not to be confused with
    – binominal quantifier expressions
      • Sp. un montón de amigas ‘a heap of friends’ (Verveckken 2015:3)
    – expressive binominal NPs
      • Eng. an angel of a child (Foolen 2004)
    – type-noun binominals
      • Fr. une espèce de baleine ‘a kind of whale’ (Mihatsch 2016)

• Informally:
  
  Noun-noun compounds and their functional equivalents
The onomasiological perspective

• Equivalent to Type 3 (Štekauer 1998)
  – forms with a complex mark in which the determined (actional) element is not present

• Major consequence #1
  – synthetic compounds are out of scope
    • e.g. English truck-driver
    • because the determined element is present (DRIVE is not a thing-root)
  – NVN constructions are out of scope
    • e.g. Vietnamese būa ăn sang [meal eat morning] ‘breakfast’
    • again, the determined element (here: EAT) is present
Advocates uniform treatment of affixal and non-affixal morphemes
– No fundamental difference between -er of banker and -man of bankman
– Such affixes develop out of nouns (e.g. Booij 2005)
– Affixoids and bound nouns suggest a continuum rather than a sharp distinction between lexemes and affixes

Major consequence #2
– denominal nominalizations are in scope
  • Slovak želez.n.ica [iron.ADZ.NMLZ] ‘railway’
– noun classifier constructions are also in scope
  • Bora túú.heju [nose.CM(hole)] ‘nostril’
A different way to cut the cake

• Traditional compartmentalization of grammar
  – “Compounding forms a subsystem of grammar somewhat distinct from ordinary syntax and morphology” (Jackendoff 2010)
• Binominals offer an alternative perspective…
Seven languages, seven strategies

- Seven languages, all but one of them spoken in Europe

- Seven strategies for combining the meanings 'road' and 'iron' in order to denote the meaning 'railway':
  1. [Mod Head]
  2. [Mod.LE.Head]
  3. [Head PREP Mod]
  4. [Mod.ADJZ Head]
  5. [Head.PERT.Mod]
  6. [Mod.Head.POSS:3SG]
  7. [Mod.ADJZ.NMLZ]

RQ: What else is out there?

Binominal word-formation in the world’s languages

Project goal
To investigate the strategies employed by the languages of the world to create complex denotations by combining two nominal (or nominalizing) elements
Research questions

This is a data-driven, exploratory project, so the research questions are necessarily rather general

✓ What is the extent and diversity of binominal word-formation in the world’s languages?

✓ In what ways might binominal constructions be classified typologically?

✓ What generalizations can be made regarding binominal constructions and how can these be explained?

✓ How do the preference patterns of individual languages correlate with areal, genetic and typological features?
Theoretical framework

• Project conception
  – Cognitive Linguistics, esp. Construction Grammar (Croft 2001; Goldberg 2006; Langacker 2008; Booij 2010)
  – Contemporary typological theory (e.g. Haspelmath 2010)

• Grammatical descriptions
  – Basic Linguistic Theory (Dixon 2010/2012)

• Explanatory part
  – Cognitive Linguistics

Steve Pepper

Binominal word-formation in the world’s languages
Compounds
Prefiguring “binominals”
Lexical typology
Other

PREVIOUS RESEARCH
Compounding

• Typological studies
  – Bauer (2001) – well-balanced 36 language sample
    • few generalizations to report
  – Guevara & Scalise (2009) – 80,000 compounds, 21 languages
    • scales of preference only
  – Štekauer, Valera & Körtvélyessy (2012) – 70 language sample
    • scope includes derivation
    • main concern: presence or absence of different types of compound

• Other studies
  • Classification of compounds (Bisetto & Scalise 2005)
  • Compounding in individual languages (innumerable studies)
  • Compounds vs. phrases, compounding vs. derivation (various studies)
  • Semantic of compounding (e.g. Hacken 2016)
“Binominals”

- Some studies come close to the concept of binominals

  - Levi (1978)  
    ‘complex nominals’  
    - nominal compounds  
    - nominalizations  
    - NPs with non-predicating adjectives
  
  - Bauer & Tarasova (2013)  
    ‘adnominal nominal modification’  
    - noun-noun compounds  
    - associative adjectives  
    - possessive constructions  
    - neo-classical compounds
  
  - Rainer (2013)  
    ‘relational adjectives and their competitors’  
    - relational adjectives  
    - genitive  
    - noun-noun compounds  
    - prepositional phrases  
    - attributivizers  
    - derivational patterns
Lexical typology, etc.

