

University of Potsdam, 2022 June 28

Is cliticization an intermediate stage between free lexeme and affix status?

MARTIN HASPELMATH

Max Planck Institute for Evolutionary Anthropology (Leipzig)

Overview:

- the idea of a **continuous scale** (synchronic and diachronic) from lexemes to affixes via clitics is a widespread stereotype, but it has not been tested systematically
- we need **clear (and simple) definitions** of the terms *lexeme*, *clitic*, and *affix*
- clitics differ from affixes primarily in that they are **promiscuous** (= do not exhibit word-class selectivity), not in that they are “**less phonologically attached**”
- clitics are usually **unstressed**, and thus even less lexeme-like than affixes
- clitics may be **welded** (= show segmental interaction), like affixes
- most types of **clitics do not become affixes**; and most types of **affixes did not pass through an intermediate clitic stage**
- clitics may be an intermediate stage on the path toward affixes, but this development is **unusual**

1. The stereotype: lexeme > clitic > affix

Typical textbook accounts of grammaticalization claim that cliticization is an intermediate stage between free lexeme and affix status (e.g. Hopper & Traugott 2003: 7; Wiemer 2014: 442; Narrog & Heine 2021: 279).

(9) content item > grammatical word > clitic > inflectional affix (> Ø)

Spencer & Luís (2012: §2.1): “Inflections are generally thought to arise in historical language change through a stage of cliticization”

For example, free personal pronouns are said to become clitic pronouns first before they turn into person-number affixes on verbs (“Givón’s Cycle”, Seržant 2021).

schematically:

Jane, she comes.
 > *Jane (,) she=comes.*
 > *Jane she-comes.*

The processes of cliticization and affixation are sometimes collapsed as “coalescence” (e.g. Haspelmath 2011b), suggesting that they constitute a single **gradual** process.

2. A continuous scale?

Is there a continuous scale (or “cline”) from lexicality to affixhood?

(or even discourse > syntax > morphology > zero; Givón 1979)

This may make sense at an intuitive level, but in order to test it scientifically, we need ways of **measuring an item’s position** on the scale. Rigorous quantitative testing requires clear qualitative comparative concepts.

So far, no attempt has been made at quantifying grammaticalization changes, as far as I know.

Few attempts at defining “affix” and “clitic” have been made (Zwicky & Pullum 1983 provide a set of “diagnostic symptoms”, which are not really criteria; cf. Haspelmath 2015).

Many authors are pessimistic about defining *clitic*, e.g.

“Because the phenomenon itself appears to lack uniformity and cohesion, it may be difficult to provide clear, unambiguous, and theory-neutral definitions... A formative is a clitic to the extent it deviates from the accepted properties of affixes or words.”
(Nevis 2000: 389)

The problems with defining *word* are well-known, too (Haspelmath 2011a; Tallman 2020). Clearly, a wide range of different phenomena are involved in “tightness of bonding” or “coalescence”, and multiple dimensions need to be taken into account (Börjars et al. 2008).

Very recently, attempts at quantifying basic linguistic forms have been made (Tallman & Auderset 2022). I will not address these here, but instead focus on improving the conceptual clarity of the key notions.

3. Defining affixes and clitics

For the great majority of cases, the following definitions give results that correspond to linguists’ intuitions about the meanings of these terms:

(1) **affix** (Haspelmath 2021)

An affix is a bound morph that is not a root and that always occurs on roots of the same class (nouns, verbs, adjectives; Haspelmath 2022)

(2) **clitic** (Haspelmath 2023)

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

A free form differs from a bound form (a clitic or an affix) in that it can **occur in isolation** (e.g. Zwicky & Pullum 1983: 502).

- It is clear that when a free form (e.g. a noun or a verb or a demonstrative) grammaticalizes, it finally becomes a **bound form** – “grammatical status” (Boye & Harder 2012) is much the same as “bound status”.

e.g.	Old Church Slavonic	<i>vlast ta</i>	‘that power’	
		<i>ta</i>	‘that one’	
	> Bulgarian	<i>vlast-ta</i>	‘the power’	
		<i>*ta</i>	‘that one’	(<i>tazi</i>)

- It is also clear that grammaticalization involves shortening. Grammaticalized forms are **bound** and **short**.

- But does a free form generally become a *clitic* before it becomes an affix?
(Apparently not.)

4. Clitics differ from affixes in being promiscuous

4.1. Focusing clitics

(3) Russian

a. *Pročita-la li Anna knigu?*
read-PST PQ Anna book
‘Did Anna read a book?’

b. *Knigu li Anna pročita-la?*
book PQ Anna read-PST
‘Did Anna read a BOOK?’

c. *Včera li Anna čita-la?*
yesterday PQ Anna read-PST
‘Did Anna read YESTERDAY?’

