

Typology: Lecture V

Morphology

Christian Bentz
University of Tübingen

May 23, 2017



WORDS BONES GENES TOOLS

Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past

EVOLAEMP

LANGUAGE EVOLUTION: THE EMPIRICAL TURN

OVERVIEW

INTRODUCTION

THREE WAY MORPHOLOGICAL TYPOLOGY

- Formatives

- Fusion

- Flexivity

- Exponence

- Combinations

- Synthesis

MEASURING MORPHOLOGICAL COMPLEXITY

- Inventory-Based Accounts

- Distribution-Based Accounts

RECENT RESEARCH

- Qualitative Studies

- Quantitative Studies

MORPHOLOGICAL BASICS

Morpheme

Minimal unit which carries meaning or has a grammatical function

e.g.: *un-believe-able-ly*

schema: *prefix-root-suffix-suffix*

Note: **affixes** can be **prefixes**, **suffixes**, **infixes**, or **circumfixes**

MORPHOLOGICAL BASICS

Morphemes can be classified as **free** or **bound**

- ▶ **bound morphemes:** depend phonologically and/or syntactically on another element, i.e. appear together with a **root** and often combine with other bound morphemes to turn a **lexeme** into a **word form**
e.g.: lexeme: *run* → *run, runs, ran, (running)*
- ▶ **free morphemes:** are phonologically and/or syntactically independent of other elements, e.g. prepositions such as *with, to, after* etc. in English
- ▶ Problem: clitics like the English genitive 's are in between bound and free morphemes
e.g. *the [book]'s cover* but *the [book I bought yesterday]'s cover*

MORPHOLOGICAL BASICS: INFLECTION VS. DERIVATION

- ▶ bound morphemes are either **inflectional** or **derivational**

- ▶ **inflection:**

- ▶ applies to every word of a category, i.e. part of speech (noun, verb, adjective, etc.)
- ▶ does not change the category
- ▶ has predictable semantics

e.g.: *house* → *house-s* (noun.SG → noun-PL)

- ▶ **derivation**

- ▶ does not apply systematically to every word of a category
- ▶ can change the category
- ▶ is semantically less predictable

e.g.: *sing* → *sing-er* (verb → noun)

but: *jump* → *jump-er* (person or object)

MORPHOLOGICAL TYPOLOGY

(1) Hawaiian (Austronesian)

A **ua** **olelo aku o** Ioane **ia ia** [...]

Then PERF say.to SUBJ.Johan he.DAT [...]

"Then Johan said to him [...]"

(2) Iñupiatun (Eskimo-Aleut)

Aglaan Jesus-ngum **itnaḡnigai** [...]

But Jesus-ERG this.say.report.3S.to.3PL

"But Jesus said to them [...]"



EDWARD SAPIR'S MORPHOLOGICAL TYPOLOGY

Edward Sapir (1921) systematizes morphological typology along **two dimensions**:

1. **degree of fusion** at the level of **morphemes**:
isolating - agglutinative - fusional
2. **degree of synthesis** at the level of **words**:
analytic - synthetic - polysynthetic

[Sapir, Edward (1921). *Language: An introduction to the study of speech*. New York: Harcourt.]

BICKEL'S AND NICHOLS' MORPHOLOGICAL TYPOLOGY

More recently, Bickel and Nichols (2005, 2007) give a **three way distinction** at the level of morphemes:

1. **Fusion:** phonological merging of formatives with their hosts
2. **Flexivity:** Degree of allomorphy, i.e. extent of declension and conjugation classes
3. **Exponence:** degree to which different categories, e.g. number and case, are expressed by the same formative

[Bickel, Balthasar and Johanna Nichols (2007). Inflectional morphology. In: Shopen, Timothy. *Language typology and syntactic description* (revised Second Edition). Cambridge: Cambridge University Press, 169-240.]

FORMATIVES

“**Formatives** are the markers of **inflectional information**. [...] They are different from words in that they cannot *govern* or be *governed* by other words [i.e. determine the inflectional category of another word], cannot require or undergo *agreement*, and cannot *head phrases*: formatives are *morphological* entities, words *syntactic*.”