• Data-driven, onomasiological studies involving semantic and morphological complexity
  – Haspelmath & Tadmor (2009)
    • *World Loanword Database* http://wold.clld.org
  – Urban (2012)
    • *Analyzeability and semantic associations in referring expressions: a study in comparative lexicology*

• Relevant to particular parts of the analysis
  – Studies of possession and the genitive (e.g. Koptjevskaja-Tamm 2002)
  – Studies of attribution (e.g. Rießler 2016)
  – ...
  – (TBD)
Onomasiological method
Meaning sample
Data sources
Language sample

DATA AND METHODOLOGY
Onomasiological methodology

• Define a set of meanings
  – Initial set: 201 “complex” meanings
  – Reduced set: 100 (92% retention of variety)

• Collect translation equivalents
  – online database(s)
  – dictionaries
  – questionnaires

• Annotate formal and semantic structure
  – Identify binominals
  – Identify construction types

• Analyse, classify, generalize, explain
Selecting the meanings

- Based on data in WOLD
  - Each meaning classified by semantic category
    - Noun, Verb, Adjective, Adverb, Function word
  - Each form classified as **analysable** (i.e. complex) or **unanalysable**

- SQL query
  - Select all meanings with semantic category **Noun** and sort them in descending order according to the number of languages in which they are represented by complex (**analysable**) forms

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id</td>
<td>meaning</td>
<td>langs</td>
</tr>
<tr>
<td>2</td>
<td>4-231</td>
<td>the nostril</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>14-45</td>
<td>the midday</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>4-213</td>
<td>the eyelid</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>17-26</td>
<td>the pupil</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>3-819</td>
<td>the spider web</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>2-71</td>
<td>the stepfather</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 159 meanings were complex in 16 more languages
- Supplemented by 42 meanings after testing against French, Welsh, Japanese, Turkish and Nizaa
Initial meaning sample (201)

NOSTRIL, MIDDAY, EYELID, PUPIL, SPIDER WEB, STEPFATHER, FISHERMAN, MERCHANT, PARENTS, STEPMOTHER, THUMB, DEFENDANT, EARLOBE, BREAKFAST, EARTHQUAKE, MURDER, NIPPLE OR TEAT, SKULL, SPINE, WATERFALL, EARWAX, POTTER, STEPSON, YOLK, CAPTIVE OR PRISONER, DIVORCE, DRINK, EYELASH, PLAINTIFF, BARK, SHORE, STEPDaugHTER, TAILOR, TOE, ANCESTORS, BEGGAR, FOOD, KID, MARRIED WOMAN, NATIVE COUNTRY, STRANGER, TEACHER, THIEF, ANXIETY, BAD LUCK, DAWN, HERDSMAN, QUARREL, SUPPER, WRIST, YOUNG MAN, ARSON, BEEHIVE, BEESWAX, BEGINNING, BIRTH CERTIFICATE, BLACKSMITH, BRACELET, DINNER, DOORPOST, EARRING, OLD WOMAN, PALM OF HAND, PITY, SORCERER OR WITCH, WHETSTONE, WHIRLPOOL, WIDOWER, ANKLE, DARKNESS, DESCENDANTS, GLOVE, HOSPITAL, HOST, LUNCH, MARRIED MAN, MEAL, PROSTITUTE, REMAINS, SCULPTOR, SHOEMAKER, SHOULDERBLADE, SWELLING, WOMB, 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, CALF, CHIEFTAIN, CROWD, DEFEAT, ENVY OR JEALOUSY, FARMER, FIREPLACE, FISHING LINE, FLAME, FOOTPRINT, GRIEF, JUDGMENT, LAMB, MARE, NIECE, OLDER SISTER, SCREWDRIVER, RAILWAY, SIBLING, SPECTACLES/GLASSES, SUNDAY, TWINS, VEIN OR ARTERY, YOUNGER SISTER, ANGER, BABY, EAST, ELECTION, END(2), FOAL OR COLT, GIRL, HANDKERCHIEF OR RAG, IDEA, INTENTION, ITCH, NEIGHBOUR, POSTCARD, PRAISE, QUEEN, RAINBOW, RAZOR, RIB, SCHOOL, VICTORY, WEAPONS, WEDNESDAY, WEST, YOUNG WOMAN, YOUNGER BROTHER, SPRING OR WELL, THATCH, TREE TRUNK, ARCTIC LIGHTS, TEAR, BICYCLE, EYEBROW, STABLE OR STALL, TRAIN, VINE, BEE, CAVE, ARMPIT, NECKLACE, POSTAGE STAMP, RAPE, NUT, BACKPACK, BICYCLE PUMP, BOW TIE, CAPITAL CITY, DAIRY COW, FLEA MARKET, GOLD RING, HAND BRAKE, HANDBAG, HORSESHOE, HUMMINGBIRD, KEYWORD, FIREWOOD, LIPSTICK, MAIL BOX, MILKY WAY, PADDLE WHEEL, FREEMAN, STONE BRIDGE, SUGAR CANE, TOILET PAPER, TOOLBOX, TOOTHBRUSH, WATER PUMP, WINDMILL
Reduced meaning sample (100)