(4) Lak (Kazenin 2002: 293)

a. *Na q̄atri d-ullali-ṣa =ra.*
I house(G4) G4-build.DUR-PTCP=1SG
‘I am building a house.’

b. *Na =ra q̄atri d-ullali-ṣa.*
I =1SG house(G4) G4-build.DUR-PTCP
‘The one who is building a house is me.’

c. *Na q̄atri =ra d-ullali-ṣa.*
I house(G4) =1SG G4-build.DUR-PTCP
‘What I am building is a house.’

4.2. 2nd-position clitics

(5) Tagalog (Kaufman 2010: 10)

a. *Na-túto =siya nang= wika =ng Intsik.*
 AV-learn =3SG.NOM GEN= language =LNK Chinese
 ‘She learned Chinese.’

b. *Hindí =siya na-túto nang= wika =ng Intsik.*
 NEG =3SG.NOM AV-learn GEN= language =LNK Chinese
 ‘She didn’t learn Chinese.’

(6) Serbo-Croatian (Bošković 2001: 8)

a. *Zašto smo mu je predstavili juče?*
 why 1PL.SBJ him.DAT her.ACC introduced yesterday
 ‘Why did we introduce her to him yesterday?’

b. *Predstavili smo mu je juče.*
 why 1PL.SBJ him.DAT her.ACC yesterday
 ‘We introduced her to him yesterday.’

4.3. Epiphrasal clitics (= clitics occurring at the periphery of a phrase)

(7) English (postphrasal)

- a. *my friend’s house*
- b. *the lady I met yesterday’s offer*
- c. *the boy I like’s new bike*

(8) German (prephrasal)

- a. *nur [mein Hund]*
- b. *nur [in ihrem Haus]*
- c. *sie will nur [spielen]*

4.4. Other kinds of clitics

(9) Colloquial German

a. *Willst de se haben?*
 want.2SG you them have
 ‘Would you like to have them?’ (Standard: *Willst du sie haben?*)

b. *Wenn de se haben willst...*
 if you them have want.2SG
 ‘If you want to have them...’ (Standard: *Wenn du sie haben willst...*)

4.5. “Phonological attachment”?

It is sometimes thought that clitics differ from affixes in that they are “phonologically attached” to a lesser degree. But is this the case? What does “attachment” mean?

6. Most types of clitics do not become affixes

- **discourse particles** rarely become affixes

- (16) a. *However, our ambitious proposal failed.*
 b. *Our ambitious proposal, however, failed.*
 c. *Our ambitious proposal failed, however.*

- **2nd position clitics** rarely become affixes

- (17) Ancient Greek (Lowe 2016: 377)
Ho dè Kúros ekselaúnei epì tòn potamón.
 DEF but Cyrus march.army to DEF river
 ‘But Cyrus marches his army to the river.’

- **coordinator clitics** rarely become affixes

- (18) Latin
 a. *ménsa* ‘the table’
 b. *mensá=que* ‘and the table’

- **question particles** rarely become affixes

- (35) Turkish
geldi mi? ‘did she come?’
öldü mü? ‘did he die?’
Ali mi? ‘Ali?’
diin mü? ‘yesterday?’

However:

It is true that person indexes do seem to become affixes in **some** cases, e.g. in the Romance languages. Old Portuguese and Old Spanish seem to attest a **clitic stage** of the weak person forms (Pescarini 2021).

- (36) Old Portuguese
*logo **lhe** el-rrei taxava que ...*
 soon 3.DAT the king ordain.IPFV that
 ‘immediately the king imposed on him as punishment that...’

- (37) Old Spanish
*Sy el físico **la** bien connoçe*
 if the physician 3.F.ACC well know.3SG
 ‘if the physician knows it well’

Modern Romance languages do not allow this sort of “interpolation”: Object clitics must **immediately precede** the verb, or **immediately follow** the verb (i.e. they have affixal status).

In Polish, the development from 2nd position clitics to verbal person markers seems to be a very typical case of gradual grammaticalization.

Andersen (1987): *Nigdy=śmy tego nie myśleli.* >
Nigdy tego nie myśleli-śmy. ‘We never thought that.’

(5)	Total no. of examples	Deviations from Wackernagel's rule		Agglutination to preterite form	
1500's	580	12	2%	130	23%
1600's	1303	64	4%	649	49%
1700's	1439	62	4%	994	68%
1800's	1988	308	15%	1395	80%
1900's	3325	503	15%	2817	84%
exp. prose	569	usual		525	92%

7. Most types of affixes did not pass through a clitic stage

The most typical affixes:

verbs: tense-aspect and person affixes
nouns: case, number and gender affixes

Most of these do not derive from elements that are usually called clitics.

The grammaticalization origin of **number** and **gender** affixes is often unclear, though **articles** seem to play an important role. Articles may be clitics, but they may also be affixes (= occur always on the noun).

