

# Types of clitics in the world's languages

MARTIN HASPELMATH

MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

Submitted: 19/12/2022    Revised version: 02/08/2023

Accepted: 02/08/2023    Published: 27/12/2023



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

This paper offers and discusses a simple definition of the term *clitic* from a comparative perspective: A clitic is a bound morph that is neither an affix nor a root. It gives examples of several semantic and positional types of clitics from a wide range of languages, and it discusses some typical phonological effects associated with clitics. In the proposed definition, the crucial contrast between affixes and clitics is that affixes are class-selective (occurring always on nouns, on verbs, or on adjectives), while clitics do not exhibit word-class selectivity. In the stereotypical view of clitics, they are “prosodically deficient” in some way, but the phonological effects are quite diverse and cannot serve as a basis for a definition. As clitics are defined as kinds of minimal forms (or morphs), they cannot be nonsegmental, and they cannot interrupt another minimal form (so that there cannot be endoclitics by definition). Finally, I note that the object person indexes of the Romance languages, which have very often been called *clitics*, are actually affixes in the modern languages, although they must go back to earlier clitics.

**Keywords:** clitic; affix; prosodic deficiency; word-class selectivity; object indexing.

## 1. Overview

This paper gives an overview of clitics in human languages, with an emphasis on clear conceptual distinctions and straightforward terminology. In addition to exemplifying a range of clitics from a wide variety of languages, I will discuss some of the earlier conceptual and terminological distinctions, and I will say how the choices made here

relate to the earlier literature. This paper thus has a clear methodological focus and does not claim to make an empirical or explanatory contribution. I begin with the definition of *clitic* as a comparative concept in (1), which is simple and clear.

(1) **Clitic**

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

As a first illustration of clitics, consider the forms in boldface in (2)-(5), which are typical examples of clitics.

(2) English (Indo-European, Germanic)<sup>1</sup>

*my friend = 's house*

(3) Russian (Indo-European, Balto-Slavic)

*Pročita-la = **li** Anna knigu?*

read-PST = PQ Anna book

'Did Anna read a book?'

(4) Persian (Indo-European, Indo-Iranic; Samvelian & Tseng 2010: 215)

*Ru-ye miz = **aš** gozâšt-im.*

on-EZ table = 3SG.OBJ put.PST-1PL.SBJ

'We put it on the table.'

(5) Tagalog (Austronesian, Malayo-Polynesian; 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.'

---

<sup>1</sup> The genealogical classification of the languages mentioned in this paper has been retrieved on Glottolog.

Traditionally, clitics have often been defined as prosodically deficient elements, and/or as forms that are somehow intermediate between affixes and independent words. Below in Section 5, I will explain why the definition in (1) is preferable, even though it does not conform fully to some linguists' intuitions about the nature of clitics.

To understand the definition adopted here, we need to understand the concepts of *bound morph* and of *affix*. Briefly, a morph is a minimal form (Haspelmath 2020), a bound form is a form that cannot occur in isolation (Bloomfield 1933: 160), and an affix is a bound morph that is not a root and that always occurs on roots of the same class (i.e. always on nouns, on verbs, or on adjectives; see Haspelmath 2021).

The clitics in the initial examples given above are evidently not roots (i.e. contentful morphs denoting an object, an action or a property; Haspelmath 2023), and they are not affixes either because they may occur adjacent to different classes of forms, as illustrated in (6) and (7) for English (eng; Indo-European, Germanic) and Russian (rus; Indo-European, Balto-Slavic) (and also in (5a) and (5b) above for Tagalog - tgl; Austronesian, Malayo-Polynesian).

(6) English (Indo-European, Germanic)

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

(7) Russian (Indo-European, Balto-Slavic)

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

This property of clitics is also called NONSELECTIVITY, contrasting with the WORD-CLASS SELECTIVITY of affixes.<sup>2</sup> An affix such as a Latin genitive suffix must always occur on a noun (contrasting with English 's, which can also occur on adverbs and verbs), and an affix such as a German person-number suffix must always occur on a verb (contrasting with Tagalog *siya*, which can occur on negation markers as well).

It is thus their nonselectivity that picks out clitics in the definition that I use here (see the further discussion in Section 6). Some linguists might prefer a characterization of clitics that makes reference to their phonological properties, but it seems impossible to define clitics phonologically (this is discussed in Section 7 below).

In the next three sections (Sections 2-4), we will see examples of various types of clitics from a wide range of languages, before we move on to a discussion of the definition of the term *clitic* (Section 5). Then I will discuss the lack of word-class selectivity (Section 6), before examining the phonological properties of clitics (Section 7). One consequence of the present definition is that clitics are concrete forms, so that *nonsegmental clitics* or *endoclitics* cannot exist (Section 8). Finally, I will say a few things about Romance *object clitics*, which have played a big role in the literature, but which turn out to be affixes rather than clitics (Section 9).

## 2. Semantic types of clitics

### 2.1. Content words and function words

Linguists often distinguish between CONTENT WORDS and FUNCTION WORDS. Content words are nouns, verbs, and adjectives, and function words are most other types of words.<sup>3</sup> They cannot be easily characterized positively, but there is widespread agreement that the most important classes of function words are relators (adpositions, subordinators), linkers (complementizers, coordinators), articles, tense-aspect auxiliaries, and various kinds of discourse markers. What most function words share (also with affixes) is that the information they convey is discursively secondary (Boye

---

<sup>2</sup> Instead of *nonselectivity*, the literature often uses the term *promiscuous attachment* or *promiscuity* (from Zwicky 1987: 136). I used the latter term in the past, but it seems better to replace it with a term that does not have unwanted associations. (Another possibility that I considered was *indiscriminacy*.)

<sup>3</sup> Some kinds of words, such as numerals and interjections, do not readily fit into this classification.

& Harder 2012). The current section will illustrate various kinds of clitics, most of which are function words.

## 2.2. Person indexes (= bound person forms)

Perhaps the best-known types of clitics are person indexes (Haspelmath 2013), and especially the object indexes of the Romance languages have been discussed extensively (in the wake of Kayne 1975). For example, Spanish *te* ‘you’ is a weak object person form and contrasts with a strong form *ti* ‘YOU’.

(8) Spanish (Indo-European, Italic)

*Te*                    *quiero*.

you.ACC          love.1SG

‘I love you.’ (contrasting with *quiero a ti* ‘I love YOU’)

Below in Section 9 we will see that the Romance object person forms are not really clitics, even though they are usually called *clitics*. But many other languages have subject and/or object person clitics, e.g. Serbo-Croatian (hbs; Indo-European, Balto-Slavic), where subject forms (for past tense constructions) like *smo* and object forms like *mu* and *je* must occur in the second position.

(9) Serbo-Croatian (Indo-European, Balto-Slavic; Bošković 2016: 28)

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*.  
introduced    1PL.SBJ                    him.DAT                    her.ACC                    yesterday

‘We introduced her to him yesterday.’

German (deu; Indo-European, Germanic) has a few clitic subject and object forms in the colloquial language, illustrated in (10).<sup>4</sup>

---

<sup>4</sup> In this connection, English colloquial object forms are sometimes mentioned as well (e.g. *hit 'em* ‘hit them’). But these English forms always follow the verb directly, so they are affixes rather than clitics

## (10) Colloquial German (Indo-European, Germanic)

*Willst de se heute haben?*  
 want.2SG you them today have

‘Would you like to have them today?’ (Standard: *Willst du sie heute haben?*)

In Lak (lbe; Nakh-Daghestanian, Daghestanian), a subject person index usually follows the verb, but when the focus is on an argument (the subject in (11b), the object in (11c)), it follows this argument.

## (11) Lak (Nakh-Daghestanian, Daghestanian; 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.’

Clitic person indexes were also seen above in (4) (Persian) and (5) (Tagalog), and more examples are given below: Halkomelem (hur; Salishan, Coast Salish) (14); Bulgarian (bul; Indo-European, Balto-Slavic) (33), (52), (71b); Ancient Greek (grc; Indo-European, Graeco-Phrygian) (42); Wambaya (wmb; Mirndi, Ngurlun) (43); Polish (pol; Indo-European, Balto-Slavic) (58a); Kugu Nganhcara (wua; Pama-Nyungan, Paman) (72); Udi (udi; Nakh-Daghestanian, Daghestanian) (84).

**2.3. Tense-aspect forms**

In many languages, tense-aspect meanings are expressed by verb-like auxiliaries, and these are commonly bound non-affixal forms, i.e. clitics. For example, the French

---

(if one wants to treat them as distinct from the full forms). By contrast, German shortened forms like *se* can occur both postverbally and in the first position of the Middle Field (e.g. *wenn de se heute haben willst* ‘if you want them today’), so they are not class-selective and are thus clitics.

auxiliary *avoir* ‘have’ is used to form a past tense, and it is a (nonaffixed) bound form rather than a free form, as it must always cooccur with a verb. In (12), for example, the second verb cannot be omitted.<sup>5</sup> This contrasts with English, where the perfect auxiliary *have* can occur without the verb and is thus not a bound form.

(12) French (Indo-European, Italic)

*J’ai changé et mon mari a changé aussi.*  
I have changed and my husband has changed too  
‘I have changed and my husband has  $\emptyset$ , too.’ (\**J’ai changé et mon mari a  $\emptyset$  aussi.*)

Two more examples of tense-aspect clitics come from Garrwa (wrk; Garrwan) and, Halkomelem.

(13) Garrwa (Garrwan; Mushin 2012: 206–207)

- a. *Jungku yal =i bangkulu-na.*  
stay 3PL =PST prison-LOC  
‘They stayed in the prison.’
- b. *Najba =yi bula.*  
see =PST 3DU.NOM  
‘They two saw him.’

(14) Halkomelem (Salishan, Coast Salish; Gerds & Werle 2014: 251)

*Nem’ =ʔə =č =ceʔ qʷat-ət tə sti:č?*  
go =QM =2SG.SBJ =FUT wait-TR DET bus  
‘Are you going to wait for the bus?’

More tense-aspect clitics are illustrated below: Wambaya (43), English (59), Italian (ita; Indo-European, Italic) (63), Bulgarian (71b).

---

<sup>5</sup> Note that the French auxiliary *avoir* is not prefixal as it need not occur immediately before the verb (e.g. *il a probablement changé* ‘he has probably changed’).

## 2.4. Articles

Definite and indefinite articles accompany nouns and cannot occur on their own, so they are bound forms. In some languages, they always occur directly on nouns and are thus affixes, e.g. in Swedish (Indo-European, Germanic) (*kung-en* [king-DEF] ‘the king’). In other languages, they occur in a peripheral position and noun modifiers may intervene, e.g. in Italian, Basque (eus; Isolate, Eurasia) and Haitian Creole (hat; French-based creole, North America). Such articles are clitics, regardless of their spelling (in Basque, the article is written as if it were a suffix: *etxea*, *etxe berria*).

(15) Italian (Indo-European, Italic)

- a. *il*        *libro*  
       the        book
- b. *l'*        *altro libro*  
       the        other book

(16) Basque (Isolate, Eurasia)

- a. *etxe*        = **a**  
       house     = ART  
       ‘the house’
- b. *etxe*        *berri* = **a**  
       house     new    = ART  
       ‘the new house’

(17) Haitian (French-based creole, North America; Fattier 2013)

- M*        *wè*    *ti*        *nèg*    *ki*        *frekan*        = **an**  
 1SG        see    little    man    REL    impertinent    = DEF  
 ‘I saw the boy who is impertinent.’

Cases like Haitian Creole, where articles occur even outside of relative clauses, are uncommon, but clitic articles are widespread. English *the* and *a(n)* are subminimal (lacking a full vowel), so that they are recognized as clitics even by authors who do not rely on the nonselectivity criterion (e.g. Dixon 2007).



Examples of a few more clitic articles are given below: Welsh (cym; Indo-European, Celtic) (37), Bulgarian (45, 53), Italian (63).

## 2.5. Question and negation particles

Polar question markers are typically clitics, because they can often be associated with a variable focus of the question (as seen in (7a-c) for Russian, and in (54) below for Turkish - tur; Turkic, Common Turkic). In languages where a question particle must occur in a peripheral position, it is still often a clitic because the initial or final expression of the question is not always of the same class. For example, in Mauwake (mhl; Nuclear Trans-New Guinea, Madang), the polar question clitic =*i* may occur on verbs or on nominals.

(18) Mauwake (Nuclear Trans-New Guinea, Madang; Berghäll 2015: 226)

- a. *Sira nain piipua-inan =i?*  
habit that leave-FUT.2SG = QM  
'Will you give up that habit?'
- b. *Nobonob ikiw-e-man nain, owowa eliwa =i?*  
Nobonob go-PST-2PL that village good = QM  
'You went to Nobonob, is it a good village?'

In Mapudungun (arn; Araucanian, Mapudungun), the question marker =*am* may occur in polar questions or constituent questions, and it can occur after different kinds of words.

(19) Mapudungun (Araucanian, Mapudungun; Zúñiga 2014: 165)

- Nepe-le-y ñi püñeñ =am?*  
wake.up-RES-IND my child = QM  
'Is my child awake?'

Languages with polar question markers as verbal affixes are not uncommon either, even though they are less frequent than question particles (Dryer 2005 finds 600 languages with question particles, and 179 languages with "interrogative verb

morphology”). In these languages, the polar question marker by definition always occurs on the verb.<sup>6</sup>

Clitic negation particles are quite common, too, and are illustrated below in (52) for Bulgarian and in (58) for Polish.

## 2.6. Adpositions

In the general literature on clitics, adpositions are not very prominent, though Dixon (2007) notes that several English prepositions (*at, for, to*, etc.) are very much like English auxiliaries and person forms in that they exhibit full forms and shortened forms. But crucially in the present context, adpositions are bound forms which indicate a nominal's semantic role and which do not always occur on the noun, so they are clitics by definition. Two examples are given in (20)-(21).

