Karsten Schmidtke-Bode and Natalia Levshina (Leipzig University)

Supplementary material SM4: On the distribution of DOM

In §4.2 of our paper, we briefly considered the world-wide distribution of differential object marking (DOM), asking whether DOM may be universally preferred over non-split P-marking. Since this question is not central to the major argument of §4.2 but nevertheless interesting in the present context, we have relegated it to this supplementary material.

The typological distribution of DOM is analyzed by both Sinnemäki (2014) and by Bickel et al. (2015) [henceforth BWZ, as in the paper]. Again, BWZ (p. 30–31) use Family-Bias estimations to calculate the likelihood of having split versus non-split P-marking systems in their sample. Their analysis reveals certain peaks for DOM in Eurasia and Sahul (just as for A-marking) and significantly fewer P-splits elsewhere. Furthermore, their Figure 2 (p. 31) suggests that the majority of taxa is biased towards non-split systems, i.e. no differential object marking in the first place, or at least that significant biases towards DOM are in the minority *everywhere*. Importantly, however, these estimations seem to take all sample languages into account; that is, languages without case marking are counted as non-split systems. In this sense, BWZ's figures are not comparable to those of Sinnemäki (2014), who asks for the likelihood of DOM if a language has developed an object case to begin with. In the following, we first present Sinnemäki's results and then re-examine BWZ's data from the same perspective as Sinnemäki.

Sinnemäki (2014) investigates a world-wide sample of 744 languages, which were selected in such a way that at least 75% of all known stocks from each macro area would be represented in the survey. In a nutshell, Sinnemäki finds that about 80% of the languages that actually have case marking for P work on a differential basis and do not extend case marking to all P arguments; in terms of taxa, this figure slightly reduces to 74% but still constitutes a clear majority. And since, crucially, this bias towards split-P marking holds similarly across all macro areas (Sinnemäki 2014: 298), he considers this strong evidence for the hypothesis that differential P-marking is actually universally preferred over consistent, i.e. non-split P-marking.

One may rightly point out here that Sinnemäki's figures are based on a broader understanding of DOM than in BWZ, insofar as several kinds of conditioning factor (i.e. referential properties of NPs as well as things like word order) are considered.[1] Therefore, when narrowed down to just animacy and definiteness (the focus of Sinnemäki's study), the number of DOM-languages governed by these two conditions (either alone or in conjunction) shrinks again (to 69%); but even when tested on taxa rather than languages, Sinnemäki still finds a highly significant tendency for DOM to be affected by (the combined effects of) animacy and definiteness, especially also in areas other than Eurasia and the Sahul. Furthermore, considering that Sinnemäki only takes splits in full NPs into account, neglecting possible pronominal splits, it is likely that his

---

[1] Sinnemäki uses the term "restricted object marking" for this broader understanding of DOM, but we will continue to employ the term "differential object marking" here even when it is about splits that are not conditioned by referential properties.

results still underestimate the cross-linguistic dominance of referentially conditioned splits in P-marking. In sum, then, Sinnemäki's figures lead us to suspect that, if a language develops object marking, differential flagging is more likely than non-split flagging.

Let us now see if we can find support for this in BWZ's data. Here is our modus operandi: Firstly, we sifted the data in such a way that all languages in which P is *generally* "unmarked" (i.e. across all referential categories) were discarded. In this way, we effectively excluded all languages without case marking for P. (Since markedness in BWZ refers to alignment sets rather than overt coding material, it is theoretically possible that a P-argument coded as generally "unmarked" actually has an overt case marker, but one that is shared with S (e.g. a non-zero Absolutive case). These were discarded along with the zero ones but since our major goal here is to compare the results to Sinnemäki (2014) and he proceeded in the exact same way, this is actually an advantage.) Secondly, the remaining languages (i.e. all those with "marked" Ps somewhere in their system) were coded for whether they work differentially – exhibiting referentially conditioned P-splits – or consistently – applying P-marking across all referential categories coded in the database. Thirdly, we examined the distribution of differential versus consistent marking at the level of taxa, basically asking: "If a taxon has P-marking languages anywhere, do these show consistent P-marking, differential P-marking or a mixture of both?" Table 1 below displays the results:

Table 1. Taxa with consistent, differential or mixed object marking

	Eurasia	Africa	Americas	Sahul	Total
Consistent	2	6	9	3	20
Differential	4	6	11	7	28
Both	7	3	2	2	14
Total	13	15	22	12	62

Note that our category "both" does not say anything about the internal distribution of the taxa in question, i.e. they may be either "biased" or "diverse" in BWZ's sense. If we take the majority value of those taxa into account (i.e. a simple majority rather than a significant bias in BWZ's sense), the following picture emerges (Table 2):

Table 2. Taxa with *preferentially* consistent, differential or mixed object marking

	Eurasia	Africa	Americas	Sahul	Total
Majority consistent	3	6	10	3	22
Majority differential	9	6	11	8	34
50:50 distribution	1	3	1	1	6
Total	13	15	22	12	62

Alternatively, one could also apply Bickel's Family Bias Method to the data. That is, we can carry out the same analysis as in BWZ, but discard the languages without case marking. The result is shown in Table 3:

Table 3. Taxa with significant *family biases* for consistent or differential object marking

	Eurasia	Africa	Americas	Sahul	Total
Bias: consistent	3.09	3.77	5.2	2.56	14.6
Bias: differential	8.35	3.75	5.76	5.77	23.6
No bias (i.e. diverse)	1.56	7.48	11	3.67	23.8
Total	13	15	22	12	62

The counts in Tables 2 and 3 show that BWZ's original estimations change if one asks the question from Sinnemäki's perspective: As compared to BWZ, it is no longer the case that consistent (i.e. unsplit) systems are in the majority everywhere. On the contrary, in Eurasia and Sahul, there is a clear preference for DOM over consistent marking, and there is no longer a strong dispreference for DOM in the other areas (especially when compared to BWZ's Figure 2).

Taken together, the above counts suggest that for the case-marking languages in BWZ's sample, DOM is always at least as frequent as consistent, unsplit object marking. We saw above that Sinnemäki's (2014) sample is much larger still, containing almost 300 more languages than in BWZ, but also that his counts are not exactly comparable: On the one hand, they include some cases of what Witzlack-Makarevich and Seržant (this volume) call "predicate-triggered DAM" and thus case splits in a much wider (and actually different) sense. On the other hand, his counts include genuine case splits induced by factors that are not considered in BWZ (e.g. where the trigger is word order rather than a referential property); and as we pointed out above, he does not consider pronouns at all, thus excluding a large number of referential splits. Overall, then, it is very likely that our Tables 2 and 3 above are actually still too conservative, underestimating the preponderance of DOM. Therefore, we are currently inclined to believe, with Sinnemäki (2014), that *when languages develop object case, there is a universal preference for it to be of the differential type.* Incidentally, Haig (2017) has made a similar proposal for object indexation; according to him, differential object indexing is a "cross-linguistic attractor state".

References

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko (2015). Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej L. Malchukov & Marc Richards (eds.), *Scales and hierarchies: A cross-disciplinary perspective*, 7–43. Berlin & Boston: De Gruyter Mouton.

Haig, Geoffrey (2017). Differential object indexing as an attractor state in the grammaticalization of agreement. Plenary talk at the *Vielfaltslinguistik* conference, Leipzig University, March 2017.

Sinnemäki, Kaius (2014). A typological perspective on Differential Object Marking. *Linguistics* 52.2: 281–313.