[Bickel & Nichols (2007), pp. 172-173]

FORMATIVES

can be **inflections (1)**, **particles (2)** or **clitics (3)**:

(1) German (Indo-European)

ein-**e** gut-**e** Lehrer-in
 a-NOM.SG.FEM good-NOM.SG.FEM teacher.NOM.SG-FEM
 "a good (female) teacher"

(2) Lai Chin (Tibeto-Burman)

Tsew Máŋ **ni?** ʔa-ka-thoʔŋ
 Tsew Mang ERG 3SG.A-1SG.P-hit
 "Tsew Mang hit me"

(3) Turkish (Turkic)

sen=**mi**; yarın=**mı**; gör-dü-n=**mü**
 me=Q; tomorrow=Q; see-PAST-2SG=Q
 "me?"; "tomorrow?"; "did you see?"

[Bickel & Nichols (2007), pp. 170-175]

Fusion

isolating

formatives are
phonologically
independent

concatenative

formatives are
phonologically
bound

nonlinear

formatives are not
phonologically
segmentable

Flexivity

non-flexive

formatives have no allomorphs

flexive

formatives have allomorphs

Exponence

cumulative

one formative has several
grammatical functions

separative

each formative has a different
grammatical function

FUSION: **ISOLATING** < CONCATENATIVE < NONLINEAR

Formatives are **free, phonologically unbound** morphemes

(4) Hawaiian (Austronesian)

A **ua** olelo aku **o** Ioane ia ia [...]

Then PERF say to SUBJ Johan he.DAT

"Then Johan said to him [...]"

FUSION: ISOLATING < **CONCATENATIVE** < NONLINEAR

Formatives are **phonologically bound** and need some other word for their realization

Turkish (Turkic)

ev	→ house (house.NOM.SG)
ev- ler	→ houses (house.NOM- PL)
ev- i	→ his/her house (house.NOM.SG- POSS)
ev- ler-i	→ his/her houses (house- PL-POSS)
ev- den	→ out of the house (house.SG- ABL)
ev- ler-den	→ out of the houses (house- PL-ABL)

Note: There can still be *vowel harmony* (e.g. **ad-lar** “name-PL”), *assimilation* (e.g. **git-ti** “go-PAST” versus **gel-di** “come-PAST”), **dissimilation** (inverse process of assimilation), and **elision** (dropping of phonetic material).

FUSION: ISOLATING < CONCATENATIVE < **NONLINEAR**

Formatives are **not segmentable** into linear strings

Modern Hebrew (Semitic)

g-d-r → enclose (enclose.INF)

g**a**dar → he enclosed (enclose.**PAST.3SG.M.A**)

g**u**dar → he was enclosed (enclose.**PAST.3SG.M.P**)

gd**o**r → enclose it! (enclose.**FUT.2SG.IMP**)

FLEXIVITY: **NON-FLEXIVE** < FLEXIVE

Non-flexive formants are *invariable across the lexicon*, i.e. do not display different *allomorphs* according to declension and conjugation classes.

Examples:

- ▶ The ergative formative *ni?* in Lai Chin in example (4) is not influenced by the lexical properties of the part of speech it modifies
- ▶ Turkish formatives such as the question clitic in *sen=mi*, *yarın=mi*, and *gör-dü-n=mi* of example sentence (5) are also considered **non-flexive**. The variance is here explained by a general phonological process (vowel harmony) rather than a strict division of the lexicon into inflectional classes

FLEXIVITY: NON-FLEXIVE < **FLEXIVE**

Flexive formatives display *allomorphy*, i.e. have different variants according to conjugation and declension classes. Note that conjugation and declension classes are *lexically* triggered phenomena.

	1st declension	2nd declension	
	stems in -a	stems in -o	
	feminine	masculine	neuter
singular			
nom.	puell-a (<i>girl</i>)	domin-us (<i>master</i>)	bell-um (<i>war</i>)
gen.	puell-ae	domin-ī	bell-ī
dat.	puell-ae	domin-ō	bell-ō
acc.	puell-am	domin-um	bell-um
abl.	puell-ā	domin-ō	bell-ō
plural			
nom.	puell-ae	domin-ī	bell-a
gen.	puell-ārum	domin-ōrum	bell-ōrum
dat.	puell-īs	domin-īs	bell-īs
acc.	puell-ās	domin-ōs	bell-a
abl.	puell-īs	domin-īs	bell-īs

[Morwood (1999). *A Latin Grammar*. Oxford University Press.]

EXPONENCE: CUMULATIVE < SEPARATIVE

Cumulative exponence means that different grammatical categories (e.g. number and case) are encoded in the *same* *formatives*.

- In Latin (see above), the formative in *puell-am* encodes both **SG** and **ACC**.
- Note that this can also be the case in *isolating languages*:

(5) Hausa (Afroasiatic)

Mūsā yā tàfi Bicì
M. 3SG.MASC:COMPL go B.