(20) Korean (Koreanic; Chae 2020: 133)

*Wuli-nun siktang = eyse achim pap = ul mek-ess-ta.*  
 we-TOP restaurant = in morning meal = ACC eat-PST-DECL  
 ‘We ate breakfast in a restaurant.’

(21) Ts'ixa (Khoe-Kwadi, Khoe; Fehn 2016: 108)

*Maá. | àm tsá gére táùn = m ʔò kúũ?*  
 when 2SG.M FUT town = SG.M ALL go  
 ‘When will you go to town?’

A few more examples of adpositional clitics are given below: Sri Lanka Malay (sci; Malay-based creole, Eurasia) (28), Russian (55), Fwe (fwe; Atlantic-Congo, Volta-Congo) (70), as well as in (47). I do not exemplify the nonselectivity of all these markers here, but all of them occur not only next to nouns, but also next to noun modifiers, as seen in (47a) and (70b) (see also fn. 19 for Japanese postposed flags).

---

<sup>6</sup> Dryer does not say so explicitly, but it appears that his *particles* are always nonselective (occurring adjacent to different types of words), while his *affixes* (or other morphology) are always verb-specific. He notes the nonselectivity requirement only for particles that are called *clitics* by the language describers: “Interrogative clitics, which attach to some word, but which exhibit freedom as to the category of word they attach to, are treated here as question particles.” (Dryer 2005: 470)

## 2.7. Subordinators

Like adpositions, subordinators are generally bound forms, and they are often nonselective as well, so that they are clitics rather than suffixes. Their clitic nature is particularly salient when there is phonological reduction, as with English *that* [ðət], French *que/qu'* (reduced before a vowel), or in the Chadic language Makary Kotoko (mpi; Afro-Asiatic, Chadic) *gí/g-* (reduced before a vowel):

(22) Makary Kotoko (Afro-Asiatic, Chadic; Allison 2020: 268)

*Ā*            *gə ən*    *g=*            *ú*    *sī*        *klayaská.*  
2SG.M.COMPL say 1SG    COMP =        1SG take    young.woman  
'He told me to take a young woman (as wife).'

In languages with predominant verb-final order, subordinators are often verbal suffixes (attaching only to verbs), but Turkish *=(y)ken* 'when' is a clitic as it attaches both to verbs and to nouns (Erdal 2000: 42).

(23) Turkish (Turkic, Common Turkic; Göksel & Kerslake 2005: 416)

- a. *Orman-da dolaş-ır =ken bir tilki gör-dü-m.*  
forest-LOC walk-AOR =when a fox see-PST-1SG  
'While walking in the forest, I saw a fox.'
- b. *Ahmet o kitab-ı öğrenci =yken oku-muş.*  
Ahmet that book-ACC student =when read-PRF  
'Ahmet read that book as ('when') a student, it seems.'

A few more examples of subordinator clitics are given in (48) below.

## 2.8. Coordinators

Coordinators meaning *and* or *or* are typically clitics, because they do not occur on their own and combine with forms of different word classes.

- (24) Amharic (Afro-Asiatic, Semitic; Demeke & Meyer 2008: 616)

*šew-u = mm      nəgus-u = mm      dəneggeṭ-u.*  
 man-DEF = CONJ    king-DEF = CONJ    be.surprised-3PL

‘The people as well as the king were both surprised.’

- (25) Tsimshian (or Sm’algyax; Tsimshian; Stebbins 2003: 395)

*Łgu      Hayda      ʼyuuta gwaʼa    daʼal    aam    wila    Smʼalgyax-t.*  
 small    Haida      man    this    but    good    FOC    Smʼalgyax-3.SBJ

‘This young man is Haida but he speaks good Smʼalgyax.’

A few more examples of coordinator clitics are given in (49) below, as well as in (31b) (English *and* = ), (56) (Latin = *que*), and (83) (Andi = *lo* = ).

## 2.9. Information-structural and discourse markers

Topic and focus markers are often clitics, as illustrated in (26)-(28), where they follow an adverb or a postposition.

- (26) Karbi (Sino-Tibetan, Kuki-Chin-Naga; Konnerth 2020: 466)

*Pini = ke              etūm              àn              chō-ràp-pèt-sināng.*  
 today = TOP            1PL.INCL            rice            eat-together-all-HORT

‘Today, let us eat together.’

- (27) Bunaq (Timor-Alor-Pantar; Schapper 2022: 174)

*Neto    Hulul    gene = na    zol.*  
 1SG    Hulul    LOC = FOC    originate

‘It is Hulul where I come from.’

- (28) Sri Lanka Malay (Malay-based creole, Eurasia; Nordhoff 2009: 275)

*TV = ka = jo              anà-kuthumung.*  
 TV = LOC = EMPH            PST-see

‘It was on TV that we saw it.’

Further examples of clitic discourse markers are English *however* in (39) and German *já* in (78).

### 3. Positional types of clitics

With respect to their position, we can distinguish several subtypes of clitics: at the most general level, we can distinguish enclitics and proclitics as well as ambiclitics (Section 3.1) and interclitics (Section 3.2). Using further kinds of criteria, we can also define second-position clitics (Section 3.3), clustering clitics (Section 3.4), and epiphrasal clitics (which occur at the edge of a clause, a nominal, or an adverbial; Section 3.5). These classes are not necessarily mutually exclusive.

#### 3.1. Enclitics and proclitics

The two best-known types are enclitics and proclitics, defined as in (29)-(30). (More transparent terms would be *postclitic* and *preclitic*, but two terms *enclitic* and *proclitic* are fairly old and are based on Greek prefixes, not on Latin *post-/pre-*.)<sup>7</sup>

##### (29) Enclitic

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

##### (30) 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 (31a), and the English coordinator *and* can occur at the beginning of the elliptical expression *and her dog*, as in (31b).

- (31) 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

---

<sup>7</sup> The simple (prefixless) term *clitic* is actually fairly new (going back to Nida 1946; Stockwell et al. 1965). In earlier grammatical descriptions of the classical languages, we mostly find *enclitic*, but *proclitic* was already used in the 19th century.

\**Lee and*), so the reduced free forms in (31a-b) are the basis for classifying 's as an enclitic and *and* as a proclitic.

The nonclitic word preceding an enclitic, and the nonclitic word following a proclitic, will be called its ANCHOR in this paper. According to another very common terminology, the element with which a clitic can occur in such contexts is its HOST, and an equals sign serves as a boundary symbol linking a clitic to its host, as in (32). In many or most cases, *anchor* and *host* refer to the same word.

- (32) a.        *my friend = s bik*  
           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 typically unclear what exactly this means. (In this paper, I do not attach particular significance to the notation with the equals sign. It can always be replaced by a space.)

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” (Halpern 1998: Section 1):

Clitics which form a prosodic unit with a host on their left are enclitics, while those forming a unit to their right are proclitics.

However, as will be discussed further in Section 7 below, it is often unclear how to identify the relevant prosodic units. Consider the Bulgarian object person index *ja* ‘her’, which occurs postverbally when the verb is clause-initial (as in (33a)), but preverbally when there is another preverbal constituent (as in (33b)). It cannot occur preverbally in initial position (see (33c)).

(33) Bulgarian (Indo-European, Balto-Slavic; Avgustinova 1994: 31)

- a.    *Vidjax ja.*  
       I.saw her  
       ‘I saw her.’

- b. *Otnovo ja vidjax.*  
 again her I.saw.  
 ‘I saw her again.’
- c. \**Ja vidjax.*  
 her I.saw  
 (‘I saw her.’)

According to the prosodic-unit criterion, should we say that *ja* in (33b) is proclitic to the verb, or that it is enclitic to the adverb? This is unclear and may not even be decidable, so it is better to define *enclitic* and *proclitic* with respect to their peripheral occurrence in free forms. According to this criterion, *ja* is an enclitic, because (33a) is possible as a free form, but (33c) is not. Thus, *otnovo* is the anchor in (33b), but it may not be the prosodic host.

A clear case of divergence between the criterion of occurrence in a free form and the prosodic criterion comes from Czech (ces; Indo-European, Balto-Slavic). Here too, the cognate person clitics are enclitic as that they cannot (in the formal standard variety) occur at the beginning of a free form. But according to Toman (1996), they lean prosodically to the following word when they occur after a long nominal phrase, as in (34), that forms a separate prosodic constituent.

(34) Czech (Indo-European, Balto-Slavic; Toman 1996: 506)

*Knihy, které tady vidíte, se dnes platí zlatem.*  
 books which here see.PRS.2PL REFL today pay.PRS.PL gold.INS  
 ‘The books you can see here are paid for with gold today.’

There are sometimes clear segmental effects depending on a neighbouring form, e.g. regressive voicing assimilation, as in English (34a) vs. (34b), or otherwise alternating forms, like Tagalog *ng* (after a vowel) vs. *na* (after a consonant), as in (35a) vs. (35b).

- (34) a. *my friend = [z] car*  
 b. *my bike = [s] brakes*

- (35) a. *wika ng Ingles* ‘English language’ (cf. also (5a))  
 b. *Ingles na wika* ‘English language’

However, the directionality of these segmental effects need not correspond exactly to the prosodic associations. For example, in English, it may appear that the copula forms *is* and *are* form a unit with the subject, as in (36a-b). However, the segmental effect in (36c) depends on the following word: In non-rhotic varieties of English, the reduced form *'re* is pronounced [r] only when the following word begins with a vowel (Dixon 2007: 583).

- (36) a. *The girl's here.* [z]  
b. *The kids're here.* [ə]  
c. *The kids're outside.* [r]

Even more strikingly, the shape of the Welsh definite article *y/yr/'r* depends both on the preceding and on the following word, as seen in (37a-c). The shape *yr* occurs in a C\_V context, the shape *y* in a C\_C context, and *'r* in a V\_ context (both V\_C and V\_V) (Hannahs & Tallerman 2006).

(37) Welsh (Indo-European, Celtic)

- a. *Prynodd y bachgen yr unig lyfr.*  
bought the boy the only book  
'The boy bought the only book.'
- b. *Prynodd y bachgen y llyfr.*  
bought the boy the book  
'The boy bought the book.'
- c. *Brynaist ti 'r llyfr?*  
bought.2SG you the book  
'Did you buy the book?'

The Welsh definite article is a proclitic (as it occurs preceding a noun in isolation, e.g. *y llyfr* 'the book'), but its shape is crucially determined by the preceding word when it does not occur initially.

For these reasons, the definitions in (29) and (30) do not make reference to prosodic units, but only to occurrence in free forms (i.e. as part of forms that can occur in isolation). We will see below in Section 7 that occurrence in isolation is not the same as absence of phonological deficiency.



Most clitics are enclitics or proclitics, but there are two other possibilities, ambiclitics and interclitics (Section 3.2). Ambiclitics are clitics that may look like a proclitic or like an enclitic, e.g. English *however*.

**(38) 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.

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

Another example is the German adposition *entgegen* ‘against’, which can be used prepositionally or postpositionally (*entgegen meinem Rat* ‘against my advice’ or *meinem Rat entgegen*, van Gijn & Zúñiga 2014: 150).

**3.2. Interclitics**

Some languages have clitics that must occur between two other forms. These are called *interclitics* here, defined as in (40).

**(40) 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 (35a-b) above. Further examples are in (41a-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 “backward-leaning” phonological behaviour of =*ng*=.

## (41) Tagalog (Austronesian, Malayo-Polynesian)

- a. *Malaki* = **ng** = *bahay*  
 big = LNK = house  
 'big house'
- b. *bahay* = **na** = *malaki*  
 house = LNK = big  
 'big house'

An even better-known example of an interclitic is the Persian Ezâfe form = (y)e =, as in *lebâs = e = zibâ = ye = Maryam* [dress =EZ= beautiful =EZ= Maryam] 'Maryam's beautiful dress' (Samvelian 2007: 608). This form occurs before an adnominal modifier, but only when a noun or another modifier precedes. It never occurs at the beginning or end of a free form.<sup>8</sup> Another example of an interclitic is the shortened English copula ('s, 're), illustrated in (36a-c) above.

Interclitics could be said to have two anchors, and the same might be said about ambiclitics when they occur in medial position (as in 39b). But it is unclear how to treat ambiclitics in general, so it may be best to restrict the term *anchor* to enclitics and proclitics for simplicity.

**3.3. Second-position clitics**

A number of languages have clitics which must occur in the second position in a clause (see Bošković 2016 for a survey). We saw a Tagalog example in (5a-b), a Russian example in (7a-c), and a Serbo-Croatian example in (9a-b). Clitics of this type were first identified for Ancient Greek, illustrated in (42).

## (42) Ancient Greek (Indo-European, Graeco-Phrygian; Sappho 118.3, Wackernagel 2020: 60)

<i>Aithiopíai</i>	<i>me</i>	<i>korai</i>	<i>Latoûs</i>	<i>anéthēken</i>
Ethiopian.DAT	me.ACC	girl.DAT	Leto.GEN	dedicated

<sup>8</sup> Samvelian (2007) treats the marker = (y)e = as a suffix, but it occurs both after nouns and adjectives, so it is not class-selective and hence a clitic.

*Arista.*

Arista.NOM

‘Aristas dedicated me to Leto’s Ethiopian daughter.’

Wackernagel’s (1892) paper made second-position clitics famous by pointing out that they occur in a number of ancient Indo-European languages (see Walkden 2020 for some background). Clitics of this type have also been found in various other parts of the world, including Uto-Aztec languages, Panoan languages, and a number of Australian languages such as Wambaya.

