

Affirmative markers: A cross-linguistic study of overspecification

TIM ZINGLER

UNIVERSITY OF INNSBRUCK

Submitted: 16/05/2025 Revised version: 18/09/2025

Accepted: 07/10/2025 Published: 23/7/2026



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

Based on a worldwide sample of 604 languages from 179 families, this study shows that affirmative markers, the functional opposite of standard negators, are highly infrequent and possibly ‘rara.’ This confirms previous impressionistic claims. Nine languages with one affirmative marker each were found. Their geographical distribution (a major cluster in eastern Africa), exponence (five of the nine markers are prefixes), and possible diachronic emergence are discussed. The existence of affirmative markers is explained against the background of ‘overspecification,’ operationalized here as the combination of a high degree of verbal synthesis and a low number of zero markers throughout the system. Languages with affirmative markers generally have a large number of morphological categories/slots and also show a tendency to overtly mark other meanings often zero-marked cross-linguistically (e.g., nominative, present tense, third person). The results suggest that some languages prefer explicitness to economy, which informs different aspects of typology.

Keywords: Morphology; morphosyntax; rarum; polarity; affixation; negation.

1. Introduction

One of the primary motivations underlying linguistic typology is the recognition that languages express certain meanings much more commonly than others as part of their grammatical systems. The traditional approach within which these asymmetries were

framed was the concept of markedness, which, among other things, assumes a correlation between semantic load and formal coding. Roughly speaking, the most basic value of a grammatical category is expected to receive the least amount of coding. That value is thus ‘unmarked,’ while the other members of the paradigm are ‘marked.’ However, the empirical facts underlying markedness have always been on a firmer footing than the actual theoretical account. For example, Greenberg (2005 [1966]: 65–69), in arguably the definitive work on typological markedness, already pointed out that what is conceived of as the unmarked value within a category is typically just the most token-frequent one. This account has since been elaborated upon by, for instance, Fenk-Oczlon (1991) and Haspelmath (2006, 2021). These authors argue that, as far as explanations go, token frequency is superior to markedness and the often vague semantic assumptions that it is based on. This is because the link between token frequency and formal shortness was already clearly established by Zipf (1965 [1935]). To the extent that this line of argument holds, then, markedness is best reduced to a descriptive term for linguistic forms. That is, a nominal form with an overt plural morph is marked, whereas the corresponding singular form without an overt singular morph is unmarked, regardless of whether and why the notion of singularity might be more cognitively basic than the notion of plurality. Unless stated otherwise, the present work will use the terms ‘marked’ and ‘unmarked/zero-marked’ in this formal sense. That is, I will not offer any claims about the potential psychological underpinnings of certain semantic values, and thus I will not try to provide an argument for markedness and/or against frequency.

Of all the marking asymmetries within grammatical categories, none might be more salient than the one found in the domain of polarity, i.e., the contrast between negative and affirmative (or positive). Dixon (2012: 90) makes the strongest possible case by claiming that “[n]egation is a term in a two-member polarity system {positive, negative}. ‘Positive’ is always the formally unmarked term, with zero realization.”¹ If true, this would mean that no language has overt affirmative markers. This would be remarkable since absolute universals are known to be rare, and fieldwork on previously undocumented languages continues to reveal structures thought not to exist (and occasionally fails to find structures thought to exist universally). On the other hand, the absence of affirmative markers from the languages of the world would

¹ An anonymous reviewer points out that this conception of polarity, which does not consider interrogative or conditional constructions (etc.), would be considered too narrow nowadays. Since it is sufficient for my argument, however, I will adopt it for the purposes of this paper.

be understandable in that the binary structure of polarity (Miestamo 2005: 2, 39) paves the way for the efficient covert coding of one of the two values of the dichotomy.

Dryer (2005: 454) points out more explicitly than Dixon (2012: 90) that all languages have negative morphs (not including intonation, word order, etc.).² This claim is essentially also made by Bybee (2006: 180), who lists the availability of a negative construction as one of the few absolute universals that have been securely established; cf. also Dahl (1979: 79), Zeijlstra (2007: 499) and Shcherbakova et al. (2023: 2) for similar statements. Greenberg's (2005 [1966]: 50) view is more nuanced. He posits that negators are universally available but points out that there are also some languages with dedicated affirmative markers.³ This description has been echoed by Haspelmath (2021: 614). With respect to the affirmative function, Lehmann (2004: 1845) suggests that it is “normally unmarked,” whereas Wurzel (2001: 184) states that it would be inconceivable to have a language with a zero-marked negator but an overt affirmative marker (cf. also Givón 1991: 350; Bybee 2011: 140). This assumption hints at an implicational hierarchy such that each language that gives overt expression to the unmarked value as traditionally understood (affirmative) will also do so with respect to the marked one (negative).

Based on the above, there appears to be a broad consensus concerning the structure of polarity systems, with the only difference coming down to whether affirmative markers are very rare or non-existent. In this context, it should be mentioned that the category of polarity seems to confirm the predictions of the frequency-based approach to coding asymmetries. That is, negated clauses are much less token-frequent than affirmative ones (cf. Givón 2001: 43, 373 for English data). This is presumably because communication is more efficient if it focuses on the few states and events that do, did, or will hold rather than on the countless states and events that do, did, or will not (cf. Hopper & Thompson 1980: 287).

Yet, while the notion of token frequency may serve as an explanation in typological projects, the empirical groundwork will have to establish the type frequencies of the

² This is not contradicted by certain zero negator constructions in some Dravidian languages since the relevant languages have overt negators in other constructions (Miestamo 2010; Pilot-Raichoor 2010).

³ He mentions Vietnamese (vie; Austroasiatic, Vietic) as one of the languages with an overt affirmative marker, but I found no such element in any of the three Vietnamese grammars that I consulted (Thompson 1987; Nguyễn 1997; Ngo 2021). Cf. Section 2 for my definition of affirmative markers and Section 3 for the database.

pattern(s) at issue. As such, it is noteworthy that this question seems not to have been addressed for affirmative markers before. For instance, if they are indeed rare, one would expect them to be discussed in Frans Plank's *Grammatisches Raritätenkabinett*.⁴ However, the only entry that includes the term 'affirmative' in that database is for "negation expressed negatively, by omission of material present in affirmative clause (marked realis vs. unmarked irrealis mood inflection)." This is not the sense intended here, where an affirmative marker expones a value of the category of polarity and is an overtly coded means that expresses the opposite of standard negation (cf. Section 2 for details). This disregard for affirmative markers could mean different things. On the one hand, they might be too common to count as a rarity, pace the intuition of most linguists. On the other hand, the complete lack of affirmative markers might be taken for granted, following Dixon (2012: 90) cited above. The latter idea finds anecdotal support in that the relevant chapters in reference grammars and textbooks are typically called 'negation' rather than 'polarity,' which suggests that negators are the only overt expressions expected in polarity systems.

The present work takes the lack of previous research on the worldwide distribution of affirmative markers as its starting point and conducts a cross-linguistic study based on a large sample of reference grammars. That is, it will investigate if, where, and how commonly affirmative markers are (claimed to be) found, which will help to clarify whether the absolute universal posited by Dixon (2012: 90) is valid. If that claim turns out to be false, affirmative markers might instead have to be classified as 'rara,' which have also drawn attention in the typological literature (cf. Section 2).

However, the interest pursued here goes beyond the empirical dimension in that this study will also investigate a specific hypothesis. As with the typological distribution of any phenomenon, the question arises as to why languages show it at all. This, of course, is difficult to answer in any case, and it might be argued that an explanation would be even harder to come by in the case of a (presumably) rare structure. Yet, there is a strand of morphological typology that happens to be instructive in the present context. McWhorter (2005) discusses how the structure of languages whose cultural transmission was never interrupted differs from that of creoles, which on his account do derive from scenarios of interrupted transmission.

⁴ I have not found an official way to cite this resource. It can be accessed at <https://typo.uni-konstanz.de/raraalt/intro/> (last access: March 2, 2026).

In the process, he develops the concept of ‘overspecification,’ which he defines as follows:⁵

Any given language gives overt manifestation to particular underlying semantic or syntactic distinctions that are left unmarked in many other languages. The fact that many or most languages operate without marking the distinctions in question entails that to mark the feature at all is ornamental, rather than necessary, to human language. As such, in a given area, one language can be considered *overspecified* in comparison to another. (McWhorter 2005: 268; emphasis in the original)

Given that languages without affirmative markers are well attested, it might be said that languages with affirmative markers are overspecified in the area of polarity. Crucially, this is more than a rechristening. McWhorter (2005) understands the concept as a gradient such that languages that have been spoken uninterruptedly for millennia, but have never had a notable number of L2 speakers, will show a relatively stronger bent for overspecification. It is those languages, then, that are particularly likely to have overt structures for functions that are left implicit in most other languages (including creoles); cf. also Berdicevskis & Semenuks (2020: 291, 293; 2022: 18).

The basic idea underlying the notion of overspecification is thus that some languages come closer to exhausting their morphosyntactic options than others. This premise has also featured in other theoretical accounts, and it should be pointed out that the terminology is less critical to the present work than the content. As such, one might also refer to overspecification as ‘verbosity,’ which Dahl (2004: 54) defines as the “cross-linguistic dispensability” of a category or meaning and which he himself considers essentially equivalent to overspecification (Dahl 2004: 80, fn. 2). More generally, one might subsume the concept under the phenomenon of ‘informativeness,’ where the idea is that highly frequent values (such as ‘affirmative’ within the category of polarity) are less communicatively informative than infrequent ones. Regardless of the terms, the strong prediction based on virtually all previous

⁵ Note that McWhorter’s approach to creole languages has engendered considerable controversy (e.g., DeGraff 2003; cf. also Bisnath et al. 2026). However, this is of no concern here since there are no creoles in my sample and no claims about creoles will be made. Rather, McWhorter’s concept of ‘overspecification’ simply happens to also be useful if taken in isolation, and I consider the adoption of an existing term preferable to coining a synonym. Cf. also below.

work in morphosyntactic typology is that languages will prefer to encode the less frequent and subsequently more informative functions overtly. It follows that the expression of the more frequent values due to overspecification/verbosity, wherever it might occur, is a phenomenon of typological interest.

If overspecification is a tendency that emerges from long periods of exclusive L1 acquisition, it should result in a rather large number of grammatical paradigms overall. Couched within a grammaticalization approach, the prediction would be that languages that create relatively many categories over time also tend to express the exponents of those categories by means of bound morphology. To the extent that that is the case, the profile of overspecifying languages might be '(poly)synthetic' in traditional terms. While there is nothing in McWhorter's (2005) account or in the present approach that limits overspecification to the morphological (as opposed to the syntactic) level, it will be seen in Section 3 that the morphological domain is indeed the relevant one for affirmative markers.

The evaluation of the hypothesis that affirmative markers cluster in overspecifying languages will be at the heart of Section 4, where I will also describe how the concept of overspecification is operationalized in this work. As is the case with all typological predictions, however, the present one will need to be understood in probabilistic terms. That is, the hypothesis is not that it would be possible to classify any potential language with affirmative markers as overspecifying. Similarly, no overspecifying language will necessarily have affirmative markers. More generally, the notion of overspecification is only one way to approach affirmative markers. It is chosen here as an explicitly tentative explanatory angle, mainly because no other theoretical account suggests itself. As such, the present work will hopefully inspire further research on the factors that lead to the emergence of rare types of markers.

To sum up, affirmative markers would be of interest to typology because their overarching category is binary, and zero-marking the affirmative function is unlikely to be misinterpreted given that the only value it contrasts with is always overtly marked. Hence, the category of polarity is sufficiently specified with only negators, and the baseline expectation is that languages will not also have ('uninformative') affirmative markers (cf. Diessel 2019: 228). Conversely, languages that do have affirmative markers would be overspecified in the above sense. In conjunction with their presumably high token frequencies, affirmative markers would thus be compelling evidence for the idea that languages may prefer explicitness to economy, at least locally.

Section 2 is dedicated to the methodological decisions that shaped the database, the latter of which will be analyzed in Section 3. Meanwhile, Section 4 will discuss the results against the background of overspecification and general findings in morphological and diachronic typology. Section 5 will conclude and suggest avenues for further research.

2. Method and definitions

Cysouw & Wohlgemuth (2010: 5) cite the idea that linguistic phenomena that occur in 1-5% of the world's languages should be considered 'rara.' If 'affirmative' is indeed marked infrequently, the relevant morphemes might thus qualify for *rarum* status on this (admittedly arbitrary) criterion. This, however, would also mean that affirmative markers cause the same empirical issues as other *rara*: they cannot be searched for systematically and require large samples as well as a laborious data collection process (Cysouw & Wohlgemuth 2010: 5). In view of this situation, I searched for affirmative markers in reference grammars of 604 languages from 179 families (where each isolate counts as a family), as defined by Glottolog (Hammarström et al. 2026).⁶ Strictly speaking, this collection constitutes a convenience sample because it is made up of the languages for which I had access to grammatical descriptions. Yet, I am not aware of any serious biases in that sample, and it is hoped that its sheer size renders it more reliable than the average convenience sample. The 179 families (including isolates) represented in my sample correspond to 41.6% of the 429 families (including 183 isolates) listed in Glottolog as of this writing.⁷ Table 1 lists the number of families and languages surveyed for this project, sorted by the six geographical macro-areas that underlie the classification in Glottolog. The actual languages and families are provided in the Appendix.

In each grammar, I performed keyword searches (where possible), scoured the table of contents (sometimes complemented by the subject index) for all potentially relevant passages, and scanned the sections on verbal morphology as well as any information about particles and clitics.

⁶ In the interest of consistency and comparability, I also rely on Glottolog for all names of languages and families mentioned here, including where they diverge from the glossonym employed by the relevant sources and/or where they might be rejected by the speech communities at issue.

⁷ <https://glottolog.org/glottolog/glottologinformation> (last access: June 22, 2026).

Macro-area	Number of families represented	Number of languages surveyed
Africa	28	101
Australia	16	51
Eurasia	25	134
North America	33	113
Papunesia	40	120
South America	37	85
Total	179	604

Table 1: Number of language families in sample, sorted by macro-area.