ANKLE, ARCTIC LIGHTS, BACKPACK, BEE, BEEHIVE, BEESWAX, BICYCLE, BICYCLE PUMP, BLACKSMITH, BOY, BRACELET, BREAKFAST, CARPENTER, CHIEFTAIN, COCK/ROOSTER, COLLARBONE, COOKHOUSE, DAIRY COW, DINNER, DOORPOST, EARLOBE, EARRING, EARWAX, EYEBROW, EYELASH, EYELID, FARMER, FIREPLACE, FISHERMAN, FISHING LINE, FLAME, FLEA MARKET, FOAL OR COLT, FOOTPRINT, GIRL, GLOVE, GOLD RING, HAND BRAKE, HANDBAG, HANDKERCHIEF, HERDSMAN, HOSPITAL, HOST, KEYWORD, KID, LAMB, LICENSE PLATE, LUNCH, MAGIC, MAIL BOX, MARE, MARRIED WOMAN, MIDDAY, MILKY WAY, MOTHER-IN-LAW, NATIVE COUNTRY, NEIGHBOUR, NIECE, NIPPLE OR TEAT, NOSTRIL, PADDLE WHEEL, PALM OF HAND, POSTCARD, POTTER, QUEEN, RAILWAY, RAINBOW, RIB, SHOEMAKER, SHORE, SHOULDERBLADE, SKULL, SORCERER, SPECTACLES, SPIDER WEB, SPINE, STABLE ORSTALL, STONE BRIDGE, SUGAR CANE, SUNDAY, SUPPER, TEAR, THATCH, THUMB, TOE, TOILET, TOILET PAPER, TOOL, TOOLBOX, TOOTHBRUSH, TRAIN, TREE TRUNK, VEIN OR ARTERY, VINE, WATER PUMP, WEDNESDAY, WIDOWER, WINDMILL, WRIST, YOLK, MOTHER-IN-LAW (OF A MAN), NURSE, PERJURY, PESTLE, ROOF, SERVANT, TOILET, TOOL, WEDDING, WIDOW, CALF, CHIEFTAIN, CROWD, DEFEAT, ENVY OR JEALOUSY, FARMER, FIREPLACE, FISHING LINE, FLAME, FOOTPRINT, GRIEF, JUDGMENT, LAMB, MARE, NIECE, OLDER SISTER, SCREWDRIVER, RAILWAY, SIBLING, SPECTACLES/GLASSES, SUNDAY, TWINS, VEIN OR ARTERY, YOUNGER SISTER, ANGER, BABY, EAST, ELECTION, END(2), FOAL OR COLT, GIRL, HANDKERCHIEF OR RAG, IDEA, INTENTION, ITCH, NEIGHBOUR, POSTCARD, PRAISE, QUEEN, RAINBOW, RAZOR, RIB, SCHOOL, VICTORY, WEAPONS, WEDNESDAY, WEST, YOUNG WOMAN, YOUNGER BROTHER, SPRING OR WELL, THATCH, TREE TRUNK, ARCTIC LIGHTS, TEAR, BICYCLE, EYEBROW, STABLE OR STALL, TRAIN, VINE, BEE, CAVE, ARMPIT, NECKLACE, POSTAGE STAMP, RAPE, NUT, BACKPACK, BICYCLE PUMP, BOW TIE, CAPITAL CITY, DAIRY COW, FLEA MARKET, GOLD RING, HAND BRAKE, HANDBAG, HORSESHOE, HUMMINGBIRD, KEYWORD, FIREWOOD, LIPSTICK, MAIL BOX, MILKY WAY, PADDLE WHEEL, FREEMAN, STONE BRIDGE, SUGAR CANE, TOILET PAPER, TOOLBOX, TOOTHBRUSH, WATER PUMP, WINDMILL...
Data sources

• World Loanword Database (WOLD)
  – 41 vocabularies, one per language
  – 1,460 meanings
  – morpheme glosses of analysable words

• Dictionaries
  – e.g. Brindle (2016)