(43) Wambaya (Mirndi, Ngurlun; Nordlinger 1998: 139, 140)

- a. *Darranggu-nu ngiyi = ng = a irrijabi.*  
 stick-LOC 3SG.NONMASC.ERG = 1.OBJ = NONFUT scratch  
 ‘The stick scratched me.’
- b. *Dagama gini = ng = a ngirra.*  
 hit 3SG.MASC.ERG = 1.OBJ = NONFUT us.EXCL.ACC  
 ‘He hit us.’
- c. *Guyala ngurr = uji ngajbi irra.*  
 NEG 1PL.INCL.ERG = IRR.PRS see them.ACC  
 ‘We have never seen them.’

Nordlinger (1998) calls the second-position clitic clusters *auxiliary* and writes them as one word, but they are not different in nature from the Tagalog or Serbo-Croatian clusters.

Second-position clitics are usually enclitics, but it may be that some of them are interclitics (requiring another form to follow). I define this type as in (44).

(44) **Second-position clitic**

A second-position clitic is a clitic that must occur (possibly as part of a clitic cluster) directly after the first word or nominal or adverbial expression of a clause, or after the first word of a nominal.

Most second-position clitics occur after the first nominal or adverbial of a clause, or after an initial verb or particle (as in (43b-c)), and it is rare to find such clitics after

the first word when this is part of a nominal. An example of such a *nominal-internal* clausal clitic is (42) from Lesbian Greek.

For the definition in (44), one might want to use the more general formulation “must occur after the first constituent”, but the terms *nominal (expression)* and *adverbial (expression)* are very clear (and *word* is fairly clear, too), while *constituent* is more abstract and may not be so clear. In the literature, there are rich discussions of the precise conditions under which second-positions occur in particular languages, often involving prosodic conditions (see, e.g., Bošković 2016). However, the vast majority of what have been called second-position clitics fall under the comparative concept in (44).

Second-position clitics within nominals can be illustrated by the Bulgarian definite article in (45).

(45) Bulgarian (Indo-European, Balto-Slavic; Halpern 1995: 153)

- a. *kniga* = **ta**  
 book = DEF  
 ‘the book’
- b. *xubava* = **ta** *kniga*  
 nice = DEF book  
 ‘the nice book’
- c. *moja* = **ta**      *xubava*      *kniga*  
 my = DEF      nice      book  
 ‘my nice book’

Not only articles, but also adpositions may occur in second position, though this is rare. Dryer (2005: 211) calls such clitics *inpositions* and gives an example from Yawuru (ywr; Nyulnyulan, Eastern Nyulnyulan).

(46) Yawuru (Nyulnyulan, Eastern Nyulnyulan; Hosokawa 1991: 81, 383–384)

- a. *dyungku* = **gun**  
 fire = LOC  
 ‘in the fire’

- b. *bika* = **gun**     *larrkadi*  
 shade = LOC     bottle.tree  
 ‘in the shade of a bottle-tree’
- c. *nyamba* = **gun**     *maya*  
 this     = LOC     house  
 ‘in this house’

### 3.4. Epiphrasal clitics

Many languages have clitics that provide information on a phrase’s relationship with its environment and occur peripherally, either in a nominal expression (i.e. adpositions), or in a clause (i.e. subordinators) (as we already saw in Sections 2.6-7). They may be proclitics or enclitics.

#### (47) Examples of adpositional clitics

- |    |           |             |       |  |
|----|-----------|-------------|-------|--|
| a. | English:  | <i>to</i>   | ‘to’  | <i>to our house</i>                      |
| b. | French:   | <i>pour</i> | ‘for’ | <i>pour notre maison</i> ‘for our house’ |
| c. | Hebrew:   | <i>le-</i>  | ‘to’  | <i>le = David</i> ‘to David’             |
| d. | Japanese: | <i>no</i>   | ‘of’  | <i>Hanako = no</i> ‘of Hanako’           |

#### (48) Examples of subordinator clitics

- |    |          |             |         |                                 |                         |
|----|----------|-------------|---------|---------------------------------|-------------------------|
| a. | German:  | <i>als</i>  | ‘when’  | <i>als wir träumten</i>         | ‘when we were dreaming’ |
| b. | Persian: | <i>ke</i>   | ‘that’  | <i>ke âmadi</i>                 | ‘that you came’         |
| c. | Arabic:  | <i>iðaa</i> | ‘if’    | <i>iðaa kun-ta hunaaka</i>      | ‘if you-are there’      |
| d. | Chinese: | <i>de</i>   | ‘which’ | [ <i>lái de</i> ] <i>nánhai</i> | ‘the boy [who came]’    |

Another class of clitics that commonly occur peripherally to a phrase is coordinator clitics (as already seen in Section 2.8). These are often interclitics.

#### (49) Examples of coordinator clitics (Fortescue 1984: 120)

- |    |          |           |       |   |
|----|----------|-----------|-------|---|
| a. | Spanish: | <i>y</i>  | ‘and’ | <i>guerra y paz</i> ‘war and peace’       |
| b. | Lezgian: | <i>ni</i> | ‘and’ | <i>buba = ni died</i> ‘father and mother’ |

- c. Russian: *ili* 'or' *zdes' ili tam* 'here or there'  
 d. Greenlandic: = *lu* 'and' *ingilluni = lu* 'and she sat down'

Many or most languages also have focus clitics, especially particles meaning 'only' or 'also', which typically occur epiphrasally, as in (50a-c).

(50) Focus clitics

- a. Polish: *tylko* 'only' *tylko dzisiaj* 'only today'  
 b. M. Greek: *ke* 'too' *ke i mitéra mu* 'my mother, too'  
 c. Hungarian: *is* 'too' *én is* 'me too'  
 d. Japanese: *mo* 'too' *watashi mo* 'me too'

However, especially clitics that mean 'also' can occur in a *floating* position, not immediately adjacent to their focus. Example (51) shows the stressed clitic *auch*, which can occur in a preverbal position, not adjacent to its focus. Thus, *AUCH* is a clitic, but it is not epiphrasal.

(51) German<sup>9</sup> (Indo-European, Germanic)

*Meine Schwester ist heute = AUCH gekommen.*  
 my sister has today = also come  
 'My sister came today, too (= 'also my sister' or 'also today').'

A non-peripheral position is not common for the other semantic classes illustrated above, though we sometimes find second-position coordinators (e.g. in Latin, ex. (56) below), and occasionally second-position adpositions (e.g. Yawuru, ex. (46) above).

Epiphrasal clitics are very common, but they are not prominent in the general-theoretical literature on clitics. This is probably because they present no particular problem of analysis, and not because they would not fall under the usual clitic concept.

---

<sup>9</sup> German stressed *AUCH* is an enclitic, because it can occur at the end of a free form (e.g. *heute = AUCH* 'today, too'). There is also a fully synonymous unstressed *auch* (used in a more formal register), which is a proclitic, because it can occur at the beginning of a free form (*auch = HEUTE* 'also today'). If these two instances of *auch* were treated as the same form, this form would be an ambiclitic.

### 3.5. Clustered clitics

Clitics sometimes occur in clusters with rigid internal ordering. Such clusters may occur in second position (as we already saw for Tagalog, Serbo-Croatian, and Wambaya, Section 3.3), but they can also be proclitic, as in the Bulgarian example in (52). Here the clitics occur in a rigid position: *da* – NEG – AUX – DAT – ACC (see Avgustinova 1994: 32 for details).<sup>10</sup>

(52) Bulgarian (Indo-European, Balto-Slavic; Spencer & Luís 2012: 125)

<i>Da</i>	<i>ne</i>	<i>si</i>	<i>mu</i>	<i>go</i>	<i>dala</i>	<i>poveče!</i>
that	not	2SG.AUX	3SG.DAT	3SG.ACC	give	more

‘Don’t give it to him any more, or else!’

For syntactic elements, this kind of rigid ordering is not expected, and this has led some linguists to think of clustered clitics more in *morphological* terms. More generally, Spencer & Luís (2012: 126) note that:

elements that are traditionally called clitics may exhibit a good many features normally associated with affixes ... when they combine into clusters: a fixed order, idiosyncratic alternations in ordering, haplology, idiosyncratic allomorphy, and accidental gaps, not to mention multiple exponence and cumulation.

This is another reason for caution in attributing the properties of types of linguistic forms to larger architectures (“morphology vs. syntax”, “grammar vs. lexicon”, etc.). At present, our understanding of the reasons for these behaviours is quite limited.

## 4. Phonological types of clitics

In this section, I describe and illustrate three types of clitics based on their phonological properties: welded clitics (Section 4.1), stress-affecting clitics (Section 4.2), and shortened clitics (Section 4.3).

---

<sup>10</sup> Above in (33), we saw that single object clitics are enclitics, but (52) shows that when they occur in a cluster that begins with *da*=, the entire cluster is proclitic. So perhaps we should say that object clitics are ambiclitics, because they can occur both at the end of a free form and (as part of a cluster) at the beginning of a free form.

#### 4.1. Welded clitics

A welded clitic is a clitic that interacts with a neighboring form in a segmental way, either by causing segmental change or by undergoing segmental change. For example, the Bulgarian enclitic definite article (*-ət/-ta/-to* in the singular, see (45)) may cause a segmental change in the preceding noun, as seen in (53); the Turkish question particle *mU* shows vowel harmony, harmonizing with the last vowel of the preceding word, as seen in (54); and several Russian prepositions (e.g. *v(o)*, *k(o)*, *o(b)*) have somewhat different shape variants depending on various properties of the next word, as seen in (55).

(53) Bulgarian (Indo-European, Balto-Slavic): definite article

<i>kniga</i>	[book]	<i>kniga-ta</i>	[book-DEF]
<i>vol</i>	[ox]	<i>vol-ət</i>	[ox-DEF]
<i>grək</i>	[Greek]	<i>gərək-ət</i>	[Greek-DEF] (with stem change)

(54) Turkish (Turkic, Common Turkic): polar question particle

<i>geldi mi?</i>	‘did she come?’
<i>öldü mü?</i>	‘did he die?’
<i>Alı mı?</i>	‘Ali?’
<i>dün mü?</i>	‘yesterday?’

(55) Russian (Indo-European, Balto-Slavic): prepositions *v(o)* ‘in’, *s(o)* ‘with’

<i>v nužde</i>	‘in need’	<i>vo vrede</i>	‘in harm’ (* <i>v vrede</i> )
<i>s radost’ju</i>	‘with joy’	<i>so straxom</i>	‘with fear’ (* <i>s straxom</i> )

In English, the difference between the two variants of the indefinite article *a(n)* (e.g. *a tree* vs. *an old tree*) is a striking case of welding in a proclitic.

For welded clitics, we might distinguish between BACKWARD-WELDED clitics (whose segmental shape interacts with the shape of a preceding form) and FORWARD-WELDED clitics (which interact with a following form). In general, backward-welded clitics are enclitics, e.g. the Turkish polar question particle *mi/mü/mu/ml*, and forward-welded forms are proclitics, e.g. the Russian prepositions *v(o)* and *s(o)*. But proclitics may also be backward-welded, as seen in the Welsh definite article *yr/’r* in (37) above, and enclitics may be forward-welded



(e.g. Kukama = *pura*, discussed by Zingler 2020: 266). Quite generally, it seems that backward-welding is more common than forward-welding (Himmelmann 2014). The interclitics that we saw in Section 3.2 are backward-welded (Tagalog *na/ng*, Persian *e/ye*).

Welding does not seem to happen very often with clitics, and it has in fact been suggested that “morphophonological idiosyncrasies” are symptomatic of affixes, but not of clitics (Zwicky & Pullum 1983: 504; Nevis 2000: 389).<sup>11</sup> If phonological interaction plays no role in the definition (as in the present proposal), then the lack (or scarcity) of segmental interactions between clitics and adjacent words becomes an interesting testable prediction that we can make.

#### 4.2. Stress-affecting clitics

Most clitics do not interact suprasegmentally (with respect to their stress or tone properties) with adjacent words, and we can call them *suprasegmentally inert*. For example, while most Turkish words have final stress, the question clitic (e.g. *geldi mi?* ‘did she come?’ in (54) above) is not part of the stress domain and thus does not carry stress. But some clitics are suprasegmentally active in that they are relevant for the stress or tone properties of their anchor words (or perhaps for other adjacent words). Here we will briefly consider clitics which affect the stress of their anchor word.

We can distinguish two types of stress-affecting clitics. First, STRESS-SHIFTING CLITICS are clitics which induce a shift of the stress pattern of their anchor word. For example, the Latin conjunctive clitic = *que* induces a stress shift to the final syllable of the anchor word that it annexes to (see Plank 2005).

(56) Latin (Indo-European, Italic)

- a. *ménsa* ‘the table’
- b. *mensá* = **que** ‘and the table’

---

<sup>11</sup> Conceivably, this could be because clitics are combined “postlexically” with their anchors, while affixes are combined “lexically” and thus undergo “lexical phonological” processes (e.g. Anderson 2005). But on such a view, it is puzzling that phenomena such as Bulgarian *grək ~ gərək-ət* as in (53) are attested at all (see also Halpern 1995 on “lexical clitics”, and Spencer & Luís 2012: Section 4.4.3).

A similar effect is found in Modern Greek (ell; Indo-European, Graeco-Phrygian), where enclitics (such as *mas* ‘our’, *mu* ‘my’) are unstressed but induce an additional stress on anchor words that are stressed on the antepenultimate syllable, as illustrated in (57b). Words that are stressed on the penultimate or ultimate syllable are unaffected, as seen in (57c-d).

