# Kaius Sinnemäki\* Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach

https://doi.org/10.1515/jhsl-2019-1010

Abstract: This paper brings together typological and sociolinguistic approaches to language variation. Its main aim is to evaluate the relative effect of language internal and external factors on the number of cases in the world's languages. I model word order as a language internal predictor; it is wellknown that, for instance, languages with verb-final word order (that is, languages in which both nominal arguments precede the main lexical verb) tend to develop complex case systems more often than languages with SVO word order do. I model population size and the proportion of second language speakers in the speech community as sociolinguistic predictors; these factors have been suggested recently to influence the distribution of the number of cases in the world's languages. Modelling the data with generalized linear mixed effects modelling suggests an interaction between the number of cases, word order, and the proportion of second language speakers on the one hand, and between the number of cases, word order, and population size, on the other. This kind of complex interactions have not been previously reported in typological research wherefore they call for more complex explanations than previously suggested for cross-linguistic variation.

**Keywords:** case marking, language typology, morphological complexity, population size, second language speakers, word order

# **1** Introduction

Up until recently sociolinguistics and language typology have been largely separate sub-fields in linguistics with little cross-fertilization and only a few attempts at bridging the two fields (on dialectology and typology, see Kortmann 2008; on 9

**<sup>\*</sup>Corresponding author: Kaius Sinnemäki,** University of Helsinki, Helsinki, Finland, E-mail: kaius.sinnemaki@helsinki.fi. https://orcid.org/0000-0002-6972-5216

Open Access. © 2020 Kaius Sinnemäki, published by De Gruyter. Correction This work is licensed under the Creative Commons Attribution 4.0 International License.

variationist typology, see Torres Cacoullos and Travis 2019).<sup>1</sup> Sociolinguistic research generally aims at understanding the societal factors that underpin linguistic variation, such as the speakers age, gender, or social class, or more macro-level factors, such as language policies. Typological research generally aims at understanding the degree and limits of linguistic diversity on a global scale. Traditionally it has focused on the internal systematicities in the languages of the world, being perhaps best known for statistical language universals, such as word order correlations (e.g. Dryer 1992). While language external factors have been important in typology, they have been mostly discussed in the context of sampling to control for the confounding effect of language contact.

Things have started to slowly change during the past 20 years as researchers have begun to ask whether there may be some systematicities in how sociolinguistic factors may influence cross-language diversity (e.g. Bisang 2004; Trudgill 1998, 2011). The first set of empirical testing has provided some evidence that, for instance, morphological complexity correlates with the number of native speakers (henceforth, L1 speakers) or with the proportion of second language speakers (henceforth, L2 speakers) (see the reviews by Ladd et al. 2015 and Nettle 2012 and references there).

Probably the most influential paper in this research, Lupyan and Dale (2010), used linguistic data from the *World Atlas of Language Structures* (henceforth *WALS*; Dryer and Haspelmath 2013; an earlier version of the atlas was used by the authors) to demonstrate that the smaller the number of native speakers was, the more complex the morphological systems of the languages tended to be. Bentz and Winter (2013) focused on just one aspect of morphological complexity, the number of morphological cases in a language, and found evidence that the greater the proportion of L2 speakers was in a speech community, the smaller its case system was as well. Sinnemäki and Di Garbo (2018) further argued that morphological complexity depends on both population size and the proportion of L2 speakers and crucially when both factors are built in the same model. This result corroborated

<sup>1</sup> This paper has benefited greatly from discussions with the editors and authors of the special issue on the rate of language change. Earlier versions of some parts of the paper were presented at the Diversity Linguistics Seminar at the University of Leipzig on 27 September 2017 and at the 93rd Annual Meeting of the Linguistic Society of America on 4 January 2019. I would like to thank the audiences of these meetings for helpful comments, and Martin Haspelmath for inviting me to present at the Leipzig seminar. I am also grateful to Christian Bentz, Sonja Dahlgren, Johanna Nichols, Jarno Porkka, and Max Wahlström for helpful comments and to Sakari Sarjakoski and Nora Fabritius for drawing the maps (Maps 1 and 2, and Map 3, respectively). This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 805371) and partly from the Academy of Finland (grant agreement No 296212).

earlier claims by Trudgill (2011) who argued that multiple sociolinguistic factors may affect typological distributions.

In this paper my aim is to assess the relative strengths of language external and language internal factors on linguistic diversity. This approach is the first step in addressing the criticism against the recent attempts (of e.g. Peter Trudgill) at bridging typology with sociolinguistics. Danylenko (2018) has criticized these attempts for being overly mechanistic and for overlooking language-internal causes. This worry about overlooking language-internal causes is justified in that if we focus only on testing the possible effect of sociolinguistic factors on linguistic patterns, we are in danger of producing spurious results by not having accounted for how linguistic patterns may interact among themselves across languages. From this perspective language-internal factors can be thought of as confounding factors whose effects need to be addressed in the research design. In this paper I put to test the predictive power of sociolinguistic and language internal factors by building them into the same research design.

I consider the division into external and internal factors as a useful descriptive tool but insufficient as a theoretical explanation (Farrar and Jones 2002). If we consider language internal features, especially morpho-syntactic properties do not influence one another directly but perhaps more so by affecting the cognitive or communicative preferences that influence the development, loss, and stability of other features (Sinnemäki 2014). For instance, verb-final word order does not directly affect the development of case marking but it may increase the probability that case marking develops in language, because in verb-final contexts case is preferred from the perspective of cognitive processing (e.g. Fedzechkina et al. 2017). In a similar way, sociolinguistic factors do not influence linguistic structures directly but may affect the societal contexts in which language is used and learned. Different types of societal contexts may condition preferences in language use and learning and may lead to different linguistic structures being preferred in different contexts, for instance, in terms of learnability. In this sense the effect of both system internal and sociolinguistic factors can be explained by functional theories that refer to (typically universal) cognitive and communicative principles that bias the emergence and stability of linguistic structures across languages (Bickel 2015).

I model the cross-linguistic distribution of one linguistic variable, the number of morphological cases in a language. Earlier research has suggested that population size (Lupyan and Dale 2010) and the proportion of L2 speakers (Bentz and Winter 2013) may influence its variation. Earlier typological research has also suggested that verb-final languages have a higher rate of developing case systems than those with SVO word order (e.g. Bentz and Christiansen 2013, Greenberg 1966, Siewierska and Bakker 2008). The number of cases thus provides fertile testing ground for assessing the relative strength of system internal (word order) and sociolinguistic factors.

However, testing language external and internal influences on typological distributions may not provide direct evidence for language change. Typical work in historical sociolinguistics uses historical corpus data to analyse sociolinguistic variation at different temporal stages and uses predefined sociolinguistic categories, such as gender, age, and social class of the individual language users. Yet patterns of (socio)linguistic variation may become observable on different levels of analysis: at the level of the speech community as a whole, its subgroup, or the individual (cf. Bergs 2005: 5). Typological data is based on grammar descriptions and archival data written over a long period of time and concern usually the variety of the whole speech community. It provides thus important corollary evidence to historical corpus data on individual and group-level variation.

Here I do not directly estimate the rates at which case systems change across languages, but the results can function as a springboard for further research that will scrutinize rates of change more directly using, for instance, historical corpus data or phylogenetic regression on typological data. However, I try to do justice to the well-evidenced observation that rates of change can vary across language families and geographical areas (e.g. Nichols 2003) by using mixed effects regression modelling. Modelling language families and geographical areas as random slopes in these models help assessing the family-specific and area-specific variation in effect sizes and thus make preliminary conclusions about rates of change as well. If systematic cross-linguistic patterns are found, then we may assume that the rate at which the complexity of case systems change are statistically sufficiently similar across the sampled language families and geographical areas.

The working hypothesis of this paper is: when modelled as competing predictors, both sociolinguistic factors (population size, proportion of second language speakers) and system internal factor (word order) affect variation in the number of cases. The rationale for this hypothesis stems from the fundamental observation that linguistic phenomena are affected by multiple causes and therefore an ideal research design includes multiple competing predictors (Gries 2003). In addition to attempting to re-evaluate these two earlier studies, I build population size and the proportion of L2 speakers as predictors in the same model to evaluate their relative strength on the number of cases in the spirit of Sinnemäki and Di Garbo (2018).

The rest of the paper is structured as follows. In Section 2 I show how the idea of varying rates of change is implemented in language typology to assess systematic cross-linguistic patterns. In Section 3 I describe the materials and methods and the statistical modelling used. Section 4 presents the results of the case

studies, Section 5 provides a general discussion, and Section 6 ends the paper with brief conclusions.

# 2 On rates of change, transition probabilities, and language universals

Rate of language change has become an increasingly important issue in linguistic theorizing. It has been well-known, especially in historical linguistics that languages change at different rates. Already Jespersen (1922: 259–261) speculated how different social disturbances, such as wars and plagues, may have caused periods of rapid change. However, for long linguists were unable to implement this knowledge to testing, for instance, the relative stability of linguistic structures in different language families or geographical regions.