There do not seem to be clear cases of affixal articles deriving from clitic articles (the Nordic suffixed article does not derive from a clitic; Börjars et al. 2008)

Tense-aspect affixes from **auxiliary verbs** (e.g. Bybee & Dahl 1989), but auxiliaries are very often non-promiscuous

e.g. Spanish *mis amigos han llamado*
 my friends have called

Case affixes are generally thought to derive from **auxiliary (relational) nouns** (e.g. König 2011), but such “auxiliary nouns” are very often non-promiscuous

e.g. Lezgian *t'ara-n k'an-ik* (Haspelmath 1993: 220)
 tree-GEN bottom-LOC
 ‘under the tree’

But it is true that **person affixes** may sometimes go back to promiscuous person forms (see above for Romance and Polish).

8. Conclusion

- Many linguists have the intuition that there are degrees of bondedness – more “**tightly bonded**” grammatical forms and more “**loosely bonded**” forms
- Often linguists seem to assume that the tightness of bonding (or “coalescence”) is of a phonological nature (they often talk about “**phonological boundness**”) – but there is no clear conceptual basis for this
- Comparative grammar research needs clear comparative concepts, and *clitic* can be defined clearly only with respect to their **promiscuous** behaviour (contrasting with class-selective affixes)
- It is of course plausible that grammaticalization involves not only **shortening** (= fewer segments) and **boundness** (= non-occurrence in isolation), but also some kind of “phonologically close association”, but this has not been formulated clearly yet.
- At present, the well-known scale “**lexeme > clitic > affix**” is not more than a stereotype, and we need to replace it by a more rigorous formulation.

References

- Andersen, Henning. 1987. From auxiliary to desinence. In Harris, Martin & Ramat, Paolo (eds.), *Historical development of auxiliaries*, 21–52. Berlin: Mouton de Gruyter.
- Boye, Kasper & Harder, Peter. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1). 1–44.
- Bybee, Joan L. & Dahl, Östen. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1). 51–103.
- Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.
- Haspelmath, Martin. 2011a. 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. 2011b. The gradual coalescence into “words” in grammaticalization. In Narrog, Heiko & Heine, Bernd (eds.), *The Oxford handbook of grammaticalization*, 342–355. Oxford: Oxford University Press.
- Haspelmath, Martin. 2015. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In Błaszczak, Joanna & Klimek-Jankowska, Dorota & Migdalski, Krzysztof (eds.), *How categorical are categories? New approaches to the old questions of noun, verb, and adjective*, 273–304. Berlin: De Gruyter Mouton.
- 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. 2022. Word class universals and language-particular analysis. In van Lier, Eva (ed.), *Oxford handbook of word classes*. Oxford: Oxford University Press (to appear).
- Haspelmath, Martin. 2023. Types of clitics in the world’s languages. (*in preparation*).
- Hopper, Paul J. & Traugott, Elizabeth Closs. 2003. *Grammaticalization: 2nd edition*. Cambridge: Cambridge University Press.
- Kaufman, Daniel. 2010. *The morphosyntax of Tagalog clitics: a typologically driven approach*. Ithaca: Cornell University.
- Kazenin, Konstantin I. 2002. Focus in Daghestanian and word order typology. *Linguistic Typology* 6(3). 289–316. (doi:10.1515/lity.2003.002)
- König, Christa. 2011. The grammaticalization of adpositions and case marking. In Narrog, Heiko & Heine, Bernd (eds.), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.

- Lowe, John J. 2016. Clitics: Separating syntax and prosody. *Journal of Linguistics*. Cambridge University Press 52(2). 375–419. (doi:10.1017/S002222671500002X)
- Monachesi, Paola. 1999. *A lexical approach to Italian cliticization*. Stanford: CSLI Publications.
- Narrog, Heiko & Heine, Bernd. 2021. *Grammaticalization*. Oxford: Oxford University Press.
- Pescarini, Diego. 2021. *Romance object clitics: Microvariation and linguistic change*. Oxford: Oxford University Press.
- Samvelian, Pollet & Tseng, Jesse. 2010. Persian Object Clitics and the Syntax-Morphology Interface. 212–232. (<https://web.stanford.edu/group/cslipublications/cslipublications/HPSG/2010/abstr-samvelian-tseng.shtml>)
- Seržant, Ilja A. 2021. Cyclic changes in verbal person-number indexes are unlikely. *Folia Linguistica* 55(s42–s1). 49–86. (doi:10.1515/flin-2021-2014)
- Spencer, Andrew & Luís, Ana R. 2012. *Clitics*. Cambridge: Cambridge University Press.
- Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14(2). e12364. (doi:10.1111/lnc3.12364)
- Tallman, Adam J. R. & Auderset, Sandra. 2022. Measuring and assessing indeterminacy and variation in the morphology-syntax distinction. *Linguistic Typology*. De Gruyter Mouton. (doi:10.1515/lingty-2021-0041)
- Wiemer, Björn. 2014. Quo vadis grammaticalization theory? Why complex language change is like words. *Folia Linguistica* 48(2). 425–468. (doi:10.1515/flin.2014.015)
- Zwicky, Arnold M. 1994. What is a clitic? In Nevis, Joel A. & Joseph, Brian D. & Wanner, Dieter & Zwicky, Arnold M. (eds.), *Clitics: a comprehensive bibliography 1892-1991*, xii–xx. Amsterdam: Benjamins.
- Zwicky, Arnold M. & Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English n't. *Language* 59(3). 502–513.