(57) Modern Greek (Indo-European, Graeco-Phrygian; van Oostendorp 2012: 1166)

- |    |                        |                 |                      |
|----|------------------------|-----------------|----------------------|
| a. | <i>o jítonas</i>       | ‘the neighbour’ |                      |
| b. | <i>o jìtonáz = mas</i> | ‘our neighbour’ |                      |
| c. | <i>i stafíða = mu</i>  | ‘my raisin’     | ( <i>i stafíða</i> ) |
| d. | <i>i ayorá = mas</i>   | ‘our market’    | ( <i>i ayorá</i> )   |

Stress-shifting clitics seem to be rare, but they have been prominent in the literature, because Greek and Latin are such important languages in Western culture, and the term *enclitic* was originally used for stress-shifting Greek clitics.

The second type is STRESS-INTEGRATED CLITICS. These are clitics which are part of the anchor word’s stress domain and carry stress when the general stress rule would assign stress to them. An example is the Polish negator *nie*, which is stressed when it occurs with a monosyllabic verb form.

(58) Polish (Indo-European, Balto-Slavic; Rubach & Booij 1985: 317)

- |    |                   |                      |
|----|-------------------|----------------------|
| a. | <b><i>nie</i></b> | <i>wiedziáta = m</i> |
|    | NEG               | knew = 1SG           |
|    |                   | ‘I did not know’     |
| b. | <b><i>nié</i></b> | <i>wie-m</i>         |
|    | NEG               | know-1SG             |
|    |                   | ‘I do not know’      |

In this regard, the negator *nie* contrasts with the polar question particle *czy*, which is a clitic, too (*czy wiém?* ‘do I know?’), but which behaves like most clitics in that it is

outside the stress domain.<sup>12</sup> Another clitic that is inside the stress domain is Mapudungun = *ám* (seen in example (19) above), which receives stress as consonant-final words have final stress (Zúñiga 2014: 165).<sup>13</sup>

#### 4.3. Shortened clitics (vs. length-invariant clitics)

Some clitics are closely related to formally similar counterparts and can be regarded as abbreviated variants of them. This type is very well known from English, illustrated in (59).

##### (59) English (Indo-European, Germanic)

FULL	SHORTENED
<i>will</i>	<i>'ll</i>
<i>would</i>	<i>'d</i>
<i>is</i>	<i>'s</i>
<i>are</i>	<i>'re</i>

##### (60) German (Indo-European, Germanic) (see (10) above)

FULL	SHORTENED	
<i>sie</i>	<i>se</i>	'she, her, they, them'
<i>du</i>	<i>de</i>	'you'
<i>es</i>	<i>s</i>	'it'
<i>er</i>	[ <i>e</i> ]	'he'

---

<sup>12</sup> Rubach & Booij (1985: 317) suggest that *nie* is a prefix, but since it occurs nonselectively (e.g. *nie dzisiaj* 'not today'), it is a clitic on the current definition.

<sup>13</sup> Very rarely, a clitic may be part of the stress domain of a preceding word that is not its anchor: In Chamicuro (ccc; Arawakan, Southern Maipuran), definite articles may be part of the stress domain of the preceding verb, noun or demonstrative (e.g. *aná? = na čmežóna* [this = DEF man] 'this man', Parker 1999: 554). This "backward-leaning" behaviour is similar to backward-welded proclitics which were mentioned briefly in Section 4.1 above.

This clitic type was called *simple clitics* by Zwicky (1977), and this term has become very well known in the literature.<sup>14</sup> However, it is not well-defined, so I prefer the new term SHORTENED CLITIC, defined as in (61).

**(61) Shortened clitic**

A shortened clitic is a clitic that has the same semanticosyntactic function as another form from which it appears to have been abbreviated and by which it can be replaced in the same position.

This notion does not seem to be particularly important in the world's languages, because such clitics are not common (see also Zingler 2020: 337–338). They have been prominent in the literature primarily because English has several such pairs where both a full form and a shortened form occur in the standard spelling.

The great majority of clitics are length-invariant, i.e. they do not occur in pairs such as (59)-(60). Of course, all languages have fast-speech phenomena, and variant forms of function words are extremely common. But there is no good reason to associate the shortened forms with cliticness, because the full counterparts of shortened clitics are very often bound forms and thus clitics, too (as with the English forms in (59)).<sup>15</sup> And the existence of pairs of full forms and shortened forms does not seem to be characteristic of clitics as opposed to affixes, because affixes often have full and reduced variants, too (e.g. German genitive suffix *-es* or *-s*, Italian 3PL suffix *-on* or *-ono*). Thus, the very notion of “simple clitics” seems to be primarily based on the peculiarities of the English spelling.<sup>16</sup>

---

<sup>14</sup> It is sometimes said that the full forms are free forms (e.g. “Simple clitics are unaccented variants of free morphemes”, Anderson 2005: 10), but this not the case for these English forms. The full forms are non-deficient in that they contain a full vowel (not just a single consonant, or just a schwa), but they are bound forms, too, like the reduced forms.

<sup>15</sup> It seems that person forms are different from function words in this regard: Independent person forms can occur on their own and are not clitics, but when shortened person forms arise from these (like German *se* and *de*), these are typically bound and are thus clitics.

<sup>16</sup> Some authors do not include the requirement that “simple clitics” must have full-form or free-form counterparts, e.g. Halpern (1998: Section 1): “An unstressed word which is otherwise unexceptional is known as a simple clitic, after Zwicky (1977)”. Given the general unclarity surrounding Zwicky's terms, it is surprising how popular they were for a few decades.

## 5. Defining clitic

The present definition of *clitic* (repeated below from Section 1 above) is somewhat unusual in that it has no direct antecedent in the literature. However, we will see in this section that it accords well with the way the term has been used in the past.

### (1) Clitic

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

Everyone agrees that clitics are bound forms (incapable of occurring in isolation), and nobody would suggest that a root (a morph denoting an object, an action or a property) can be a clitic.<sup>17</sup> It is also clear that a clitic stands in contrast with an affix, as there are a large number of works that aim to distinguish between affixes and clitics in a language (often following the lead of Zwicky & Pullum's famous 1983 paper). There are thus mainly two points where one might want to opt for a modification of the definition in (1), and one might question (62a) or (62b) (or both):

(62)

- a. clitics are monomorphic
- b. clitics are defined as not exhibiting word-class selectivity (= as non-affixal)

That a clitic is a single morph is not something that has often been said, but it is easy to see that this is the case for the great majority of elements that have been called clitics. Almost all question or negation particles, discourse particles, short adpositions, subordinators and coordinators are monomorphic, and so are many person indexes. It is true that auxiliaries and articles are not uncommonly multimorphic, as illustrated in (63) by the Italian article *le* and the auxiliary *hanno*, which can be analyzed into smaller constituent morphs (*l-* + *-e*, *ha-* + *-nno*).

(63) Italian (Indo-European, Italic)

*L-e*      *donn-e*      *ha-nno*      *lavorato*.  
DEF-F.PL   woman-PL   AUX.PRF-3PL   worked  
'The women have worked.'

---

<sup>17</sup> Exceptionally, Chae (2020: 105) discusses "clitic nouns" in Korean (kor; Koreanic), but this way of talking is very unusual. The elements he discusses are normally treated as derivational suffixes.

However, the constituent morphs of *le* and *hanno* can be treated as individual clitics, so that *l=e=* and *ha=nno=* can be seen as clitic clusters. It is true that this is a nontraditional way of describing these forms, but there does not seem to be a good alternative. If a clitic could consist of a sequence of morphs (i.e. if there could be such a thing as “a composite clitic”), then all clitic clusters could be composite clitics. Thus, it is best to specify that clitics must be monomorphic by definition.

The other somewhat nontraditional component of the definition is (62b), the lack of word-class selectivity. This criterion (also called *promiscuity* for short) will be discussed further below in Section 6.

In addition to boundness (= non-independence) and nonselectivity, quite a few authors mention phonological criteria, especially *phonological dependence* or *deficiency*, as in (64).

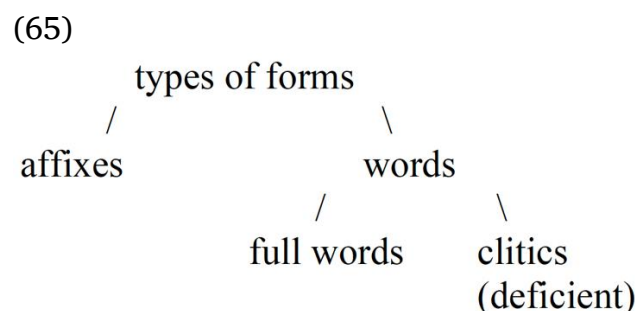
(64)

- a. “[clitics:] grammatical elements which themselves bear no stress and which make up a phonological word with a host item (that bears stress) which the clitic either precedes (it is then a proclitic) or follows (an enclitic)” (Dixon 2007: 574)
- b. “The best way to define the special status of clitics is that in terms of prosodic deficiency: they are words in the morpho-syntactic sense, but not in the phonological sense.” (Booij 2012: 290)
- c. “The most prominent property of clitics is their deficiency. Most often this deficiency is attributed to the phonological status of clitics: clitics are defective in their phonological representation and therefore have to prosodically combine with an adjacent non-clitic word.” (Ionova 2020: 22)
- d. “Clitics are function words that lack independent stress.” (Pescarini 2021: Section 1.1)

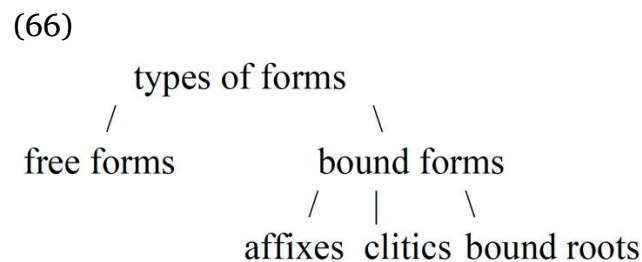
But is such an additional phonological criterion necessary? As Anderson (2011: 4) notes, phonological deficiency or dependence does not distinguish clitics from affixes:

With relatively few exceptions, the affixes found within words as formal markers of derivational and inflectional structure also lack an autonomous organization into prosodic constituents at or above the level of the [p-word].

Thus, the one criterion that clearly distinguishes clitics from affixes is the lack of word-class selectivity (or *promiscuity* of attachment; Section 6). Even though linguists often express the intuition that clitics are “phonologically attached” to their host (resulting in the Greek term *enklitikón* ‘leaning’), this criterion would be significant only if we already knew that clitics are words. If we could initially divide the types of (non-phrasal) forms into affixes and words, as in (65), then the prosodic deficiency of clitics would be relevant.



But in fact, clitics and affixes are very similar in that they are prosodically deficient, and there is no simple phonological criterion distinguishing affixes from words. Thus, the subdivision in (66) below is much more straightforward, as it is based on the simple criterion of boundness (non-occurrence in isolation).



Note that in addition to being bound forms, affixes and clitics must be defined as nonroot forms, because roots may be bound, too. For example, English requires an article or a plural marker with count noun roots (e.g. A: *What do you want to buy?* B: *A book/Book-s/\*Book*), and it requires an object with many transitive verbs (e.g. A:

*What will you do with it?* B: *Replace it/\*Replace*). Thus, many more roots are bound forms than most English-speaking linguists seem to realize.<sup>18</sup>

By specifying that a clitic is not a root in the definition in (1), we distinguish it from simple nouns, verbs and adjectives without making reference to phonological criteria. As the phonological criteria are complex and difficult to apply (see Section 7), this is a clear advantage of this definition.

But the most important way in which the definition in (1) is superior to many other views is that it relies on a small set of clear criteria that could easily be used in textbooks. By contrast, the earlier literature has made reference to a wide range of diagnostics, and the heterogeneity of the criteria is reflected in quotations such as (67).

(67)

- a. “a serious problem which prevails in much of the work on clitics... is that there is no criterial definition, but rather a list of tendencies, general characteristics, and typical features...” (Klavans 1985: 116)
- b. “the various elements which are called clitics form a heterogeneous bunch, at least superficially, and exactly what is meant by “clitic” varies from study to study” (Halpern 1998: Section 1)
- c. “It is extremely difficult to come up with an explicit set of characteristics that may be used to identify clitics cross-linguistically, because the parameters involved vary from case to case” (Stebbins 2003: 385)

Perhaps the most problematic aspect of the earlier approaches (in the tradition of Zwicky & Pullum 1983) is not that they are complex and that the criteria are heterogeneous, but that they may not always point in the same direction and that their application is often subjective (Haspelmath 2015: Section 3). Subjectiveness is a frequent problem in linguistics when “test batteries” are applied and there is no clear rule for how to proceed when the diagnostics do not all point in the same

---

<sup>18</sup> Traditionally, the term “bound root” has been applied especially to roots that occur only in compounds, e.g. Mandarin Chinese *-gōng* ‘worker’, which only occurs in compounds such as *mù-gōng* [wood-worker] ‘carpenter’ and *diàn-gōng* [electricity-worker] ‘electrician’ (Arcodia 2012: 91). English roots such as *book* are often treated as “free morphemes” in textbooks, but by the criterion of independent occurrence, they are not free.



direction (Croft 2010: Section 2.10; Tallman 2020). This problem affects not only the definition of *clitic*, but also the definition of the subtype *special clitic*, which is often said to involve *special syntax*: “A special clitic is ... a “little word” whose syntax is not assimilable to that of full words that might seem to be syntactically parallel” (Anderson 2005: 79). But if there are no limits on the ways in which special clitics might differ from full forms, then this definition cannot be applied objectively.

Taylor (1995: 181) explicitly argues for “graded membership” in the categories *word*, *clitic* and *affix*, but he provides no generally applicable method for measuring degrees of wordhood, cliticness or affixhood. He may well be right that a modular view of grammar (where morphology and syntax are two different modules) is inappropriate, but this does not entail the conclusion that categories like *word* and *clitic* must exist and must have a prototype structure. Maybe these concepts are primarily tools used by linguists and play no role in speakers’ mental grammars. In any event, unless we know exactly what someone means by an affix, a clitic, or word, it is very difficult to evaluate their statements.