In traditional approach to typology it is customary to emphasize that samples should include only languages that are independent from one another to control away the possible confounding effects of faithful transmission and horizontal transmission. Closely related languages are not sampled because similarities in those languages could not be attributed to possible universal (e.g. cognitive or communicative) effects but rather to the features having been faithfully transmitted from the ancestor population. Similarly, languages that have been in intense contact with one another are not sampled because their similarities are attributable to borrowing.

While this traditional approach is in line with the requirements of classic statistical tests, it leads to observing synchronic frequency distributions. Synchronic distributions are ultimately unable to demonstrate universal preferences (cf. Bickel 2013, Maslova 2000). Already Greenberg (1978) viewed language universals not as mere synchronic frequency distributions of features across the world's languages but as probabilities that languages change from one state (type) to another.<sup>2</sup> In this approach the focus is on comparing the transition probability that a language shifts from type A to type B ( $p_{AB}$ ) and from type B to type A ( $p_{BA}$ ). If the probability  $p_{AB}$  is greater than the probability  $p_{BA}$ , there is then evidence that type B is preferred over type A across languages and that in the long run type B would dominate over type A.

**<sup>2</sup>** While it may be possible to characterize languages and their change from a more comprehensive perspective – English, for one, has changed first from a synthetic type towards a more analytic type and then back to greater syntheticity again (cf. Szmrecsanyi 2016) – the term 'type' is here understood as referring to local features only, not to language as a whole.

To observe the probability of change, it is necessary to have either historical data from multiple languages or to compare patterns within attested families. Since historical data is available only for a limited number of languages, probabilities of change are estimated in typology by comparing synchronic data from closely related languages to one another to evaluate recurring patterns of change (or stability) across families. This idea has been implemented in different ways, for instance, by observing systematic biases within families (Bickel 2013) or by using phylogenetic comparative methods to assess probably of change within families (e.g. Dunn et al. 2011).

Probability of language shifting from one type to another is clearly related to rate of language change. In general, low probability of change in a given time-frame translates to low rate of change and high probability of change translates to high rate of change. These ideas have also been implemented in research on the relative stability of linguistic structures across languages (see Wichmann 2015). It has been a major area of interest to linguists whether structural stability of features depends on geographical areas and language families. Dediu and Cysouw (2013) compare seven different approaches that have been proposed in the literature for approaching stability. Based on the comparison of the metrics on *WALS* data they find support for some structures being overall more stable than others and thus more persistent against change. The most stable features involved phonology and word order and the least stable ones the different grammatical categories of the noun (e.g. the coding of definiteness and case).

In these studies data come from structural features, language families, and geographical areas. No information is used on sociolinguistic context. But if we think more about it, information on families and areas is used for assessing the effects of faithful cultural transmission and horizontal diffusion. Both these phenomena are inherently social. Cultural knowledge, such as language, is faithfully passed on in the population from the older generation to the younger when the social conditions are reasonably stable and under such circumstances the result is only small degrees of change (Thomason and Kaufman 1988: 9–10). If the social conditions become more unstable because of war, disasters, massive population movements, or similar very disruptive events, the dynamics in the population change, possibly affecting transmission of cultural knowledge and introducing more effects from dialect contact or perhaps resulting in the population shifting to another language. Increased contact effects may mean that the population affecting children's language learning increases in diversity, as it may have incorporated new people from different linguistic backgrounds. As an example, many researchers have attributed the loss of case marking on nouns in Scandinavian languages to the Hanseatic traders who spoke Low German and came to Scandinavia in large numbers from the 13th until the 15th centuries (e.g. Kusters

2003, Olthof 2017). Contact with other languages may also be stable if the social conditions are stable, but under abrupt social changes the dynamics of language contacts may also change (e.g. Thomason 2001: 15–26).

The changing stability of social conditions and their influence on linguistic diversification has been discussed prominently from the perspective of the punctuated equilibrium model of language change (Dixon 1997; see also Nevalainen et al. and Leiwo, this issue). Dixon proposed that the history of many languages can be described as long periods of relative stability which are occasionally disturbed by large natural or social events called punctuations that cause changes in speaker populations and thereby also in the languages they speak. For my purposes Dixon's model and its recent elaboration by Operstein (2015) underline the close-knit relationships between social conditions, cultural transmission, diffusion, and rate of change. Hypotheses about social conditions influencing rates of change have also appeared in evolutionary linguistic research and researchers have begun to test whether rates of change depend on sociolinguistic conditions, such as population size (e.g. Bowern 2010, Greenhill et al. 2018, Grollemund et al. 2015). In research on cultural evolution, demographic factors, such as population size, play also crucial role since they may influence the rate of cultural innovations, including linguistic ones (Richerson et al. 2009).

I bring together many of these strands concerning rate of change. First, I model the effect of sociolinguistic factors on a structural feature, namely, the number of cases. This aspect of the research design models the possible effect of societal factors on language structures. Second, I include in the same research design the effect of word order on the number of cases. This design allows me to weigh how much the number of cases is affected by the selected system internal and sociolinguistic factors. Third, while I do not directly estimate rates of change, I use mixed effects regression to estimate the effect sizes in different families and areas. This research design allows me to evaluate whether the effect sizes vary a lot across families and thus to make indirect inferences about rates of change, stability, and language universals as well.

## 3 Materials and methods

#### 3.1 Case marking

I define case here following Blake (2001: 1) as 'a system of marking dependent nouns for the type of relationship they bear to their heads' (also Iggesen 2013, Witzlack-Makarevich 2019). Case thus marks a relation between a noun and its head, which can be a (lexical) verb or an adposition, among others. When linguists have traditionally talked about case, they have meant inflectional Indo-European-like cases. In such systems case markers are phonologically more or less bound to the stem, even completely fused to the host noun. If not fused, the stem and/or the case markers may show allomorphic variation. But the syntactic relation of the noun to its head may also be marked by clitics, particles, or adpositions. Especially particles and adpositions are often considered separate phonological and/or grammatical words and typical clitics are also more weakly bound to the stem compared to affixes. However, since affixal case forms are sometimes invariant, such as the accusative suffix *-ta* in Imbabura Quechua in (1), it is very difficult in practice to analyse reliably across languages what is a bound morpheme, a clitic, or an adposition (cf. Haspelmath 2008). We also know from historical development of morphology that affixes originate as free forms that become increasingly bound to the stem over time, possibly leading to wholesale fusion at later stages of grammaticalization.

 (1) Chai warmi(-ca) ñuca-man aswa-ta cara-wa-rca-mi that woman(-TOP) 1SG-DAT beer-ACC serve-1SG.OBJ-PST.3-VAL 'That woman served me some beer.' (Jake 1983: 21)<sup>3</sup>.

Given the difficulty of separating affixes, clitics, and adpositions from one another, it is hard to develop a water-tight definition of morphological case. In this article I use Iggesen's (2013) data on the number of morphological cases instead of taking the more time-consuming path of reanalysing a completely new sample from scratch. This choice also enables me to provide a point of comparison to Lupyan and Dale (2010) and Bentz and Winter (2013), who used the same data. I thus follow Iggesen's (2013) analyses but acknowledge some of the analytical problems.

Iggesen (2013) delimits the analysis of case to productive case on nominals, excluding case on pronouns, and focuses on inflectional case marking. He includes clitics and adpositions provided that there is sufficient degree of phonological integration between them and the host noun in what he calls 'basic syntactic construction.' These constructions are noun phrases that contain the head noun but no modifiers. A similar approach to phrasal clitics has been taken by Dryer (2013a). Including such clitics may be justified on further grounds: a strict focus on just inflectional case could reflect Euro-centricity and leave large geographical areas as having no languages with case. As an example, no African

**<sup>3</sup>** The following is a list of abbreviations used in this paper: 1 first person, 3 third person, acc accusative, aux auxiliary, dat dative, erg ergative, obj object, pl plural, pst past, sg singular, top topic, val validator.

Feature value	Feature description	Feature coding
1	no cases	0
2	2 cases	2
3	3 cases	3
4	4 cases	4
5	5 cases	5
6	6–7 cases	6
7	8–9 cases	7
8	10 or more cases	8
9	borderline case marking	0

 Table 1: The feature values and descriptions for the number of cases (lggesen 2013) and the feature coding used in this paper.

language has been reported to having an inflectional case system according to Creissels (2000: 247).

With regard to certain analytical choices Iggesen (2013) is more liberal compared, for instance, to Dryer (2013a). As an example, he includes the marking expressed by particles in Bawm (Kuki-Chin, Sino-Tibetan) in (2). In this paper I do not address such discrepancies between the analysts but confine myself to making them explicit for the reader.

(2)	а	pâ	nih	fâ	tlâ	а	vuak	nâ	
	the	father	ERG	child	PL	3sg	beat	3pl	
	'the	father be	ats the	e childre	en'.				
	(Rei	chle 1981	: 28; g	lossing	follow	vs that	in Com	rie 201	3).

