

Vielfaltslinguistik Graz, 14.-15. Juni 2024

Reihenfolge statt Komplexifizierung als Erklärung für morphologische Komplexität

MARTIN HASPELMATH

Max-Planck-Institut für evolutionäre Anthropologie

1. Has morphological complexity research gone astray?

- Morphological complexity has been a fashionable topic over the last 20 years
- A widespread idea has been that **complexity is sociohistorically conditioned**: second-language acquisition is a bottleneck for morphological complexity (Trudgill 2011), so creoles and semi-creoles tend to have little morphology (McWhorter 2001)
- This is not implausible, but what exactly is morphological complexity? “**Complexity**” is a vague notion that can be applied to morphology in a wide variety of ways (Arkadiev & Gardani 2020) (§2)
- Most stereotypically, a morphologically complex language is one whose words consist of many morphs – but **what exactly is a word?** (Haspelmath 2011; 2023a) (§3)
- In this talk, I point out that **wordhood uncertainty** leads to **complexity uncertainty**, and more specifically, that **affix status** is closely related to **element order**
- Affixes are very similar to **clitics**, which are words – they differ only in selectivity, and selectivity is closely related to word order (§5); compare:

(1) Old Spanish

Sy el físico la bien connoçe
 if the physician 3.F.ACC well know.3SG
 ‘if the physician knows it well’

(clitic, non-selective)

(2) Modern Spanish

Si el médico la-conoce bien
 if the physician 3.F.ACC-know.3SG well
 ‘if the physician knows it well’

(affix, selective)

- Thus, we need to understand **element order regularities** better if we want to understand “morphological complexity” (§6)
- This conclusion holds regardless of what one thinks about the nature of clitics – even if one is unhappy with my definition of *clitic* (Haspelmath 2023b), one needs to address the issue of wordhood uncertainty
- Thus, **we need to revisit our ideas of “complexity”**, and it may be that the Trudgill/McWhorter view relies far too much on a stereotypical and Eurocentric view of wordhood

2. Why is morphology “complex” and “difficult”?

The causal mechanism of **simplification through adult L2 acquisition** seems plausible at first – adults acquiring a language find “morphology” difficult.

This idea has a venerable history, e.g.

“Bei Völkern ohne Geschichte gewahren wir ... nicht selten ein wahres Wuchern der sprachlichen Form, einen Rand und Band überschreitenden Sprachtrieb, der Bildungen hervorruft, die durch übermäßige Fülle den Gedankenaustausch mit fremden Völkern erschweren und so als Hemniß der Cultur erscheinen. Dieß gilt vor allem von den meisten Indianersprachen Amerikas.” (Schleicher 1860: 36)

“Among peoples without history we often observe a true proliferation of linguistic form, an unconstrained linguistic drive that creates constructions which, through their overabundance, make the exchange of ideas with foreign peoples difficult and so seem as an impediment to culture. This is true especially of the majority of the Indian languages of America.”

But in what way is morphology “complex”, other than by *making words long*?

Typischerweise haben polysynthetische Sprachen eine große Anzahl an gebundenen Morphemen. Insbesondere haben sie eine sehr hohe Anzahl von Morphemen pro Wort (hoher **Synthesegrad**), z. B. **Yupik**:

tuntussuqatarniksaitengqiggtuq
tuntu -ssur -qatar -ni -ksaite -ngqiggte -uq
 Rentier -jagen -FUT -sagen -NEG -nochmal -3SG.IND
 „Er hat nicht nochmal gesagt, dass er Rentiere jagen geht.“

Wikipedia, *Polysynthese*

Two examples of “polysynthetic” words (from Fortescue et al. 2017):

- (3) Purepecha (Mexico; Chamoreau 2017: 687)

thiri-ra-a-x-ka
 eat-CAUS-3PL.OBJ-AOR-ASS.1
 ‘I gave to eat to them.’

- (4) Adyghe (Caucasus; Lander & Testelelets 2017: 953)

w-jə-mə-ke-λeβ^wəž’-ew
 2SG.DAT-3SG.ERG-NEG-CAUS-see-while
 ‘while it does not let you see [it]’

These “complex words” consist of individual morphs that occur in a specific order:

- (5) *thiri ra a x ka*
 eat make them AOR I.ASS (‘I gave to eat to them.’)
- (6) *w jə mə ke λeβ^wəž’ ew*
 to.you he.ERG not let see while (‘while it does not let you see [it]’)

If these morphs are not written as complex words, **they do not look “complex” anymore** – in fact, these “polysynthetic” languages would look like **isolating** languages (cf. Haspelmath 2018).