## 6. Lack of word-class selectivity

A crucial component of the definition in (1) is the lack of word-class selectivity of clitics, because it is in this way that clitics differ from affixes, which are bound nonroot morphs as well. That clitics contrast with affixes in this way is fairly widely recognized, as is shown by the quotations in (68).

(68)

- a. “Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.” (Zwicky & Pullum 1983: 503)
- b. “It is largely because of their freedom to attach to practically any part of speech that clitics are recognized as a special linguistic unit.” (Taylor 1995: 180)
- c. “[clitic:] a unit that is not a word in a prototypical sense, but with fewer selectional restrictions than a grammatical affix” (Hildebrandt 2015: Section 1)
- d. “an element is only considered a clitic if it has a non-selective distribution but is dependent on a host domain with respect to one or more parameters of phonological wordhood” (Zingler 2022: 9)

In the literature, we also often read that clitics differ from (standard) affixes in that they are associated with phrases rather than words; in fact, clitics are often called “phrasal affixes” (e.g. Anderson 1992: Ch. 8). And as we saw in Section 3.4 above, some clitics (called *epiphrasal clitics*) are clearly associated with nominal (or adverbial) expressions, or with clauses, and they occur strictly on the periphery of such phrases (e.g. English prepositions, or words for ‘also’ in many languages). These elements are clitics and not affixes because the phrases that they occur with do not always have the same kind of word at the periphery; for example, English prepositions are epiphrasal clitics and not prefixes, because they can occur adjacent to nominal modifiers, not only to nouns (as in (47a) above: *to our house*).

The definition of clitics as *nonroot bound forms* that are not affixes has the advantage that it does not rely on a notion of *phrase*, but only on the nature of adjacent forms. While nominal (and adverbial) phrases may be easy to identify in most languages, this is much less straightforward with verb phrases and all kinds of other phrases. Moreover, not only clitics, but also many affixes have a *phrasal* distribution in a certain sense: Tense affixes and case affixes occur on verbs and nouns, respectively, but semantically and functionally they combine with (verb and noun) phrases.

Linguists often treat case affixes and adpositions together (as *flags*, Haspelmath 2019), and also tense affixes and tense clitics (as *tense markers*), because they do not differ except in their position with respect to the noun or verb. And sometimes we find minimal pairs which appear to show that clitics are not really different in nature from affixes. For example, in Egyptian Arabic (arz; Afro-Asiatic, Semitic) *prepositions* such as *mafa-* ‘with’ are prefixes, as illustrated in (69).

(69) Egyptian Arabic (Afro-Asiatic, Semitic; Gary & Gamal-Eldin 1982: 63, 86)

- a. *mafa- xaal-i*  
 with- uncle-1SG  
 ‘with my uncle’
- b. *mafa- l-bint di*  
 with- DEF-girl this  
 ‘with this girl’

They always occur immediately adjacent to the noun because all nominal modifiers follow the noun, including demonstratives. By contrast, their Standard Arabic counterparts are proclitics, because they occur before the demonstratives and these can occur on their own (and are thus not prefixes), e.g. *maʕa = haad̥aa l-kitaab* ‘with this book’.

Similarly to Standard Arabic, the Bantu language Fwe has a preposition-like connective proclitic (*bo =* or *ye =* in (70)) that can occur not only adjacent to nouns, but also to prenominal demonstratives which are not prefixes (Gunnink 2022: 189).<sup>19</sup>

(70) Fwe (Atlantic-Congo, Volta-Congo; Gunnink 2022: 186, 189)

- a. *bàntù bò = kúmùnzì*  
people CONN = village  
‘the people of the village’
- b. *Èmìsì yè = cí cìshámù mùshámù.*  
roots CONN = this tree COP.medicine  
‘The roots of this tree are medicine.’

Moving on to tense markers, an instructive pair of languages is Greek and Bulgarian. The Modern Greek future-tense marker *tha-* seen in (71a) is a prefix, while the very similar Bulgarian *šte =* in (71b) is a clitic. Its preverbal position is rigid, but it precedes the clitic *go*, and so both must be proclitics. In Modern Greek, the spelling may suggest that *tha-* is a clitic, too, but the Greek object indexes behave differently from Bulgarian: They are always adjacent to the verb and are thus affixes, which means that the future-tense marker *tha-* is a prefix (see also Joseph 2002).

(71)

- a. Modern Greek (Indo-European, Graeco-Phrygian)  
*tha- to- páro*  
FUT it take.1SG  
‘I will take it’

---

<sup>19</sup> In Japanese, the reason for treating postpositions as clitics rather than case suffixes is similarly subtle: as Nakamura (2018: 249) notes, the phrasal clitic =*dake* ‘only’ may precede the clitic postpositions, e.g. *Hanako = dake = ga* ‘only Hanako (NOM)’. See also Chae (2020: 39–40, 140) for a discussion of the similar situation in Korean.

## b. Bulgarian (Indo-European, Balto-Slavic)

**šte** = **go** = *vzema*  
 FUT it take.1SG  
 'I will take it'

The term *nonselectivity*, or the older term *promiscuity*, may suggest that clitics are completely indifferent as to the words they are adjacent to, but this is not the case. Clitics contrast with affixes, which must be class-selective (occur always with roots of the same class), so any deviation from full class selectivity means that the element in question must be a clitic rather than an affix. In Standard Arabic, prepositions can only occur with nouns and demonstratives, so they are fairly choosy, but they are not affixes. In some of the quotations in (68), the authors assume degrees of selectivity, but a definition of a term like *clitic* must be clear-cut.<sup>20</sup>

It should be noted that bound nonroot forms are affixes also when they are “mobile” in that they may occur on either side of the root (Haspelmath 2021: 19). This means that some class-selective elements which have often been regarded as clitics because of their phonological properties are not clitics. For example, the person-number indexes in Kugu Nganhcara (boldfaced in (72)) are verbal affixes because they always occur next to the verb.

(72) Kugu Nganhcara (Pama-Nyungan, Paman; Klavans 1985: 104; Smith & Johnson 1985: 103–104)

- a. *Nhila pama-ng nhingu pukpe-wu ku?a wa: (=)-ngu.*  
 DET.NOMman-ERG DET.DAT child-DAT dog give 3SG.DAT  
 'The man gave the dog to the child.'
- b. *Nhila pama-ng nhingu pukpe-wu ku?a (=)ngu- wa:*  
 DET.NOMman-ERG DET.DAT child-DAT dog 3SG.DAT give  
 'The man gave the dog to the child.'

<sup>20</sup> Peter Arkadiev (p.c.) has expressed the intuition that clitics are perhaps better characterized as being completely nonselective, whereas affixes may be somewhat flexible with respect to word classes (e.g. number suffixes attaching both to nouns and adjectives). This would be a logical alternative, but I think that it is much easier to determine that a bound morph is fully selective (i.e. occurs only on one class) than to determine that it is fully nonselective. As a result, the definition of *clitic* is fairly broad here, including all bound nonroot morphs that are not fully selective.

- c. *Nhila pama-ng nhingku kuʔa (=)ngku- wa:*  
 DET.NOMman-ERG you.DAT dog 2SG.DAT give  
 ‘The man gave a dog to you.’

They may occur postverbally (as in 72a) or preverbally (as in 72b-c), but they are always verb-adjacent and thus count as affixes (specifically, they are ambifixes; see Arkadiev & Lander 2020). The literature has often regarded them as clitics (Klavans 1985: Section 2.2; Cysouw 2005), and indeed they are phonologically peculiar in that they have a phonotactic shape that excludes them from initial occurrence. In this sense, one might say that a verbal affix such as *ngku-* (in 72c) leans onto the element preceding it. But few clitics are restricted in this way, and alternations between preverbal and postverbal position of person indexes may be conditioned by a variety of factors (see (33) above for Bulgarian, and (87) below for Portuguese - *por*; Indo-European, Italic).

A language that is similar to Kugu Nganhcara in that its person indexes are most often directly postverbal or preverbal is Persian, illustrated in (73) (example (73b) is repeated from (4) above).

(73) Persian (Indo-European, Indo-Iranic; Samvelian & Tseng 2010: 215)

- a. *Ru-ye miz gozâšt-im = aš*  
 on-EZ table put.PST-1PL.SBJ = 3SG.OBJ  
 ‘We put it on the table.’
- b. *Ru-ye miz = aš gozâšt-im.*  
 on-EZ table = 3SG.OBJ put.PST-1PL.SBJ  
 ‘We put it on the table.’

That these Persian forms are clitics (and not ambifixes as in Kugu Nganhcara) can be seen in examples like (74b), where the preverbal element is extracted and fronted, and the object clitic *= aš* must be fronted with it.<sup>21</sup>

---

<sup>21</sup> Here one might want to object that the verb is not *kon* ‘do’, but *bâz* ‘open’. Indeed, deciding which part of the complex predicate *bâz kardan* [open do] ‘to open’ is the verb requires that we say something in addition, but this is beyond the scope of the present discussion.

(74) Persian (Indo-European, Indo-Iranic; Samvelian & Tseng 2010: 221)

- a. *Mi-xâh-i fardâ bâz = aš bo-kon-i.*  
 IMPF-want-2SG tomorrow open = 3SG.OBJ SBJV-do-2SG  
 ‘You want to open it tomorrow.’
- b. *Bâz = aš agar mi-xâh-i fardâ bo-kon-i..*  
 open = 3SG.OBJ if IMPF-want-2SG tomorrow SBJV-do-2SG  
 ‘If you want to open it tomorrow...’

Another question is how we treat person indexes that occur both on nouns and verbs. For example, van Gijn & Zúñiga (2014: 154) note that “many languages of the Americas, (part of) the verbal person markers are isomorphic with the nominal (possessive) person markers. For some language analysts, this is reason to regard them as clitics”, and they give the example in (75).

(75) Plains Cree (Algic, Algonquian-Blackfoot; Wolfart 1996: 412, 420)

- a. *ni-wâpam-â-w*  
 1-see-DIR-3  
 ‘I see him.’
- b. *ni-sîsîp-im*  
 1-duck-POSS  
 ‘my duck’

If the argument indexes and the person indexes were thought to be the same affixes (i.e. to have the same meaning), then they would indeed be clitics. But if they are semantically different (which seems more reasonable to say), then they are homophonous sets of affixes.

A reviewer observes that derivational affixes in European languages sometimes occur on bases of different classes, e.g. English *dis-* in *dis-honest* (adjective), *dis-order* (noun), and *dis-avow* (verb). Again, if these were treated as having the same meaning, they would be clitics. This would be an unintuitive result, but such unexpected effects at the fringes are often unavoidable.

## 7. Phonological “deficiency” and “dependence”

As I noted in Section 5, it is often said in the literature that clitics are “phonologically deficient”, or “prosodically dependent” on an adjacent form, and introductory works typically mention the etymology of the original Greek term *enklitikón* (‘element that leans on another element’). However, there seem to be no proposals for characterizing the phonological properties of clitics in such a way that they can be applied uniformly to all languages, and for this reason, phonological properties play no role in the definition in (1). It should also be noted that the notion *bound* that is part of the definition is a purely syntactic notion; there is no such thing as *phonological boundness*. In this section, I discuss a number of ways in which “deficiency” or “dependence” have been characterized, and I explain briefly why these notions are not suitable for defining clitics in a general way.

### 7.1. Unpronounceability

Clitics do not have independent stress, and it is for this reason that they are typically thought to be in need of a “host” (Bonet 2019):

One consequence of clitics being prosodically defective is that they cannot be the sole element of an utterance, for instance as an answer to some question; they need to always appear with a host.

But what exactly does “prosodically deficient (or defective)” mean? If it just means that a clitic “cannot be the sole element of an utterance” (= that it is a bound form), then Bonet’s statement is tautological.

Now it is sometimes suggested that such elements are “unpronounceable” by themselves: “In order to be pronounced, a formative (word, affix, etc.) needs to be part of an accentual unit” (Halpern 1998: 101). However, stress (or accent) is an abstract property that is very much dependent on the surrounding material. If a monosyllabic morph is an utterance by itself (e.g. English *here!*, or *yes*), the question of stress does not arise because a syllable can be unstressed only in relation to an adjacent syllable. “Lack of stress” is thus not a property that can lead to unpronounceability.

We sometimes observe that clitics are deficient in that they are subminimal, i.e. they have fewer segments or moras than a minimal free form needs to have. Some languages have nonsyllabic clitics, as seen in many cases above (e.g. Amharic =*mm* in (24), English 's, Italian *l'* in (15b), Tagalog =*ng* in (41a)), and some clitics have reduced or short vowels that are not sufficient for minimal free forms. Clitics with schwa [ə] were seen above in (10) (German *de, se*). These clitics could be said to be “unpronounceable in isolation”, but most clitics are not of this sort – clitics like those in (76) are phonologically perfectly complete and just happen to require the syntactic cooccurrence with another adjacent form.

(76) Russian:	<i>li</i>	polar question marker (ex. (2))
Greek:	<i>mas</i>	1st person plural adposessive index (ex. (57b))
Tagalog:	<i>siya</i>	3rd person pronoun (ex. (5))
German:	<i>als</i>	temporal subordinator (ex. (48a))

## 7.2. Stresslessness

Clitics are typically outside the stress domain of their anchor word (Section 4.2) and stressless, and they are sometimes defined as elements that lack stress (e.g. by Dixon 2007; see (64)). But we saw earlier that stress-integrated clitics may carry stress (e.g. Polish *nie wiem* ‘I do not know in Section 4.2), so some authors have added the specification that clitics do not carry *independent* stress. However, even that is not entirely true, as some languages have clitics that are inherently stressed, e.g. English *tóo* and German admonitive *já* (see also Section 8):

(77) *He found the house wonderful, and she liked it, tóo.*

(78) *Komm já rechtzeitig nach Hause!*  
 come ADM in.time to home  
 ‘(I admonish you that) you come home in time!’