The feature values for the number of cases in Iggesen (2013) are presented in Table 1. The feature-values from zero to five represent the actual number of cases but from 6 upwards the feature-values represent a range of counts (e.g. feature-value 7 represents languages with 8–9 cases). Regardless of this conflation of values I treat this variable as a count variable in the spirit of Lupyan and Dale (2010) and Bentz and Winter (2013) who also used this data in the same way. I also follow Hewson and Bubenik (2006: 364) analysing languages with borderline case marking as having no case marking.<sup>4</sup> Lastly, seven languages were excluded, because they have become extinct after the grammatical description has been written. The resulting dataset contains 254 languages (see the supplementary material).<sup>5</sup>

**<sup>4</sup>** I include languages with borderline case marking in case study 1 which provides a point of contrast to Lupyan and Dale (2010). Bentz and Winter (2013) excluded those languages from their sample.

#### 3.2 Word order

Typologists have different opinions about how word order should be approached and defined. The first issue concerns whether to focus on the main lexical verb or the finite element of the predicate. For instance, in Kisi (Mel, Niger-Congo) the main lexical verb tends to occur between the arguments in simple tensed clauses (3a) but after them in compound tensed clauses (3b), which mark tense, aspect, and mood categories by a finite auxiliary. If research focused on the position of the finite element of the predicate, Kisi would be analysed as an SVO language, but if the focus was on the position of the main lexical verb, Kisi would be analysed as having no dominant word order (SVO/SOV). In this paper I focus on the position of the main lexical verb, following Dryer (2013b).

- (3) a. sáá kíndá dìòó. Saa close door 'Saa closed the door.' (Childs 1995: 218).
  - b. fàlà có léɛŋndó yìkpàá.
     Fallah AUX machete to.sharpen
     'Fallah is sharpening the machete.'
     (Childs 1995: 250).

Another important issue with regard to word order is whether to focus on the so-called basic word order or, for instance, on the most frequent word order (see Song 2012). In this paper I emphasize usage frequency, since usage-phenomena offer the most productive connecting point between sociolinguistics and language typology. More precisely I follow Dryer (2013b) in defining word order as the dominant order of the subject, object, and verb, where the verb is the main lexical verb, and subject and object are the arguments of a transitive verb. A word order variant is analysed as dominant if its text count adds up to two thirds or more in all transitive clauses, otherwise no variant is analysed as dominant and the language is classified as having no dominant word order. See Dryer (2013c) and Sinnemäki (2010) for more details and for discussion of borderline cases.

Word order can be classified into types in different ways. Based on earlier research it is well-known that verb-final languages tend to develop complex case systems, thus suggesting a binary factor 'verb-final word order.' On the other hand,

**<sup>5</sup>** An anonymous reviewer asked about the role of head marking in explaining the results. Case is by definition a matter of dependent marking, but head marking has also been shown to correlate with dependent marking and word order (e.g. Nichols 1992, Sinnemäki 2008). As a result, it may be justified to include head marking in the model as well. However, inclusion of head marking in the research design must be left for future research, because its inclusion would have increased the complexity of the research design and called for some other methodological approach.



Figure 1: Boxplot of the number of cases over word order types.

languages with SVO word order tend to develop complex case systems much more rarely, suggesting an alternative binary factor 'SVO word order.' Rather than predefining how word order should be classified into types for modelling purposes, I turn this into an empirical question. In the spirit of Gries and Hilpert (2010: 299–304) I identify meaningful groups of word order that share a high degree of similarity within the group but a low degree of similarity across groups.

For this purpose, I plot the number of cases over five different word order types in a boxplot: verb-final, verb-initial, SVO, OVS and no dominant order. The rationale for these initial types is the following. For instance, SOV and OSV orders seem to behave alike with regard to other word order properties but also in respect to case marking (Dryer 1997). Moreover, some languages, such as Imonda (Seiler 1985: 179), can be characterized as dominantly verb-final even if they cannot be classified as dominantly SOV or OSV. Focusing on the more general verb-final word order has an added methodological advantage: it avoids the risk of idiosyncratic analyses of supposedly OSV languages that in later analysis have turned out to be better analysed as dominantly SOV or verb-final or having no dominant word order. Warao, for one, has been analysed as OSV (Dryer 2013b, Romero-Figeroa 1985), as SOV (Mosonyi and Mosonyi 2000: 128), and as verb-final (Herrmann 2004). The analysis of VOS and VSO orders in verb-initial orders is analogical, but there is no reason to classify SVO and OVS together (see Dryer 1997 and Sinnemäki 2010: 876 for further details).

The boxplot in Figure 1 suggests that no dominant, OVS, and verb-final types are very similar with one another: their first and third quantiles (0 and roughly 8–9 cases, respectively) are almost identical and their differences in terms of median number of cases (3, 3.5 and 5, respectively) are not too large either. These types thus group together quite clearly. On the other hand, SVO and verb-initial types group together in a different way: their median number of cases is zero and their third quantile is just two cases. Since in these latter two types the object follows the verb

(VO order), it appears that the most promising way to modelling word order would be to binomially contrast VO order with non-VO order (including verb-final, no dominant order, and OVS). To focus on VO order makes sense also from the typological perspective of case systems: languages typically leave especially the nominative S unmarked and rather mark some other arguments or adjuncts with cases, roughly following the case hierarchy proposed by Blake (2001: 155–160).

Overall, my data for word order come from Dryer (2013b) and from my own data collection in which I have followed Dryer's principles of analysis (see the supplementary material).

#### 3.3 Sociolinguistic variables

I use two sociolinguistic variables to model the effect of language external factors on linguistic diversity, namely, population size and the proportion of L2 speakers.<sup>6</sup> Population size has become a widely used variable especially in research on language evolution and linguistic diversification (e.g. Lupyan and Dale 2010; see also the excellent review in Greenhill et al. 2018 and the references there). Proportion of L2 speakers is not as widely used as a variable owing largely to limitations in data availability, but it is a more direct proxy for effects from intensive language contact on language structure. Both variables have their problems and some of those are discussed here (see also Sinnemäki and Di Garbo 2018, Kempe and Brooks 2018, Koplenig 2019).

One issue concerning both population size and the proportion of L2 speakers is related to rate of change. Here I discuss population size but the same issue concerns the proportion of L2 speakers as well. When we correlate a grammatical feature with population size, we are making an underlying assumption that there is plausible causality between language structure and a sociolinguistic feature: namely, we are assuming that these factors may change at reasonably comparable rates. Presumably, a change in population size would be roughly matched with a change in a grammatical feature – otherwise it would not make sense to correlate them in the first place.<sup>7</sup>

**<sup>6</sup>** As for L2 speakers, it is possible that their level of education may condition their influence on the target language's structures. Here I pay no attention to education level, since accommodation by more competent speakers to less competent speakers may be a more important mechanism of change than the idiosyncratic changes produced by the L2 speakers themselves (Atkinson et al. 2018).

<sup>7</sup> Note that in this paper I test a particular causality, namely, that sociolinguistic feature (population size or proportion of L2 speakers) has an effect on linguistic structure. The other direction of causality, that is, a sociolinguistic feature depending on a linguistic feature, is a logical possibility but rarely discussed as a real possibility. As suggested by an anonymous reviewer, there may be at least some room, albeit presumably very limited, for advancing this idea further. We could

However, we do not know whether the rate at which population size changes differs from the rate at which grammar changes. They probably due to some extent. It is possible that population size may change faster than grammar in some contexts, but if so, that would require those contexts to be known – which we don't. For many languages we simply do not know whether the figures we have for population size really reflect the situation that held at the time when the grammatical features developed that are now coded in descriptions (cf. Sinnemäki 2009: 131–132). But this is something we must accept and assume that the number of current speakers provide estimations at least in terms of relative number of speakers in the situations that held at the time when the grammatical features under study emerged.

I define population size as the number of L1 speakers. Since population size varies from a handful to more than a billion speakers, I transform it using the base-10 logarithm to better scale them (following Lupyan and Dale 2010). Languages with fewer than 50 speakers were coded as having 50 speakers (cf. Lupyan and Dale 2010); this may be considered an absolute minimum for a language to be viably passed to the next generation.<sup>8</sup>

I define the proportion of L2 speakers following Bentz and Winter (2013) as the proportion of non-native speakers in the whole speech community. The size of the whole speech community includes both native and non-native speakers, and thus, the proportion of L2 speakers is calculated as in (4).