In addition to the reference grammars, I performed less thorough searches in about 75 sketch grammars and about 90 books (edited volumes and monographs) on specific language families or areas, some of which contained sketch grammars themselves. Few of these additional resources described language families not covered by the grammars, and none yielded novel data. Therefore, these languages are not part of the above-mentioned sample, and, with one major exception to be discussed in Section 3.1, the data and information in the next section thus come from reference grammars.

While the concept of negation covers a range of functions, the markers aimed at here were only those that are the semantic opposite of “standard negators” as defined by Miestamo (2005). These markers, like German *nicht* or Spanish *no*, occur in declarative clauses, have predicate scope, and are not inherently emphatic. The restriction to standard declarative negation rules out noun phrase negators (*no idea*), negative pronouns (*nobody*), and negative adverbs (*never*); cf. Dryer (2005: 454). Meanwhile, the condition of predicate scope excludes negators with lexical scope (as in *unkind*). On the basis of that definition, it follows that the affirmative markers of interest will also have to occur in, and have scope over, declarative predicates, while not being emphatic. Their main function is thus to present the predicate of the clause as factual. While the overt expression of such a notion might appear unusual, the affirmative value itself falls out from the semantic structure of a polarity paradigm in the same way that ‘third person’ falls out from a person paradigm, etc.

In examples (1a) and (1b), from Anal (anm; Sino-Tibetan, Kuki-Chin), the kind of affirmative marker relevant to this study is illustrated by contrasting it with the corresponding negator. The elements at issue here and throughout will be highlighted

in boldface. It can be seen that the clauses are minimal pairs, contrasting only in terms of polarity, and that the two exponents occur in the same post-verbal slot. Further details of this affirmative marker will be discussed in Section 3.1.⁸

(1) Anal (Sino-Tibetan, Kuki-Chin; Ozerov 2018: 715)

- a. *tùŋ-kà-hín-nín*
dry-AFF-PL-1
'We (EXCL) are dry.'
- b. *tùŋ-má-hín-nín*
dry-NEG-PL-1
'We (EXCL) are not dry.'

Given that polarity is binary, the most straightforward situation would be for affirmative markers and negators to form a polarity paradigm whose members are mutually exclusive (cf. Malchukov 2011: 263). In such a scenario, affirmative markers would occur in exactly those clauses in which standard negators do not appear, and each clause would feature the affirmative marker or the negator. Yet, as will be seen in the next section, the occurrence of affirmative markers is usually subject to further constraints, and so a perfectly complementary distribution of the two exponents is rarely found. As such, the most important distributional criterion for the present study is that affirmative markers and negators cannot co-occur at the clause level. This is because no clause can predicate something and simultaneously negate that same predicate. If a negator and an affirmative marker are claimed to occur in the same clause, (at least) one of the two markers must actually perform another function (cf. Vázquez Álvarez 2011: 296–297 on Chol⁹ for one such case). Any instances of putative affirmative markers occurring alongside negators were disregarded and not counted toward the database. Finally, the focus on declaratives simply entails that affirmative markers *can* occur in that clause type; it does not mean that affirmative markers are *restricted* to declaratives. In fact, affirmative markers that occur in interrogative

⁸ I will homogenize glossing abbreviations across sources and examples. In particular, I will gloss all elements analyzed as affirmative markers as 'AFF.' Beyond that, I will follow the Leipzig Glossing Rules (Comrie et al. 2015) to the extent possible.

⁹ Chol; ctu; Mayan, Western Mayan.

clauses crucially aid in distinguishing affirmative polarity from declarative mood (cf. below).

Despite the rigid criterion of mutual exclusivity, it needs to be mentioned that explicit information or evidence concerning the compatibility of negators and alleged affirmative markers is not provided in every case. Thus, some of the analyses below might be subject to debate. Furthermore, a strict paradigmatic organization of the affirmative marker and the corresponding negator(s) would mean that the former cannot be optional because standard negators are not optional either (Dahl 2004: 137). As will be seen, however, this criterion is violated by some of the otherwise credible affirmative markers.

As suggested above, there are certain functions that an affirmative marker might be confused with. First and foremost, the affirmative function needs to be distinguished from that of the declarative. The two will often co-occur given that affirmative declaratives are the most frequent kind of clause (Velupillai 2012: 346), but a declarative clause can be negated, unlike an affirmative clause. In addition, a declarative marker is likely to be in competition with other speech act or mood markers such as interrogatives, whereas the only semantic contrast for an affirmative marker is a (set of) negator(s).

Having said that, the data below will show that there seem to be cumulative affirmative/declarative markers. Similarly, affirmative is not equivalent to indicative, the latter of which should contrast with subjunctive markers (etc.) and also be compatible with negators. Analogous arguments hold for other typically zero-marked values in the verbal domain, such as realis mood. All of these categories are semantically compatible with negation and most likely stand in paradigmatic opposition to exponents other than negators. So, while the semantic contribution of declarative, indicative, and realis markers is difficult to pinpoint in isolation, they crucially differ from affirmative markers on distributional criteria, which will thus be invoked in the analyses below. Only once these functions are separated out does the idea that affirmative markers are rare truly suggest itself. In fact, Miestamo (2007: 305) questions the existence of affirmative markers because every item described as such in his sample is either emphatic or an expression of indicative mood. This shows that such adjacent functions are highly relevant to the present work, and the case studies below will discuss them to the extent that the sources permit.

With respect to the formal dimension, affirmative markers had to be grammatical items, incapable of acting as utterances on their own. Hence, affirmative markers as

understood here are not particles corresponding to ‘yes,’ nor interjections, question tags, or ‘positive polarity verbs’ (cf. Mihas & Pérez 2015: 562 on Ashéninka Perené).¹⁰ Another way to think of the phenomenon is that affirmative markers will resist any attempt at translation into English, even by means of circumlocution, since English does not have standard (i.e., non-emphatic) affirmative forms (cf. [1a]).

It should be clear from the discussion so far that the present work will only deal with dedicated affirmative markers, i.e., elements that explicitly express the function of ‘affirmative.’ This is to distinguish them from well-known morphosyntactic effects that polarity distinctions might have elsewhere in the clause, especially in the tense-aspect-mood (TAM) system. That is, the fact that there is often a more limited range of TAM functions in negated than in affirmative clauses¹¹ does not mean that the additional TAM markers available in affirmative clauses are affirmative markers themselves. Note that by focusing on elements that express ‘affirmative,’ I will not always be able to determine from the description whether that is the only function of the relevant unit. In light of this, I will cast a rather wide net and also include certain items for which it is stated that ‘affirmative’ is one of their functions. Details will be discussed in the relevant empirical sections. On a general level, such an inclusive approach seems appropriate in an exploratory study given that it will be easier to discard false positives than to retrieve false negatives.

One domain in which both polarity functions may be marked is that of copulas, which often differ (suppletively) in affirmative and negative contexts. Affirmative copulas will not be considered affirmative markers for the purposes of this project because standard negation, the benchmark for the definition of affirmative markers, has a verbal predicate (Miestamo 2005: 1) and thus lacks a copula (i.e., a semantically bleached item linking to non-verbal predicates). Beyond that, affirmative markers of all overt kinds were considered for this project; that is, there was no a priori restriction to morphological and syntactic units. So, tonal affirmative markers, for example, would have been included. Yet, one of the main conclusions from the database is that all affirmative markers are segmentally instantiated affixes (or clitics).

¹⁰ Ashéninka Perené; prq; Arawakan, Southern Maipuran.

¹¹ This effect is itself a manifestation of markedness as traditionally conceived. Greenberg (2005 [1966]: 29) calls it “defectivation.”

3. Affirmative markers

In this section, I will discuss items that can be analyzed as affirmative markers given the description in the source. However, the amount of information and exemplification provided for the (putative) affirmative markers is never extensive. Therefore, not all of the criteria discussed in Section 2 can reliably be confirmed for every instance, even crucial questions will occasionally remain open, and the individual sub-sections are rather brief. Note in particular that the absence of affirmative markers from negated clauses cannot be demonstrated by any finite number of examples. Hence, statements concerning such an inability in the relevant sources will be taken at face value. Overall, the case studies presented below form a gradient from more to less straightforward instances, and the first three cases are the most convincing ones. Section 3.10 will provide an interim summary.

3.1. *Anal (anm; Sino-Tibetan, Kuki-Chin)*

Information on the Anal affirmative marker comes from Ozerov (2018, 2019) and Pavel Ozerov (p. c.). Note that Anal is the only language discussed here for which the facts and data were not drawn from a reference grammar. Since Ozerov (2018) is dedicated to tone and Ozerov (2019) discusses person indexation, information on the affirmative marker is generally sparse. However, Pavel Ozerov is the primary fieldworker on the language, and he confirmed that the affirmative marker, which he had independently analyzed as such previously, meets both the semantic and formal criteria stipulated in Section 2.

The affirmative marker in question is the suffix *-ka*, whose interaction with tonal marking will be ignored here. Evidence for the idea that this suffix marks the ‘affirmative’ function is copious: it only occurs in affirmative contexts (cf. Ozerov 2019: 37), appears in the same slot as the negator (Ozerov 2019: 40), and is thus mutually exclusive with it. Those aspects were discussed for examples (1a) and (1b) above. In addition, the affirmative marker can occur in interrogative verb forms (cf. Ozerov 2019: 40) and thus does not show the behavior of a declarative marker. Note further that an inability to occur in negated constructions not only argues against a declarative analysis but poses a similar problem for the idea that the element might be an indicative marker. Both declarative and indicative markers should be compatible with negators. On the other hand, it should be mentioned that the

affirmative marker does not occur in all affirmative contexts. Specifically, it is absent from transitive non-future clauses, where the affirmative function has been taken over by a nominalization construction (Ozerov 2019: 37, 39–40, 49). Example (2a) shows the affirmative suffix in a transitive future construction, while (2b) provides crucial evidence for its occurrence in an interrogative clause.

(2) Anal (Sino-Tibetan, Kuki-Chin; Ozerov 2019: 40)

- a. *abũ-tçú-to* *tçá-ná-kà*
 tree(type)-fruit-ABS eat-FUT-AFF
 ‘They will eat fruits.’
- b. *dá:²* *à-:-pé:-ká-nín*
 what 2-NON.AG-give.FUT.SAP-AFF-1
 ‘What shall I give you?’

Idiosyncratic restrictions such as the inability to occur in transitive non-future contexts will be seen for each of the markers discussed in this section. Rather than trying to account for them in the individual case studies, I will make an attempt at a general explanation in Section 4.2. There, I will also address the tentative diachronic link that Ozerov (2019: 49) suggests for the affirmative marker in Anal and the copula in closely related Monsang Naga (nmh; Sino-Tibetan, Kuki-Chin). For the time being, the evidence for an affirmative analysis of the Anal suffix is straightforward on both formal and functional grounds. As such, the suffix is the most compelling piece of counterevidence to the aforementioned ideas that affirmative markers either do not exist at all or are declarative/indicative markers instead.

3.2. *Isimjeega-Rootigaanga* (–; Nilotic, Southern Nilotic)

In *Isimjeega-Rootigaanga*, the affirmative and the negative prefix occur in the same slot of the verbal structure (Griscom 2019: 122) and therefore tend to be mutually exclusive – with the exception of some complicating factors concerning auxiliary constructions and their lexicalization (Griscom 2019: 186). Both the function and the distribution of the affirmative marker seem straightforward given that the “affirmation of a proposition is typically coded by the affirmative prefix” (Griscom 2019: 187). Furthermore, while the prefix does not appear in the imperative (Griscom

2019: 188), it is clearly not a declarative marker because it does occur in interrogative verb forms (Griscom 2019: 75). Generally speaking, the absence of any marker from imperative verb forms is unremarkable since imperatives are commonly bare forms (cf. Becker 2024: 397). Therefore, the inability of affirmative markers to occur in imperatives is not a counterargument to an affirmative analysis. As far as distributional restrictions are concerned, the affirmative marker is barred from conditional contexts (Griscom 2019: 188).

Example (3a) shows one of the two allomorphs of the affirmative marker, which are conditioned by a complex set of principles (Griscom 2019: 187). Meanwhile, example (3b) illustrates the occurrence of the affirmative prefix in an interrogative verb form. This form is actually marked for ‘affirmative’ in two different slots, once by each prefix allomorph (cf. Griscom 2019: 188 for the details).

(3) Isimjeega-Rootigaanga (Nilotic, Southern Nilotic)

- a. *q-à-bár ɲàn bòn-ér-d*
 AFF-3-hit earth people-PS.SG-SS.SG
 ‘People farmed.’ (Griscom 2019: 188)
- b. *g-àd-q-à-más*
 AFF-PSI-AFF-3-agree
 ‘They still agree?’ (Griscom 2019: 75)

There is little additional information about the affirmative prefix in the source. Overall, though, this is one of the most convincing cases discussed here. The relevant marker is not only semantically compatible with an affirmative analysis but is also in (more or less) complementary distribution with the negator. Taken together, this is strong support for the idea that these two items express opposing values of a single polarity paradigm.

3.3. Kabardian (*kbd*; Abkhaz-Adyghe, Circassian)

Colarusso (1992: 125) states that Kabardian “appears to be unique in the world in having a distinct mood mark for simple positive declaratives (in all but the present active tense).” Apart from the fact that Applebaum (2013: 60) provides a different analysis of the distribution of this item, this claim requires some clarification. First,

it can be seen throughout this section that Kabardian is not in fact unique in this regard. Second, the affirmative marker is apparently described as an exponent of mood because it is limited to declarative clauses. Therefore, the question arises as to whether it is more properly analyzed as a declarative or an affirmative marker. The strongest argument for classifying it as an affirmative marker comes from Applebaum (2013: 62), who states that the ‘declarative’ marker (which is also the term she uses) cannot co-occur with negators. It follows that the affix in question only occurs in affirmative contexts, which are therefore the delimiting factor. As argued above, if this item were a declarative marker, its inability to occur in negated clauses would not be accounted for. It should also be noted here that the negator and the affirmative marker occur in the same suffix slot (Applebaum 2013: 60), and the resulting incompatibility of the two markers further argues for an affirmative and against a declarative function. Additional evidence for the affirmative analysis comes from the Russian literature, which has traditionally described this item as an affirmative marker (e.g., Kumakhov 1989, 2006).¹²

However, the suffix does not occur in interrogative clauses (Colarusso 1992: 125, 126), which might support the declarative analysis. The bottom line is that every clause in which the suffix appears is affirmative, while the suffix does not occur in all declarative contexts given its absence from negated clauses. It thus seems defensible and more descriptively economical to consider it an affirmative marker, though with a potentially significant declarative component. Such an interaction is generally plausible given the frequent overlap of those two functions, and further evidence for this idea will be provided in following sub-sections.