• Language experts
  – 201 meaning questionnaire completed by 20 contributors
  – 100 meaning questionnaire now ready for distribution
Language sample

• Types of language sample
  – probability (stratified)
    • areally and genetically as unbiased as possible
    • permits the broadest generalizations and predictions
    • as described by Dryer (1989)
  – variety
    • optimized in order to illustrate diversity as fully as possible
    • as described by Miestamo et al. (2016)
  – convenience (opportunity)
    • based primarily on the availability of data

• Current sample
  – 50 languages in DB, 16 languages received, 60 promised
  – Goal is 200 languages
# 59 languages (already collected)

## EURASIA
- Irish (Celtic)
- Welsh (Celtic)
- Dutch (Germanic)
- English (Germanic)
- German (Germanic)
- Old High German (Germanic)
- Greek (Greek)
- Italian (Romance)
- Romanian (Romance)
- Selice Romani (Romani)
- Kildin Saami (Finnic)
- Czech (Slavic)
- Lower Sorbian (Slavic)
- Polish (Slavic)
- Russian (Slavic)
- Lithuanian (Baltic)
- Hindi (Indo-Aryan)
- Archi (Lezgic)
- Bezhta (Avar-Andic-Tsezic)
- Oroqen (Tungusic)
- Sakha (Turkic)
- Ket (Yeniseian)
- Japanese (Japanese)

## AUSTRALIA-NEW GUINEA
- Gurinji (Pama-Nyungan)
- Ceq Wong (Aslian)
- Malagasy (Barito)
- Manange (Bodic)
- Äiwo (Malayo-Polynesian)
- Hmong Daw (Hmong-Mien)
- Thai (Kam-Tai)
- Indonesian (Malayan)
- Hawaiian (Oceanic)
- Takia (Oceanic)
- Sangir (Sangiric)
- Chinese (Sinitic)
- Vietnamese (Vietic)

## S-E ASIA AND OCEANIA
- Saramaccan (English-based)
- Seychelles Creole (French-based)

## SOUTH AMERICA
- Mapudungun (Araucanian)
- Galibi (Carib)
- Murui (Huitotoan)
- Wichi (Matacoan)
- Hupdé (Nadahup)
- Quechua (Quechuan)

## AFRICA
- Swahili (Bantu)
- Tarifiyt Berber (Berber)
- Bandial (Atlantic)
- Bainouk Gubeeher (Atlantic)
- Hausa (Chadic)
- Gawwada (Cushitic)
- Iraqw (Cushitic)
- Datooga (Nilotic)
- Kanuri (Saharan)
- Amharic (Semitic)
- Maltese (Semitic)

## NORTH AMERICA
- Q'eqchi' (Mayan)
- Tzotzil (Mayan)
- Otomi (Otomanguean)
- Yaqui (Uto-Aztecan)

(Boldface = WOLD language)
17 more languages signed up at SLE

**EURASIA**
- Balkan Gagauz Turkish (Turkic)
- Ancient Greek (Greek)
- Piedmont Sintí (Indo-Aryan)
- Spanish (Romance)
- Slovak (Slavic)
- Basque (Basque)
- Estonian (Finnic)
- Finnish (Finnic)
- Hill Mari (Mari)
- Moksha (Mordvin)
- Kamas (Permic)

**AFRICA**
- Aramaic (Semitic)

**AUSTRALIA-NEW GUINEA**
- Bunak (Timor-Alor-Pantar)

**SOUTH AMERICA**
- Cabécar (Chimchan)
- Bora (isolate)
- Harakmbut (isolate)
- Moseten (isolate)
# 28 more languages signed up at ESSLT

**EURASIA**  
Latvian (Baltic)  
Assamese (Indo-Aryan)  
Nepali (Indo-Aryan)  
Ingrian (Finnic)  
Hungarian (Hungarian)  
Khanty (Khantyic)  
Turkish (Turkic)  

**AFRICA**  
Barein (Chadic)  
Hadiyya (Cushitic)  
Sidamo (Cushitic)  
Hebrew (Semitic)  
Chuwabu (Bantu)  
Minianka (Kwa)  
Bambara (Mande)  
Kupsabiny (Southern Nilotic)  