The data for population size comes largely from the nineteenth edition of the *Ethnologue* (Lewis et al. 2016a; see the supplementary material). For the purpose of this paper, I use the number of speakers in all countries. The *Ethnologue* is sometimes criticized for overestimating the number of speakers for small languages. My experience in using the *Ethnologue* now for roughly 15 years is that they have improved in reliability over the years, for instance, now providing data sources more systematically. For the number of L2 speakers I

hypothesize, for instance, that a prohibitively difficult morphology, perhaps coupled with a language ideology against error prone talk of second language learners, might contribute to a language having only a few second language speakers. Yet it is unclear, and intuitively unlikely, whether a high degree of complexity alone could lead to a particular sociolinguistic outcome. Space does not permit to explore this issue more here, but a more principled and nuanced treatment of the direction of causality would be beneficial in the future. The CHIELD database (Roberts et al. 2020), for instance, could offer initial and useful ideas in that regard.

<sup>8</sup> I owe to Harald Hammarström for suggesting this idea to me.

use the data in Bentz and Winter (2013). Their data come primarily from the *Ethnologue*, the Rosetta project, and the UCLA Language Materials Project.<sup>9</sup> Data on the number of L2 speakers is often difficult to come by especially outside Eurasia. Actually, most of the available data for L2 speakers come from Eurasia and Africa. This is good to keep in mind when interpreting the results. For a more thorough discussion of the problems related to data on the number of L2 speakers, see Sinnemäki and Di Garbo (2018: 6–8).

#### 3.4 On statistical modelling

The main research question of this paper is whether sociolinguistic factors and word order have an effect on the number of cases. A fundamental distinction in the domain of case marking is whether languages have a case system to begin with. A corollary research question here is thus whether sociolinguistic factors and word order have an effect on the presence vs. absence of cases. I thus model these two research questions separately, following Bentz and Winter (2013). The null hypothesis is that sociolinguistic and grammatical factors have no effect on the number or the presence of cases in the world's languages. I test these hypotheses using generalized linear mixed effects modelling (henceforth, GLMM). For recent application of this method to typological data, see Bentz and Winter (2013), Jaeger et al. (2011), Roberts et al. (2015) and Sinnemäki and Di Garbo (2018). The central idea in these modelling techniques is that the value of the dependent variable is predicted based on the predictor variable(s) and using a grouping structure (that is, random structure) in the modelling to adjust the variables of interest.

In the research design I use the number (or presence) of cases as the response variable whose distribution is modelled based on the predictors and the random structure. Two types of predictors will be used, namely, word order as a grammatical predictor and population size and the proportion of L2 speakers as sociolinguistic variables. Theoretically speaking, the model uses the distribution of word order and the distribution of the sociolinguistic factors to predict the outcome of the number of cases given the random structure. Word order was coded as a binary variable with values 'yes' for 'dominant VO word order' and 'no' for 'no dominant VO word order.' Population size was coded as the log<sub>10</sub> number of native speakers and the proportion of L2 speakers as percentage of L2 speakers in the whole population.

I used genealogical affiliation and geographic location of the sample languages as random intercepts to adjust the estimates for the number of cases. I modelled genealogical affiliation using the highest level of classification in the

<sup>9</sup> See http://rosettaproject.org and http://www.lmp.ucla.edu (service closed currently).



Map 1: The 24 areas of the AUTOTYP (Bickel et al. 2017) database illustrated on a world map.

*WALS* genealogical taxonomy, namely, families. For geographic location of languages, I followed the *AUTOTYP* (Bickel et al. 2017) and classified languages into 24 areas in which they are primarily spoken. The areas are illustrated in Map 1.

As for the random slopes, it is sometimes argued that so-called maximal models that include all possible theoretically motivated random slopes should be used whenever possible (e.g. Barr et al. 2013). However, even though maximal models may converge, some of the random structure in them can be negligibly small and should thus be simplified. I used random slopes for word order over families but not for the sociolinguistic factors over families, since it is more likely that word order is inherited from a common ancestor. Instead I used random slope for the sociolinguistic factors and for word order over geographic areas, but only in Case study 1. in Case study 2 I tried using random slopes for the sociolinguistic factors or word order over area and/or for word order over families, but owing to convergence issues or too small random variances I ended up using only random intercepts for these models. If a model did not converge or if any of the random variances approached zero (in the range of 10<sup>-7</sup>), I simplified the model by removing the respective random factors from it (cf. Matuschek et al. 2017). However, if the random intercepts over area or language family approached zero, I retained them in the model because of their theoretical importance. I do not explain the simplification process separately for each model in the main text; see the supplementary R script for some further information.

For hypothesis testing I used GLMMs in the R programming environment (R Core Team 2018).<sup>10</sup> About half of the world's languages have no cases at all and this

causes problems in modelling a count response variable. For this reason, I used the package glmmADMB (Fournier et al. 2012, Skaug et al. 2016) which offers ways to dealing with zero inflation; this is also what Bentz and Winter (2013) did. Parameters are estimated in glmmADMB by maximum likelihood ratio using Laplace approximation. Following Sinnemäki and Di Garbo (2018), this approximation was further improved by using so-called importance sampling, providing the argument impSamp with values greater than 0 (Skaug and Fournier 2006).<sup>11</sup>

Since the number of cases is discrete count data, it would be appropriate to use Poisson regression in the modelling. Poisson distribution assumes that the sample mean is identical with the sample variance. However, in all the models the dispersion ratios were significantly different from 1 (p < 0.05), which means that using Poisson modelling is not justified for the data. In case of overdispersion of this kind it is possible to use instead the negative binomial distribution which relaxes the assumption of overdispersion. All the reported results on the number of cases are thus based on negative binomial regression. The presence vs. absence of cases, on the other hand, is a binary variable. It requires a logistic regression model, which uses here the predictors to model the probability of case in a language.

I evaluated each variable's effect with maximum likelihood ratio test using nested models, that is, by comparing a model with the variable of interest to a simpler model without the variable of interest in additive GLMMs. Using nested models helps extract maximum likelihood ratio, and its significance, just for the variable of interest. I thus started with the simplest model that contained only the intercept and the random structure and added the fixed effects in (roughly) stepwise fashion (cf. Baayen 2013).

For evaluating goodness-of-fit, I compared the model's Akaike Information Criterion (AIC), or more precisely its small sample equivalent AICc that is corrected for bias (Burnham and Anderson 2002). AIC is widely used for evaluating the importance of a predictor by considering the extent to which adding a fixed effect reduces AIC: lower values improve the model's fit and thus the greater the reduction in AIC is, the more important the predictor is. Burnham and Anderson (2002: 70–71) provide rough guidelines for interpreting this reduction: if the reduction in AIC is smaller than 2, no significant difference exists between the models; a reduction between 4 and 7 suggests that the difference is important; if the reduction is 10 or

**<sup>10</sup>** The packages MuMIn (Bartoń 2018), sjPlot (Lüdecke 2018), and ggplot2 (Wickham 2016) were further used for graphics and statistics in R. Other R packages are referred to in the main text.

**<sup>11</sup>** It is unclear how high values are reasonable for impSamp. Ben Bolker, the author of the function admbControl for the glmmADMB package, uses values up to 200 but values as high as 5,000 have been used (Macneill and Lajeunesse 2019: 761). I used values between 0 and 300, preferring small values.

greater, there is no support for the model that has higher AIC value. Competing models' AIC values can further be compared with Akaike weights which scale the differences in AIC to a scale of 1 and thus provide an easy and effective way for model comparison.<sup>12</sup>

#### **4** Results

Results of two case studies are reported in this section. In Case study 1, I used word order and population size as predictors, and the sample contains data in 254 languages. In Case study 2, I used word order, population size, and the proportion of L2 speakers as predictors; in that case study the sample is 66 languages.

The histogram distribution of the variables is provided in Figure 2. The distribution of the number of cases is somewhat bimodal, although there are many languages that have no cases. On the other hand, the bimodal nature of this distribution is split according to VO word order: 61% of the languages with no cases have VO word order, while 82% of the languages with at least two cases have non-VO word order. Population size is rather normally distributed around a mean of roughly 100,000 speakers (log10(100,000) = 5). Proportion of L2 speakers is



**Figure 2:** Frequency histogram and the superimposed density estimates in the sample languages for number of cases (A), log number of L1 speakers (B), and proportion of L2 speakers (C). Since no language has exactly one case in (A), their absence was smoothed over by using biased cross-validation for bandwidth in density estimation.

**<sup>12</sup>** An anonymous reviewer recommended using the C-index for evaluating goodness-of-fit. In logistic regression modelling, goodness-of-fit is often evaluated using the index of concordance (C-statistic, C-index) between the predicted probability and the observed response which is equal to the area under the Receiver Operating Characteristic curve (ROC curve; e.g. Hosmer and Lemeshow 2000: 160-164). However, Barth and Kapatsinski (2018) show that a real predictor may fail to contribute to a model's predictive capacity measured by C-index and it may instead be inflated by the random structure. For this reason, I do not evaluate goodness-of-fit using C-index (see also Sinnemäki and Di Garbo 2018).

again skewed to the right, namely, most languages have less than 50% share of L2 speakers. The areal distribution of the sample languages and their number of cases



Map 2: Distribution of number of cases on a world map.