“Musa went/has gone to Bichi”

yârā **sun** ga macìjî-n?
children 3PL.COMPL see snake -ART.PL

"Did the children see the snakes"

EXPONENCE: CUMULATIVE < **SEPARATIVE**

Separative exponence means that different grammatical categories (e.g. number and case) are encoded in *separate formatives*.

Example:

In the Turkish example from above *ev-**ler-den*** “out of the houses” (house-**PL-ABL**), number and case are encoded with two separate formatives

COMBINATIONS

FLEXIVITY AND FUSION

Traditionally, flexive languages like Latin have been automatically associated with higher degrees of fusion, and isolating languages with lower degrees of fusion. However, any conceivable combination of flexivity and fusion is possible and attested (though they are more or less common):

flexive-isolating: Yidj (Pama-Nyungan)

flexive-concatenative: Latin (Indo-European)

flexive-nonlinear: Modern Hebrew (Afroasiatic)

nonflexive-isolating: Lai Chin (Sino-Tibetan)

nonflexive-concatenative: Turkish (Turkic)

nonflexive-nonlinear: Kinyarwanda (Atlantic-Congo, Bantu)

[see Bickel & Nichols (2007): 186-188]

COMBINATIONS

EXPONENCE AND FLEXIVITY/FUSION

Exponence is generally independent of both *flexivity* and *fusion*. However, there is some tendency for nonflexive concatenative (“agglutinative”) morphology to be separative (as in Turkish), and for flexive formatives to be cumulative (as in Latin and Russian). Likewise, cumulative exponence is most likely found with bound morphology, though not necessarily.

[see Bickel & Nichols (2007): 188-189]

SYNTHESIS

While *fusion*, *flexivity* and *exponence* are relevant for formants (i.e. at the level of the morpheme), **synthesis** is a category that plays out at the **word level**. Here, the traditional cline holds:

analytic < synthetic < polysynthetic

[see Bickel & Nichols (2007): 189 pp.]

SYNTHESIS

“[...] [is] measured by the **number of formatives and lexical roots** that are bound together in one word: one or very few formatives and at most one root in the case of **analytic words**, a moderate number of formatives together with one root in **synthetic words**, and an abundant mixture of formatives and lexical roots in **polysynthetic words**.”

[see Bickel & Nichols (2007): 189]

MEASURING MORPHOLOGICAL COMPLEXITY

Inventory-Based Accounts

Estimate the *inventory size* of morphological categories, i.e. count the number of different case markers, tense/aspect markers, etc.

Advantages: expert judgement, robust, cross-linguistically comparable

Disadvantages: coarse grained, does not reflect actual language production, often opaque and non-reproducible

MEASURING MORPHOLOGICAL COMPLEXITY

Distribution-Based Accounts

Measure the actual *distribution* of different word types/markers over the given morphological categories in language production

Advantages: fine-grained, reflects language production, reproducible

Disadvantages: corpus dependence, question of extrapolation/cross-linguistic comparability

INVENTORY-BASED ACCOUNTS

Inflectional Synthesis of the Verb (WALS, Chapter 22)

Values

●	0-1 category per word	5
●	2-3 categories per word	24
○	4-5 categories per word	52
●	6-7 categories per word	31
●	8-9 categories per word	24
●	10-11 categories per word	7
●	12-13 categories per word	2

Number of Cases (WALS, Chapter 49)

Values

○	No morphological case-marking	100
○	2 cases	23
●	3 cases	9
●	4 cases	9
●	5 cases	12
●	6-7 cases	37
●	8-9 cases	23
●	10 or more cases	24
◇	Exclusively borderline case-marking	24

DISTRIBUTION-BASED ACCOUNTS

Example: Word for “brother” in the Bible

► **Latin**

01004008 Dixitque Cain ad Abel **fratrem** suum [...]

01004009 Ubi est Abel **frater** tuus?

01004011 [...] suscepit sanguinem **fratris** tui de manu tua!

► **Italian**

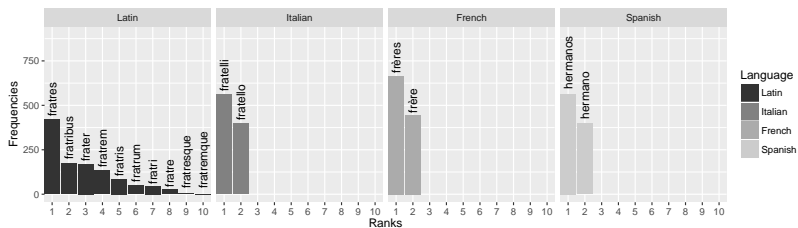
01004008 Caino disse al **fratello** Abele [...]