**S-E ASIA AND OCEANIA**  
Bunun (Formosan)  
Rukai (Formosan)  
Saisiyat (Formosan)  
Thai (Kam-Tai)  

**AUSTRALIA-NEW GUINEA**  
Ngarnga (Mirndi)  
Wik-Mungkan (Pama-Nyungan)  
Baining (Baining)  
Tuwari (Walio)  

**SOUTH AMERICA**  
Ye'kwana (Guianan)  
Saliban (Jodi-Saliban)  
Ticuna (Ticuna-Yuri)  
Guarani (Tupi-Guarani)  

**NORTH AMERICA**  
Mixtec (Otomanguean)  
Seri (isolate)
Language sample as of 9/2016
Cross-linguistic frequency
Cross-linguistic variation (EYELID)
Conceptual variation (EYELID, TRAIN)
Regional variation (Europe, S-E Asia, Central Asia, Africa, S. America)
Genetic resemblances (Germanic, Romance, Slavic)
Intra-language competition (Lower Sorbian)

PRELIMINARY FINDINGS
Preliminary findings

• Based on WOLD 41 (+ German, Polish, French, Italian)

• Cross-linguistic frequency of binominals
  – from Mandarin to Selice Romani

• Cross-linguistic variation (eyelid)

• Conceptual variation (eyelid, train)

• Regional variation
  – Europe, South-East Asia, “Central Asia”, Africa, South America

• Genetic resemblance (and contact)
  – Germanic, Romance and Slavic

• Intra-language competition (Lower Sorbian)
Cross-linguistic frequency

Binominals as a percentage of complex nominals
# Cross-linguistic variation \textbf{EYELID} (1)

12 head-initial binominals (various constructions)

<table>
<thead>
<tr>
<th>language</th>
<th>word [gloss]</th>
<th>c1</th>
<th>c2</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceq Wong</td>
<td>hoʔ mět [skin eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>lihilihi maka [rim eye]</td>
<td>rim</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Indonesian</td>
<td>kelopak mata [sheath eye]</td>
<td>sheath</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Seychelles Creole</td>
<td>lapo lizye [skin eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Thai</td>
<td>plùak.taa [skin-eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>White Hmong</td>
<td>tawv muag [skin eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Kanuri</td>
<td>fěrtè shîm.bè [root eye.GEN]</td>
<td>root</td>
<td>eye</td>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>Tarifiyt Berber</td>
<td>aʕrua n tit’t’ [back PREP eye]</td>
<td>back</td>
<td>eye</td>
<td>Head PREP CON.Mod</td>
</tr>
<tr>
<td>Lower Sorbian</td>
<td>lapka na woku [flap PREP eye]</td>
<td>flap</td>
<td>eye</td>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>Hausa</td>
<td>fáatà.r ídòo [skin.LK eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head.LK Mod</td>
</tr>
<tr>
<td>Malagasy</td>
<td>hodi.màso [skin eye]</td>
<td>skin</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
<tr>
<td>Otomi</td>
<td>xi-da [body_hair-eye]</td>
<td>body hair</td>
<td>eye</td>
<td>Head Mod</td>
</tr>
</tbody>
</table>
# Cross-linguistic variation \textbf{EYELID} (2)

13 head-final binominals (various constructions)

<table>
<thead>
<tr>
<th>language</th>
<th>word [gloss]</th>
<th>c1</th>
<th>c2</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hup</td>
<td>kǝwǝg b’ók [eye skin]</td>
<td>eye</td>
<td>skin</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Quechua</td>
<td>ñawi kara [eye skin]</td>
<td>eye</td>
<td>skin</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Kali’na</td>
<td>enu[lu].pipo [eye.skin]</td>
<td>eye</td>
<td>skin</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Saramaccan</td>
<td>wójo.kókóo [eye.shell]</td>
<td>eye</td>
<td>shell</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Takia</td>
<td>mala.n sklo.n [eye.3SG skin.3SG]</td>
<td>eye</td>
<td>skin</td>
<td>Mod.3SG Head.3SG</td>
</tr>
<tr>
<td>Bezhta</td>
<td>häyš ƛ’äq’e [eye:GEN roof]</td>
<td>eye</td>
<td>roof</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>Ket</td>
<td>dēs-d-iŋolt [eye-POSS-skin]</td>
<td>eye</td>
<td>skin</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>Dutch</td>
<td>oog.lid [eye.lid]</td>
<td>eye</td>
<td>lid</td>
<td>Mod.Head</td>
</tr>
<tr>
<td>English</td>
<td>eye.lid [eye.lid]</td>
<td>eye</td>
<td>lid</td>
<td>Mod.Head</td>
</tr>
<tr>
<td>Kildin Saami</td>
<td>čall’m.rūmtas [eye.brim]</td>
<td>eye</td>
<td>brim</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Mandarin</td>
<td>yan3.pi2 [eye.skin]</td>
<td>eye</td>
<td>skin</td>
<td>Mod Head</td>
</tr>
<tr>
<td>Archi</td>
<td>lul-li-n dorki [eye.OBL.GEN top]</td>
<td>eye</td>
<td>top</td>
<td>Mod.OBL.GEN Head</td>
</tr>
<tr>
<td>Wichí</td>
<td>tot.telhu t’oj [POSS:IND.eye skin/hide]</td>
<td>eye</td>
<td>skin/hide</td>
<td>POSS:IND.Mod Head</td>
</tr>
</tbody>
</table>
### Cross-linguistic variation EYELID (3)