Examples (4a) and (4b) illustrate the relevant affirmative item in the Russian and in the Turkish variety, respectively. Note that Colarusso (1992: 125) remains non-committal with regard to the exact shape of the suffix.

(4) Kabardian (Abkhaz-Adyghe, Circassian)

a. *sə-λaaʒa-ś(a)*

1SG-work-AFF

‘I work.’ (Colarusso 1992: 125)

¹² I am deeply indebted to Ludger Paschen for finding and summarizing the Russian sources for me.

- b. *zə ʈɒpq ʃə-ʔv-s*
 one tribe OBL-exist-AFF
 ‘There was a tribe.’ (Applebaum 2013: 197)

3.4. *Bambassi (myf; Blue Nile Mao)*

M. Ahland (2012: 387) states that his analysis of the affirmative prefix in Bambassi is ‘tentative’ because the item has a ‘perplexing’ distribution. Among other things, it is optional and seems to interact with mood, clause type, and person (M. Ahland 2012: 390–398). Crucially for the present analysis, the marker is prohibited in all negative contexts, which is a necessary condition for affirmative markers but not for declarative and/or indicative markers. It can also occur with certain interrogatives, which suggests that it is not a declarative marker (M. Ahland 2012: 387, 388). Interestingly, the prefix had previously been analyzed as a declarative marker, but this runs counter to its incompatibility with negative contexts (M. Ahland 2012: 388). This shift in analysis also gains support from the fact that there is at least one actual declarative affix in the language, which can be seen in (5a). The parentheses in that example are taken over from the source, where they presumably signal the optional status of the affirmative prefix. Meanwhile, example (5b) illustrates the use of the affirmative marker in a question.

(5) Bambassi (Blue Nile Mao)

- a. *(ha-)tí-pè:f-á*
 (AFF-)1SG-slap-DECL
 ‘I slapped (it).’ (M. Ahland 2012: 390)
- b. *íʃ mùnts’-ìʃ hì-àn ha-hój-à:*
 DEF woman-SBJ 2SG-COM AFF-go-INTR
 ‘Did the woman go with you?’ (M. Ahland 2012: 474)

Based on the description by M. Ahland (2012), there is no major argument against the affirmative analysis of the Bambassi prefix, but of course the marker is insufficiently understood. That said, its optional status and restriction to certain

contexts are potentially indicators of its recent grammaticalization.¹³ If so, this might suggest that affirmative markers grammaticalize after negators, which would be in accordance with the notion of overspecification outlined in Section 1. This issue will be discussed further in Section 4.2.

3.5. *Laz (lzz; Kartvelian, Zan)*

There are several affirmative ‘preverbs’ in Laz. These will be treated as prefixes here, following Lacroix (2019: 840), who shows that they have fixed slots in the verbal structure. These markers never occur in negated clauses, relative clauses, with non-initial verbs in cases of coordination, or with certain middle verbs (Lacroix 2009: 433, 437; 2019: 843). In addition, they are sensitive to information structure and involved in TAM marking (Lacroix 2019: 844–845). The inability of the markers to occur in negated contexts lends credence to the affirmative analysis, whereas the other restrictions at least do not argue against it. Note that Lacroix’s (2009, 2019) own analysis of the matter is equivocal; this will also be reflected in the present case study.

While Lacroix (2009) describes in detail which of the affirmative markers are chosen with which kinds of verbs, that choice seems to depend on the semantics of the surrounding verb rather than on that of the affirmative items themselves. I will thus limit myself to an illustration of the affirmative marker *do-*, for which the source provides the most clear-cut evidence of affirmative usage. None of the other ‘preverbs’ are claimed to be affirmative elements here. The prefix *do-* occurs with the majority of verbs that do not also bear a spatial preverb (Lacroix 2009: 435). The only example that explicitly shows the (possible) affirmative function is reproduced in (6). Note that both the glossing and the translation in the original source are in French and were translated into English by the present author.

(6) Laz (Kartvelian, Zan; Lacroix 2009: 436)

badi-k ncumu-ši oxoi do-k’od-u
 old.man-ERG salt-GEN house AFF-build-AOR.3SG
 ‘The old man built a house of salt.’

¹³ This assessment especially suggests itself on the basis of the short texts in M. Ahland (2012: 713–725), where the affirmative marker fails to occur in many of the contexts where it would seem to be semantically eligible.

Overall, then, the claim that the above prefix is an affirmative marker is not entirely secure, even though Lacroix (2009: 438) emphasizes that the incompatibility of the affirmative items with negative contexts is one of the few traits that the prefixes share. He also states that they are very frequent (Lacroix 2009: 433), which should be the case given the high token frequency of affirmative clauses. On the other hand, he highlights that there is a single example in his corpus where an allegedly affirmative preverb occurs in a negated clause, specifically in a prohibitive construction (Lacroix 2009: 438). While this would be a sufficient reason to discard the element from the current database, the prefix in the prohibitive construction is not the one seen in (6). Hence, there is no strong evidence to suggest that *do-* violates the definition of an affirmative marker used here. Clearly, though, a better understanding of the many restrictions to which it is subject might eventually challenge that analysis.

Lacroix (2019: 844) calls the affirmative preverbs “one of the most intricate issues in the grammar of Laz.” Analogous statements are also found for some of the other languages (cf. Bambassi above), which suggests that the contribution of (putative) affirmative markers is cognitively challenging. Perhaps, however, this impression mainly emerges because languages with affirmative markers tend to be described by grammarians whose native languages do not have them.

3.6. Northern Gumuz (*guk*; *Gumuz*, *Nuclear Gumuz*)

The initial affix in the Northern Gumuz verb structure is called an ‘affirmative’ marker because it does not occur with interrogatives or negatives (C. Ahland 2012: 206, 209). While the former restriction would be compatible with a declarative function, the latter suggests that the element is more likely to be a cumulative affirmative/declarative marker. In addition, the item is not found with all affirmative verbs in complex clauses (C. Ahland 2012: 206). Note that C. Ahland (2012) refers to the category that the affirmative prefix instantiates as ‘mood,’ but this is presumably because of the heterogeneous elements with which it competes for the initial verb slot, i.e., subordinate markers and the jussive (C. Ahland 2012: 207). Example (7) illustrates the affirmative prefix.

(7) Northern Gumuz (*Gumuz*, *Nuclear Gumuz*; C. Ahland 2012: 207)

d-á-gáχ

AFF-3SG.INTR-be.good

‘It was good.’

Since the function of the affirmative marker is not further elaborated upon in the source, the analysis provided there will be accepted as is. It is also of interest that Northern Gumuz has an affirmative affix but no negator affixes (C. Ahland 2012: 184). On the face of it, this would seem to contradict the development hinted at in Section 3.4, according to which affirmative markers grammaticalize after negators. On the other hand, it might of course be that negators in Northern Gumuz have been subjected to more erosion and/or renewal than the affirmative item, perhaps as part of the well-known negative cycle (van Gelderen 2008). To the extent that the current negator has only been established recently, this would help to explain its lower degree of fusion. This general point will be discussed in Section 4.2.

Interestingly, the affirmative prefix in Southern Gumuz (-; Gumuz, Nuclear Gumuz), which is a separate language according to Glottolog, can occur in questions. This element is segmentally different from the one in Northern Gumuz but occurs in the same verb slot (cf. C. Ahland 2012: 206). The prefix in an interrogative verb form is shown in (8).

(8) Southern Gumuz (Gumuz, Nuclear Gumuz; C. Ahland 2012: 208)

b-á^h-tsá-gá = áj

AFF-2SG.INTR-go-N.FUT = Q

‘Did you go?’

This distribution suggests that the Southern Gumuz item is less of a declarative and more of an affirmative marker than the one in Northern Gumuz shown in (7). However, the description by C. Ahland (2012: 207) suggests that the interrogative usage in (8) is exceptional, and hence this is difficult to evaluate. I will return to this matter as part of the diachronic discussion in Section 4.2.

3.7. Karo (Ethiopia) (*kxh*; South Omotic)

In Karo, there is a marker of affirmative/declarative function that Belete (2018: 177) calls a suffix. However, since it can occur on verbal and non-verbal predicates alike, it might best be analyzed as a clitic instead (Zwicky & Pullum 1983; Spencer & Luís 2012; Zingler 2022). More importantly, the claim that the element marks the declarative function is based on the fact that it never co-occurs with imperative, hortative, or interrogative morphs (Belete 2018: 179). It also does not occur in

dependent clauses (Belete 2018: 179). Meanwhile, there is another postposed element with the same restrictions that occurs in negated declaratives following the negator affix (Belete 2018: 179). In sum, then, the dichotomy between the affirmative/declarative marker and the negative/declarative marker seems straightforward. As such, the Karo marker is more clearly cumulative than the ones in Kabardian and Northern Gumuz sketched above. The former element is illustrated in (9a), and its negative counterpart is given in (9b).

(9) Karo (Ethiopia) (South Omotic)

- a. *wuti zoot-ite-na wo = do*
 1PL swim-PROG-AFF.DECL 1PL = exist
 ‘We are/were swimming.’ (Belete 2018: 177)
- b. *na ni-ma-mo*
 3SG come-NEG-NEG.DECL
 ‘She will not come.’ (Belete 2018: 178)

The marker in (9a) is arguably the best evidence for the close relationship between the affirmative and the declarative function. That is, while this interaction was already brought up for some of the other examples, none of the languages discussed so far expresses these overlapping categories as systematically as Karo. In addition, it might be cumulative elements like these that limit the distribution of pure affirmative markers (e.g., if the cumulative markers become pure declarative rather than pure affirmative markers). This will be taken up in Section 4.2. At the same time, the two markers show the crucial difference between affirmative and declarative markers. If *-na* in (9a) were merely a declarative, it would also be expected in (9b).

3.8. Male (Ethiopia) (*mdy*; *Ta-Ne-Omotic*, *North-West Ometo*)

The situation in Male closely resembles that described for Karo above. There is a cumulative affirmative/declarative marker that is in complementary distribution with the negative/declarative marker, neither of which co-occurs with interrogative, optative, or imperative morphs (Amha 2001: 148). Here too, then, there is enough evidence to suggest that the marker expresses both ‘affirmative’ and ‘declarative.’ It is furthermore of interest that the affirmative/declarative marker is limited to

independent clauses and to what Amha (2001: 148) calls ‘neutral’ statements, which she distinguishes from emphatic contexts. This sufficiently proves that the marker at issue meets another functional requirement spelled out in Section 2, namely, the inability to be used emphatically. A major difference compared to Karo is that the marker always occurs on the verb, including when the verb does not appear clause-finally (Amha 2001: 149–150). Unlike the Karo element, then, this marker appears to be a bona fide suffix. The affirmative/declarative marker is shown in (10a), and its negative equivalent is exemplified in (10b).

(10) Male (Ethiopia) (Ta-Ne-Omoti, North-West Omoti; Amha 2001: 148)

- a. *ʔ-atsí* *ziginó* *mukk-é-ne*
 person-M.NOM yesterday come-PFV-AFF.DECL
 ‘The man came yesterday.’
- b. *ʔ-atsí* *máári* *mazz-uwá-se*
 person-M.NOM house.ABS build-IPFV.NEG-NEG.DECL
 ‘The man is not building a house.’

The obvious formal, functional, and distributional similarities between the affirmative markers in Karo and Male raise the question of whether this overlap has come about by chance. While language contact is one possible alternative, it should be pointed out that South Omoti and Ta-Ne-Omoti are both branches within the Afro-Asiatic family in the *World Atlas of Language Structures* (WALS; Dryer & Haspelmath 2013). This differs from the classification in Glottolog, where the two units are separate families. In this case, then, the WALS classification might offer a more straightforward explanation of the parallels between the two markers, but the present argument does not hinge on this issue.

3.9. Southern Yukaghir (*yux*; *Yukaghir*, *Kolymic*)

The Southern Yukaghir verb is not usually marked for positive polarity (Maslova 2003: 144). In other words, the affirmative prefix found in the language is optional. Beyond that, the marker also has distributional preferences in that it is used most frequently in the irrealis and can only occur with ‘main’ forms, which seems to mean those that are not subject focus or object focus forms (cf. Maslova 2003: 139, 144). An additional complication is that the marker can be used to express emphatic

affirmation (Maslova 2003: 144), which might violate a major criterion specified in Section 2. However, the examples provided throughout the grammar do not suggest that the marker generally expresses emphasis, and it will therefore be interpreted as capable of encoding the pure ‘affirmative’ meaning. One apparently non-emphatic example, in realis mood, is given in (11a). Meanwhile, (11b) is relevant for two different reasons. On the one hand, it shows that the item can occur in questions and is thus not a declarative marker. On the other hand, it co-occurs with an irrealis marker, thus illustrating the semantic difference between affirmative polarity and realis mood.

(11) Southern Yukaghir (Yukaghir, Kolymic)

- a. *me-met-ahurpe-š-čē*
 AFF-REFL-suffer-CAUS-1SG.INTR
 ‘I make myself suffer.’ (Maslova 2003: 145)
- b. *tiŋ korobka-ge m-et-šøw-jemet*
 this box-LOC AFF-IRR-enter-2PL.INTR
 ‘Would you enter this box?’ (Maslova 2003: 172)

Apart from the exact functional contribution, there is another factor complicating the analysis of the above element. The standard negator in the language is also a verbal prefix, but it is not assigned to the same slot as the affirmative marker (cf. Maslova 2003: 139), unlike in Anal or Isimjeega-Rootigaanga. Hence, there is at least the possibility that the affirmative marker and the negator can co-occur, which would critically undermine the idea that the former is an affirmative marker. Then again, Malchukov (2011: 264) points out that exponents of complementary functions can co-occur to produce non-compositional meanings. Such a scenario would presumably not affect the present analysis.

