A diachronic approach to the formal idiosyncrasies of indexes

Tim Zingler

Affiliation TBD

1 Introduction

This work builds on the insight that indexes (i.e., argument-indexing agreement markers and/or pronouns) show a wider range of formal and distributional idiosyncrasies than the exponents of other inflectional categories (e.g., Julien 2002: ch. 5; Fuß 2005: 62-67). In order to support this claim, I will discuss indexes that are extrametrical with respect to reduplication and "mobile" affixes that can occur in different slots of otherwise identical words. In addition, I will illustrate indexes that can freely occur on either member of a phrase-level construction as well as indexes that behave like full-fledged affixes in one type of context but like clear-cut words in another. The claim that the range of these traits is unique to indexes is based on a larger project by the author, which in addition to indexes also investigates exponents of definiteness, case, and tense. The latter are all more homogeneous in their behavior than the indexes. The explanations for this discrepancy rest on the different diachronic pathways to which indexes are subject and which themselves constitute an important topic for further research. The overall database comprises 60 languages, which belong to 60 WALS genera and are evenly distributed across five macro-areas.

2 Data

For the purposes of this contribution, I will focus on the three phenomena that most clearly distinguish indexes from the other exponents: extrametricality, mobility, and duality. Each of these will be defined in the relevant sub-sections below. Section 3 will then argue that these data require diachronic explanations and will suggest such explanations for each of the patterns.

2.1 Extrametricality

The definition of extrametricality that I employ here is wider than usual. Whereas the concept is traditionally applied to strings that are irregular in that they fall outside a stress domain, I will also use this term here for strings that are irregular in that they fall outside a domain of reduplication. There are two indexes in my sample that fail to undergo reduplication processes that otherwise apply to all morphological items in the relevant position of the verb, whereas none of the definiteness, case, or tense markers show such behavior. This is illustrated below with data from Fwe (Atlantic-Congo; fwe; Gunnink 2018), where locative arguments are indexed on the verb via markers for the noun classes 16-18. The data of interest are given in (1) and (2).

(1) ndi-a-endí-end-i=ko
1SG-PST-RDP-go-PST=LOC$_{17}$
'I kept going there.'

(2) ndì-ngòngòt-á=hò
1SG-knock-FV=LOC$_{16}$
'I knock on it.' (Gunnink 2018: 272)

Pluractional reduplication in Fwe targets the verb stem, which includes the root as well as all inflectional and derivational suffixes (Gunnink 2018: 199-200, 249). As can be gleaned from (1), then, the class 17 index is not a suffix for the purposes of this process because the reduplicant *endí* consists of the root and the past tense morph but excludes the following locative index *ko*. Meanwhile, in (2), the high tone that usually co-expresses present tense on the final mora is on the penultimate mora because the verb is in clause-final position (Gunnink 2018: 272). Since the final mora corresponds to the class 16 locative index *ho*, however, the latter must be part of the phonological word for the purpose of tone assignment. Note that while Gunnink (2018: 271-272) classifies the locative indexes as enclitics, they always take up the final slot in the verb template and thus have the syntagmatic distribution of affixes. Finally, while the Fwe example centers on postposed elements, most other extrametrical items in my sample are preposed. The relevance of this distributional fact will be addressed in Section 3.

2.2 Mobility

Indexes also show the ability to take up one of several slots in both morphological templates and phrase-level syntactic constructions without bringing about a semantic difference between the resulting alternatives. It is this behavior for which I suggest the umbrella term "mobility" here. There are two clear cases of "mobile affixes" among the indexes in my sample, twice as many as there are among the remaining data combined. In addition, there are also two indexes that show syntactic mobility, which I did not find for any exponent of the other three categories. (Note that *multiple* case marking across an NP is a kind of concord, not mobility as defined here.)

A mobile affix as defined here can be found in San Francisco del Mar Huave (Huavean; hue; Kim 2008). The second-person marker can occur on either side of a subordinate verb, and the two orderings are explicitly described as equally acceptable (Kim 2008: 346). These two options are contrasted in (3) and (4) below.

(3)	m-e-chutu-r	(4)	chutu-m-ia-r	
	SB-2-sit-2.INTR		sit-SB-2-2.INTR	
	'that you (SG) sit'		'that you (SG) sit'	(Kim 2008: 347)

The segmental variation between the indexes follows from allophonic principles. That is, diphthongization affects vowels that precede a tautosyllabic plain consonant, as in (4), but not those in open syllables, as in (3); cf. Kim (2008: 53). The fact that the index is subject to such processes suggests that it is part of a larger phonological word. However, it constitutes an idiosyncrasy because it undermines the idea that phonological words map onto grammatical words whose constituents follow a rigid order (cf. Dixon & Aikhenvald 2003; Haspelmath 2011).