Map 3: Distribution of VO (verb object) and non-VO word orders on a world map.

is provided in Map 2, while Map 3 shows the distribution of word order on a world map.

#### 4.1 Case study 1: Word order and population size

In the first case study I additively build word order and population size as competing predictors in the same model. In this way the effect of word order is compared to a sociolinguistic factor to evaluate the relative effect of language internal vs. language external effects (cf. Danylenko 2018). I assess the relative strength of these factors on the number of cases in negative binomial models and on the presence of cases in binary models.<sup>13</sup>

To evaluate the importance of different factors, I compare a sequence of nested models, that is, by comparing a model with the variable of interest to a simpler model without the variable of interest in additive GLMMs. Response in these evaluations is either number or presence of cases. I start with the simplest one and increase the model complexity in step-wise fashion (Baayen 2013). In the second and third models there is just one predictor in each to evaluate each predictors effect in isolation (comparing these models individually to the bare intercept model). I then add population size and the interaction of population size and word order to a model that already contains word order. The sequence of models is the following:

- Response ~ 1
- Response ~ 1 + VO
- Response ~ 1 + log\_L1
- Response  $\sim 1 + VO + log_L1$
- Response  $\sim 1 + VO + \log_L 1 + VO:\log_L 1$

This sequence is assessed statistically by the sequential likelihood ratio tests presented in Table 2. In this sequence the second model is compared to the first, the third model to the first, the fourth model to the second, and the fifth model to the fourth. The table also lists the AICc, deviance, number of parameters, difference in AICc, and the Akaike weights for each compared model.

According to the likelihood ratio tests population size had a significant effect on the number of cases (p = 0.018) (negative binomial model), but only when modelled in isolation, not when modelled together with word order in the same

**<sup>13</sup>** Because multiple models are compared in each case study, this would call for a correction for multiple testing, such as Bonferroni correction. My main source for additive GLMMs, however, does not use correction for p-values (Baayen 2013; see also Tomaschek et al. 2019). In addition, since my general aim is to re-evaluate the results of Lupyan and Dale (2010) and Bentz and Winter (2013), I prefer not to use a more stringent approach than they did. Overall, given the exploratory nature of the statistical testing in this paper, I prefer letting readers use their own judgment in weighing the conclusions.

Negative binomial       (Intercept)       933.2         model <sup>a</sup> $1 + VO = yes$ 919.3       16.3         model <sup>a</sup> $1 + VO = yes$ 919.3       5.6 $1 + VO = yes$ $929.8$ 5.6 $1 + VO = yes + log_L L1$ $919.6$ 1.8 $1 + VO = yes + log_L L1 + VO:log_L L1$ $920.2$ 1.6         Binary model <sup>b</sup> $I(ntercept)$ $274.4$ 16.1 $1 + VO = yes$ $10g_L L1 + VO:log_L L1$ $920.2$ 1.6         Binary model <sup>b</sup> $I(ntercept)$ $274.4$ 16.1 $1 + VO = yes$ $10g_L L1$ $276.5$ 0.0	AICc	Deviance	N of param.	<i>p</i> -value	Difference in AICc	Akaike weight
model <sup>a</sup> $  + VO = yes$ 919.3       16.3 $  + \log L1$ 929.8       5.6 $  + VO = yes + \log L1$ 919.6       1.8 $  + VO = yes + \log L1$ 919.6       1.8 $  + VO = yes + \log L1$ 919.6       1.8         Binary model <sup>b</sup> $  (ntercept)$ 274.4       16.1 $  + VO = yes$ $  og_L L1 + VO: log_L L1$ 920.2       1.6 $  + VO = yes + log_L L1 + VO: log_L L1$ 920.2       1.6 $  + VO = yes + log_L L1 + VO: log_L L1$ 274.4       16.1 $  + VO = yes + log_L L1$ 276.5       0.0 $  + VO = yes + log_T L1$ 260.4       16.1	933.2		9			0.000
I + log_L1       929.8       5.6         I + VO = yes + log_L1       919.6       1.8         I + VO = yes + log_L1 + VO:log_L1       920.2       1.6         Binary model <sup>b</sup> [(ntercept)       274.4       16.1         I + VO = yes       log_L1       276.5       0.0         I + VO = yes + log_11       276.5       0.0	919.3	16.1	7	0.0000	-13.9	0.401
I + VO = yes + log_L1       919.6       1.8         I + VO = yes + log_L1 + VO:log_L1       920.2       1.6         Binary model <sup>b</sup> I(ntercept)       274.4       16.1         I + VO = yes       10g_L1       276.5       0.0         I + VO = yes + log_11       276.5       0.0	929.8	5.6	7	0.0181	-3.5	0.002
I + VO = yes + log_L1 + VO:log_L1       920.2       1.6         Binary model <sup>b</sup> I(ntercept)       274.4         I + VO = yes       260.4       16.1         I + VO = yes       276.5       0.1         I + VO = wes + log_L1       262.4       0.1	Jg_L1 919.6	1.8	8	0.1802	0.3	0.339
Binary model <sup>b</sup> I(ntercept) 274.4 I + VO = yes 260.4 16.1 I + log_L1 276.5 0.0 I + VO = ves + log 11 262.4 0.1	0g_L1 + V0:log_L1 920.2	1.6	6	0.2053	0.5	0.238
I + V0 = yes     260.4     16.3       I + log_L1     276.5     0.0       I + V0 = ves + log 11     267.4     0.3	274.4		9			0.000
+ log_L1 276.5 0.0   + VO = ves + log 11 262.4 0.0	260.4	16.1	7	0.0000	-14.0	0.386
I + VO = ves + log 1 262.4 0.1	276.5	0.0	7	1.00	2.1	0.000
	Jg_L1 262.4	0.1	8	0.7746	2.1	0.139
I + V0 = yes + log_L1 + V0:log_L1 260.0 4.6	Jg_L1 + V0:log_L1 260.0	4.6	6	0.0317	-2.5	0.475

<sup>b</sup>impSamp = 1.

model (p = 0.18). This is corroborated by the difference in AICc: when compared to the null model, the model containing just population size reduced AICc by 3.5 but when adding population size to a model that contained word order AICc actually increased by 0.3, meaning that the model became worse. The interaction term between word order and population size was also non-significant (p = 0.21). Among the six competing negative binomial models the one that contains only word order had 40% chance of being the best one.

In the binary models word order had a significant effect on the presence of cases when word order was modelled in isolation (p < 0.001). Adding word order to the null model also reduced AICc by 14. Population size had no significant effect on the presence of cases whether modelled in isolation or in the same model with word order: adding population size to the null model or the one including word order increased AICc by 2.1, thus making the model worse. However, the interaction term between word order and population size was significant (Deviance = 5.23; df = 1; p = 0.032).<sup>14</sup> The Akaike weights suggest that the model containing the interaction term has 48% chance of being the best model among the six competing models.

Table 3 presents coefficients for the predictors' effect in the best-fitting models, that is, in the negative binomial model that contained only word order and in the binary model that contained the interaction term between word order and population size. In the negative binomial model the coefficient for word order is -1.221

	Parameters	Estimate	Std. error	<i>z</i> -value	<i>p</i> -value
Negative binomial model <sup>a</sup>	(Intercept)	1.740	0.047	37.32	0.0000
	VO = yes	-1.221	0.334	-3.66	0.0003
Binary model <sup>b</sup>	(Intercept)	0.273	0.834	0.33	0.7430
	VO = yes	0.356	1.676	0.21	0.8317
	log_L1	0.222	0.187	1.19	0.2357
	VO = yes:log_L1	-0.704	0.357	-1.97	0.0490

Table 3: Coefficients for the best-fitting models in Case study 1.

<sup>a</sup>impSamp = 36.

<sup>b</sup>impSamp = 8.

**<sup>14</sup>** Since in these tests the p-value was very close to the critical 0.05, it would be good to assess its validity with parametric bootstrap using, for instance, the package pbkrtest in R (Halekoh and Højsgaard 2014). However, it is not yet possible to implement this method to regression models computed by glmmADMB. Alternative models computed with other packages (e.g. lme4) are unable to address zero inflation, thus producing unreliable results, too, wherefore I refrain from using them.

and its inverse logarithm is 0.295. This means that languages with VO word order have about 70.5% fewer cases than those with non-VO word order.

As for the interaction term in the binary model, their interpretation is difficult as the interaction coefficient is a ratio of the log odds. However, it is possible to interpret the parameters separately for VO and non-VO languages (following Jaccard 2001 and UCLA Statistical Consulting Group (no date)). The reference level for word order is non-VO, so for these languages, a one-unit increase in population size results in log odds of 0.222 and an inverse logarithm of 1.248 (a one-unit change from e.g. 4 to 5 corresponds to a change from log(10,000) to log(100,000)). This means that for non-VO languages a one-unit increase in population size increases the odds of having cases by 25%. For VO languages, on the other hand, a one unit increase in population size results in log odds of -0.481 (-0.704 + 0.222) and an inverse logarithm of 0.618. This means that for VO languages a one-unit increase in population size **decreases** the odds of having cases by 38%. This opposite behaviour of non-VO and VO languages is visually depicted in Figure 3 (plot B). The random structure in Table 4 suggests that the variances are not too close to zero.