Such stressed particles are not usually called *clitics* in the literature, but they fall under the definition in (1) as they are bound morphs and are not class-selective. They are focus and discourse particles and thus fall in the semantic range of forms that are often clitics. Similar reports of stressed clitics are occasionally found in the literature,



e.g. Lowe (2014) on accented clitics in Vedic Sanskrit (san; Indo-European, Indo-Iranic), and Aissen (2017) on a stressed deictic clitic in Tsotsil (tzo; Mayan, Core Mayan).

It should also be noted that not all languages have word stress or word accent, and that it is not even clear how to define *stress* in such a way that the notion can be applied to all languages (Hyman 2014). Stress or stresslessness is thus not suitable as a general criterion for identifying clitics.

### 7.3. Phonological wordhood

Clitics are often discussed in the context of phonological words (or “prosodic words”; see Hall 1999; Hildebrandt 2015), and it is often said that clitics “do not constitute independent prosodic words, and lean on adjacent lexical heads to form prosodic words” (Elordieta 2014: 19; see also Aikhenvald et al. 2020: 12). The term *host* is usually used for the adjacent form on which a clitic *leans*.<sup>22</sup>

However, there is a wide range of criteria that have been used to identify phonological words, and it has been found that these criteria do not always give the same results even within one language. For example, stress domains and vowel harmony domains are different in Turkish (see (54) above), and in German, the criteria of coordination deletability and vocalic minimality conflict in the case of the diminutive suffix *-chen* (Hall 1999: 18). More generally, languages often show conflicting criteria for phonological domains and wordhood (Bickel et al. 2009; Tallman 2020). It should also be noted that phonological words are widely agreed to be partially isomorphic with morphosyntactic words, but in the absence of any kind of agreement on how to identify morphosyntactic words, it is quite impossible to identify phonological words in general.<sup>23</sup> Thus, the notion of phonological wordhood cannot be said to be well-established, despite its relative popularity since the 1990s.

---

<sup>22</sup> The term *host* was first used by Zwicky (1977). Recall from Section 3.1 above that it is not always clear which form a clitic is associated with prosodically, so in this paper, I use the term *anchor* for the word that is adjacent to a proclitic or an enclitic.

<sup>23</sup> This problem is briefly mentioned by Newell et al. (2017: 2), though without drawing any conclusions from it: “Phonologists can [...] give us some information about word domain. Phonology per se, however, lacks a theory of how the ‘word’ comes to be, and phonologists generally look to morphologists or syntacticians to derive this construct. The latter two groups, however, don’t know, and are often content with the fact that phonologists, at least, can tell them that something is a word, when it is.”

Moreover, even if it were clear how we identify phonological words, this would not be sufficient to identify clitics, because clitics could be related to phonological words in three different ways: (i) they could be integrated into the same phonological word ( $\omega$ ) as their host; (ii) they could form a recursive phonological word that also includes the host as an embedded phonological word; or (iii) they could be adjoined to their host and form a phonological phrase ( $\varphi$ ) with it. These three possibilities are illustrated in (79)-(81). It should be noted that these are hypothetical descriptions, and there is no consensus in the literature about any of these analyses.

(79) Integration: Dutch (Indo-European, Germanic; Booij 1996: 219)

*Jan kocht het boek.*  
 (jan) $_{\omega}$  (kɔx-tət) $_{\omega}$  (buk) $_{\omega}$   
 Jan bought the book  
 'Jan bought the book.'

(80) Recursive phonological word: English (Indo-European, Germanic; Selkirk 1995: 458) (see also fn. 3)

*need 'm*  
 ((nid) $_{\omega}$  əm) $_{\omega}$   
 'need him'

(81) Adjoined to host, forming a phonological phrase: Spanish (Indo-European, Italic; Elordieta 2014: 31)<sup>24</sup>

*leyé-ndo-te-la*  
 ((lejendo) $_{\omega}$  tela) $_{\varphi}$   
 read-GER-2SG-3SG.F  
 'reading it to you'

Quite similar options exist for affixes, which may be COHERING (integrated) or NON-COHERING (adjoined; e.g. Dixon 2020; Raffelsiefen 2023), so phonological wordhood does not seem to be helpful in distinguishing clitics from affixes. Moreover, the motivations for these prosodic analyses are very diverse and cannot be easily

---

<sup>24</sup> We will see in Section 9 below that the Spanish object indexes are affixes, not clitics, but in the literature on prosodic domains, they have often been treated as clitics.

generalized across languages. Some authors have highlighted the special situation of “backward-leaning” proclitics like Dutch *het* in *het boek* ‘the book’ in (79),<sup>25</sup> and such elements have even been called “enclitics” because of their phonological properties. However, as these phonological properties are not uniform, they cannot be the basis for the demarcation of clitics from affixes, or for the distinction between proclitics and enclitics (as we already saw in Section 3.1 for segmental effects).

Clitics and affixes are already distinguished from other forms by being neither free forms nor roots, so it appears that their phonological properties are not needed to single them out. Intuitively, many linguists feel that phonological dependence is part of the nature of clitics, but as this notion is vague and cannot be applied without many additional assumptions, it is better to rely on boundness and on the distinction between roots and nonroots.<sup>26</sup>

## **8. Clitics as concrete forms (morphs)**

In the definition proposed here, a clitic is a morph, i.e. a concrete continuous segmental form. This means that there can be no *zero clitics*, that clitics cannot be tonal or otherwise non-segmental (Section 8.1), and that morphs cannot occur inside other morphs, so that there can be no *intraclitics* or *endoclitics* (Section 8.2).

### **8.1. There are no “tonal clitics” or “accentual clitics”**

While suprasegmental effects such as stress and tone often show similarities with grammatical forms, they are not forms themselves. Forms are continuous segment sequences, which excludes the possibility of “tonal morphs” (Haspelmath 2020: Section 4). This also means that there can be no tonal clitics, as has occasionally been suggested (e.g. Van de Velde 2009). There cannot be accentual clitics either, as was sometimes discussed for Tongan (ton; Austronesian, Malayo-Polynesian; e.g. Anderson 2005: 94–101). In this language, definiteness is marked by a stress shift to

---

<sup>25</sup> See Cysouw (2005) on “ditropic clitics” (“backward-leaning” proclitics and “forward-leaning” enclitics), and Anderson (2005: Section 2.2) on “backward-leaning” proclitic determiners in Kwakwaka (kwk; Wakashan, Northern Wakashan).

<sup>26</sup> The absence of phonological definitional criteria makes it possible to state generalizations about the phonological properties of clitics, e.g. that they are overwhelmingly stressless. This is not so by definition, but it can be treated as a testable empirical claim.

the final mora of the nominal, regardless of which word occurs at the end. This accentual marking shares the property of nonselectivity with clitics, but it cannot be a “processual special clitic” (Anderson 2005: 95) if a clitic is defined as a kind of morph.<sup>27</sup> Similarly, saying that “stress is a proclitic” in Modern Greek (van Oostendorp 2012) is not compatible with the definition of a clitic as a kind of morph.

## 8.2. There are no “intraclitics” or “endoclitics”

Some authors have suggested that languages may have *intraclitics*, i.e. clitics that occur between two morphs of a word-form, or *endoclitics*, i.e. clitics that occur inside a root, just as infixes are often thought of as affixes that occur inside a root. However, if we adopt the definition of *clitic* in (1) and the definition of *affix* in Haspelmath (2021), this is not possible. An affix cannot occur outside of a clitic (because affixes by definition occur next to roots or affixes), and a root cannot be “broken up” by an “infix” or an “endoclitic” (because roots by definition are segment sequences).

An example of a Russian intraclitic might be the preposition *v* = in the reciprocal construction in (82), and an example of an Andi intraclitic might be the additive marker =*lo* in (83). The supposed morph-internal status is shown by the angle brackets in *-do⟨lo⟩sub*.

(82) Russian (Indo-European, Balto-Slavic; Arkadiev 2016: 331)

*Oni razočarovalis' drug v drug-e.*  
 they were.disappointed each in other-LOC  
 ‘They were disappointed by each other.’

(83) Andi (Nakh-Daghestanian, Daghestanian; Maisak 2021: 21)

*Men ru-⟨lo⟩sub, qwar-⟨lo⟩sub.*  
 you say-PROH⟨ADD⟩ write-PROH⟨ADD⟩  
 ‘Neither talk, nor write!’

---

<sup>27</sup> One could imagine a definition of clitic that includes tonal or processual or other kinds of abstract elements, but I would not know how to do this. Nonsegmental effects are based on certain correspondences between two forms, which are often treated as operations or transformations, and they cannot be readily compared with segmental forms.

Arkadiev (2016) and Maisak (2021) regard these forms as *intraclitics*,<sup>28</sup> but this is not possible in the current conceptual framework. If the preposition *v* ‘in’ is not a prefix but a proclitic (Arkadiev 2016: 327), then the first element *drug* in (82) cannot be a prefix or compound member, but must be something else. Affixes cannot stand outside a clitic, and compounds cannot have a clitic inside them either, by definition. This means that the Andi additive marker *=lo* is an enclitic, and so is the element *=s:ub* that follows it in (83). The gloss is thus [say-PROH1 =ADD= PROH2] rather than [say-PROH<ADD>], i.e. there are two prohibitive markers (a suffix preceding the enclitic *=lo*, and an enclitic following the enclitic).

The best-known case of a supposed root-internal endoclititic has been reported from Udi (another Nakh-Daghestanian language), for which Harris (2000, 2002) provides extensive documentation and discussion. She regards bound person forms like *=z=* in (84) as endoclititics because she treats verbs like *a...-qʔ-* ‘receive’ as single roots (*aqʔ-*).

(84) Udi (Nakh-Daghestanian, Daghestanian; Harris 2000: 598)

<i>Kayuz-ax</i>	<i>a-z-qʔ-e.</i>
letter-ACC	receive1-1SG-receive2-AOR
‘I received the letter.’	

What is surely unusual about Udi is that there are quite a few short bipartite verbs like *a-qʔ-* ‘receive’ (Harris lists 27 such verbs), but they must be treated as consisting of two different morphs, and thus somewhat analogous to English bipartite verbs like *take part* or *make headway*. The proper gloss of (84) is thus [receive1 = 1SG = receive2-AOR2], showing that the person index is a clitic, preceded by the first morph of the verbal expression and followed by its second morph.

When I say here that the Udi *endoclititics* are not clitics, I am not making a substantive claim. According to the definition in (1), these elements cannot be clitics, because a clitic is a type of form, and root-internal segment sequences cannot be forms. As I noted, this follows from the definition of a FORM (Haspelmath 2020: Section 4), and of a ROOT: a form is a sequence of segments that has a linguistic

---

<sup>28</sup> They actually call them “endoclititics” and do not make the distinction between *intraclitics* (between two morphs) and *endoclititics* (inside a root) that I take from Plungian (2000).

function,<sup>29</sup> and a root is a minimal form that denotes an object, an action or property. While one can imagine that a root could be “broken up” by some material that occurs “inside” it, this would not be in line with the definition of a root as a kind of minimal form. “Breaking up” a root is an abstract operation, similar to operations “deleting” or “transforming” forms, or movement operations, or zero elements. Such abstract operations and elements are often useful for language-particular analysis, but they cannot be used in comparative concepts.<sup>30</sup> For example, if we were to relax the definition of a form as a sequence of segments, then all kinds of non-continuous sets of elements could be said to constitute a single form (e.g. English *take ... part*). The continuity requirement is thus crucial and fundamental to our general concepts of grammar.

### 9. Romance object “clitics” as affixes

So far in this paper, I have hardly touched upon object indexes in the Romance languages, even though these kinds of elements are more prominent in the literature on *clitics* than any other type. I left the discussion of these forms until the end because they are not clitics, but affixes. Consider examples such as (85)-(88) (examples from Spanish were cited above in (8) and (81)).

(85) French (Indo-European, Italic)

*Mon frère la connaît.*  
 my brother her knows  
 ‘My brother knows her.’

(86) Italian (Indo-European, Italic; Monachesi 2005: 55)

*Martina te lo spedirà.*  
 Martina you.DAT it.ACC send.FUT.3SG  
 ‘Martina will send it to you.’

---

<sup>29</sup> Since a form is a sequence of segments, “circumfixes” and “circumclitics” cannot be types of forms. One may talk about a “circumfixing construction” (one that includes a prefix and a suffix), or about a “circumcliticizing construction” (one with a proclitic and an enclitic occurring simultaneously), but these constructions must contain two forms.

<sup>30</sup> The reason for this is that comparative concepts must be defined in the same way in all languages (as noted in fn. 27). This is generally impossible for abstract operations. Languages can be readily compared in terms of their forms, but not in terms of their abstract operations and elements.

(87) (European) Portuguese (Indo-European, Italic; Luís & Kaiser 2016: 215, 217)

- a. *Ontem chamou-me.*  
yesterday she.called-me  
'Yesterday she called me.'
- b. *Porque me chamou?*  
why me she.called  
'Why did she call me?'

(88) Romanian (Indo-European, Italic; Monachesi 2005: 44)

- Mihai nu-l așteaptă.*  
Mihai not-him waits  
'Mihai doesn't wait for him.'

These elements are not clitics according to the definition in (1) because they are bound forms that always occur on the verb, whether preverbally or postverbally.<sup>31</sup> That they are affixes rather than clitics is actually fairly widely accepted in the literature (Miller & Sag 1997; Luís 2004; Monachesi 2005: Section 3.3; Bermúdez-Otero & Payne 2011).