Morphology may be “complex” and/or “difficult” in other ways (e.g. by deviating from one-to-one correspondences, or by being “autonomous”, Arkadiev & Gardani 2020) – but here the focus will be on **morphological synthesis**.

3. Complexity and the synthetic index

– Greenberg (1960) was the first author to actually measure morphological synthesis, producing the following table:

	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Synthesis.....	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
Agglutination.....	.09	.11	.34	.30	.51	.6703
Compounding.....	1.13	1.00	1.03	1.00	1.02	1.00	1.07	1.00
Derivation.....	.62	.20	.10	.15	.35	.07	.00	1.25
Gross inflection.....	.84	.90	.39	.53	.82	.80	.00	1.75
Prefixing.....	.16	.06	.01	.04	.00	1.16	.00	.00
Suffixing.....	1.18	1.03	.49	.64	1.15	.41	.00	2.72
Isolation.....	.16	.15	.52	.75	.29	.40	1.00	.02
Pure inflection.....	.46	.47	.29	.14	.59	.19	.00	.46
Concord.....	.38	.38	.19	.11	.12	.41	.00	.38

Lupyan & Dale (2010): “Linguistic Niche Hypothesis”

complexity is largely equated with **morphological** as opposed to lexical marking, e.g.

languages with fewer (L2) speakers tend to be (languages)

- with grammatical markers showing a greater **degree of fusion** to the stem
- with more **case** markings
- with more grammatical categories **marked on the verb**
- more likely to possess noun/verb **agreement**
- more likely to make possibility and evidentiality distinctions using **inflections such as affixes**
- more likely to encode negation using using **inflections (affixes)**
- more likely to encode the future tense **morphologically**
- more likely to express pronominal subjects **morphologically** rather than lexically
- ...

But what is „morphological“ marking other than marking by means of **forms written together** with a verbal or nominal root?

4. Wordhood uncertainty

- There is no clarity around the notion of “word” (Haspelmath 2011) – we do not have a good basis for the distinction between morphology and syntax
- Many linguists seem to think that
 - the solution is to work with multiple word notions
 - it does not matter because all concepts are fuzzy anyway
 - or both

But phonological wordhood is not independent of morphosyntactic wordhood (Newell 2017), and no general proposals for wordhood distinctions have been successful. Lexicalism is dead (Bruening 2018).

- One aspect of wordhood uncertainty is cliticness uncertainty – the “clitic” notion is not any better defined than the “word” notion.

A new proposal for defining words: Haspelmath (2023a)

word

A word is (i) a free morph, or (ii) a clitic, or (iii) a root or a compound possibly augmented by nonrequired affixes and augmented by required affixes if there are any.

A new proposal for defining clitics: Haspelmath (2023b)

clitic

A clitic is a bound morph that is neither an affix nor a root.

These definitions are not fuzzy, but the definition of ‘word’ is decidedly **unnatural**.

Words are not defined by a single criterion, but as the set of **four different kinds** of forms that do not clearly have much in common.

(7) examples of four kinds of words:

- | | |
|----------------------------|---|
| a. free morph | <i>nice, work, now, ouch</i> |
| b. clitic | <i>the, to, 's</i> |
| c. root (plus affixes) | <i>tree, nice-r, go-ing, re-work, re-place-ment-s</i> |
| d. compound (plus affixes) | <i>flower-pot, wind-shield, dog-sit, flower-pot-s</i> |

Clitic seems to be defined more naturally, as a bound morph that is neither a root nor an affix, but ‘affix’ is not defined in a simple way:

A new proposal for defining affixes: Haspelmath (2021)

affix

An affix is a bound morph that is not a root, that occurs on a root, and that cannot occur on roots of different root classes.

With these new definitions, we may have achieved “wordhood clarity”, but much uncertainty remains:

- How significant are the notions ‘affix’, ‘clitic’, ‘word’?
- Do they have any role for learnability by adults?

(I suspect: **Grammatical markers** in general are not very salient, so that they tend to **get lost in creolization** – but this applies to clitics and affixes in the same way; see Haspelmath & Michaelis 2017. Wordhood seems to be irrelevant.)