Figure 3 presents the effect plots for the predictors in the best-fitting negative binomial and binary models. In the plots the predictors' values are presented on the *x*-axis and the predicted values of the response (number of cases or the probability of case) on the *y*-axis. For the negative binomial model (plot A) the effect plot suggests that the predicted number of cases is close to 6-7 cases in languages with non-VO word order but roughly 2 cases in VO languages. The difference between the word orders here is obvious. The difference is clear within language families, too. For instance, there are nine Afro-Asiatic languages in the sample and all those that are



**Figure 3:** Effect plot for word order on the number of cases in plot A and marginal effect plot for the interaction term between word order and population size in plot B. (Note that in the plots for marginal effects the black dots are concentrations of several grey dots in roughly the same position.)

		Variance	Std. dev.
Negative binomial model	Group = area		
	(Intercept)	0.00004	0.0059
	VO = yes	1.0864	1.0423
	Group = family		
	(Intercept)	0.0160	0.1264
Binary model	Group = area		
	(Intercept)	1.6870	1.2988
	log_L1	0.0204	0.1427
	VO = yes	0.1644	0.4050
	Group = family		
	VO = yes	1.6607	1.2890
	(Intercept)	4.9047	2.2150

Table 4: Random effect variances for the best-fitting models in Case study 1.

non-VO have two or more cases (e.g. Beja and Amharic) with an average of three cases, while more than half of the Afro-Asiatic VO languages in the sample have no cases at all (e.g. Hausa and Egyptian Arabic), averaging to no cases.

In the binary model (plot B), the predicted probability of case is quite similar in small languages whether the word order is VO or non-VO (roughly 70% in non-VO languages and roughly 50% in VO languages with about 100 speakers; log10 (100) = 2). The predicted probability of case decreases as population size increases – as in Lupyan and Dale (2010) – but only in VO languages: the predicted probability of case drops to about 10% in VO languages with 1,000,000 speakers (log10(1,000,000) = 6). On the other hand, the predicted probability of case increases as population size increases – pace Lupyan and Dale (2010) – but this happens only in languages with non-VO word order: the predicted probability of case goes up to about 90% in non-VO languages with 100,000,000 speakers (log10(100,000,000) = 8). All in all, the effect of word order on the presence of cases seems conditioned by population size: large VO languages are likely to have no cases while large non-VO languages are likely to have cases, but as population size becomes smaller, the probability of case becomes much more similar in VO and non-VO languages.

# 4.2 Case study 2: Word order, proportion of L2, and population size

In the second case study I additively build word order, proportion of L2 speakers, and population size as competing predictors in the same model. The relative effect

of the sociolinguistic factors is compared to one another in the spirit of Sinnemäki and Di Garbo (2018) and their effects are also compared to word order to evaluate the relative effect of language internal vs. language external effects (cf. Danylenko 2018). I assess the relative strength of these factors on the number of cases in negative binomial models and on the presence of cases in binary models.

To evaluate the importance of different factors in the additive models, I compare a sequence of nested models, starting with the simplest one that contains only the intercept and increasing the model complexity in roughly step-wise fashion. In the second, third and fourth models there is just one predictor in each to evaluate each predictor's effect in isolation (comparing these models individually to the bare intercept model). I then add proportion of L2 speakers and population size to a model that already contains word order. The sequence of models is the following:

- Response ~ 1
- Response ~ 1 + VO
- Response  $\sim 1 + \text{prop}_{L2}$
- Response  $\sim 1 + \log_{L1}$
- Response  $\sim 1 + VO + prop_L2$
- Response  $\sim$  1 + VO + prop\_L2 + log\_L1

This sequence is assessed statistically by the sequential likelihood ratio tests presented in Table 5. In the negative binomial modes, the likelihood ratio test and the reductions in AICc suggest that only the proportion of L2 speakers is an important predictor of the number of cases (reduction in AICc is roughly 6) but word order and population size are not (AICc increases when adding these to simpler models). Based on the Akaike weights, the model that includes only the proportion of L2 speakers has 63% chance of being the best among the six competing negative binomial models. In the binary models, the likelihood ratio test (p < 0.001) and the reductions in AICc (>10) suggest that both word order and the proportion of L2 speakers are important predictors of the presence of case when compared to the null model. Adding proportion of L2 speakers to the same model with word order results in proportion of L2 speakers having a significant effect in this model as well (p = 0.004). Adding population size to this model or to the null model increases AICc, which means that the models become worse. Based on the Akaike weights, the model that includes both word order and the proportion of L2 speakers has 71% chance of being the best among the six competing binary models.

Based on these results, the proportion of L2 speakers appears at least an equally important predictor compared to word order. For model comparison I again use AICc, keeping in mind that the lower its value is, the better the model is. In the binary models, AICc was reduced by 12.4 when the model containing the

							y III case staat 2.
	Model structure	AICc	Deviance	N of param.	<i>p</i> -value	Difference in AICc	Akaike Weight
Negative binomial	l(ntercept)	247.3		5			0.038
model <sup>a</sup>	l + V0 = yes	249.6	0.1	9	0.7642	2.3	0.012
	l + prop_L2	241.7	8.0	9	0.0046	-5.6	0.626
	l + log_L1	249.6	0.1	9	0.7164	2.3	0.012
	$I + VO = yes + prop_L2$	243.7	8.5	7	0.0036	-6.0	0.233
	I + VO = yes + prop_L2 + log_L1	245.8	0.4	8	0.5179	2.2	0.079
Binary model <sup>b</sup>	l(ntercept)	81.5		£			0.000
	l + V0 = yes	69.6	14.1	4	0.0002	-11.9	0.025
	l + prop_L2	69.0	14.7	4	0.0001	-12.4	0.034
	l + log_L1	82.5	1.2	4	0.2721	1.1	0.000
	$I + VO = yes + prop_L2$	62.9	9.0	5	0.0027	-6.1	0.713
	I + VO = yes + prop_L2 + log_L1	65.2	0.1	9	0.7089	2.3	0.228

<sup>a</sup>impSamp = 2. <sup>b</sup>impSamp = 1.

DE GRUYTER MOUTON

proportion of L2 speakers was compared to the null model but when the model containing word order was compared to the null model, the reduction in AICc was 11.9. The reduction in AICc is thus of the same range for the proportion of L2 speakers compared to word order. On the other hand, in the negative binomial model the proportion of L2 reduced AICc by 5.6 (when compared to the null model) but word order increased it by 2.3; in these models the proportion of L2 was clearly a better predictor of the number of cases. AIC thus provides a useful way of comparing the importance of structural and sociolinguistic effects to one another and the initial results suggest rather unexpectedly that the latter factors may be at least as important as the former.

Table 6 presents coefficients for the predictors' effect in the best-fitting models, that is, in the negative binomial model that contained only proportion of L2 speakers and in the binary model that contained word order and proportion of L2 speakers. In the negative binomial model the coefficient for the proportion of L2 speakers is –2.187 and its inverse logarithm is 0.112. This means that languages spoken by communities with 100% L2 speakers have about 89% fewer cases than those with no L2 speakers. In the binary model the coefficient for the proportion of L2 speakers is –5.568 and its inverse logarithm is 0.0038. This means that languages spoken by communities with 100% L2 speakers have about 99.6% lower

	Parameters	Estimate	Std. error	z-value	<i>p</i> -value
Negative binomial model	(Intercept)	1.930	0.120	16.07	0.0000
	prop_L2	-2.187	0.630	-3.47	0.0005
Binary model	(Intercept)	3.502	1.152	3.04	0.0024
	VO = yes	-2.595	0.985	-2.63	0.0084
	prop_L2	-5.568	2.214	-2.52	0.0119

**Table 6:** Coefficients in the best-fitting models in Case study 2.

 Table 7: Random effect variances in the best-fitting models in Case study 2.

		Variance	Std. dev.
Negative binomial model	Group = area		
	(Intercept)	0.0061	0.0780
	Group = family		
	(Intercept)	0.000005	0.0021
Binary model	Group = area		
	(Intercept)	3.2420	1.800
	Group = family		
	(Intercept)	0.0428	0.2070



**Figure 4:** Effect plots for the proportion of L2 speakers on the number of cases in plot A, on the probability of case in plot B, and for word order on the probability of case in plot C. Marginal effect plot for the proportion of L2 speakers and word order on the probability of case in plot D.

odds of having cases than those with no L2 speakers. The coefficient for word order is –2.595 and its inverse logarithm is 0.075. This means that languages with VO word order have about 92.5% lower odds of having cases than those with non-VO word order. Table 7 presents variances for the random effects in these model.