3.10. Interim summary

This section discussed nine plausible affirmative markers from as many languages. While I know of no major arguments to question the classification of the items in Anal, Isimjeega-Rootigaanga, or Kabardian, the other items are of a more dubious status. This is either because they are (possibly) not ‘pure’ affirmative markers (Northern Gumuz, Karo, Male) or because the functions of the putative affirmative

markers are generally somewhat unclear (Bambassi, Laz, Southern Yukaghir). Overall, though, this database is sufficient to show that affirmative markers exist in at least some languages. This result thus disproves Dixon’s (2012: 90) strong claim that affirmative markers are categorically absent from languages. As will be argued below, the database lends credence to the idea that cross-linguistic rarities should neither be unduly “normalized” nor unnecessarily “exoticized” (cf. Anderson et al. 2025: 8–9).

Going back to the notion of ‘rara’ introduced in Section 2, the database suggests that affirmative markers may indeed qualify for *rarum* status. The sample underlying this study contains 604 languages, of which no more than nine contain affirmative markers. This corresponds to 1.5% and would thus be safely within *rarum* range. It should also be noted that, if three or more of the less secure cases were discarded, the ratio would fall to or below 1%, thus making affirmative markers a ‘*rarissimum*’ on the same arbitrary quantitative criterion (Cysouw & Wohlgemuth 2010: 5). Given these uncertainties, it seems most prudent to forgo a strict definition of *rara* and/or *rarissima*. Rather, the present study demonstrates beyond doubt that affirmative markers are a minority option, and polarity stands a good chance of being the most lopsided grammatical category in terms of marking asymmetries. Another fact that underscores the rarity of affirmative markers is that they do not even seem to be common at the family level. That is, none of the languages that are in my sample and related to the ones discussed in this section are claimed to have affirmative markers (but bear in mind the problematic issues surrounding ‘Omotic’ and Gumuz). This suggests that affirmative markers arise only when a very specific set of conditions is met, which causes their distribution to be rather sporadic. Section 4.2 will spell out some ideas along those lines.

Table 2 shows the affirmative markers discussed in Sections 3.1–3.9, summarizing their form, type of exponence, and position relative to the head (i.e., the verb in every language except Karo, where the marker can also follow other word classes).

Language	Form	Type of marker	Position of marker
Anal	<i>ka</i>	Affix	Postposed
Isimjeega- Rootigaanga	<i>q/g</i>	Affix	Preposed
Kabardian	<i>ś(a)/s</i>	Affix	Postposed
Bambassi	<i>ha</i>	Affix	Preposed
Laz	<i>do</i>	Affix	Preposed

Language	Form	Type of marker	Position of marker
Northern/Southern Gumuz	<i>d/b</i>	Affix	Preposed
Karo (Ethiopia)	<i>na</i>	Clitic	Postposed
Male (Ethiopia)	<i>ne</i>	Affix	Postposed
Southern Yukaghir	<i>me</i>	Affix	Preposed

Table 2: Overview of affirmative markers.

As can be seen, all markers save one are affixes, and even the outlier is described as an affix in the source. This is noteworthy because free affirmative markers were also searched for. Equally remarkably, five of the eight affixes are prefixes. The relevance of these patterns will be discussed and substantiated in Section 4.2.

Another pattern that emerges from the present section is the geographical concentration of the languages with affirmative markers. All nine languages come from the macro-areas of Eurasia or Africa, as defined by Glottolog. Even more remarkably, more than half of the database derives from a salient cluster of languages in eastern Africa, while another possibly relevant niche is found in the Caucasus. This distribution is shown in Figure 1, created with the GlottoshinyR tool by Börstell (2021).

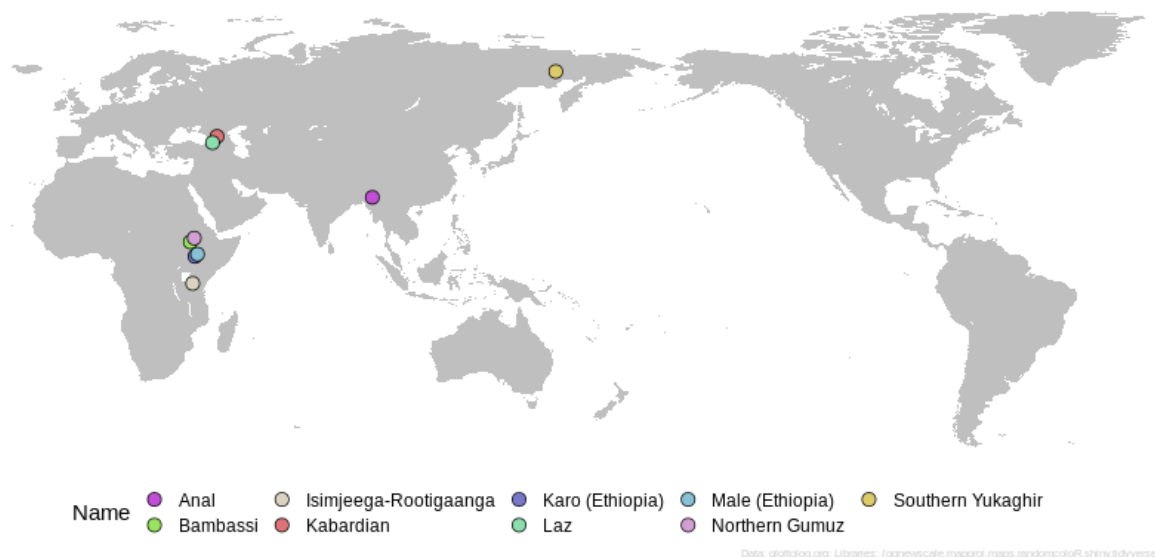


Figure 1: Geographical distribution of languages with affirmative markers.

Note that of the 179 families investigated here, only 53 are Eurasian or African (cf. Table 1). Hence, more than 70% of the families are from the other four macro-areas, and so the results cannot plausibly be attributed to a biased sample. While the geographical pattern suggests that affirmative markers might diffuse via contact, none of the markers likely to have been involved in borrowing for areal reasons are actually close in form, as can also be gleaned from Table 2 (this excludes Karo and Male, which are possibly related). The areal component will be taken up again in Section 4.1.

4. Discussion

This discussion aims to shed light on the empirical patterns described in Section 3. In so doing, it will focus on the question of why affirmative markers might occur in the languages in which they are found and why they show a slight prefixing preference. Note, however, that these proposals are necessarily speculative and do not claim to identify sufficient criteria for the existence of overspecification or affirmative markers. The alternative would be to skirt explanatory attempts entirely and to chalk up all *rara* to mere chance. In that sense, then, the tack chosen here is more ambitious and less fail-safe. The backdrop for the first part of the discussion will be McWhorter's (2005) concept of overspecification, which will be operationalized and applied in Section 4.1. Meanwhile, Section 4.2 will address major typological and diachronic issues raised by the data.

4.1. Affirmative markers as a product of overspecification

For the purposes of this discussion, overspecification will be broken down into two logically independent and falsifiable parameters: degree of synthesis and overt marking of other usually unmarked categories. Below, I will refer to the latter criterion as the 'systemic' parameter. Note that McWhorter (2005, 2007: 21–28) does not explicitly name these parameters as indicators of overspecification because he generally does not provide a rigid definition of this concept (cf. also Berdicevskis & Semenuks 2020: 284). Instead, the criteria that he employs are tailored to the specific languages that he investigates (often in pairwise fashion). However, the parameters suggested here are in tune with McWhorter's overall approach and mostly differ from

his in that they are more abstract so as to be applicable to a larger number of languages at once.

The first parameter concerns the degree of synthesis.¹⁴ The idea is that, if a language overspecifies, it will have a relatively large number of grammatical categories that it expresses synthetically. Synthesis is traditionally discussed with respect to the verb, and this (usually implicit) bias is presumably due to the fact that verbs tend to have richer morphology than nouns (e.g., Bybee et al. 1990: 4). However, potential effects of this bias are irrelevant to the present study since affirmative markers modify the meaning of the predicate and, as was seen above, are marked on the verb almost without exception. Hence, the empirical results suggest that, if anything, it is the verbal morphology that is part of the explanation.

The degree of synthesis in a language is often represented in reference grammars by means of a ‘template,’ which indicates the number of affix paradigms that can be manifested on a single finite verb form. Of course, such templates only indicate the logically possible maximum, and most verb tokens in languages with large templates do not approximate this maximum. Nevertheless, such templates are instructive when compared to languages like English, which has been argued to have a ‘monosuffix constraint,’ possibly the last stage before a fully analytic structure (Aronoff & Fuhrhop 2002: 485). Thus, a template for English verbs would be limited to a single inflectional slot, and templates in languages going beyond a single slot can thus be considered increasingly synthetic. It is this gradient understanding of synthesis that will underlie the analysis offered below.

I found explicit statements regarding the maximum verbal template for six of the nine languages that have affirmative markers, and I secured the template for Anal from Pavel Ozerov (p. c.). The two languages for which I do not have templates are the ‘Omotic’ ones, Karo and Male.¹⁵ Table 3 presents the number of prefix and suffix slots in the seven languages. In cases where the source gave several templates due to a difference conditioned by verb type, I chose the least marked one on the traditional definition. For Bambassi, this means the realis instead of the irrealis template, and for Northern Gumuz this means the non-future instead of the future template.

¹⁴ I will not deal with the potential difference between ‘synthesis’ and ‘polysynthesis,’ which is at best difficult to operationalize and possibly conceptually flawed (cf. Haspelmath 2018 for discussion).

¹⁵ I also did not find relevant statements/templates in any of the related languages for which I have grammars.

Language	Prefix slots	Suffix slots	Source
Anal	5	6	Pavel Ozerov (p. c.)
Isimjeega- Rootigaanga	3	5	Griscom (2019: 122, 126)
Kabardian	13	5	Applebaum (2013: 41)
Bambassi	2	7	M. Ahland (2012: 350)
Laz	4	7	Lacroix (2019: 840)
Northern Gumuz	7	6	C. Ahland (2012: 184)
Southern Yukaghir	3	4	Maslova (2003: 139)

Table 3: Number of affix slots in the verbal template of languages with affirmative markers.

The patterns in Table 3 are rather remarkable. The average number of prefix slots is 5.3, and the average number of suffix slots is 5.7. (If Karo and Male had no affixes at all, which is not the case, the average would be 4.1 for prefixes and 4.4 for suffixes.) As such, the average sum of prefix and suffix slots in the languages is roughly between 9 and 11, depending on the exact structure of Karo and Male. This puts languages with affirmative markers toward the highly synthetic end of the worldwide spectrum since Bickel & Nichols (2013) only find two languages with more than eleven templatic slots and only nine languages with more than nine templatic slots in their 145-language sample.¹⁶ Based on that evidence, the idea that affirmative markers are found in languages with a high degree of synthesis has been confirmed. More specifically, languages with affirmative markers not only show a strong bent for grammaticalization by creating exponents for a rare function in the first place, but also by expressing an above-average number of functions by means of bound morphology.

An additional aspect of interest is the relationship between prefixes and suffixes. On the one hand, there are essentially as many prefix slots as suffix slots on average, and on the other hand, there is not a single language without prefixes (though Karo seems to belong in that category; cf. Belete 2018: ix–xi). If Greenberg (1966: 92) is

¹⁶ Note that the classification adopted by Bickel & Nichols (2013) is not identical to the one used for the present project. In particular, on their counting procedure, English is assigned two inflectional slots because the verb can mark agreement and past tense (but not in the same word form). Hence, their approach seems to count paradigmatic options while mine counts syntagmatic options. Nevertheless, their study is sufficiently close in aim to the present argument, and it is the most suitable point of comparison that I am aware of.

correct in claiming that exclusively suffixing languages are fairly common, it can be concluded that languages with affirmative markers are (largely) drawn from the subset of languages that do have prefixes. The relevance of this finding will be addressed in Section 4.2.

The second parameter for the operationalization of overspecification is the number of other cross-linguistically zero-marked meanings that are overtly expressed in the languages with affirmative markers. The idea behind this approach is that, if overspecification is a general tendency within certain languages, most grammatical meanings in those languages might be encoded by overt material. I will not offer any ideas as to why such an effect might manifest itself in any given language except that analogy is one of the driving forces of language structure and change (e.g., Bybee 2010: ch. 4). That is, once a language overtly expresses one typologically unmarked function, this principle might be extended to other relevant meanings via a ‘transparadigmatic’ process (or, less formally, a ‘snowball effect’). Such an extension of overt marking might ultimately be triggered by a preference for ‘iconic’ structures, which supply every meaning with an overt form (cf. Wurzel 1987: 92; Croft 2003: 102; Cristofaro 2014: 289). Based on that reasoning and the improbability of affirmative marking, the systemic approach pursued here could also be phrased as a rough implicational hypothesis: a language that has affirmative markers will also have markers for other typically unmarked functions. That is the claim to be tested here.

In the interest of comparability, I will limit the investigation to the cross-linguistically common categories of case and number on nouns and tense and agreement (i.e., person marking, indexation) on verbs. For languages that do not have tense, I chose aspect. Based on the lists compiled by Croft (2003: 156–157) and Bybee (2015: 100), the typologically unmarked members of those categories are nominative and absolutive for case, singular for number, present for tense (or perfective for aspect), and first and/or third person (singular) for indexation. Hence, the prediction based on the systemic assumption is that languages with affirmative markers will also express (some of) those values overtly.

Table 4 lists the resulting patterns. Empty cells indicate either lack of evidence for non-zero marking or counterevidence to non-zero marking. All person markers are to be understood as singular. Where there are multiple paradigms of person markers, I chose the least ‘marked’ one on the traditional approach (e.g., indicative instead of subjunctive). The numbers in parentheses indicate the pages where the relevant information can be found in the corresponding sources used in Section 3. The page

numbers for Kabardian refer to Applebaum (2013). Information on Anal is from Ozerov (2019) or from Pavel Ozerov (p. c.).