5. Affixes differ from clitics in selectivity

Since Zwicky & Pullum (1983), it has been widely accepted that affixes exhibit host-class **selectivity** (“non-promiscuity”):

They occur **only** on verbs, or **only** on nouns, or **only** on adjectives.

By contrast, clitics are non-selective or “promiscuous”, e.g.

(8) English

- | | |
|--|----------------------|
| a. <i>my friend's house</i> | (adjacent to noun) |
| b. <i>the lady I met yesterday's offer</i> | (adjacent to adverb) |
| c. <i>the boy I like's new bike</i> | (adjacent to verb) |

(9) Russian

- | | |
|--|-------------------------|
| a. <i>Pročita-la li Anna knigu?</i>
read-PST PQ Anna book
'Did Anna read a book?' | (= 3; adjacent to verb) |
| b. <i>Knigu li Anna pročita-la?</i>
book PQ Anna read-PST
'Did Anna read a BOOK?' | (adjacent to noun) |
| c. <i>Včera li Anna čita-la?</i>
yesterday PQ Anna read-PST
'Did Anna read YESTERDAY?' | (adjacent to adverb) |

Many types of elements can occur either as affixes or as clitics, e.g.

articles	English <i>the + cup</i> <i>the + new + cup</i>	Swedish <i>kopp + en</i> <i>den nya kopp + en</i>
flags	English <i>the girl + s</i> <i>the girl I like + s</i>	German <i>des Mädchen + s</i> <i>des Mädchen + s, das ich mag</i>
plural markers	German <i>Kind + er</i> <i>große Kind + er</i>	Haitian Creole <i>liv + yo</i> ‘books’ <i>machchet plat + yo</i> ‘straight machetes’

tense markers Russian
opis + yva + l
 describe + IMPF + PST
 ‘was describing’

Jamaican (Creole) (Farquharson 2013)
Di pikni wehn + de + kil di flowaz.
 the child PST IMPF kill the flowers
 The child was killing the flowers.

Dis-ya pikni when + aalwiesz + ogli.
 this-here child PST always ugly
 ‘This child was always ugly.’

But aren’t clitics defined as “phonologically dependent” elements?

NO:

- Both clitics and affixes are **bound forms**, i.e. they cannot occur on their own (Bloomfield 1933; Haspelmath 2021)
- Both clitics and affixes tend to be short and may show **segmental interaction** with their host (“welding”), e.g.

English	<i>the cup</i>	Swedish	<i>kopp-en</i>	‘the cup’
	<i>th[i] other cup</i>		<i>flicka-n</i>	‘the girl’

- Both clitics and affixes tend to be unstressed, but may sometimes be stressed

The notion of “phonological dependence” has never been made precise in such a way that it can be used to distinguish uniformly between clitics and affixes (Haspelmath 2023b: §7).

6. Different orders lead to differences in selectivity and thus affixhood

If we adopt root selectivity as the key criterion, then it is clear that the affix status of a grammatical marker is co-determined by **element order** in a wide range of contexts.

Again compare:

(1) Old Spanish

Sy el fisico la bien conosco
 if the physician 3.F.ACC well know.3SG
 ‘if the physician knows it well’

(clitic, non-selective)

(2) Modern Spanish

Si el médico la-conoce bien
 if the physician 3.F.ACC-know.3SG well
 ‘if the physician knows it well’

(affix, selective)

In Modern Spanish, **the order of the adverb changed**, from preverbal to postverbal, and as a result, the object clitic has become an object prefix.

More contrasting examples:

Preceding flags are **prefixes** when the adjective follows, but **prepositions** when the adjective precedes the noun:

- | | |
|--|--|
| (A) <i>flag</i> + NOUN (+ adjective)
<u>Hebrew</u> <i>be+vatim gdolim</i> | <i>flag</i> (+ adjective) + NOUN
<u>English</u> <i>in+ big houses</i> |
|--|--|

Following flags are **suffixes** when the adjective precedes, but **postpositions** when the adjective follows the noun:

- | | |
|--|--|
| (B) (adjective +) NOUN + <i>flag</i>
<u>Turkish</u> <i>büyük evler+de</i> ‘in big houses’ | NOUN (+ adjective) + <i>flag</i>
<u>Basque</u> <i>etxe handi + etan</i> |
|--|--|

Following plural markers are **suffixes** when the adjective precedes, but **plural words** when the adjective follows the noun:

- | | |
|--|--|
| (C) (adjective +) NOUN + <i>plural</i>
<u>Hungarian</u> <i>nagy fá+k</i>
big tree+PL | NOUN (+ adjective) + <i>plural</i>
<u>Kimaghama</u> <i>do mamu + ragha</i>
tree big PL (Dryer 1989: 883) |
|--|--|

Preceding aspect markers are **prefixes** when the adverb precedes, but **auxiliaries** when the adverb follows the noun:

- | | |
|--|--|
| (D) <i>aspect</i> + VERB (+ adverb)
<u>Papiá Kristang</u> <i>eli ta+les buku agora</i>
she ASP+read book now
(Baxter 1988: 128-129) | <i>aspect</i> (+ adverb) + VERB
<u>English</u> <i>she is+ now reading a book.</i> |
|--|--|

Following relativizers are **suffixes** (forming **participles**) when the object precedes the verb, but **particles** when the object follows it:

- | | |
|---|--|
| (E) (object +) VERB + <i>relativizer</i>
<u>Turkish</u> <i>pencere-yi aç+a+n kadın</i>
window-ACC open+REL woman
‘the woman who opened the window’ | VERB (+ object) + <i>relativizer</i>
<u>Mandarin</u> <i>dākāi chuānghù +de nǚrén</i>
open window rel woman |
|---|--|

Following complementizers are **suffixes** (forming **nominalizations**) when the object precedes the verb, but preceding complementizers are **particles** when the subject precedes the verb:

- | | |
|---|--|
| (F) subject + verb + <i>complementizer</i>
<u>Basque</u> <i>etxe-ra doa+la</i>
‘that she is going home’ | <i>complementizer</i> + subject + verb
<u>Indonesian</u> <i>bahwa+ dia akan pulang</i>
that she PROG go.home |
|---|--|

Thus, **affixhood is dependent on element order.**

It is true that two kinds of *morphological marking* are independent of element order:

(i) root change	German	<i>Garten</i>	<i>Gärten</i>	‘garden(s)’
(ii) reduplication	Indonesian	<i>anak</i>	<i>anak-anak</i>	‘child(ren)’

But root changes are not common (typically restricted to a smallish number of roots), and neither root changes nor reduplication have played a role in the discussion of morphological complexity.

(On the contrary, one may get the impression that creole languages have more reduplication than non-creoles, e.g. Kouwenberg (ed.) 2003.)

7. Conclusion

Sociolinguistic typologists have made some very interesting claims which now doubt have some prima facie plausibility (e.g. Kusters 2003; McWhorter 2001; 2007; 2016; Lupyan & Dale 2010; Trudgill 2011).

The basic idea is actually quite old, going back at least to 1818:

“Mais cette transition au système analytique a lieu bien plus rapidement, et, pour ainsi dire, par secousses, lorsque, par l’effet de la conquête, il existe un conflit entre deux langues, celle des conquérans et celle des anciens habitans du pays. Voilà ce qui a eu lieu dans les provinces de l’empire occidental, conquises par les peuples germaniques, et en Angleterre lors de l’invasion des Normands. De la lutte prolongée de deux langues, dont l’une étoit celle de la grande masse de la population, l’autre celle de la nation prépondérante, et de l’amalgame final des langues et des peuples, sont issus le provençal, l’italien, l’espagnol, le portugais, le françois et l’anglais.”

“But this transition to the analytic system took place more rapidly, and, so to speak, by jolts, when, due to conquest, a conflict between two languages arises, the language of the conquerors and the language of the earlier inhabitants of the country. This took place in the provinces of the Western Roman empire which were conquered by Germanic peoples, and in England after the Norman invasion. The extended struggle between two languages, one of which was the language of the great majority of the people, and other other the language of the ruling group, and the eventual merging of the languages and the peoples gave rise to Provençal, Italian, Spanish, Portuguese, French and English.”

However, these ideas were to a large extent based on **synthesis** (= number of morphs per word) and thus on a ‘word’ notion. Because of **wordhood uncertainty**, they need to be approached in a more rigorous fashion in the future.

A recent quantitative study was unable to confirm the claims by McWhorter and Lupyan & Dale (see Shcherbakova et al. 2023), but this may have several reasons, e.g.

- wordhood unclarity in Grambank features (Skirgård et al. 2023)
- insufficient sociolinguistic information
- sociohistorical effects not reflected in synchronic grammars

The correlations are certainly not as clear as it may appear when one cherry-picks individual cases (such as Latin and French), but it still appears plausible that L2 learners will simplify some patterns of languages.