Mobility at (roughly) the level of a phrase can be seen in Lillooet (Salishan; lil; van Eijk 1997), where the third-person plural marker *wit* can occur either on the auxiliary or on the lexical verb of an otherwise identical construction. These possibilities are juxtaposed in (5) and (6).

(5)	waʔ-wit-ás=maɬ=ƛ'uʔ	ʔíƛ'əm	
	AUX-3PL-SBJV=HORT=well	sing	
	'Let them sing/they might as well sing.'		(van Eijk 1997: 153)

(6)	waʔ-as=máɬ=ƛ'uʔ	ʔíƛ'əm-wit	
	AUX-SBJV=HORT=well	sing-3PL	
	'Let them sing/they might as well sing.'		(van Eijk 1997: 153)

Here too, the index is not simply a free word because it crucially falls within a larger domain in terms of stress assignment. Primary stress (marked by an acute accent) usually falls on the first syllable whose nucleus is not a schwa, but it can also be marked on the third syllable provided this is not the last syllable of the phonological word (van Eijk 1997: 14, 17). Hence, the primary stress on the third syllable in (5) can only be explained if the preceding index accounts for the second syllable. In contrast to the San Francisco del Mar Huave item, however, the Lillooet index is an even more drastic idiosyncrasy because its freedom extends to the syntactic level. Note also that neither the Huave nor the Lillooet marker is adequately classified as a clitic. This is because neither element is limited to second position in the clause or to a specific position with respect to a phrase, whereas both indexes *are* limited in terms of their possible hosts/stems.

2.3 Duality

By "duality," I refer to the fact that some indexes behave like full words in some contexts but like prototypical affixes in others. As such, they clearly differ from clitics understood as "syntactic affixes" (cf. Anderson 2005), which always interact phonologically with a phrasal host and thus show the same behavior in all contexts. Note that while duality is also found among markers of the other categories, languages often have multiple paradigms of indexes (e.g., Cardinaletti & Starke 1999). This increases the likelihood that duality is more common in indexation than in other grammatical domains. Yet, further research on all aspects of this issue is required.

One index that clearly shows duality comes from the Umari Norte variety of Hup (Naduhup; jup; Epps 2008). The third-person singular marker typically occurs before the verb, but since it can also appear in other syntagmatic contexts it is not simply a prefix (cf. Epps 2008: 285, 755-756). The relevant contrast is illustrated in (7) and (8), which show the unmarked preverbal and the marked clause-final position of the index, respectively.

(7) “hɨ́t tã=hám-ã??” tɔ̃=nɔ-máh-ãh
where 3SG=go-Q 3SG=say-REP-DYN
‘“Where did he go?” he said.’ (Epps 2008: 135)

(8) maŋgă tá?-ay hɨ́d-ăn yamhidɔ?-nɨ́h tɨ́h?
Margarita RI-INCH 3PL-OBJ sing-NEG 3SG
‘What about Margarita, didn’t she sing to them?’ (Epps 2008: 172)

In (7), both tokens of the index undergo consonant cluster reduction, due to which they lose their final /h/, and vowel harmony, due to which their vowel qualities are assimilated to those of the following vowels. Both of these processes are limited to the phonological word (cf. Epps 2008: 103-104), which lends credence to the affix analysis of these variants. By contrast, the underlying form seen in (8) is not subject to any phonological process and can be freely placed within the clause in a way that is typically taken to define grammatical words. Hence, the behavior of the third-person index differs on both the phonological and the syntactic dimension depending on its syntagmatic context.

3 Explanation and discussion

What unites all the indexes analyzed above is that they behave like bound elements on some set of criteria and/or in some contexts but like free words on another set of criteria and/or in other contexts. These mismatches constitute the idiosyncrasies of interest here, and it will be assumed in this contribution that they ultimately come about because the indexes at issue are in the process of grammaticalizing from free pronouns to pronominal/agreement affixes (cf. Bybee 2015: 152-153). Yet, the different types of idiosyncrasies illustrated here must nevertheless have resulted from different diachronic trajectories, and these will be sketched in this section.

With respect to indexes that are extrametrical in terms of reduplication, it is important to bear in mind that free pronouns are typically emphatic in nature. Since this emphatic status is defined by segmental weight and the ability to bear prominence, pronouns will first have to lose the properties they share with phonological words before they can begin to integrate into another word domain (i.e., the verb). Meanwhile, markers of case and tense often derive from adpositions and auxiliaries, respectively, and both of these form classes are typically already phonologically reduced. Hence, indexes apparently have to go through more diachronic stages to become full-fledged affixes than do exponents of other categories, and this makes indexes more likely to show idiosyncratic behavior at any given point in time.