Language	Number	Case	Person	Tense/aspect
Anal		Absolutive	First (27)	Present
Isimjeega- Rootigaanga			First (175)	
Kabardian		Absolutive (29)	Third (42)	Present (57)
Bambassi				Present (460)
Laz				Present (340–341)
Northern Gumuz		Nominative (147)	First, third (123– 126)	
Karo (Ethiopia)				
Male (Ethiopia)		Nominative, absolutive (56)		Perfective (114)
Southern Yukaghir			First, third (140)	

Table 4: Overt marking of typologically unmarked functions in languages with affirmative markers.

The results displayed in Table 4 are not as readily interpretable as those in Table 3¹⁷. In other words, a tendency to mark all members of a paradigm throughout the system is not as obviously a concomitant of affirmative marking as is verbal synthesis. First and foremost, there is no support from the category of number given that no language that marks affirmative also marks singular. This undermines the systemic idea in that number is relatively close to being a binary feature like polarity, unlike person and

¹⁷ Note the following complications with regard to Table 4. First, case in Kabardian is cumulated with definiteness: absolutive is unmarked when indefinite but marked when definite. Neither Croft (2003) nor Bybee (2015) list marking asymmetries for definiteness, and I did not find relevant statements in Greenberg (2005 [1966]) either. This is presumably because markedness values in the domain of definiteness are context-dependent (Givón 1991: 336). However, Becker (2024: 397) establishes that indefiniteness shows the strongest tendency to be zero-marked across the nominal, verbal, and adjectival categories she tested. If that means that ‘indefinite’ is the ‘unmarked’ value in a definiteness paradigm, the Kabardian pattern is unremarkable. Second, Kabardian also has zero allomorphs for third person and present tense. Finally, first person in Southern Yukaghir is unmarked for transitive indicative verbs but marked for intransitive indicative verbs; cf. example (11a) above.

case. On the other axis, Karo does not overtly mark any of the relevant values. As such, it provides counterevidence to the implicational hypothesis proposed above.

In addition, much of the positive evidence comes from person indexation, where the situation regarding coding asymmetries is unclear due to the general issue of whether first or third person is less marked. While Bybee (2015: 100) treats both persons as equally unmarked compared to second person, Croft (2003: 156) ranks third person as less marked than first. Hence, an evaluation of the category of person not only presents empirical problems but also theoretical ones. Finally, given that a language cannot have both a prototypical nominative and a prototypical absolutive due to functional overlap, it stands to reason that Male has at most one case marker supporting the overall idea investigated here. In sum, only the domains of case and tense/aspect provide any evidence for the systemic nature of overspecification.

There are some mitigating factors to explain these mixed results. Apart from the fact that Table 4 selected a subset of categories, the above analysis also introduced nominal morphology into the picture. Of course, it might be that languages only tend to overspecify the verbal domain, if any. That would at least line up with the fact that verbal morphology is usually more synthetic than its nominal equivalent. In fact, number and case are not marked at all in Isimjeega-Rootigaanga (Griscom 2019: 78), and Bambassi lacks case (M. Ahland 2012: 194). So, the comparison in Table 4 is not straightforward. Overall, 14 of the 36 cells in Table 4 are filled and 10 of the 18 cells for the two verbal categories. In order to assess whether these ratios are at all typologically meaningful, one would need analogous studies for different sets of languages and categories, but I know of no research addressing precisely this issue.¹⁸

In conclusion, it can be argued that synthetic verbal morphology is a necessary (if by no means sufficient) condition for the presence of affirmative markers, while the overt expression of other cross-linguistically unmarked functions is a lesser (though still suggestive) predictor. Thus, to the extent that overspecification was appropriately operationalized here, the hypothesis that affirmative markers tend to be found in overspecifying languages was broadly supported.

¹⁸ However, the worldwide marking patterns collected in Grambank (Skirgård et al. 2023) are still instructive in this regard. The data there show that present tense markers occur in 32.2% of all sample languages (771/2,392; GB082), singular markers occur in 17.3% of all sample languages (425/2,442; GB042), and core case marking on non-pronominal noun phrases, including nominatives and absolutives, occurs in 32.7% of all sample languages (777/2,377; GB070). Hence, all these unmarked functions are overtly expressed much more commonly than ‘affirmative’ is at 1.5%, which suggests that the systemic link between the former set and the affirmative markers might be very weak.

The idea that overspecification can spread across grammatical sub-systems also gains support from the fact that most of the languages with affirmative markers are spoken in eastern Africa. It is precisely languages from that area that commonly have marked nominatives (König 2006; Handschuh 2014: 221–223; cf. also examples [10a] and [10b] above). Of course, one would need to investigate if those items instantiate the prototypical definition of nominatives. Furthermore, there is the issue that languages with marked nominatives often have unmarked accusatives, so the presence of marked nominatives does not automatically entail overspecification. Nevertheless, the present study leads to the conclusion that the overt expression of typically unmarked functions is more likely to be found in eastern Africa than in other areas. Possible causes of this pattern will hopefully be the focus of future research.

As mentioned in Section 1, McWhorter’s (2005) argument is based on the assumption that overspecification is most likely in languages that have had few non-native speakers. As such, the sociohistorical profile of the languages in the database might provide ancillary support for the claimed link between affirmative marking and overspecification. While admittedly crude, the best way to approximate this issue might be to establish the number of speakers for each language. The idea is that small speech communities are more likely to be socially homogenous and/or geographically isolated. Therefore, the relevant languages are more likely to have had few L2 speakers and should thus provide the prerequisites for overspecification. So, the prediction is that languages with affirmative markers are spoken by small communities.

The following figures have to be understood against the background of the global trend of language loss. That is, the numbers given below may have greatly reduced by now. Apart from that, however, they seem to bear out the overall idea. Bambassi, Karo, Isimjeega-Rootigaanga, and Southern Yukaghir were each spoken by 5,000 people or fewer at the time or before the relevant grammar appeared (M. Ahland 2012: 13; Belete 2018: 17; Griscom 2019: 1; Maslova 2003: 19). Anal has 20,000 speakers (Ozerov 2019: 26), while Male has 46,000 (Amha 2001: 3). C. Ahland (2012: 3) gives 200,000 as the number of Gumuz people, which means that the number of speakers of (Northern) Gumuz might be considerably lower. Lacroix (2019: 830) concedes that the number of Laz speakers is unknown but cites figures as high as 250,000, whereas Applebaum (2013: 4) mentions 647,000 speakers of Kabardian.¹⁹

¹⁹ Daniel & Lander (2011: 125) give roughly similar numbers for the two Caucasian languages: “several dozens of thousands” in the case of Laz and “about one million” for Kabardian.

Overall, then, the sociohistorical aspect seems to tally with the idea that affirmative markers are an outgrowth of overspecification, at least if community size is an adequate operationalization.

4.2. *Typological and diachronic issues*

It was mentioned in Section 3 that the affirmative markers show a prefixing preference. This is unexpected on the assumption that they behave like morphological negators, which are more often suffixal than prefixal (Dryer 2013; Berg 2020: 376).²⁰ In fact, a juxtaposition of the affirmative markers and the standard negators in the corresponding languages shows that there is no symmetry in this regard. (As expected, all the languages with overt affirmative markers also have overt negators.) The relevant patterns are shown in Table 5.

Language	Type of affirmative marker	Type of standard negator
Anal	Suffix	Suffix
Isimjeega- Rootigaanga	Prefix	Prefix
Kabardian	Suffix	Prefix, suffix
Bambassi	Prefix	Suffix (multiple)
Laz	Prefix	Word (preverbal)
Northern Gumuz	Prefix	Enclitic
Karo (Ethiopia)	Enclitic	Suffix
Male (Ethiopia)	Suffix	Suffix
Southern Yukaghir	Prefix	Prefix

Table 5: Type of affirmative marker and standard negator.

Only six of the nine languages consistently position their affirmative markers and negators on the same side, and only four of the nine languages use the exact same type of form for their affirmative marker and their standard negator (unless the Karo

²⁰ Despite their overall suffixing preference, morphological negators are relatively more likely to be prefixed compared to affixal exponents of other categories. This is borne out by the data in Berg (2020) and Asao (2015: 24, 25). In both databases, indexation and possession are essentially the only domains that show a stronger proclivity for prefixation than does negation.

negator is an enclitic for the same reason that its affirmative marker was analyzed as one). At the very least, this positional variation seems to mean that affirmative markers and negators do not reliably show a strict paradigmatic organization. On the whole, these patterns argue against the idea that affirmative markers are created on the model of previously existing negators. Rather, as briefly suggested in Section 3.6, the possibly distorting effect of the negative cycle is not to be discounted here. That is, negators come and go more often than affirmative markers, and their age presumably impacts their degree of fusion. At the same time, degree of fusion also correlates with position in that grammatical markers generally show a more pronounced postposing bias on the morphological than on the syntactic level (Bybee et al. 1990; Asao 2015; Berg 2020). The heterogeneous pattern of the negators above is thus the more or less expected effect of the various forces that shape them.

Before the question of why the affirmative markers are predominantly prefixes will be addressed, it should be highlighted that this type of exponence is doubly unexpected. On the one hand, it is remarkable that the affirmative markers tend to be morphologically bound at all. In Berg's (2020: 376) sample, the negator is a free word in 300 languages and an affix in only 173 languages. This further illustrates the formal differences between negators and affirmative markers. The second discrepancy is the violation of the general suffixing preference. For instance, the worldwide 71-language sample in Bybee et al. (1990: 4) produced 1,236 verbal suffixes and only 426 verbal prefixes. Other things being equal, there should thus be about three times as many affirmative suffixes as prefixes in the database, contrary to the actual results. A juxtaposition of Berg's (2020) negator data and the affirmative markers in the present study yields a statistically significant difference at the 0.01 level on a Fisher's exact test, while an analogous test of Bybee et al.'s (1990) affix data and the affirmative markers is marginally significant at the 0.1 level.²¹

Based on the above, the main issue remaining is why affirmative markers show a prefixing bias. While the database might appear too small to speak of any patterns and to permit general inferences, it is unclear how much larger it could grow given the rarity of the phenomenon. Therefore, an attempt at explanation does not seem premature. One possible avenue is that the affirmative markers take the whole

²¹ Note that the Karo clitic required slight adjustments for both tests. For the comparison with the negator data, I classified it as a syntactic item, even though clitics combine properties of morphological and syntactic elements. In the second test, it was accommodated by translating Bybee et al.'s (1990) dichotomy between prefixes and suffixes into a difference between prefixes and non-prefixes.

relevant template as their model. That is, the affirmative markers might have become prefixes because the corresponding template already contained a sufficient number of prefixes (and/or prefix slots) to accommodate them via some kind of analogy (cf. Weir 1986: 314–315; Grossman et al. 2018: 533 for that line of argument; but cf. also Enrique-Arias 2002: 14–15 for the opposite claim). This idea is based on the reasoning that languages are more likely to build on the kinds of structures that they already have than they are to create new ones (cf. the concept of ‘system-congruity’ in Wurzel 1987: 62–76). This would be compatible with the idea that affirmative markers are formed predominantly once the process of overspecification is already under way, which, in principle, is supported by the patterns in Table 4. Further evidence for this chronological argument is that some of the affirmative markers are optional, which argues against an advanced stage of grammaticalization (Lehmann 2015: 148–150).

It should be mentioned that prefixes also emerge because of highly particular, language-specific factors (Mithun 2003; Grossman & Polis 2018). Yet, it is unclear why five different languages should all at some point have met the specific structural requirements for the formally and functionally unlikely development of affirmative prefixes. Thus, a more general explanation such as the one above seems preferable. On the other hand, the fact remains that the mere availability of prefixes would not *explain* why the affirmative markers became prefixes as well, not least because some of them actually became suffixes regardless. It follows that a full explanation of the distribution of the affirmative markers will need to be multifactorial, incorporating both typological and language-specific facts. Note that the focus here was on prefixation because the existence of affirmative suffixes can always be attributed to the factors responsible for the general suffixing preference (Bybee et al. 1990; Himmelmann 2014; Asao 2015; Berg 2020).

Another relevant point concerns the fact that four of the five languages with affirmative prefixes have more suffix slots than prefix slots, which dovetails with the finding that prefixes are also paradigmatically less elaborated than suffixes (Berg 2022: 392–393). One consequence of this difference is that prefixes are more predictable, which might support them even though they occur in an otherwise disadvantaged position for affixes (cf. Asao 2015: 121). The notion of predictability should be of even greater interest in the case of affirmative markers because the minimal semantic oppositions of which they are a part should render them highly predictable, i.e., less informative (cf. Section 1). Yet, since they do not necessarily occur in the same structural positions as the relevant negators and may even share a

slot with semantically unrelated markers (e.g., Northern Gumuz), it is not obvious that they are more predictable and/or easier to process than members of other paradigms.

The notions of overspecification and systemic analogy implicate the diachronic dimension in that they rely on the idea that (elements like) affirmative markers emerge as part of broader historical processes. Unfortunately, most of the languages with affirmative markers are not sufficiently documented to permit any strong claims about the history of the affirmative markers, and the relevant sources do not attempt any explicit reconstructions. It is also telling that the comprehensive *World Lexicon of Grammaticalization* (Kuteva et al. 2019) does not contain a single entry on affirmative markers, either as sources or as targets. As such, the kind of overspecification account attempted here largely has to rely on synchronic data, with all the pitfalls that follow from that method. Once that is accepted, though, some ideas about how affirmative markers might grammaticalize suggest themselves.

The data collection process for this project unearthed a number of elements that might be marginal affirmative markers but for which the description was insufficiently explicit.²² The most relevant set of markers of this sort subsumes the ‘particles’ in Southeastern Tepehuan (stp; Uto-Aztecan, Tepiman; Willett 1991: 158), Halh Mongolian (khk; Mongolic-Khitani, Eastern Mongolic; Janhunen 2012: 258), and Koasati (cku; Muskogean, Alabaman-Koasati; Kimball 1985: 459–460). These elements might lend credence to Bybee et al.’s (1994: 19–21) ‘Parallel Reduction Hypothesis,’ according to which grammaticalizing markers undergo semantic and formal reduction in lockstep. This is because these items, unlike the affirmative markers in Section 3, are either free or lightly bound, and they are also ‘heavier’ than the affirmative markers in terms of their function. For instance, the Southeastern Tepehuan particle still seems to carry some modal force, while the Koasati marker might still be emphatic to some degree.