But future research must be more rigorous, and must take into account the dependence of the affix–clitic distinction on element order.

References

- Arkadiev, Peter & Gardani, Francesco (eds.). 2020. *The complexities of morphology*. Oxford: Oxford University Press.
- Baxter, Alan N. 1988. *A grammar of Kristang (Malacca Creole Portuguese)*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, Australian National University. (<http://dx.doi.org/10.15144/PL-B95>)
- Bruening, Benjamin. 2018. The lexicalist hypothesis: Both wrong and superfluous. *Language* 94(1). 1–42. (doi:10.1353/lan.2018.0000)
- Dryer, Matthew. 1989. Plural words. *Linguistics* 27. 865–896. (doi:10.1515/ling.1989.27.5.865)
- Farquharson, Joseph T. 2013. Jamaican structure dataset. In Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (<https://apics-online.info/contributions/8>)
- Fortescue, Michael D & Mithun, Marianne & Evans, Nicholas (eds.). 2017. *The Oxford handbook of polysynthesis*. Oxford: Oxford University Press.
- Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. (doi:10.1515/flin-2017-1005)
- Haspelmath, Martin. 2018. The last word on polysynthesis: A review article. *Linguistic Typology* 22(2). 307–326. (doi:10.1515/lingty-2018-0011)
- Haspelmath, Martin. 2021. Bound forms, welded forms, and affixes: Basic concepts for morphological comparison. *Voprosy Jazykoznanija* 2021(1). 7–28. (doi:10.31857/0373-658X.2021.1.7-28)
- Haspelmath, Martin. 2023a. Defining the word. *WORD* 69(3). 283–297. (doi:10.1080/00437956.2023.2237272)
- Haspelmath, Martin. 2023b. Types of clitics in the world’s languages. *Linguistic Typology at the Crossroads* 3(2). 1–59. (doi:10.6092/issn.2785-0943/16057)
- Haspelmath, Martin & Michaelis, Susanne Maria. 2017. Analytic and synthetic: Typological change in varieties of European languages. In Buchstaller, Isabelle & Siebenhaar, Beat (eds.), *Language variation – European perspectives VI: Selected papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015*. Amsterdam: Benjamins.
- Kouwenberg, Silvia (ed.). 2003. *Twice as meaningful: Reduplication in pidgins, creoles and other contact languages*. London: Battlebridge.
- Kusters, Wouter. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection* (LOT Dissertation Series). Vol. 77. University of Leiden: LOT.
- Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1). e8559. (doi:10.1371/journal.pone.0008559)
- McElvenny, James. 2021. Language complexity in historical perspective: The enduring tropes of natural growth and abnormal contact. *Frontiers in Communication* 6. (doi:10.3389/fcomm.2021.621712)
- McWhorter, John H. 2001. The world’s simplest grammars are creole grammars. *Linguistic Typology* 5(2–3). 125–166.
- McWhorter, John H. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.
- McWhorter, John H. 2016. Is radical analyticity normal? Implications of Niger-Congo and Southeast Asia for typology and dichronic theory. In van Gelderen, Elly (ed.), *Cyclical change continued*, 49–92. Amsterdam: Benjamins.

- Newell, Heather & Noonan, Máire & Piggott, Glyne & Travis, Lisa deMena (eds.). 2017. *The structure of words at the interfaces*. Oxford: Oxford University Press.
- Schlegel, August Wilhelm. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande.
- Schleicher, August. 1860. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: Böhlau.
- Shcherbakova, Olena & Michaelis, Susanne Maria & Haynie, Hannah J. & Passmore, Sam & Gast, Volker & Gray, Russell D. & Greenhill, Simon J. & Blasi, Damián E. & Skirgård, Hedvig. 2023. Societies of strangers do not speak less complex languages. *Science Advances*. American Association for the Advancement of Science 9(33). eadf7704. (doi:10.1126/sciadv.adf7704)
- Skirgård, Hedvig & Haynie, Hannah J. & Blasi, Damián E. & Hammarström, Harald & Collins, Jeremy & Latache, Jay J. & Lesage, Jakob et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9(16). eadg6175. (doi:10.1126/sciadv.adg6175)
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Zwicky, Arnold M. & Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English n't. *Language* 59(3). 502–513.