Figure 4 presents the effect plots for the predictors in the best-fitting negative binomial and binary models. In the plots the predictors' values are presented on the *x*-axis and the predicted values of the response (number of cases or probability of cases) on the *y*-axis. The effect plot for the best-fitting negative binomial model suggests that the predicted number of cases is about 8–9 in languages that have roughly no L2 speakers in the speech community, but it drops towards zero the more the proportion of L2 speakers approaches 100% (see plot A). As an example, Georgian has six to seven cases and Icelandic four cases and in their speech communities the proportion of L2 speakers is less than 5%. On the other hand, Amharic has two cases (fewer than many of its Afro-Asiatic cognates) and almost 25% L2 speakers, Urdu has also two cases (fewer than many of L2 speakers. The effect plot for the best-fitting binary model further suggests that the predicted probability of case is about 90% in languages that have roughly no L2 speakers in the speech community, but it drops towards zero the more the proportion of L2 speakers approaches 100% of L2 speakers. The effect plot for the best-fitting binary model further suggests that the predicted probability of case is about 90% in languages that have roughly no L2 speakers in the speech community, but it drops towards zero the more the proportion of L2 speakers.

approaches 100% (see plot B). Based on this data there is thus a clear downward trend in the number and probability of case as the proportion of L2 speakers grows in a community. As for word order, the effect plot for the best-fitting binary model suggests that the predicted probability of case is roughly 85% in languages with non-VO word order, but it drops to 30% in languages that have VO word order (see plot C). The difference between the word orders here is clear, although the confidence levels are very broad owing to the relatively small sample size (compare to the much smaller confidence levels for word order in Figure 3).

In Case study 2 the data is too small to evaluate an interaction term between the predictors in either the negative binomial or the binary models. But it is possible to evaluate the marginal effects for the number of cases separately for VO and non-VO languages. I do this here only for the binary model, since word order had a significant effect only in this model. Based on plot D in Figure 4, there is a downward trend in both VO and non-VO languages: the predicted probability of case decreases as the proportion of L2 speakers increases. However, the predicted probability of case is almost 30% higher in non-VO type compared to VO type in languages that have roughly no L2 speakers. As the proportion of L2 speakers grows up to roughly 50%, the probability of case in VO languages drops quickly to about 10%. In non-VO languages, the probability of case is still roughly 70% in languages with about 50% L2 speakers, but then it drops to 20% or less in languages with 90% or more L2 speakers. This suggests that word order may condition the effect of the proportion of L2 speakers on the probability of case so that in languages with few L2 speakers the probability of case is much lower in VO languages compared to non-VO languages, but as the proportion of L2 speakers grows, the difference becomes smaller. Part of the reason for this result is that there are no cases in languages with 60% or more L2 speakers.

Based on the results of Case study 2 the proportion of L2 speakers has a significant effect on the number and presence of cases, population size has no effect on either, and word order has a significant effect on the presence of cases. In Case study 1 and in Case study 2 it seems that population size has an effect on the number or presence of cases only when modelled in isolation (negative binomial model) or when correlated with word order (binary model). Proportion of L2 speakers, on the other hand, seems to affect the number and presence of cases whether modelled with or without word order or population size. Lastly, the effect of word order affects the number and the presence of cases, but the latter only when modelled together with either population size (Case study 1) or proportion of L2 speakers (Case study 2). The correlation between case and word order attested in earlier typological research may, therefore, need to be qualified, as it may be conditioned by sociolinguistic factors to some degree.

### **5** Discussion

In this paper I have modelled language internal and external factors in competition with one another and the results suggest a somewhat surprising interaction between the factors. The current results are preliminary and suggestive, given especially the relatively small sample size in Case study 2, but they point to interesting hypotheses that can be tested with larger datasets and different methods in future research. Here I briefly discuss the results and possible explanations.

The results of Case study 1 suggested that population size had a significant negative effect on the number of cases when modelled on its own, much as in Lupyan and Dale (2010). However, it had no effect on the number of cases when modelled as a competing predictor with word order. On the other hand, VO languages were much less likely to have cases than non-VO languages. Nevertheless, the effect of word order was shown to depend on population size when modelling the effect of their interaction on the presence of cases in the binary models. This complex interaction suggested that in small languages the predicted effect of VO and non-VO word order did not differ much from each other. However, as population size grew, the predicted probability of case decreased in VO languages but increased in non-VO languages.

These results bring two earlier generalizations together. First, according to Lupyan and Dale (2010), the number of cases decreases as population size increases, as seems to be the case for many other measures of morphological complexity as well (also Koplenig 2019). The current results agree with these results, but only when modelling population size in isolation from word order. When modelling the interaction of population size with word order, the probability of case also decreased as population size increased but only in VO languages, that is, in languages that have a dominant SVO word order or dominant verb-initial word order (VSO, VOS, and VSO/VOS). Second, according to Greenberg (1966) and many others, verb-final languages<sup>15</sup> are likely to develop a complex case system. Figure 1 suggested that complex cases systems are likely to develop also in languages with no dominant word order, and more likely in OVS languages compared to SVO and verb-initial languages. The current results agree with the earlier results but suggest that the relationship between cases and non-VO order is the clearer the larger the non-VO language is.

Lupyan and Dale (2010) propose the linguistic niche hypothesis to account the correlation between morphological complexity and population size. According to

**<sup>15</sup>** Greenberg originally referred only to SOV languages, but verb-final languages more generally behave in an analogous way.

this hypothesis, linguistic structures adapt to the environments in which they are used and learned. Thus, structures that are difficult to learn, such as morphological complexity, are not very likely to develop and they are not so easily passed to the next generation (e.g. Clahsen et al. 2010). I do not provide evidence for or against this hypothesis (see Kempe and Brooks 2018 for a critical review) but merely suggest that in the light of my results, it is not morphological complexity per se that seems to adapt to population size, but it is rather the relationship between case and word order that adapts to it. Earlier experimental research has suggested that case is more likely to develop in verb-final constructions compared to non-verb-final constructions (e.g. Fedzechkina et al. 2017). In the light of current results, this experimental research offers only partial explanation to the interaction between case and word order and may need to be rethought, taking into account also the social context of language use and non-VO word order patterns more generally.

But why should population size condition the interaction between word order and case? Here I only provide very speculative remarks. Koplenig (2019) showed that languages with more speakers tend to have greater entropy rate. What this means from an information-theoretic perspective is that languages with more speakers tend to be less redundant and thus more transparent, possibly adhering more closely to transparency (Sinnemäki 2009). In non-VO languages with more speakers this might mean that syntactic relations are not signalled via linear order but via cases, while in SVO and verb-initial languages with more speakers this might mean that syntactic relations are more likely signalled via linear order than via cases: in both circumstance, syntactic relations can be signalled nonredundantly and transparently by relying on either linear order or cases. Some evidence for these patterns may be found from verb-final Sino-Tibetan languages: Lepcha (69 000 speakers) has two cases, Ladakhi (117 000) speakers has five cases, Meithei (roughly 1.5 million speakers) has 6–7 cases, and Burmese (roughly 33 million speakers) has 8-9 cases. Counter-examples also do occur, such as Finnish which is SVO, has at least 12 productive cases, and with about 5.5 million speakers is one of the largest languages in the Uralic language family.

Explanations relying on the role of linear order beg the question why the coding of word order in the case studies did not concern the order of other nominals besides the core arguments of a transitive verb? A practical reason for this was that although Dryer and Gensler (2013) provide data on the order of object, oblique and verb, much new data should have been analysed to include the subject as well, and this was not possible within the limits of this paper. A more methodological reason was that the order of obliques is typically more variable than the order of core arguments and this would inevitably lead to inflation of the no dominant order type. Lupyan and Dale (2010) also propose to use population size as a proxy for language contact effects so that the larger the language is, the more likely it is going to have a large proportion of L2 speakers as well. Sinnemäki and Di Garbo (2018) showed that this is not necessarily the case and that it is best to consider population size and the proportion of L2 speakers as independent sociolinguistic parameters.

As for Case study 2, the results suggested that the proportion of L2 speakers had an inverse effect on the number of cases as well as on the presence of case, similarly to Bentz and Winter (2013). Word order also had an inverse effect on the presence of case but not on the number of cases. However, the predicted probability of case was greater in non-VO languages than in VO languages and this difference was the most pronounced in languages that had up to 50% of L2 speakers. On the other hand, in languages with larger proportion of L2 speakers, the predicted probability of case dropped towards zero regardless of word order. The effect of large proportion of L2 speakers thus seems to thwart the effect of word order on the probability of case.

The main explanation in the literature for the inverse correlation between morphological complexity and L2 speakers is that morphological complexity is difficult for L2 speakers to acquire and use and more often than not they restructure and simplify the morphological patterns of the language they learn (e.g. Klein and Perdue 1997, Roberts and Bresnan 2008). While this explanation is widely used in the literature, it is difficult to explain why the idiosyncrasies produced by nonnative speakers would spread to native speakers' language use as well, and likewise why word order could condition that influence.