A more diverse set of ‘near-affirmative’ items helps to shed further light on what kinds of meanings the affirmative markers might have grammaticalized from (assuming that there is at least occasional continuity between the near-affirmative

²² For instance, Sumerian (sux; isolate; Middle East) might have had two affirmative prefixes, but the descriptive situation is too uncertain to admit those items into the database (cf. Jagersma 2010: 577–579). I thank Victoria Birkner and Alexander Steiner for pointing me toward a decision on this issue.

and the actual affirmative markers).²³ For instance, in Abui (abz; Timor-Alor-Pantar, Central Alor; Kratochvil 2007: 280–281), a polysemous marker that may also express the ‘affirmative’ function is transparently derived from the verb ‘say,’ while a so-called ‘end-of-event’ marker in Mina (Cameroon) (hna; Afro-Asiatic, Chadic), which is at least implicitly affirmative, might derive from a verb/copula (Frajzyngier et al. 2005: 259). Another origin of near-affirmative markers is the particle ‘yes,’ which was indicated twice (Etxepare 2003: 538 for Basque²⁴; Linn 2001: 320 for Yuchi²⁵). Of course, none of this qualifies as definitive evidence, but some of these paths would be plausible. In particular, ‘yes’ is functionally compatible, while copulas resemble affirmative markers in their degree of semantic bleaching. Since rather many languages have affirmative copulas,²⁶ they might ultimately be the most likely origin (also recall the suggestion along those lines for Anal in Ozerov 2019: 49).

Another conceivable source of affirmative markers, which emerges from the Karo and Male data in particular, is cumulative declarative/affirmative markers, perhaps via the reanalysis from declarative via declarative/affirmative to affirmative. At a broader level, every diachronic assumption about affirmative markers made here suggests that they develop over long periods of time and are late additions to the morphological inventory of a language. In particular, the account sketched here lines up with the approach in Harris (2008, 2010) as well as with Dahl’s (2004: Ch. 6) concept of ‘maturation.’ Both suggest that the synchronic rarity of any given phenomenon owes to the fact that it only comes about after a complex (and thus long) sequence of specific diachronic stages.

Finally, it should be possible to tie the different morphosyntactic restrictions of the affirmative markers to their historical emergence. While some of the distributional patterns described in Section 3 are almost polar opposites (e.g., Isimjeega-Rootigaanga excluding affirmative markers from conditional clauses but Southern Yukaghir preferring them in irrealis contexts), there is also one low-level pattern that came to light. Specifically, several of the restrictions suggest that affirmative markers are most likely to occur in simple independent clauses. On the one hand, this might again argue for their relative recency given that subordinate clauses have been

²³ An anonymous reviewer asks whether evidential markers might be sources of affirmative markers. I did not find any indication of such a link, either among the clear-cut or the more dubious cases.

²⁴ Basque; eus; isolate; Western Europe.

²⁵ Yuchi; yuc; isolate; Oklahoma.

²⁶ In Grambank, 47.9% of all languages (1,152/2,407) have a copula with predicate nominals (GB117).

claimed to be more conservative (Bybee 2002; but cf. also Jing et al. 2023). On the other hand, the context of simple independent clauses would also render copulas a more likely origin than, for instance, verbs of saying, the latter of which are involved in complex sentences almost by definition.

5. Conclusion

On the basis of a large and balanced sample, this work furnished empirical evidence for the common claim that affirmative markers, the mirror image of standard negators, are highly infrequent in the languages of the world. At the same time, the few cases that were found are sufficient to undermine the more radical idea that affirmative markers are categorically absent from grammatical systems. The theoretical approach used to inform these data was McWhorter's (2005) concept of overspecification, understood here as the combination of a high degree of verbal synthesis and a comparatively low number of zero markers. This idea manages to account rather well for the data, but it only provides necessary rather than sufficient criteria. The database further suggests that affirmative markers show strong areal biases (eastern Africa), are confined to smaller speech communities, and grammaticalize rather late as well as from a variety of sources. Further research on all these aspects is necessary to establish why affirmative markers are found where they are and to the extent that they are. In addition, the results obtained suggest that the historical, functional, and formal links between affirmative and declarative markers merit further study, as do the syntactic/semantic restrictions of the former.

As extensions of this work, other typologically unmarked meanings would be obvious targets. Many of the relevant functions were mentioned or discussed above: declarative, active, indicative, realis, present, and perfective in the verbal domain; nominative, absolutive, and singular in the nominal domain (cf. also Mayerthaler 1987: 48). Based on the data collection for this work and the available data in Grambank, the prediction is that overt coding will not be as infrequent for any of these functions as it is for affirmative markers (no more than 1.5% of all sample languages). Meanwhile, a serious competitor for overt affirmative markers in terms of worldwide rarity might be morphemes expressing the basic, 'positive' degree of adjectives (e.g., *fast* rather than *faster* or *fastest*). Like 'affirmative,' this semantic value is rarely even thought of as a grammatical function, let alone as part of a paradigm, and neither Mayerthaler (1981: 27) nor Bobaljik (2012: 32, fn. 5) find a single

instance in their respective samples. Finally, it would be of considerable interest if creoles turned out to have affirmative markers. Based on the argument presented here, creoles are particularly unlikely to check that box given that affirmative markers only seem to arise in languages that already have large amounts of (old) morphology. However, that prediction is influenced by the specific database obtained here. That is, creoles might have syntactically free affirmative markers instead, which would then require a reconsideration of the reasoning pursued in this work. Future research will hopefully address these issues.

Assuming a broader perspective, further work on rarely marked functions would also permit a more profound assessment of overspecification. In particular, the causes of overspecification would be relevant to sub-fields beyond morphosyntax in that they seem to support the assumption that some systems might be more speaker-friendly (by requiring relatively little coding) while others might be more listener-friendly (by requiring more extensive coding). In that sense, the most interesting consequence of empirical research on such phenomena might be a new operationalization of a time-honored typological idea (cf. Bisang 2020 for some discussion).

Acknowledgments

This work benefited from the support of many people. Pavel Ozerov handed me the best example of an affirmative marker that I ended up finding and also provided crucial additional information on Anal. Ludger Paschen's selfless literature review greatly strengthened the case for Kabardian presented here. Thomas Berg and the audience at the 21st International Congress of Linguists in Poznań kindly let me know which parts of the argument to highlight and which to background. Victoria Birkner and Alexander Steiner generously discussed the complex Sumerian situation with me. Anonymous reviewers provided feedback that allowed me to make significant improvements with regard to content, structure, and argumentation. However, none of these scholars necessarily agree with the interpretation of the data or the framework adopted here, and any errors and infelicities are solely my responsibility.

Abbreviations

1/2/3 = 1st/2nd/3rd person
ABS = absolutive

AFF = affirmative
AOR = aorist

CAUS = causative
COM = comitative

DECL = declarative	LOC = locative	PROG = progressive
DEF = definite	M = masculine	PS = primary suffix
ERG = ergative	NEG = negation	PSI = persistive aspect
EXCL = exclusive	N.FUT = non-future	Q = interrogative
FUT = future	NOM = nominative	REFL = reflexive
GEN = genitive	NON-AG = non-agentive	SAP = speech act participant
INTR = intransitive	OBL = oblique	SBJ = subject
IPFV = imperfective	PFV = perfective	SG = singular
IRR = irrealis	PL = plural	SS = secondary suffix

References

- Ahland, Colleen. 2012. *A grammar of Northern and Southern Gumuz*. Eugene: University of Oregon. (Doctoral dissertation.)
- Ahland, Michael. 2012. *A grammar of Northern Mao (Màwés Aas'è)*. Eugene: University of Oregon. (Doctoral dissertation.)
- Amha, Azeb. 2001. *The Maale language*. Leiden: Universiteit Leiden. (Doctoral dissertation.)
- Anderson, Cormac, Natalia Kuznetsova & Shelece Easterday. 2025. Rarities in phonetics and phonology: Exploring rarities. In Natalia Kuznetsova, Cormac Anderson & Shelece Easterday (eds.), *Rarities in phonetics and phonology: Structural, typological, evolutionary, and social dimensions*, 3–28. Berlin: Language Science Press.
- Applebaum, Ayla Ayda Bozkurt. 2013. *Prosody and grammar in Kabardian*. Santa Barbara: University of California. (Doctoral dissertation.)
- Aronoff, Mark & Nanna Fuhrhop. 2002. Restricting suffix combinations in German and English: Closing suffixes and the monosuffix constraint. *Natural Language and Linguistic Theory* 20(3). 451–490. DOI: 10.1023/A:1015858920912
- Asao, Yoshihiko. 2015. *Left-right asymmetries in words: A processing-based account*. Buffalo: State University of New York. (Doctoral dissertation.)
- Becker, Laura. 2024. Zero marking in inflection: A token-based approach. *Journal of Language Modeling* 12(2). 349–413.
- Belete, Alemgena. 2018. *Documentation and grammatical description of Kara*. Addis Ababa: University of Addis Ababa. (Doctoral dissertation.)
- Berdicevskis, Aleksandrs & Arturs Semenuks. 2020. Different trajectories of morphological overspecification and irregularity under imperfect language

- learning. In Peter Arkadiev & Francesco Gardani (eds.), *The complexities of morphology*, 283–305. Oxford: Oxford University Press.
- Berdicevskis, Aleksandrs & Arturs Semenuks. 2022. Imperfect language learning reduces morphological overspecification: Experimental evidence. *PLoS ONE* 17(1), e0262876. DOI: 10.1371/journal.pone.0262876
- Berg, Thomas. 2020. Ordering biases in cross-linguistic perspective: The interaction of serial order and structural level. *Linguistic Typology* 24(2). 353–397. DOI: 10.1515/Lingty-2019-2031.
- Berg, Thomas. 2022. Paradigmatic consequences of the suffixing preference. *Studies in Language* 46(2). 376–401. DOI: 10.1075/SI.20019.ber
- Bickel, Balthasar & Johanna Nichols. 2013. Inflectional synthesis of the verb. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/chapter/22>. (Accessed May 14, 2025.)
- Bisang, Walter. 2020. Radical analyticity and radical pro-drop scenarios of diachronic change in east and mainland southeast Asia, west Africa and pidgins and creoles. *Asian Languages and Linguistics* 1(1). 34–70. DOI: 10.1075/Alal.00002
- Bisnath, Felicia, Hannah Lutzenberger, Marah Jaraisy, Rehana Omardeen & Adam Schembri. 2026. Deconstructing notions of morphological ‘complexity’: Lessons from creole and sign languages. *Journal of Linguistics* 62(1), 177–207. DOI: 10.1017/S0022226725000180
- Bobaljik, Jonathan. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/9069.001.0001
- Börstell, Carl. 2021. *GlottoshinyR*. Available at: <https://linguistr.shinyapps.io/glottoshinyr/>. (Accessed May 14, 2025.)
- Bybee, Joan. 2002. Main clauses are innovative, subordinate clauses are conservative. In Joan Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse*, 1–17. Amsterdam: Benjamins.
- Bybee, Joan. 2006. Language change and universals. In Ricardo Mairal & Juana Gil (eds.), *Linguistic universals*, 179–194. Cambridge: Cambridge University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. DOI: 10.1017/Cbo9780511750526

- Bybee, Joan. 2011. Markedness: Iconicity, economy, and frequency. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 131–147. Oxford: Oxford University Press.
- Bybee, Joan. 2015. *Language change*. Cambridge: Cambridge University Press. DOI: 10.1017/Cbo9781139096768
- Bybee, Joan, William Pagliuca & Revere Perkins. 1990. On the asymmetries in the affixation of grammatical material. In William Croft, Keith Denning & Suzanne Kemmer (eds.), *Studies in typology and diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*, 1–42. Amsterdam: Benjamins.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Colarusso, John. 1992. *A grammar of the Kabardian language*. Calgary: University of Calgary Press.
- Comrie, Bernard, Martin Haspelmath & Balthasar Bickel. 2015. *The Leipzig glossing rules*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>. (Accessed September 18, 2025)
- Cristofaro, Sonia. 2014. Competing motivation models and diachrony: What evidence for what motivations? In Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (eds.), *Competing motivations in grammar and usage*, 282–298. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press. DOI: 10.1017/Cbo9780511840579
- Cysouw, Michael & Jan Wohlgemuth. 2010. The other end of universals: Theory and typology of rara. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory*, 1–10. Berlin: De Gruyter.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17(1/2), 79–106.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: Benjamins. DOI: 10.1075/Slcs.71
- Daniel, Michael & Yury Lander. 2011. The Caucasian languages. In Bernd Kortmann & Johan van der Auwera (eds.), *The languages and linguistics of Europe: A comprehensive guide*, 125–157. Berlin: De Gruyter Mouton.
- DeGraff, Michael. 2003. Against creole exceptionalism. *Language* 79(2). 391–410. DOI: 10.1353/Lan.2003.0114

- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108671040
- Dixon, R. M. W. 2012. *Basic linguistic theory*. Vol. 3. Oxford: Oxford University Press.
- Dryer, Matthew. 2005. Negative morphemes. In Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 454–457. Oxford: Oxford University Press.
- Dryer, Matthew. 2013. Order of negative morpheme and verb. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <https://wals.info/feature/143>. (Accessed May 14, 2025.)
- Dryer, Matthew & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info> (Accessed September 14, 2025.)
- Enrique-Arias, Andrés. 2002. Accounting for the position of verbal agreement morphology with psycholinguistic and diachronic explanatory factors. *Studies in Language* 26(1). 1–31.
- Etxepare, Ricardo. 2003. Negation. In José Ignacio Hualde & Jon Ortiz De Orbina (eds.), *A grammar of Basque*, 516–564. Berlin: Mouton De Gruyter.
- Fenk-Oczlon, Gertraud. 1991. Frequenz und Kognition – Frequenz und Markiertheit. *Folia Linguistica* 25(3–4). 361–394.
- Frajzyngier, Zygmunt, Eric Johnston & Adrian Edwards. 2005. *A grammar of Mina*. Berlin: Mouton De Gruyter. DOI: 10.1515/9783110893908
- Gelderen, Elly van. 2008. Negative cycles. *Linguistic Typology* 12(2). 195–243. DOI: 10.1515/Lity.2008.037
- Givón, Talmy. 1991. Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language* 15(2). 335–370. DOI: 10.1075/sl.15.2.05giv
- Givón, Talmy. 2001. *Syntax: An introduction*. Vol. 1. Amsterdam: Benjamins. DOI: 10.1075/Z.Syn1
- Greenberg, Joseph. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.), *Universals of language*, 2nd edn., 73–113. Cambridge, MA: The MIT Press.
- Greenberg, Joseph. 2005 [1966]. *Language universals, with special reference to feature hierarchies*. Berlin: Mouton De Gruyter.