Authors who argued for affixal status of the Romance object indexes have typically adduced Zwicky & Pullum's (1983) diagnostic symptoms, pointing out that they occur in rigid clusters, that they sometimes show arbitrary gaps and idiosyncratic phonological behaviour, and that they tend to disallow wide scope over coordination:

(89) French (Indo-European, Italic; Miller & Sag 1997: 7)

- \*Pierre les voit et écoute.* (OK: *Pierre les voit et les écoute.*)  
Pierre them sees and hears  
'Pierre sees and hears them.'

---

<sup>31</sup> It may be unexpected to see mobile bound morphs treated as affixes, but mobile affixes are not unprecedented (e.g. Bickel et al. 2007: 43; Ryan 2010; Jenks & Rose 2015). If elements which always occur on the same type of root but show some mobility were not treated as affixes, this would have to be specified in the definition of "affix", and this definition would need to become still more complex (see Haspelmath 2021: 19).

In the present context, these additional properties of object indexes play no role, because there is only one criterion of cliticness: nonselectivity. Rigid positions in clusters are of course attested in clitics (see Section 3.5), and so is idiosyncratic phonological behaviour (see Section 4.1). An arbitrary gap is also attested in the English Genitive clitic (which does not occur after plural *-s*: *the girls'(\*s) party*, Zwicky 1987), and wide scope in coordination is sometimes even attested with derivational suffixes, so it can hardly be criterial for the clitic/affix distinction.<sup>32</sup> However, by the criterion of word-class selectivity, the object indexes are affixes, so the present conclusion conforms to that reached by Miller & Sag and those following them.

Of course, the Romance object indexes derive from personal pronouns whose position was freer in earlier times, and in medieval texts, they were not always verb-adjacent. Thus, this is a clear instance of a diachronic development from clitics to affixes, and the peculiar distribution of postverbal and preverbal object indexes in European Portuguese is a remnant of this earlier clitic stage. The situation in Modern Greek is quite similar: As we saw in (71a) above, the object person indexes, which have often been called *clitics*, are actually affixes.

## 10. Conclusion: clitics are not intermediate between words and affixes

We have seen a variety of different types of clitics in this paper, as well as a variety of different properties that are found in clitics. I showed that they can all be subsumed under the simple definition in (1) (a clitic is a bound morph that is neither an affix nor a root), but I did not claim that this definition says anything deep about their nature. It is merely a definition, after all. But it is simple and clear, and it has sharp boundaries rather than merely specifying a canon or a prototype.

The definition may seem to be broader than has often been implied, e.g. by including adpositions and subordinators (Section 2.6-7), which have not often been regarded as clitics. However, it is unclear why they should be excluded, and it may be a historical accident that they did not become prominent in the literature on clitics. The result is that most function words are clitics (Section 2.1), but only those that cannot occur on their own (= that are bound forms). Closed-class function words

---

<sup>32</sup> Erdal (2007: 178) cites the following example from Turkish, where the “professional” suffix *-cı* has scope over two nouns (i.e. allows “suspended affixation”): *kum- ve çakıl-cı geldi* [sand- and gravel-PROF came] ‘the supplier of sand and gravel has come’. Nobody would suggest that it is a clitic.



such as demonstratives, auxiliaries and response words ('yes') can often be used in isolation and are therefore not clitics.

Much work on clitics over the last few decades is motivated by the hope of explaining the behaviour of (certain kinds of) clitics by appealing to certain kinds of architectures or rule types, such as lexical vs. postlexical rules (e.g. Halpern 1995; Anderson 2005). However, no particular proposal has been widely accepted, and it appears that the possibilities of an *architectural* approach have been exhausted. Prominent authors like Zwicky (1994) and Spencer & Luís (2012) have suggested that *clitic* is no more than a name for a problem: a label for a range of linguistic expressions that do not fit readily into other classes, not a name of a theoretical construct, and not a name for a "unified class of phenomena" (Zwicky 1994: xiii).

In this paper, by contrast, I do give a definition which by its nature singles out a unified class of phenomena, and this allows the term *clitic* to be more than a name for a problem: It is a comparative concept that helps us to compare languages with respect to phenomena that we find interesting without talking past each other. But since *clitic* is defined as a comparative concept, there is no claim that it carves out part of the underlying reality of languages: Like other terms for comparative concepts, it is a METHODOLOGICAL TOOL, not a "theoretical construct". To the extent that the term allows us to formulate testable claims about the world's languages, and to the extent that these claims are supported, we will have found valuable cross-linguistic generalizations, but it may still be unclear how we can explain these generalizations.<sup>33</sup>

In addition to the architectural approach, a popular view has been that clitics are in some way intermediate between free words and affixes. Zwicky (1977: 1) initially characterized them as "presenting analytic difficulties because they are neither clearly independent words nor clearly affixes", and Nevis (2000: 389) even suggested that a form is a clitic "to the extent that it deviates from the accepted properties of affixes or words".<sup>34</sup> But just as *clitic* is not more than a comparative concept with some

---

<sup>33</sup> I do not actually expect to find robust generalizations that crucially rely on the clitic vs. affix distinction, but as we need to know what a clitic is in order to distinguish words from non-words (see Haspelmath 2023), this definition is very important for all works that make claims about words. Again, it may be that the most robust generalizations will eventually be shown to involve form classes other than words, but the 'word' concept is so central to linguistics that it is good to have a clear definition of it with sharp boundaries.

<sup>34</sup> In a non-serious mode, Sadock (1995: 260) suggested that a clitic could be defined as "an element whose distribution linguists cannot comfortably consign to a single grammatical component".

usefulness for linguists, the familiar affix vs. word distinction (and the morphology vs. syntax division in grammar) could be largely based on the orthographic word. The supposed “analytic difficulties” of clitics would then reflect the difficulties of deciding how to write them (jointly, or separately, or with a hyphen or other boundary symbol). The definition of *affix* is actually much more complex than the definition of *clitic* (once a definition of *affix* is in place), as can be seen in Haspelmath (2021), and defining *word* is not straightforward either (see Haspelmath 2023). But in whatever way we end up defining these terms, the definitions are unlikely to give us deep insights into their nature.

Clitics could be “intermediate” between free words and affixes if there were a single dimension along which they vary, a kind of “scale of coalescence”, or “tightness of bonding”. It has often been suggested that there is such a continuous scale, with a diachronic counterpart in grammaticalization (Hopper & Traugott 2003: 142):

(90) the coalescence scale

free word > clitic > affix

But no systematic way of quantifying the degrees on a scale of coalescence or of tightness of bonding has been suggested, and linguists have mostly relied on their intuitions of what constitutes *tight* or *loose* attachment. In view of the great variety of phenomena that have been cited as diagnostics, we cannot conclude that there is sufficient evidence that the scale is real. Aikhenvald (2002: 42) says that applying a wide range of criteria “suggests a scalar, or continuum-type approach – that is, some morphemes turn out to be more affix-like and others to be more word-like”. But Börjars & Harries (2008) have rightly emphasized that the different dimensions of variation need not correlate with each other, and van Gijn & Zúñiga (2014: 155) make this very concrete: They examine four such dimensions (phonological integration, rigid position, syntactic weight, and lexical class) for twelve morph types from different languages, and they do not find a clear clustering of the dimensions. There is no reason to think that there is a single scale or continuum.

There are thus many open questions that need to be addressed by future research, but I hope that by providing simple and clear definitions of terms such as *affix* (Haspelmath 2021) and *clitic* (definition (1) in Section 1), this research will be facilitated. The definitions do not answer any theoretical questions, but it should have

become clear that it is possible to have such definitions even without answers to our broader questions.

## Acknowledgements

For useful comments on an earlier version of this paper, I am particularly indebted to Peter Arkadiev, Irenäus Kulik and Tim Zingler, but also to quite a few commentators on Academia.edu. In addition, I thank two anonymous reviewers and the editors.

## Abbreviations

1 = 1 <sup>st</sup> person	DUR = durative	NONFUT = non-future
2 = 2 <sup>nd</sup> person	ERG = ergative	NONMASC = non-masculine
3 = 3 <sup>rd</sup> person	EMPH = emphatic	OBJ = object
ADM = admonitive	EXCL = exclusive	PL = plural
ACC = accusative	EZ = ezafe	POSS = possessive
ADD = additive	F = feminine	PQ = polar question
ALL = allative	FOC = focus	PRF = perfect
AOR = aorist	FUT = future	PROH = prohibitive
ART = article	G4 = gender 4	PRS = present
AUX = auxiliary	GEN = genitive	PST = past
AV = actor voice	GER = gerund	PTCP = participle
COMP = complementizer	HORT = hortative	QM = question marker
COMPL = completive	IMPF = imperfective	REFL = reflexive
CONJ = conjunction	INCL = inclusive	REL = relative
CONN = connective	IND = indicative	RES = resultative
COP = copula	INS = instrumental	SBJ = subject
DAT = dative	IRR = irrealis	SBJV = subjunctive
DECL = declarative	LNK = linker	SG = singular
DEF = definite	LOC = locative	TOP = topic
DET = determiner	M = masculine	TR = transitive
DIR = direct Form	NEG = negation	
DU = dual	NOM = nominative	

## References

- Aikhenvald, Alexandra Y. 2002. Typological parameters of clitics, with special reference to Tariana. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A cross-linguistic typology*, 42–78. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra Y., R. M. W. Dixon & Nathan M. White. 2020. The essence of “word”. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Nathan M. White (eds.), *Phonological word and grammatical word: A cross-linguistic typology*, 1–24. Oxford: Oxford University Press.
- Aissen, Judith. 2017. Special clitics and the right periphery in Tsotsil. In Claire Bowerman, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond): Structures, relations, analyses*, 235–262. Berlin: Language Science Press. (DOI:10.5281/zenodo.495449)
- Allison, Sean. 2020. The notion of “word” in Makary Kotoko. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Nathan M. White (eds.), *Phonological word and grammatical word: A cross-linguistic typology*, 260–284. Oxford: Oxford University Press.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. New York: Oxford University Press.
- Anderson, Stephen R. 2011. Clitics. In Marc Van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*. Wiley Online Library: Wiley. (<https://doi.org/10.1002/9781444335262.wbctp0084>)
- Arcodia, Giorgio F. 2012. *Lexical derivation in Mandarin Chinese*. Taipei: Crane.
- Arkadiev, Peter M. & Yu A. Lander. 2020. Ambifiksy i drugie zveri [Амбификсы и другие звери] (Ambifixes and other beasts). In Andrej A. Kibrik, Ks. P. Semënova, D.V. Sičinava, Sergei G. Tatevosov, A. Ju. Urmančieva (eds.), *VAProsy jazykoznanija: Megasbornik nanostatej*, 35–42. Moskva: Buki Vedi.
- Arkadiev, Peter M. 2016. K voprosu ob èndoklitikax v russkom jazyke [On endoclitics in Russian]. In Anton V. Zimmerling & Ekaterina A. Ljutikova (eds.), *Arxitektura klauzy v parametričeskix modeljax: Sintaksis, informacionnaja struktura, porjadok slov*, 325–331. Moskva: Jazyki slavjanskix kul'tur.
- Avgustinova, Tania. 1994. On Bulgarian verbal clitics. *Journal of Slavic Linguistics* 2(1). 29–47.

- Berghäll, Liisa. 2015. *A grammar of Mauwake*. Berlin: Language Science Press. (<http://langsci-press.org/catalog/book/67>)
- Bermúdez-Otero, Ricardo & John Payne. 2011. There are no special clitics. In Alexandra Galani, Glyn Hicks & George Tsoulas (eds.), *Morphology and its interfaces*, 57–69. Amsterdam: John Benjamins.
- Bickel, Balthasar, Goma Banjade, Martin Gaenszle, Elena Lieven, Netra P. Paudyal, Ichchha P. Rai, Manoj Rai, Novel Kishore Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83(1). 43–73.
- Bickel, Balthasar, Kristine Hildebrandt & René Schiering. 2009. The distribution of phonological word domains: A probabilistic typology. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological domains: Universals and deviations*, 47–75. Berlin: Mouton de Gruyter.
- Bloomfield, Leonard. 1933. *Language*. New York: H. Holt and Company.
- Bonet, Eulalia. 2019. Clitics and clitic clusters in morphology. *Oxford Research Encyclopedia of Linguistics*. (doi:10.1093/acrefore/9780199384655.013.519)
- Booij, Geert. 1996. Cliticization as prosodic integration: The case of Dutch. *The Linguistic Review* 13(3–4). 219–242. (doi:10.1515/tlir.1996.13.3-4.219)
- Booij, Geert E. 2012. *The grammar of words: An introduction to linguistic morphology (3rd edition)*. Oxford: Oxford University Press.
- Börjars, Kersti & Pauline Harries. 2008. The clitic-affix distinction, historical change, and Scandinavian bound definiteness marking. *Journal of Germanic Linguistics* 20(4). 289–350. (doi:10.1017/S1470542708000068)
- Bošković, Željko. 2016. Second-position clitics cross-linguistically. In Franc Lanko Marušič & Rok Žaucer (eds.), *Formal studies in Slovenian Syntax: In honor of Janez Orešnik*, 23–53. Amsterdam: John Benjamins.
- Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1). 1–44.
- Chae, Hee-Rahk. 2020. *Korean morphosyntax: Focusing on clitics and their roles in syntax*. Abingdon: Routledge.
- Croft, William. 2010. Ten unwarranted assumptions in syntactic argumentation. In Kasper Boye, & Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, 313–350. Berlin: Mouton de Gruyter.
- Cysouw, Michael. 2005. Morphology in the wrong place: A survey of preposed enclitics. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Rainer,