Appendix:

Positional types of clitics: Enclitics, proclitics, ambiclitics, interclitics

The two best-known types are enclitics and proclitics, defined as in (10)-(11).

(10) enclitic

An enclitic is a clitic that can occur at the end of a free form but not at the beginning.

(11) proclitic

A proclitic is a clitic that can occur at the beginning of a free form but not at the end.

For example, English Genitive 's can occur at the end of an elliptical answer, as in (12a), and the English coordinator *and* can occur at the beginning of the elliptical expression *and their dog*, as in (12b).

(12)a. A: *Is this your bike?* B: *No, (it's) my friend's.*

b. A: *Who is coming?* B: *My friend Lee.* C: *And her dog!*

The opposite situations are quite impossible (*my friend's bike* cannot under any circumstances be shortened to **'s bike*, and *Lee and her dog* cannot be shortened to **Lee and*), so the reduced free forms in (12a-b) are the basis for classifying 's as an enclitic and *and* as a proclitic.

Linguists often say that the element with which a clitic can occur in such contexts is its HOST, and they use an equals sign as a boundary symbol, as in (13).

- (13)a. *my friend=s bike*
 b. *Lee and=her=dog*

It is often said that “a clitic attaches to its host”, and the equals sign is generally taken as signaling this kind of “attachment”, but it is often typically what exactly this means. Most often, linguists say that clitics form a prosodic unit (such as a phonological word) with their hosts, and they generally attribute this to their phonological deficiency:

“Clitics which form a prosodic unit with a host on their left are enclitics, while those forming a unit to their right are proclitics.” (Halpern 1998: §1)

There are sometimes clear segmental effects, e.g. regressive voicing assimilation, as in English (14a) vs. (14b), or otherwise alternating forms, like Tagalog *ng* (after a vowel) vs. *na* (after a consonant), as in (15a) vs. (15b).

- (14)a. *my friend=[z] bike*
 b. *my bike=[s] wheels*

- (15)a. *wika ng Ingles* ‘English language’
 b. *Ingles na wika* ‘English language’

However, this is by no means always the case, and the effects may be quite variable:

- (16)a. *my bike=’[s] at home*
 b. *my bikes [ə] here*
 c. *my bikes [ər]=over there*

For this reason, the definitions in (10) and (11) do not make reference to prosodic units, but only to the **occurrence in free forms** (i.e. as part of forms that can occur in isolation).

Most clitics are enclitics or proclitics, but there are two other possibilities, ambiclitics and interclitics. Ambiclitics are clitics that can occur at the end or at the beginning in a free form, e.g. English *however*.

(17) **ambiclitic**

An ambiclitic is a clitic that can occur at the end of a free form or at the beginning.

It is not usual to qualify *however* as a clitic, but as it does not occur on its own and is neither a root nor an affix, it is a clitic on the definition of this paper. It can occur initially, medially, or finally in a free form.

- (18)a. ***However***, *our ambitious proposal failed.*
 b. *Our ambitious proposal, however, failed.*
 c. *Our ambitious proposal failed, however.*

Some languages have clitics that must occur between two overt forms. These are called *interclitics* here, and defined as in (19).

(19) **interclitic**

An interclitic is a clitic that can occur neither at the end of a free form nor at the beginning.

An example is the Taglog linker =*ng/na*=, which was already illustrated in (4a-b) above. Further examples are in (20a-b). This morph (with variant *ng* after a vowel, and *na* after a consonant) occurs between an attributive adjective and a noun (these two elements may occur in either order). According to the positional criterion, it is an interclitic, not an enclitic or proclitic, despite the phonological behaviour of =*ng*=.

(20)a. *malaki =ng= bahay*
 big =LNK house
 ‘big house’

b. *bahay =na= malaki*
 house =LNK big
 ‘big house’

Interclitics cannot be said to have a host, or maybe they have two hosts. (It seems better to restrict the term *host* to enclitics and proclitics, because with ambiclitics, too, it would not be clear whether any of the adjacent forms should count as hosts.)