Most extrametrical indexes in my database are preposed, and this tendency holds for both reduplication and other (supra)segmental processes. Crucially, a preposed position is likely to limit the degree of formal integration in that preposed elements fuse with following items less easily than do postposed ones with preceding items (e.g., Himmelmann 2014). Given that indexes are preposed much more frequently than exponents of most other grammatical categories (cf. Siewierska 2004: 165), indexes thus face unique obstacles on their path toward full affix-hood. Finally, reduplication processes typically include a word edge as part of their target domain, and since indexes predominantly occur at word edges (Bybee 1985), they are simply more likely to fall within a domain of reduplication. This, in turn, is a logical prerequisite for being considered extrametrical with regard to such a domain. In sum, then, the interaction of reduplication, extrametricality, and indexes can be derived from general morphological and diachronic facts.

Word-internal mobility plausibly derives from “exuberant agreement” (cf. Harris 2008), in which the same index is marked in multiple locations of a single verb form. At that stage,

language users might reanalyze the templatic position of the index as flexible. If so, this would pave the way for a period during which a given index can interchangeably be marked in any of several slots. While the nature of word-external mobility is more obscure, it might be explained by the fact that agreement ultimately references a cognitive entity rather than a morphosyntactic unit (cf. Kibrik 2019). Once this basic assumption is granted, the grammatical properties of a given referent would be compatible with any member of the predicate because each member of the predicate can be tied to the referent in some form. That indexes are usually only marked once, on the verb, might then be due to economy considerations as well as to the fact that verbs are the most frequently available collocate and ultimately reanalyzed as the only possible one.

Finally, the phenomenon of duality as defined here is unremarkable on the recognition that the diachronic development from a syntactic construction to a morphologically complex string does not necessarily involve an intermediate clitic stage (cf. Lehmann 2020: 226). That is, to the extent that the fusion of formerly independent elements is conditioned by the token frequency of their collocation (e.g., Bybee 2002), frequent combinations as in (7) will show fusion whereas infrequent ones such as in (8) will not. It follows that, once duality is clearly distinguished from clitic-hood, the former might turn out to be cross-linguistically frequent in its own right.

In sum, the idiosyncratic behavior of indexes largely seems to derive from the fact that indexation is a more important communicative function than that expressed by other inflectional categories. As such, the grammatical properties of referents are often expressed multiple times per clause and/or emphasized, and the relevant constructions may then lead to extrametrical or mobile affixes, etc. This idea obviously needs to be fleshed out in more detail, but it promises to reveal an interesting correspondence between form and meaning in that a functionally more complex category might then also be expressed by a more complex set of exponents.

References

Anderson, Stephen. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Bybee, Joan. 1985. *Morphology*. Amsterdam: Benjamins.

Bybee, Joan. 2002. Sequentiality as the basis of constituent structure. In Talmy Givón & Bertrand Malle (eds.), *The evolution of language out of pre-language*, 107–132. Amsterdam: Benjamins.

Bybee, Joan. 2015. *Language change*. Cambridge: Cambridge University Press.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 145–233. Berlin: De Gruyter.

Dixon, R. M. W. & Alexandra Aikhenvald. 2003. Word: A typological framework. In R. M. W. Dixon & Alexandra Aikhenvald (eds.), *Word*, 1–41. Cambridge: Cambridge University Press.

Epps, Patience. 2008. *A grammar of Hup*. Berlin: De Gruyter.

Fuß, Eric. 2005. *The rise of agreement*. Amsterdam: Benjamins.

Gunnink, Hilde. 2018. *A grammar of Fwe*. PhD thesis, University of Gent.

Harris, Alice. 2008. Explaining exuberant agreement. In Thorhallur Eythórsson (ed.), *Grammatical change and linguistic theory*, 265–283. Amsterdam: Benjamins.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45. 31–80.

Himmelmann, Nikolaus. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90. 927–960.

Julien, Marit. 2002. *Syntactic heads and word formation*. Oxford: Oxford University Press.

Kibrik, Andrej. 2019. Rethinking agreement: Cognition-to-form mapping. *Cognitive Linguistics* 30. 37–83.

Kim, Yuni. 2008. *Topics in the phonology and morphology of San Francisco del Mar Huave*. PhD thesis, University of California, Berkeley.

Lehmann, Christian. 2020. Univerbation. *Folia Linguistica Historica* 41. 205–252.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Van Eijk, Jan. 1997. *The Lillooet language*. Vancouver: University of British Columbia Press.