#### 7 non-binominals

<table>
<thead>
<tr>
<th>language</th>
<th>word [gloss]</th>
<th>c1</th>
<th>c2</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manange</td>
<td>2mi.2pʰi [eye.up]</td>
<td>eye</td>
<td>up</td>
<td>N + Adv</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>wente nge [over eye]</td>
<td>over</td>
<td>eye</td>
<td>Adv + N</td>
</tr>
<tr>
<td>Old High German</td>
<td>slegi.brâwa [hit.eyebrow]</td>
<td>hit</td>
<td>eyebrow</td>
<td>V + N</td>
</tr>
<tr>
<td>Oroqen</td>
<td>yɛ:ja.ŋi wu.n [eye.GEN up.3SG:POSS]</td>
<td>eye</td>
<td>up</td>
<td>N + Adv</td>
</tr>
<tr>
<td>Otomi</td>
<td>yoda [yo = ?; da = eye]</td>
<td>?</td>
<td>eye</td>
<td>? + N</td>
</tr>
<tr>
<td>Q'eqchi'</td>
<td>r.ix u [3ERG.back surface]</td>
<td>back</td>
<td>surface</td>
<td>(opaque)</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>mí [eyelid]</td>
<td></td>
<td></td>
<td>(monomorphemic)</td>
</tr>
</tbody>
</table>
Conceptual variation

EYELID
- Modifier
  - eye (100%)
- Head
  - skin/hide (52%)
  - lid/sheath/shell (20%)
  - brim/rim (8%)
  - roof/top (8%)
  - root/back (8%)
  - body hair (4%)

TRAIN
- Head
  - car, vehicle, cart (100%)
- Modifier
  - electricity (8%)
  - fire (46%)
  - ground, land (15%)
  - railway (15%)
  - smoke, steam (15%)

POI: Variation in head diversity vs. modifier diversity: salience?
# Regional variation (1) Europe (SAE)

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod Head</td>
<td>Mod Head</td>
<td>Mod.LE.Head</td>
</tr>
<tr>
<td>Mod.ADJZ Head</td>
<td>Head PREP Mod</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>Base.PL</td>
<td>Head Mod.ADJZ</td>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>Base.AGT</td>
<td>Head Mod</td>
<td>Head Mod.GEN</td>
</tr>
<tr>
<td>Base.PL</td>
<td>Head Mod</td>
<td>Head Mod.ADJZ</td>
</tr>
<tr>
<td>Base.NMLZ</td>
<td>Head Mod</td>
<td>Base.PL</td>
</tr>
<tr>
<td>Base.F</td>
<td>Head Mod</td>
<td>Base.NMLZ</td>
</tr>
<tr>
<td>Base.DIM</td>
<td>Head Mod</td>
<td>Base.F</td>
</tr>
<tr>
<td></td>
<td>Base.DIM</td>
<td>Base.DIM</td>
</tr>
<tr>
<td></td>
<td>Base.AUG</td>
<td>Base.AUG</td>
</tr>
<tr>
<td></td>
<td>Base.AGT</td>
<td>Base.AGT</td>
</tr>
<tr>
<td></td>
<td>Base.ADJZ.DIM</td>
<td>Base.ADJZ.DIM</td>
</tr>
<tr>
<td></td>
<td>NEG.Base</td>
<td>NEG.Base</td>
</tr>
</tbody>
</table>

**POI:** Strikingly different patterns in three SAE languages
Regional variation (2) South-East Asia

<table>
<thead>
<tr>
<th>Language</th>
<th>Mod Head</th>
<th>Base.SYN</th>
<th>Base.RED</th>
<th>Base.NMLZ</th>
<th>Head Mod</th>
<th>Base.NMLZ</th>
<th>Mod Head</th>
<th>Head Mod</th>
<th>Base.NMLZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>108</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>55</td>
<td>1</td>
<td>10</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>Hmong</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnamese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thai</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