- Griscom, Richard. 2019. *Topics in Asimjeeg Datooga verbal morphosyntax*. Eugene: University of Oregon. (Doctoral dissertation.)
- Grossman, Eitan & Stéphane Polis. 2018. Swimming against the typological tide or paddling along with language change? Dispreferred structures and diachronic biases in affix ordering. *Journal of Historical Linguistics* 8(3). 388–443. DOI: 10.1075/Jhl.17027.Gro
- Grossman, Eitan, Anton Antonov & Guillaume Jacques. 2018. A cross-linguistic rarity in synchrony and diachrony: Adverbial subordinator prefixes exist. *STUF* 71(4). 513–538.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2026. *Glottolog* 5.3. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://glottolog.org> (Accessed June 22, 2026.)
- Handschuh, Corinna. 2014. *A typology of marked-S languages*. Berlin: Language Science Press. DOI: 10.17169/Fudocs_Document_000000019858
- Harris, Alice. 2008. On the explanation of typologically unusual structures. In Jeff Good (ed.), *Linguistic universals and language change*, 54–76. Oxford: Oxford University Press.
- Harris, Alice. 2010. Explaining typologically unusual structures: The role of probability. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory*, 91–103. Berlin: De Gruyter.
- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70. DOI: 10.1017/S0022226705003683
- Haspelmath, Martin. 2018. The last word on polysynthesis: A review article. *Linguistic Typology* 22(2). 307–326. DOI: 10.1515/Lingty-2018-0011
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. DOI: 10.1017/S0022226720000535
- Himmelmann, Nikolaus. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960. DOI: 10.1353/Lan.2014.0105
- Hopper, Paul & Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Jagersma, Abraham. 2010. *A descriptive grammar of Sumerian*. Leiden: Universiteit Leiden. (Doctoral dissertation.)
- Janhunen, Juha. 2012. *Mongolian*. Amsterdam: Benjamins. DOI: 10.1075/Loall.19

- Jing, Yingqi & Paul Widmer & Balthasar Bickel. 2023. Word order evolves at similar rates in main and subordinate clauses: Corpus-based evidence from Indo-European. *Diachronica* 40(4). 532–556. DOI: 10.1075/Dia.20035.Jin
- Kimball, Geoffrey. 1985. *A descriptive grammar of Koasati*. New Orleans: Tulane University. (Doctoral dissertation.)
- König, Christa. 2006. Marked nominative in Africa. *Studies in Language* 30(4). 655–732. DOI: 10.1075/Sl.30.4.02kon
- Kratochvil, František. 2007. *A grammar of Abui, a Papuan language of Alor*. Utrecht: LOT.
- Kumakhov, Muchadin. 1989. *Sravnitel'no-Istoricheskaja Grammatika Adygsikx (Cherkessikx) Jazykov*. [Comparative-historical grammar of the Adyghe (Cherkess) languages]. Moscow: Nauka.
- Kumakhov, Muchadin. 2006. *Kabardino-Cherkesskij Jazyk: V Dvukh Tomakh. Tom I*. [Kabardian-Cherkess language in two volumes. Volume 1: Writing system, phonetics and phonology, morphology, syntax]. Moscow: Nalchik.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.
- Lacroix, René. 2009. *Description du dialecte Laze d'Arhavi (caucasique du sud, Turquie): Grammaire et textes*. [Description of the Laz dialect of Arhavi (South Caucasian, Turkey): Grammar and texts.] Lyon: Université Lumière Lyon 2. (Doctoral dissertation.)
- Lacroix, René. 2019. Laz. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of western Asia: An areal perspective*, 830–860. Berlin: De Gruyter Mouton.
- Lehmann, Christian. 2004. Interlinear morphemic glossing. In Geert Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas & Wolfgang Kesselheim (eds.), *Morphology: An international handbook on inflection and word-formation*, Vol. 2, 1834–1857. Berlin: De Gruyter.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*. 3rd edn. Berlin: Language Science Press. DOI: 10.5281/Zenodo.7520796
- Linn, Mary. 2001. *A grammar of Euchee (Yuchi)*. Lawrence: University Of Kansas. (Doctoral dissertation.)
- Malchukov, Andrej. 2011. Interaction of verbal categories: Resolution of infelicitous grammeme combinations. *Linguistics* 49(1). 229–282. DOI: 10.1515/Ling.2011.006

- Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir*. Berlin: Mouton De Gruyter. DOI: 10.1515/9783110197174
- Mayerthaler, Willi. 1981. *Morphologische Natürlichkeit [Morphological Naturalness]*. Wiesbaden: Athenaion.
- Mayerthaler, Willi. 1987. System-independent morphological naturalness. In Wolfgang Dressler (ed.), *Leitmotifs in natural morphology*, 25–58. Amsterdam: Benjamins.
- McWhorter, John. 2005. *Defining creole*. Oxford: Oxford University Press.
- McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.
- Miestamo, Matti. 2005. *Standard negation*. Berlin: De Gruyter.
- Miestamo, Matti. 2007. Symmetric and asymmetric encoding of functional domains, with remarks on typological markedness. In Matti Miestamo & Bernhard Wälchli (eds.), *New challenges in typology: Broadening the horizons and redefining the foundations*, 293–314. Berlin: De Gruyter.
- Miestamo, Matti. 2010. Negatives without negators. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory*, 169–194. Berlin: De Gruyter.
- Mihas, Elena & Gregorio Santos Pérez. 2015. *A grammar of Alto Perené (Arawak)*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110766301
- Mithun, Marianne. 2003. Why prefixes? *Acta Linguistica Hungarica* 50(1/2). 155–185.
- Ngo, Binh. 2021. *Vietnamese: An essential grammar*. Routledge: London.
- Nguyen, Đình Hòa. 1997. *Vietnamese*. Amsterdam: Benjamins. DOI: 10.1075/Loall.9
- Ozerov, Pavel. 2018. Tone assignment and grammatical tone in Anal (Tibeto-Burman). *Studies in Language* 42(3). 708–733.
- Ozerov, Pavel. 2019. Person indexation in Anal. *Himalayan Linguistics* 18(1). 26–53.
- Pilot-Raichoor, Christiane. 2010. The Dravidian zero negative: Conceptualisation and diachronic context of its morphogenesis. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rara and rarissima: Documenting the fringes of linguistic diversity*, 267–304. Berlin: De Gruyter.
- Shcherbakova, Olena, Susanne Maria Michaelis, Hannah Haynie, Sam Passmore, Volker Gast, Russell Gray, Simon Greenhill, Damián Blasi & Hedvig Skirgård. 2023. Societies of strangers do not speak less complex languages. *Science Advances* 9. DOI: 10.1126/Sciadv.Adf7704

- Skirgård, Hedvig, Hannah Haynie, Damián Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bower, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy Bakker, Giulia Barbos, Robert Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto Herrera Miranda, Nataliia Hübler, Biu Huntington-Rainey, Jessica Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard Kowalik, Olga Krasnoukhova, Nora Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tônia Martins, Celia Mata German, Suzanne Van Der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney Rey, Julia Rizaew, Eloisa Ruppert, Kim Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário De Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉靖婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen Levinson, Martin Haspelmath, Simon Greenhill, Quentin Atkinson & Russell Gray. 2023. Grambank reveals global patterns in the structural diversity of the world's languages. *Science Advances* 9. DOI: 10.1126/Sciadv.Adg6
- Spencer, Andrew & Ana Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press. DOI: 10.1017/Cbo9781139033763
- Thompson, Lawrence. 1987. *A Vietnamese reference grammar*. Honolulu: University of Hawai'i Press.
- Vázquez Álvarez, Juan Jesús. 2011. *A grammar of Chol, a Mayan language*. Austin: University of Texas. (Doctoral dissertation.)
- Velupillai, Viveka. 2012. *An introduction to linguistic typology*. Amsterdam: Benjamins.
- Weir, Helen. 1986. Diachronic development of certain verb prefixes in an OSV language (Nadëb). *Lingua* 68(4). 291–316.
- Willett, Thomas. 1991. *A reference grammar of Southeastern Tepehuan*. Dallas: Summer Institute of Linguistics.

- Wurzel, Wolfgang. 1987. System-dependent morphological naturalness in inflection. In Wolfgang Dressler (ed.), *Leitmotifs in Natural Morphology*, 59–96. Amsterdam: Benjamins.
- Wurzel, Wolfgang. 2001. *Flexionsmorphologie und Natürlichkeit: Ein Beitrag zur morphologischen Theoriebildung*. [Inflectional morphology and naturalness: A contribution to the formation of morphological theory]. 2nd edn. Berlin: Akademie Verlag. DOI: 10.1515/9783050080130
- Zeijlstra, Hedde. 2007. Negation in natural language: On the form and meaning of negative elements. *Language and Linguistics Compass* 1(5). 498–518. DOI: 10.1111/J.1749-818x.2007.00027.X
- Zingler, Tim. 2022. Clitics, anti-clitics, and weak words: Towards a typology of prosodic and syntagmatic dependence. *Language and Linguistics Compass*, e12453. DOI: 10.1111/lnc3.12453
- Zipf, George. 1965 [1935]. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: The MIT Press.
- Zwicky, Arnold & Geoffrey Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59(3). 502–513. DOI: 10.2307/413900

Appendix. Languages surveyed.

The tables below list the languages for which I studied reference grammars. They are alphabetically sorted by macro-area and, within macro-areas, by language family. Glossonyms in square brackets reflect the names used by the relevant source in cases where that name differs from the one used in Glottolog. Boldfaced languages are the ones discussed in Section 3.

Family	Languages
Afro-Asiatic	Buwal, Coptic, Dangaleat [East Dangla], Ghomara (Berber), Goemai, Hdi, Jibbali, Konso, Mpade [Makary Kotoko], Mehri, Mina (Cameroon), Miya, Moloko, Sukur [Sakun], Senhaja de Srair [Senhaja Berber], Sidamo [Sidaama], Somali, Standard Arabic, Tamasheq [Tamashek], Tigre, Wandala
Atlantic-Congo	Aghem, Bena (Tanzania) [Bena], Bullom So [Mani], Chakali, Chuwabu [Cuwabo], Emai-Iuleha-Ora [Emai], Etulo, Ewe, Fon [Fongbe], Fwe, Ga [Gã], Gbeya, Gyele [Gyeli], Igbo, Ikpana (Logba), Liko, Lunda, Makonde, Mambai [Mambay], Mankanya, Mundabli-Mufu [Mundabli], Mungbam, Ncane-Mungong [Nchane], Ndengereko [Ndengeleko], Ndut, Noon, Southern Kisi [Kisi], Supyire Senoufo [Supyire], Swahili, Tafi, Tigon Mbembe [Mbembe]

Family	Languages
Bangime	Bangime
Blue Nile Mao	Bambassi [Northern Mao]
Central Sudanic	Ma'di
Dizoid	Nayi [Naayì], Sheko
Dogon	Jamsay Dogon [Jamsay], Tommo So Dogon [Tommo So]
Eastern Jebel	Gaam [Gaahmg]
Furan	Fur
Gumuz	Northern Gumuz, Southern Gumuz
Kadugli-Krongo	Krongo
Khoe-Kwadi	Kxoe [Khwe], Nama (Namibia) [Nama Hottentot], Northern Tshwa [Tjwao], Ts'ixa
Koman	Uduk
Kru	Grebo
Kuliak	Ik, Nyang'i
Kxa	Amkoe [ḥHōā]
Maban	Maba (Chad)
Mande	Jalkunan, Seeku [Seenku], Yalunka [Jalonke]
Narrow Talodi	Lumun
Nilotic	Isimjeega-Rootigaanga [Datooga (Asimjeeg)] , Lango (Uganda), Lopit, Luo (Kenya and Tanzania) [Dholuo], Luwo, Teso [Ateso], Turkana
Nubian	Kenuzi-Dongola [Kunuz]
Saharan	Dazaga, Kanuri (Central)
Sandawe	Sandawe
Shabo	Shabo [Chabu]
Songhay	Koyraboro Senni Songhai [Koyraboro (Koroboro) Senni], Koyra Chiini Songhay [Koyra Chiini], Tagdal
South Omotic	Dime, Hamar-Banna [Hamar], Karo (Ethiopia) [Kara]
Surmic	Majang, Mursi
Ta-Ne-Omotic	Kachama-Ganjule-Haro [Ganta], Male (Ethiopia) [Maale] , Zayse-Zergulla [Zaysete]

Table 6: Language sample: Africa. [28 families, 101 languages].