01004009 Dov'è Abele , tuo **fratello**?

01004011 [...] ha bevuto il sangue di tuo **fratello**!

DISTRIBUTION-BASED ACCOUNTS

The word for “brother” in Bible translations



[Bentz & Berdicevskis (2016). Learning pressures reduce morphological complexity: linking corpus, computational and experimental evidence. *Proceedings of COLING 2016*.]

DISTRIBUTION-BASED ACCOUNTS

Table 1. Forms extracted with the Greek lemma *ÁNTHRŌPOS* (Strong's number 444; 553 tokens in New Testament).

ENGLISH (EARLY MODERN)

MAN: *man* 336 [tokens] *man's* 9

MEN: *men* 186 *men's* 3

HUNGARIAN

EMBER: *ember* 170 *embernek* 104 *emberek* 85 *embereknek* 36 *embert* 40 *emberektől* 16 *embereket* 15 *emberi* 10 *emberekkel* 5 *embereknél* 5 *emberből* 6 *emberben* 5 *embere* 4 *embernél* 3 *emberekhez* 3 ... (21 more types)

ZULU

NTU: *umuntu* 114 *yomuntu* 89 *muntu* 64 *abantu* 62 *kubantu* 38 *kwa-bantu* 20 *babantu* 11 *yabantu* 9 *ngabantu* 12 *kumuntu* 11 *nomuntu* 9 *bantu* 7 *okwabantu* 4 *kunabantu* 5 *zabantu* 8 *ngokwabantu* 3 *ngu-muntu* 5 ... (12 more types)

MAORI

TANGATA: *tangata* 532

[Wälchli, Bernhard (2012). Indirect measurement in morphological typology. In: *Methods in Contemporary Linguistics*, Berlin: De Gruyter Mouton, 69-92 p.]

DISTRIBUTION-BASED ACCOUNTS

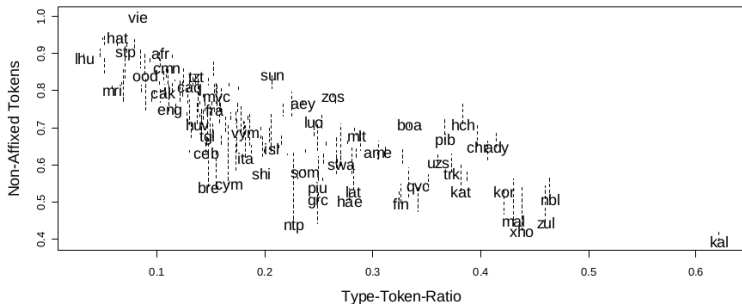


Figure 3. Synthesis measured as type-token ratio (x-axis) and as proportion of non-affixed tokens with Greek and Vietnamese as trigger (y-axis)

[Wälchli, Bernhard (2012). Indirect measurement in morphological typology. In: *Methods in Contemporary Linguistics*, Berlin: De Gruyter Mouton, 69-92 p.]

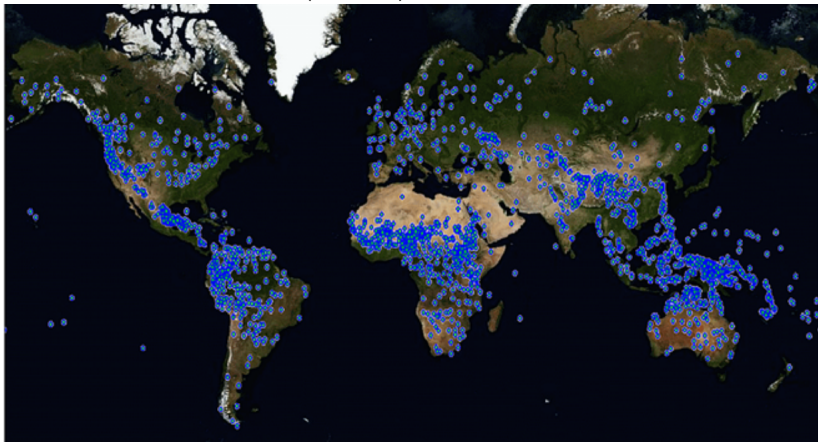


QUALITATIVE STUDIES: WRAY AND GRACE (2007)

“Languages that are used predominantly for **esoteric (intra-group)** communication tend to have features that are semantically and grammatically ‘complex’, while those used also (or even exclusively) for **exoteric (inter-group)** communication become ‘simplified’ towards rule-based regularity and semantic transparency.”