POI: Predominance of **Mod Head** (or **Head Mod**) plus some borrowing
Regional variation (3) “Central Asia”

<table>
<thead>
<tr>
<th>Archi</th>
<th>Bezhta</th>
<th>Ket</th>
<th>Sakha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod.GEN Head</td>
<td>Mod Head</td>
<td>Mod Head</td>
<td>Mod Head.Gen</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Mod.ADJZ Head</td>
<td>Mod.GEN Head</td>
<td>Mod.GEN Head</td>
<td>Mod Head</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Mod Head</td>
<td>Mod.ADJZ Head</td>
<td>Mod.PrIV Head</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Mod.ABL Head</td>
<td></td>
<td>Base.VR</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mod.SUF Head</td>
<td></td>
<td>Base.VR.NMLZ</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

POI: GEN constructions and competition (?) Mod Head vs. Mod Head.Gen
### Regional variation (4) Africa

<table>
<thead>
<tr>
<th></th>
<th>Hausa</th>
<th>Kanuri</th>
<th>Berber</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head.LK Mod</td>
<td>46</td>
<td>Head Mod GEN</td>
<td>Head OF CON.Mod 6</td>
</tr>
<tr>
<td>Base.ABSTR</td>
<td>2</td>
<td>Head Mod ASS</td>
<td>Head 3SG.OF CON.Mod 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mod Head</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.AGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.PLACE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.PLACE2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.PLACE3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.SON</td>
<td></td>
</tr>
</tbody>
</table>

**Swahili**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Head CON Mod</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base.ABSTR</td>
<td></td>
<td>Head Mod</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CM Base</td>
<td>11</td>
</tr>
</tbody>
</table>

**POI:** Variety of pertensive, genitive, construct and connective strategies
### Regional variation (5) South America

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>20</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kali’na</td>
<td></td>
<td>Mod.Poss.Head.Poss</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapudungun</td>
<td></td>
<td></td>
<td></td>
<td>Head.Mod</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hup</td>
<td>36</td>
<td></td>
<td>8</td>
<td>Mod.Head</td>
<td>Mod.Vblz.Agt.Head</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wichí</td>
<td></td>
<td>Mod.Head</td>
<td>15</td>
<td></td>
<td>Base.Loc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mod.Head.LOC</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mod.Clf.Gen.Head</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poss.Mod.Head</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poss.Base.LOC</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poss.Base.AGT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base.AGT</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**POI:** Variety of strategies, including possessives and classifiers
Genetic resemblance (and contact)

Germanic, Romance & Slavic

POI: Confirms English / French / Polish pattern across each genus
### Competition Lower Sorbian (1)

14 adjectival constructions

<table>
<thead>
<tr>
<th>meaning</th>
<th>word [gloss]</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEESWAX</td>
<td>pcołk.owy wósk [bee.ADJ wax]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>DOORPOST</td>
<td>żuri.ny słup [door.ADJ post]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>EARLOBE</td>
<td>wuch.owa lapka [ear.ADJ flap]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>FISHING LINE</td>
<td>wużeń.ska šnora [?worm.ADJ line]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>LICENSE PLATE</td>
<td>cysł.owa tofl.icka [number.ADJ plate:DIM]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>LICENSE PLATE</td>
<td>numer.owa tofl.icka [number.ADJ plate:DIM]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>MOTHER-IN-LAW</td>
<td>pśichod.na maś [future.ADJ mother]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>NATIVE COUNTRY</td>
<td>wóśc.ny kraj [ancestor.ADJ country]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>NOSTRIL</td>
<td>nos.owa żérka [nose.ADJ hole]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>POSTAGE STAMP</td>
<td>list.owa znam.ka [letter.ADJ sign.DIM]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>POSTCARD</td>
<td>post.owa kórt.ka [post.ADJ card.DIM]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>THATCH</td>
<td>słomjane kšywo [straw:ADJ roof]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>VINE</td>
<td>win.owy keř [vine.ADJ bush]</td>
<td>Mod.ADJZ Head</td>
</tr>
<tr>
<td>WRIST</td>
<td>ruc.ne zgibk [hand.ADJ joint]</td>
<td>Mod.ADJZ Head</td>
</tr>
</tbody>
</table>
Competition Lower Sorbian (2)