Family	Languages
Bunaban	Gooniyandi
Gaagudju	Gaagudju
Garrwan	Garrwa (Western)
Gunwinyguan	Bininj Kun-wok [Bininj Gun-wok], Anindilyakwa [Enindhilyakwa], Kunbarlang, Ngalakgan [Ngalakan], Rembarrnga [Rembarnga], Wubuy [Nunggubuyu]

Family	Languages
Iwaidjan Proper	Mawng
Mangarrayi-Maran	Mangarrayi
Maningrida	Guragone [Gurr-goni], Nakara [Nakkara]
Mirndi	Jingulu, Wambaya
Nyulnyulan	Bardi, Nyigina, Nyulnyul, Yawuru
Pama-Nyungan	Eastern Arrernte [Mparntwe], Dieri [Diyari], Djambarrpuyngu [Djamparrpuynu], Dyirbal, Gurindji, Guugu Yimidhirr, Jaru [Wanyjirra], Kuku-Yalanji, Mangala [Mangarla], Martuthunira, Ngardi, Ngarinman [Bilinarra], Ngiyambaa [Ngiyamba], Nhanda, Thayore [Kuuk Thaayorre], Martu Wangka [Wangkajunga], Warlmanpa, Warrongo, Wik-Mungkan, Wirangu, Yidiñ [Yidip], Yintyinka-Ayabadhu [Yintyingka]
Southern Daly	Murriny Patha, Nangikurrunggurr [Ngan'gityemerri]
Tangkic	Kayardild, Lardil [Yukulta]
Tiwi	Tiwi
Western Daly	Ami [Emmi], Marithiel [Marriithiyel]
Worrorran	Ngarinyin [Ungarinjin], Worrorra
Yangmanic	Wardaman

Table 7: Language sample: Australia. [16 families, 51 languages].

Family	Languages
Abkhaz-Adyge	Abkhaz, Kabardian
Austroasiatic	Chut [May], Ho, Jehai [Jahai], Kharia, Khmer (Central) [Cambodian], Khmu, Koho [Koho-Sre], Pacoh, Pnar, Semelai, Vietnamese
Basque	Basque
Burushaski	Burushaski (Eastern)
Chukotko-Kamchatkan	Chukchi
Dravidian	Kannada, Betta Kurumba, Kurukh [Kurux]
Great Andamanese	Akajeru
Hmong-Mien	Iu Mien, Western Xiangxi Miao [Xong]
Indo-European	Afrikaans, Bengali, Breton, Catalan, Czech, Danish, Domari, Dutch, Eastern Armenian, Eastern Panjabi [Panjabi], French, Greek, Hindi, Hittite, Icelandic, Irish, Laki, Latin, Latvian, Lithuanian, Maithili, Marathi, Northern Tosk Albanian, Norwegian, Polish, Portuguese, Rajbanshi, Romani, Romanian, Russian, Serbian-Croatian-Bosnian [Bosnian/Croatian/Montenegrin/Serbian], Sinhala, Slovenian [Slovene], Spanish, Swedish, Welsh, Western Balochi [Modern Standard], Western Farsi [Persian]
Japonic	Japanese, Miyako [Irabu]

Family	Languages
Jarawa-Onge	Jarawa (India) [Jarawa]
Kartvelian	Georgian, Laz
Koreanic	Korean, Jejueo
Kusunda	Kusunda
Mongolic-Khitani	Bonan [Bao'an Tu], Mangghuer, Halh Mongolian [Mongolian]
Nakh-Daghestanian	Hinuq, Hunzib, Ingush, Khinalug [Xinaliq], Khwarshi-Inkhoqwari [Khwarshi], Lezgian, North-Central Dargwa [Mehweb], Southwestern Dargwa [Sanzhi Dargwa]
Nivkh	Amur Nivkh
Sino-Tibetan	Ao Naga [Mongsen Ao], Bantawa, Burmese, Mandarin Chinese, Darma, Dhimal, Dumi, Dura, Ersu, Galo, Guqiong, Hills Karbi [Karbi], Japhug, Katso [Khatso], Kurtokha [Kurtöp], Lahu, Lepcha, Limbu, Manange, Manipuri [Meithei], Nocte Naga [Hakhun Tangsa], Purik-Sham-Nubra [Purik], Qiang, Sichuan Yi [Nuosu], Southern Pumi [Prinmi], Solu-Khumbu Sherpa [Sherpa], Sunwar, Tibetan, Tshangla, Western Parbate Kham [Kham], Yakkha
Sumerian	Sumerian
Tai-Kadai	Lakkia [Lakkja], Lao, Thai, Turung
Tungusic	Evenki, Manchu, Udihe
Turkic	Kazakh, Turkish
Uralic	Finnish, Hungarian, Mansi (Central), Nenets (Tundra), Nganasan, North Saami [North Sámi], Pite Saami, Skolt Saami, Vach-Vasjogan [Eastern Khanty]
Yeniseian	Ket, Yugh
Yukaghir	Northern Yukaghir, Southern Yukaghir [Yukaghir (Kolyma)]

Table 8: Language sample: Eurasia. [25 families, 134 languages].

Family	Languages
Algic	Arapaho, Cree (Plains), Cree (Woods), Ottawa [Nishnaabemwin], Malecite-Passamaquoddy [Passamaquoddy-Maliseet], Potawatomi, Shawnee, Siksika [Blackfoot], Yurok
Athabaskan-Eyak-Tlingit	Chilcotin-Nicola [Tsilhqút'in], Eel River Athabaskan [Wailaki], Hupa-Chilula [Hupa], North Slavey [Slave], Sarsi [Sarcee], Tanacross, Upper Tanana
Caddoan	Caddo
Chibchan	Pech, Rama, San Blas Kuna [Guna], Teribe
Chimariko	Chimariko
Cochimi-Yuman	Havasupai-Walapai-Yavapai [Hualapai], Maricopa, Tipai [Jamul Tiipay]
Eskimo-Aleut	Alaskan Iñupiaq (North), Aleut, Central Alaskan Yupik, Eastern Canadian Inuktitut, Kalaallisut [West Greenlandic], Seward Alaska Inupiatun [Iñupiaq]
Huavean	Huave (San Francisco del Mar), Huave (San Dionisio del Mar)

Family	Languages
Iroquoian	Cayuga, Cherokee (Oklahoma), Onondaga, Seneca, Huron-Wyandot [Wyandot, Wendat]
Karuk	Karuk [Karok]
Keresan	Western Keres [Acoma (Keresan), Laguna Keres]
Kiowa-Tanoan	Kiowa, Towa
Kutenai	Kutenai
Mayan	Chol, Itza' [Itzaj], Huastec (Southeastern, Potosino), Mam, Popti' [Jacaltepec], Sacapulteco [Sacapultec], Tz'utujil, Yucatec Maya
Misumalpan	Ulwa
Miwok-Costanoan	Mutsun
Mixe-Zoque	Chimalapa Zoque, Highland Popoluca [Sierra Popoluca], Northeast Chiapas Zoque, Tlahuitoltepec Mixe [Ayutla Mixe]
Molale	Molale [Molalla]
Muskogean	Alabama, Choctaw, Creek, Koasati, Mikasuki
Otomanguean	Atzingo Matlatzinca [Ocuilteco], Coatlán-Loxicha Zapotec, Nopala Chatino [Teotepec Eastern], Northern Pame, Sochiapam Chinantec [Sochiapan Chinantec], Tataltepec Chatino, Western Highland Chatino (Yaitepec), Zacatepec Chatino (Eastern), Zenzontepec Chatino
Pomoan	Kashaya, Southern Pomo
Sahaptian	Nez Perce, Northwest Sahaptin [Yakima Ichishkiin/Sahaptin]
Salishan	Bella Coola, Columbia-Wenatchi [Nxa'amxcin], Comox, Lillooet, Halkomelem [Musqueam], Spokane [Spokan], Squamish
Siouan	Assiniboine, Biloxi, Crow, Hidatsa, Lakota [Lakhota]
Tarascan	Purepecha
Totonacan	Filomena Mata Totonac, Huehuetla Tepehua, Tlachichilco Tepehua, Yecuatla [Misantla] Totonac
Tsimshian	Gitksan [Gitksan], Nisga'a [Nisgha], Southern-Coastal Tsimshian [Coast Tsimshian]
Uto-Aztecan	Comanche, Cupeño, Huarijio [River Warihío], Hopi, Northern Paiute, Panamint [Tümpisa Shoshone], Pipil, Tarahumara (Central) [Choguita Rarámuri], Southeastern Tepehuan, Tohono O'odham, Ute-Southern Paiute [Ute]
Wakashan	Nuu-chah-nulth
Washo	Washo
Wintuan	Wintu
Yuchi	Yuchi
Yuki-Wappo	Wappo, Northern Yukian [Yuki]

Table 9: Language sample: North America. [33 families, 113 languages].

Family	Languages
Abun	Abun
Angan	Menya
Anim	Marind [Coastal Marind]
Austronesian	Acehnese, Balantak, Belep, Bunun, Cebuano, Chamorro, Dakaka [Daakaka], Dela-Oenale [Dela], East Makian [Taba], Fijian (Boumaa), Filipino, Hawaiian, Irarutu, Kambera, Kéo, Kilivila, Kokota, Kosraean [Kusaiean], Lamaholot, Madurese, Makasar [Makassarese], Malay (Papuan), Malay (Baba), Mafea [Mavea], Manam, Maori, Muna, Nafsan [South Efate], Neverver, Paiwan, Papanana, Pohnpeian [Ponapean], Puyuma, Rapanui [Rapa Nui], Saaroa [Lha'alua], Standard Indonesian [Indonesian], Tagalog, To'abaita [Toqabaqita], Tsat [Hainan Cham], Tukang Besi, Tuvalu [Tuvaluan], Ughele, Vaeakau-Taumako, Vurës, Western Subanon
Baibai-Fas	Momu-Fas [Momu]
Baining	Mali, Qaqet
Bilua	Bilua
Border	Imonda, Kilmeri
Bosavi	Aimele [Eimela]
Bulaka River	Wanam Yelmek
Duna	Duna
East Bird's Head	Meyah, Moskona
East Strickland	Konai
Fuyug	Fuyug
Hatam-Mansim	Hatam
Inanwatan	Suabo
Keram	Pondi, Ulwa (Papua New Guinea)
Kuot	Kuot
Lakes Plain	Abawiri
Lavukaleve	Lavukaleve
Lower Sepik	Kopar, Yimas
Manubaran	Doromu-Koki
Maybratic	Maybrat
Ndu	Ambulas, Iatmul, Manambu
Nimboranic	Nimboran
North Bougainville	Rotokas
North Halmahera	Ternate, Tidore
Nuclear Torricelli	Abu' Arapesh [Abu'], Bukiyip, Kamasau, Yeri

Family	Languages
Nuclear Trans New Guinea	East Kewa [Kewapi], Fore, Huli, Kesawai [Koromu (Kesawai)], Lower Grand Valley Dani, Ma Manda, Mauwake, Mian, Oksapmin, Tauya, Usan, Wano, Yagaria [Hua], Yau-Nungon
Savosavo	Savosavo
Senagi	Dera (Indonesia) [Menggwa Dla]
Sepik	Alamblak, Awtuw, Mehek
Sko	Bauni [Barupu], I'saka, Skou, Wutung
South Bougainville	Siwai [Motuna]
Taulil-Butam	Taulil [Tulil]
Taiap	Taiap [Tayap]
Timor-Alor-Pantar	Abui, Adang, Bunak [Bunaq], Klon, Makasae-Makalero, Reta, Teiwa
West Bomberai	Kalamang
Yam	Komnzo, Ngkontar Ngkolmpu [Ngkolmpu]
Yele	Yele

Table 10: Language sample: Papunesia. [40 families, 120 languages].

Family	Languages
Araucanian	Mapudungun [Mapuche]
Arawakan	Achagua, Apurinã, Ajiúninka Apurucayali [Axininca Campa], Ashéninka Perené [Alto Perené (Arawak)], Baure, Garifuna, Lokono [Arawak Lokono Dian], Nanti, Parecís [Paresi-Haliti], Paunaka, South Ucayali Ashéninka [Ashéninka], Tariana, Yine (Piro)
Arawan	Culina [Kulina], Madi [Jarawara]
Aymaran	Southern Aymara [Aymara Muylaq'], Jaqaru
Barbacoan	Awa-Cuaiquer [Awa Pit]
Boran	Bora [Miraña]
Bororoan	Bororo
Cahuapanan	Jebero
Camsá	Camsá [Kamsá]
Cariban	Bakairí [Bakairi], Galibi Carib [Carib], Hixkaryána [Hixkaryana], Panare, Trió [Tiriyó], Waimiri-Atroari, Wayana
Chapacuran	Wari'
Chicham	Aguaruna [Aguaruna (Iiniá Chicham)], Achuar-Shiwiari [Shiwiari], Huambisa [Wampis]
Guaicuruan	Kadiwéu, Mocoví, Pilagá, Toba
Hibito-Cholón	Cholón
Huitotoan	Murui Huitoto [Murui Bue]

Family	Languages
Kakua-Nukak	Kakua
Kwaza	Kwaza
Lengua-Mascoy	Enxet Sur
Máku	Máku
Mataguyan	Nivaclé [Nivacle], Vezoz [Wichí], Weenhayek [Wichi]
Mosetén-Chimané	Mosetén-Chimané [Mosetén]
Movima	Movima
Naduhup	Hup
Nambiquaran	Mamaindé [Mamaindê], Sabanê
Nuclear-Macro-Je	Apinayé [Apinajé], Karajá, Rikbaktsa
Pano-Tacanan	Cashibo-Cacataibo [Kakataibo], Cavineña, Chácobo, Ese Ejja, Matsés [Matses], Resígaro, Shipibo-Conibo [Shipibo-Konibo]
Peba-Yagua	Yagua
Puinave	Puinave
Quechuan	Huallaga Huánuco Quechua, Yauyos Quechua
Saliban	Maco [Mako]
Ticuna-Yuri	Tikuna [Ticuna]
Trumai	Trumai
Tucanoan	Cubeo [Kubeo], Desano, Kotiria
Tupian	Akuntsu [Akuntsú], Cocama-Cocamilla [Kukama-Kukamiria], Kamayurá [Kamaiurá], Karitiâna [Karitiana], Karo (Brazil), Mundurukú, Paraguayan Guaraní [Guaraní], Tapieté [Tapiete], Tuparí
Urarina	Urarina
Yanomamic	Ninam [Yanam (Shiriana)], Yanomám
Yuracaré	Yuracaré [Yurakaré]

Table 11: Language sample: South America. [37 families, 85 languages].

CONTACT

tim.zingler@uibk.ac.at