- Franz (eds.), *Morphology and its demarcations: Selected papers from the 11th Morphology Meeting, Vienna, February 2004*, 17–37. Amsterdam: John Benjamins.
- Demeke, Girma A. & Ronny Meyer. 2008. The enclitic *-mm* in Amharic: reassessment of a multifunctional morpheme. *Linguistics* 46(3). 607–628. (doi:10.1515/LING.2008.020)
- Dixon, R. M. W. 2007. Clitics in English. *English Studies* 88(5). 574–600. (doi:10.1080/00138380701566102)
- Dixon, R. M. W. 2020. Words within words: Examples from Yidiñ, Jarawara, and Fijian. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Nathan M. White (eds.), *Phonological word and grammatical word: A cross-linguistic typology*, 25–38. Oxford: Oxford University Press.
- Dryer, Matthew S. 2005. Polar questions. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 470–473. Oxford: Oxford University Press.
- Elordieta, Gorra. 2014. The word in phonology. In Iraide Ibarretxe-Antuñano & José-Luis Mendivil-Giró (eds.), *To be or not to be a word: New reflections on the definition of word*, 6–65. Cambridge: Cambridge Scholars Publishing.
- Erdal, Marcel. 2000. Clitics in Turkish. In Aslı Göksel & Celia Kerslake (eds.), *Studies on Turkish and Turkic Languages: Proceedings of the Ninth International Conference on Turkish Linguistics, Lincoln College, Oxford, August 12-14, 1998*, 41–48. Wiesbaden: Harrassowitz.
- Erdal, Marcel. 2007. Group inflexion, morphological ellipsis, affix suspension, clitic sharing. In M.M. Jocelyne Fernandez-Vest (ed.), *Combats pour les langues du monde: Hommage à Claude Hagège*, 177–189. Paris: L'Harmattan.
- Fattier, Dominique. 2013. Haitian Creole structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (<https://apics-online.info/contributions/49>)
- Fehn, Anne-Maria. 2016. *A grammar of Ts'ixa (Kalahari Khoe)*. Köln: Universität zu Köln. (Doctoral Dissertation.) (<http://kups.ub.uni-koeln.de/7062/>)
- Fortescue, Michael. 1984. *West Greenlandic*. London: Croom Helm.
- Gary, Judith Olmsted & Saad Gamal-Eldin. 1982. *Cairene Egyptian Colloquial Arabic*. Amsterdam: North Holland.
- Gerds, Donna B. & Adam Werle. 2014. Halkomelem clitic types. *Morphology* 24(3). 245–281. (doi:10.1007/s11525-014-9241-0)

- Göksel, Aslı & Celia Kerslake. 2005. *Turkish: An essential grammar*. London: Routledge.
- Gunnink, Hilde. 2022. *A grammar of Fwe*. Berlin: Language Science Press.
- Hall, Tracy A. 1999. The phonological word: A review. In Tracy A. Hall & Ursula Kleinhenz (eds.), *Studies on the phonological word*, 1–22. Amsterdam: John Benjamins.
- Halpern, Aaron. 1995. *On the placement and morphology of clitics*. Stanford: CSLI Publications.
- Halpern, Aaron. 1998. Clitics. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 101–122. Oxford: Blackwell.
- Hannahs, S. J. & Maggie Tallerman. 2006. At the interface: Selection of the Welsh definite article. *Linguistics* 44(4). 781–816. (doi:10.1515/LING.2006.025)
- Harris, Alice C. 2000. Where in the word is the Udi clitic? *Language* 76(3). 593–616.
- Harris, Alice C. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.
- Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntax of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 197–226. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2015. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In Joanna Błaszczak, Dorota Klimek-Jankowska & Krzysztof Migdalski (eds.), *How categorical are categories? New approaches to the old questions of noun, verb, and adjective*, 273–304. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2019. Indexing and flagging, and head and dependent marking. *Te Reo* 62(1). 93–115. (doi:10.17617/2.3168042)
- Haspelmath, Martin. 2020. The morph as a minimal linguistic form. *Morphology* 30(2). 117–134. (doi:10.1007/s11525-020-09355-5)
- 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. 2023. Defining the word. *WORD* 69(3). 283–297. (doi:10.1080/00437956.2023.2237272)
- Hildebrandt, Kristine A. 2015. The prosodic word. In John R. Taylor (ed.), *The Oxford handbook of the word*, 221–245. Oxford: Oxford University Press.
- Himmelman, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960.

- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization: 2nd edition*. Cambridge: Cambridge University Press.
- Hosokawa, Komei. 1991. *The Yawuru Language of West Kimberley. A Meaning-Based Description*. Canberra: Australian National University.
- Hyman, Larry M. 2014. Do all languages have word accent? In Harry Van der Hulst (ed.), *Word Stress: Theoretical and typological issues*, 56–82. Cambridge: Cambridge University Press.
- Ionova, A. 2019. *The unbearable lightness of clitics*. Leiden: Leiden University-LOT. (Doctoral Dissertation.) (<https://hdl.handle.net/1887/83258>)
- Jenks, Peter & Rose, Sharon. 2015. Mobile object markers in Moro: The role of tone. *Language* 91(2). 269–307. (doi:10.1353/lan.2015.0022)
- Joseph, Brian D. 2002. Defining “word” in Modern Greek: A response to Philippaki-Warbuton & Spyropoulos 1999. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001*, 87–114. Dordrecht: Springer Netherlands. (doi:10.1007/978-94-017-3726-5\_3)
- Kaufman, Daniel. 2010. *The morphosyntax of Tagalog clitics: a typologically driven approach*. Ithaca: Cornell University. (Doctoral Dissertation.)
- Kayne, Richard S. 1975. *French syntax*. Cambridge: MIT Press.
- Kazenin, Konstantin I. 2002. Focus in Daghestanian and word order typology. *Linguistic Typology* 6(3). 289–316. (doi:10.1515/lity.2003.002)
- Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61(1). 95–120. (doi:10.2307/413422)
- Konnerth, Linda. 2020. *A grammar of Karbi*. Berlin: Mouton de Gruyter.
- Lowe, John. 2014. Accented clitics in the Ṛgveda. *Transactions of the Philological Society* 112(1). 5–43. (doi:10.1111/1467-968X.12013)
- Luís, Ana R. 2004. *Clitics as morphology*. Colchester: University of Essex. (Doctoral Dissertation.)
- Luís, Ana R. & Georg A. Kaiser. 2016. Clitic pronouns: Phonology, morphology, and syntax. In W. Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The Handbook of Portuguese Linguistics*, 210–233. Wiley Online Library: Wiley. (doi:10.1002/9781118791844.ch12)
- Maisak, Timur. 2021. Endoclitics in Andi. *Folia Linguistica* 55(1). 1–34. (doi:10.1515/flin-2020-2069)



- Miller, Philip H. & Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language & Linguistic Theory* 15(3). 573–639. (doi:10.1023/A:1005815413834)
- Monachesi, Paola. 2005. *The verbal complex in Romance: A case study in grammatical interfaces*. Oxford: Oxford University Press.
- Mushin, Ilana. 2012. *A grammar of Garrwa*. Berlin: Mouton de Gruyter.
- Nakamura, Wataru. 2018. Case. In Yōko Hasegawa (ed.), *The Cambridge handbook of Japanese linguistics*, 249–275. Cambridge: Cambridge University Press.
- Nevis, Joel A. 2000. Clitics. In Geert E. Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphology: An international handbook on inflection and word-formation (Volume 1)*, 388–404. Berlin: Mouton de Gruyter. (<https://doi.org/10.1515/9783110111286.1.5.360>)
- Newell, Heather, Máire Noonan, Glyne Piggott, & Lisa deMena Travis (eds.). 2017. *The structure of words at the interfaces*. Oxford: Oxford University Press.
- Nida, Eugene A. 1946. *Morphology: The descriptive analysis of words (1st edition)*. Ann Arbor: University of Michigan Press.
- Nordhoff, Sebastian. 2009. *A grammar of Upcountry Sri Lanka Malay*. Amsterdam: University of Amsterdam-LOT. (Doctoral Dissertation.)
- Nordlinger, Rachel. 1998. *A grammar of Wambaya, Northern Australia*. Canberra: Pacific Linguistics. (<http://dx.doi.org/10.15144/PL-C140>)
- Parker, Steve. 1999. On the behavior of definite articles in Chamicuro. *Language* 75(3). 552–562. (doi:10.2307/417060)
- Pescarini, Diego. 2021. *Romance object clitics: Microvariation and linguistic change*. Oxford: Oxford University Press.
- Plank, Frans. 2005. The prosodic contribution of clitics: Focus on Latin. *Lingue e Linguaggio* 4(2). 281–292. (doi:10.1418/20726)
- Plungian, Vladimir A. 2000. *Obščaja morfologija: Vvedenie v problematiku*. Moskva: URSS.
- Raffelsiefen, Renate. 2023. Morpho-phonological asymmetries in affixation. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet & Antonio Fábregas (eds.), *The Wiley Blackwell companion to morphology*. Wiley Online Library: Wiley. (<https://doi.org/10.1002/9781119693604.morphcom050>)
- Rubach, Jerzy & Booij, Geert E. 1985. A grid theory of stress in Polish. *Lingua* 66(4). 281–320. (doi:10.1016/0024-3841(85)90032-4)

- Ryan, Kevin M. 2010. Variable affix order: Grammar and learning. *Language* 86(4). 758–791. (doi:10.1353/lan.2010.0032)
- Sadock, Jerrold M. 1995. A multi-hierarchy view of clitics. *Chicago Linguistic Society* 31(2). 258–279.
- Samvelian, Pollet & Jesse Tseng. 2010. Persian object clitics and the syntax-morphology interface. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*. Université Paris Diderot, Paris 7, France, 212–232. Stanford: CSLI.
- Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics* 43(3). 605–645.
- Schapper, Antoinette. 2022. *A grammar of Bunaq*. Berlin: Mouton de Gruyter.
- Selkirk, Elisabeth O. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, 439–469. Amherst: Graduate Linguistic Student Association (GLSA).
- Smith, Ian & Steve Johnson. 1985. The syntax of clitic cross-referencing pronouns in Kugu Nganhcara. *Anthropological Linguistics* 27(1). 102–111.
- Spencer, Andrew & Ana R. Luís. 2012. *Clitics*. Cambridge: Cambridge University Press.
- Stebbins, Tonya. 2003. On the status of intermediate form classes: Words, clitics, and affixes in Smalgyax (Coast Tsimshian). *Linguistic Typology* 7(3). 383–415. (doi:10.1515/lity.2003.019)
- Stockwell, Robert P., Jean Donald Bowen & John Watson Martin. 1965. *The grammatical structures of English and Spanish*. Chicago: University of Chicago Press.
- Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14(2). e12364. (doi:10.1111/lnc3.12364)
- Taylor, John R. 1995. *Linguistic categorization: Prototypes in linguistic theory (2nd edition)*. Oxford: Clarendon Press.
- Toman, Jindřich. 1996. A note on clitics and prosody. In Aaron Halpern & Arnold M. Zwicky (eds.), *Approaching second: Second position clitics and related phenomena*, 505–510. Stanford: CSLI Publications.
- Van de Velde, Mark. 2009. Eton tonology and morphosyntax: A holistic typological approach. In Patience L. Epps & Alexandre Arkhipov (eds.), *New challenges in typology: Transcending the borders and refining the distinctions*, 35–60. Berlin: Mouton de Gruyter.
- van Gijn, Rik & Fernando Zúñiga. 2014. Word and the Americanist perspective. *Morphology* 24(3). 135–160. (doi:10.1007/s11525-014-9242-z)

- van Oostendorp, Marc. 2012. Stress as a proclitic in Modern Greek. *Lingua* 122(11). 1165–1181. (doi:10.1016/j.lingua.2012.05.006)
- Wackernagel, Jacob. 2020 [1892]. *On a law of Indo-European word order*. Berlin: Language Science Press. (Ed. Walkden, George & Sevdali, Christina & Macleod, Morgan.) (doi:10.5281/zenodo.3978908)
- Walkden, George. 2020. Introduction. In George Walkden, Christina Sevdali & Morgan Macleod (eds.), *Jacon Wackernagel's "On a law of Indo-European word order,"* 3–19. Berlin: Language Science Press. (doi:10.5281/zenodo.3978908)
- Wolfart, H. Christoph. 1996. Sketch of Cree, an Algonquian Language. In Ives Goddard (ed.), *Handbook of North American Indians, Vol. 17: Languages*. Washington: Smithsonian Institute.
- Zingler, Tim. 2020. *Wordhood issues: Typology and grammaticalization*. Albuquerque: University of New Mexico. (Doctoral Dissertation.) ([https://digitalrepository.unm.edu/ling\\_etds/71](https://digitalrepository.unm.edu/ling_etds/71))
- Zingler, Tim. 2022. Clitics, anti-clitics, and weak words: Towards a typology of prosodic and syntagmatic dependence. *Language and Linguistics Compass* 16(5–6). e12453. (doi:10.1111/lnc3.12453)
- Zúñiga, Fernando. 2014. (Anti-)cliticization in Mapudungun. *Morphology* 24(3). 161–175. (doi:10.1007/s11525-014-9244-x)
- Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 59(3). 502–513.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington: Indiana University Linguistics Club. (doi:10.5281/zenodo.7436775)
- Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23(1). 133–148.
- Zwicky, Arnold M. 1994. What is a clitic? In Joel A. Nevis, Brian D. Joseph, Dieter Wanner & Arnold M. Zwicky (eds.), *Clitics: a comprehensive bibliography 1892-1991*, xii–xx. Amsterdam: John Benjamins.

#### CONTACT

[martin\\_haspelmath@eva.mpg.de](mailto:martin_haspelmath@eva.mpg.de)