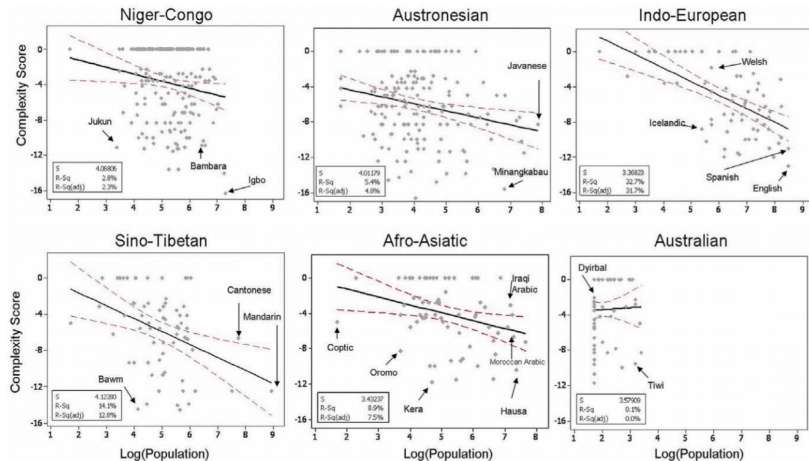
[Wray and Grace (2007) The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117, 543-578.]

QUANTITATIVE STUDIES (POPULATION SIZE): LUPYAN AND DALE (2010)



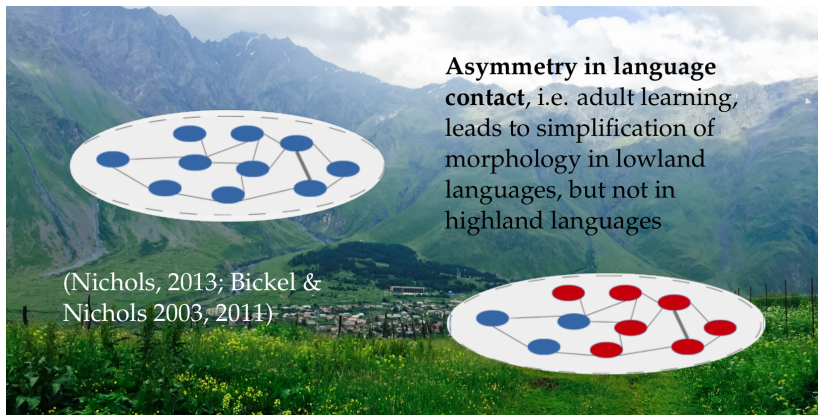
[Lupyan and Dale (2010) Language Structure is partly determined by social structure. *PLoS One*.]

QUANTITATIVE STUDIES



[Lupyan and Dale (2010) Language Structure is partly determined by social structure. *PLoS One*.]

QUANTITATIVE STUDIES (ALTITUDE): NICHOLS (2013, 2016)



Asymmetry in language contact, i.e. adult learning, leads to simplification of morphology in lowland languages, but not in highland languages

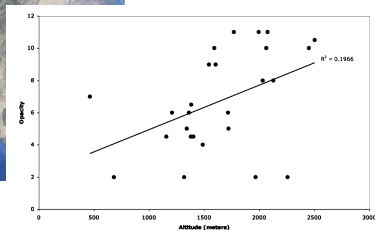
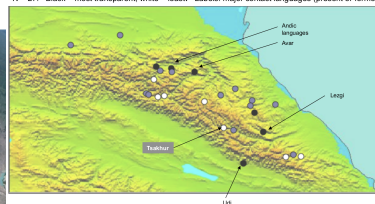
(Nichols, 2013; Bickel & Nichols 2003, 2014)

QUANTITATIVE STUDIES



Transparency (of gender classification and agreement) and altitude

$N = 27$. Black = most transparent, white = least. Labels: major contact languages (present or former)



[Nichols (2016) Complex edges, transparent frontiers: grammatical complexity and language spreads. In: Baechler, R. and Seiler, G. (eds.) *Complexity, Isolation, and Variation*. Berlin/Boston, De Gruyter.]

THANKS. SEE YOU NEXT WEEK!