2 prepositional and 2 genitival constructions

<table>
<thead>
<tr>
<th>meaning</th>
<th>word [gloss]</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>EARWAX</td>
<td>taług z wucha [wax PREP ear]</td>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>EYELID</td>
<td>lapka na woku [flap PREP eye]</td>
<td>Head PREP Mod</td>
</tr>
<tr>
<td>NIECE</td>
<td>bratša źowka [brother:GEN daughter]</td>
<td>Mod.GEN Head</td>
</tr>
<tr>
<td>NIECE</td>
<td>sotšy źowka [sister:GEN daughter]</td>
<td>Mod.GEN Head</td>
</tr>
</tbody>
</table>
# Competition Lower Sorbian (3)

12 derivations

<table>
<thead>
<tr>
<th>meaning</th>
<th>word [gloss]</th>
<th>construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PITY</td>
<td>lut.osć [sorrow:N.SUF]</td>
<td>Base.ABST</td>
</tr>
<tr>
<td>FISHERMAN</td>
<td>ryb.ař [fish.SUF]</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>GUARD</td>
<td>straž.nik [guard(duty).SUF]</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>HOST</td>
<td>gósć.inař [guest.SUF]</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>POTTER</td>
<td>gjarnc.ař [pot.SUF]</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>TEACHER</td>
<td>šula.ř [school.AGT]</td>
<td>Base.AGT</td>
</tr>
<tr>
<td>BABY</td>
<td>góle.tko [child.SUF]</td>
<td>Base.DIM</td>
</tr>
<tr>
<td>QUEEN</td>
<td>kral.owka [king.SUF]</td>
<td>Base.F</td>
</tr>
<tr>
<td>SCREWDRIVER</td>
<td>šrub.owak [screw.SUF]</td>
<td>Base.INST</td>
</tr>
<tr>
<td>FIREPLACE</td>
<td>wogni.šćo [fire.SUF]</td>
<td>Base.LOC</td>
</tr>
<tr>
<td>GLOVE</td>
<td>ruka.jca [hand.SUF]</td>
<td>Base.SUF</td>
</tr>
<tr>
<td>MERCHANT</td>
<td>wik.owa.ř [market.VBLZ.SUF]</td>
<td>Base.VRBZ.AGT</td>
</tr>
</tbody>
</table>
Formal analysis
Semantic analysis
Typological classification
Statistical analysis

ANALYSIS
Formal analysis

• Two kinds of formal analysis
  – based on language-specific descriptive categories
  – based on cross-linguistic comparative concepts

• Language-specific analysis
  – uses descriptive categories particular to the language in question
  – follows “standard” grammars
  – e.g. Hausa “linker”
    • dôoki.n kárfeè [horse.LK metal] BICYCLE ➔ Head.LK Mod

• Cross-linguistic analysis
  – uses comparative concepts applicable cross-linguistically
  – e.g. Hausa “linker” is actually a pertensive (Dixon 2010)
  – marking on the possessed (head), not the possessor (dependent)
Semantic analysis

• Focus primarily on semantic relations
• Apply an existing classification system
  – probably Bourque (2014)
• Look for further evidence in support of the “two paths” hypothesis
  – (Pepper 2010, 2016)

• Consider relevance of semantic maps
  – (Croft 2003)

• (TBD)
Typological classification

**Formal**
- Formal parameters
  - synthetic, analytic
  - head-marking, dependent
  - etc.
- Koptjevskaja-Tamm (2002)
  - possessive noun phrases in Europe

**Semantic**
- Semantic parameters
  - possession
  - part-whole
  - location
  - attribution
  - (TBD)
Statistical analysis

• Correlations with
  – constituent order
    (cf. work on Bauer36)

<table>
<thead>
<tr>
<th>Type</th>
<th>A</th>
<th>E</th>
<th>O</th>
<th>G</th>
<th>N</th>
<th>S</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPoss+NMod</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>NPoss+ModN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PossN+NMod</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PossN+ModN</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>other</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>tbd</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
</tbody>
</table>

Compound head / Poss. phrase

<table>
<thead>
<tr>
<th></th>
<th>NMod</th>
<th>ModN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPoss</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>other:7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PossN</td>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>

• Other correlations
  – (TBD)

• Other statistical applications
  – (TBD)
Conclusion

• This project is just past the halfway stage
  – The functionally motivated comparative concept of binominals seems to offer much richer and more interesting variation than the formally motivated concept of noun-noun compounds (or compounds in general)
  – Binominals cut across traditional boundaries between phrases, compounds and derivations.
  – But do they challenge the compartmentalization of grammar into syntax, morphology and lexicon, and if so, how?

• There is still much to be explored
  – Not least regarding semantic relations
  – And there is a pressing need for greater theoretical clarity regarding terminology, definitions, etc.
  – And I need more data…
Bibliography (1/4)


Bibliography (2/4)