An alternative account puts the emphasis on accommodation processes (e.g. Trudgill 2011: 56–60). In accommodation more competent speakers (native speakers and competent L2 speakers) produce simplifying adjustments, such as avoiding opacity, to less competent speakers and these adjustments are then adopted and transmitted to the next generation, leading eventually to language change. According to recent experimental research, the accommodation model may offer a key linking mechanism why, for instance, morphology is simplified if the language is learned by many L2 speakers (Atkinson et al. 2018).

Neither of these explanations, language learning difficulty and accommodation, however, address whether the degree of similarity between languages, that is, typological distance, affects linguistic behaviour in situations of L2 learning. This is an important issue for future research, since typological distance between L1 and L2 seems to be an important factor in L2 learning of complex morphological features (e.g. Schepens et al. 2013).

To understand the conditioning influence of word order on the sociolinguistic factors, it may be helpful to discuss what language universals are. In classic language typology language universals are probabilistic type frequencies between structural patterns, presented in the form of implicational universals. But a growing trend in typology is to conceptualize universals as structural pressure on how languages change over time; that is, as diachronic laws of type preference (e.g. Bickel 2013, Sinnemäki 2010: 877 and references there). From this perspective it is possible to argue that there is a universal pressure, perhaps related to cognitive and communicative preferences, for languages with non-VO word order to develop cases and for VO languages not to develop them. The sociolinguistic environment in which languages are learned and used, then, may either amplify or reduce this pressure.

The results of the case studies partly agree with those in Sinnemäki and Di Garbo (2018) who showed that for modelling the distribution of the degree of inflectional synthesis the best model contained both population size and the proportion of L2 speakers as predictors. Case study 2 does not support this conclusion directly and the available data is far too small to model a complex interaction between the three predictor's word order, population size, and the proportion of L2 speakers. However, Case study 1 and Case study 2 suggest indirectly that both population size and the proportion of L2 speakers are important predictors of at least the presence of cases, the former in complex interaction with word order and the latter as an independent predictor together with word order.

On the other hand, the results partly disagree with the conclusions of Koplenig (2019), whose data suggest that morphological complexity is affected by population size but not by the proportion of L2 speakers. The discrepancy in our results with regard to the proportion of L2 speakers may owe to two methodological issues: differences in sample size and differences in modelling the effect of L2 speakers. In my models the proportion of L2 speakers was computed directly from the number of native and non-native speakers, but the sample size was relatively small (66 languages). This small sample size makes the results only suggestive. A problematic issue is also the areal bias in the sample: roughly 80% of the languages for which it was possible to deduce the number of L2 speakers came from Africa and Eurasia. This means that the results cannot be very reliably generalized outside these continents.

To address this shortage of data Koplenig (2019) uses *Ethnologue*'s evaluations for language vitality, namely, Expanded Graded Intergenerational Disruption Scale (henceforth EGIDS; Lewis et al. 2016b) to proxy the proportion of L2 speakers in the community. EGIDS is built on Joshua Fishman's (1991) earlier scale, and its levels range from 0 (international language use) to 10 (extinct). Notably, the editors define languages in EGIDS categories 4 or lower (towards 10) as local languages, which are not expected to have any L2 speakers (Lewis et al. 2016b: 18). However, while EGIDS data is available for more than 7000 languages, a huge

increase in sample size compared to Lupyan and Dale (2010), for instance, it is not without problems to define languages with EGIDS 4 or lower as not having any L2 speakers. According to Lewis et al. (2016b: 18) roughly 91% of languages in the *Ethnologue* belong to these categories, but many of them tend also to be spoken by relatively small people groups that are often highly multilingual (Lüpke 2016). It is probably fair to conclude that until the field develops more rigorous methods for assessing contact-effects of multilingualism this kind of discrepancies in results will remain at least partly unresolved.

Evidence in this paper comes from structural data curated from grammatical descriptions and from demographic data. The former can be thought of as human expert judgements (by the linguist) on structural patterns that are often based on corpora. In that sense there is a clear link between corpora and grammar descriptions. Comparison between corpus evidence and evidence from grammar descriptions suggests these different data sources provide converging evidence at least for the interaction between linear order and morphology (Koplenig et al. 2017) as well as for measures of morphological complexity (Bentz et al. 2016). These two types of evidence thus seem relatively closely related, which means that the kinds of factors that affect variation in corpora can be meaningfully applied to typological data as well. This convergence is also at the heart of Hawkins' (2004, 2014) model, which is based on the idea that preferences within a language correspond to preferences across languages.

# 6 Conclusion

In (historical) sociolinguistics it has been long acknowledged and shown that both external and internal factors drive linguistic variation. Earlier research in sociolinguistic typology has also accumulated evidence that linguistic structures are systematically affected across languages by the sociolinguistic environment in which they are learned and spoken. Evidence has been drawn from both case studies on historical contact scenarios (e.g. Kusters 2003) as well as from quantitative typological work (e.g. Bentz and Winter 2013, Lupyan and Dale 2010, Sinnemäki 2009, Sinnemäki and Di Garbo 2018).

In this paper I bring this sociolinguistic typological research one step closer to (historical) sociolinguistics, by providing preliminary typological and speech community-level evidence that both sociolinguistic and structural factors affect linguistic variation. I propose a new more holistic approach to evaluating typological distributions, which considers system internal factors in competition with sociolinguistic factors – an approach that has to my knowledge not been adopted

DE GRUYTER MOUTON

before. The results suggest complex interactions between the sociolinguistic and system internal factors on typological distributions.

It is still quite unorthodox to research this kind of hypotheses, but this paper points to the feasibility of doing exactly so. While I do not directly assess how these different factors affect rates of change, I modelled language family as a random slope wherever possible, which is a first step in addressing the well-known fact in language typology and historical linguistics that rates of change can vary across families (Nichols 2003). The results open a set of new hypotheses about rate of change and the factors that may affect it that can be tested in future research, for instance, with phylogenetic regression methods that more directly evaluate probability of change.

# References

- Atkinson, Mark, Kenny Smith & Simon Kirby. 2018. Adult learning and language simplification. *Cognitive Science* 42. 2818–2854.
- Baayen, Harald R. 2013. Multivariate statistics. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 337–372. Cambridge: Cambridge University Press.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.
- Barth, Danielle & Vsevolod Kapatsinski. 2018. Evaluating logistic mixed-effects models of corpuslinguistic data in light of lexical diffusion. In Dirk Speelman, Kris Heylen & Dirk Geeraerts (eds.), *Mixed-effects regression models in linguistics* (Quantitative Methods in the Humanities and Social Sciences), 99–116. Cham: Springer.
- Bartoń, Kamil. 2018. MuMIn: Multi-Model Inference. R package version 1.42.1. Avialable at https://CRAN.R-project.org/package=MuMIn (accessed 14 May 2020).
- Bentz, Christian & Morten H. Christiansen. 2013. Linguistic adaptation: The trade-off between case marking and fixed word orders in Germanic and Romance languages. In Feng Shi & Gang Peng (eds.), *Eastward flows the great river: Festschrift in honor of prof. William S-Y. Wang on his 80th birthday*, 48–56. Hong Kong: City University of Hong Kong Press.
- Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3(1). 1–27.
- Bentz, Christian, Tatyana Ruzsics, Alexander Koplenig & Tanja Samardzic. 2016. A comparison between morphological complexity measures: Typological data vs. language corpora. In Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, Thomas François & Philippe Blache (eds.), Proceedings of the workshop on computational linguistics for linguistic complexity, 142–153. Osaka: The COLING 2016 Organizing Committee. Available at https://aclanthology. info/papers/W16-4117/w16-4117 (accessed 4 February 2019).
- Bergs, Alexander. 2005. Social networks and historical sociolinguistics: Studies in morphosyntactic variation in the Paston letters (1421–1503). Berlin: Mouton de Gruyter.
- Bickel, Balthasar. 2013. Distributional biases in language families. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency: in honor of Johanna Nichols*, 415–444. Amsterdam: John Benjamins.

- Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heike Narrog (eds.), *The Oxford handbook of linguistic* analysis, 2nd ed., 901–923. Oxford: Oxford University Press.
- Bickel, Balthasar, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Rießler, Lennart Bierkandt, Fernando Zúñiga & John B. Lowe. 2017. The AUTOTYP typological databases. Version 0.1.0. Available at https://github.com/autotyp/ autotyp-data/tree/0.1.0 (accessed 28 November 2017).

Bisang, Walter. 2004. Dialectology and typology – an integrative perspective. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 11–45. Berlin: Mouton de Guyter.

Blake, Barry J. 2001. Case. Cambridge: Cambridge University Press.

Bowern, Claire. 2010. Correlates of language change in hunter-gatherer and other 'small' languages. *Language and Linguistics Compass* 4(8). 665–679.

- Burnham, Kenneth P. & David R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed. New York: Springer.
- Childs, G. Tucker. 1995. *A grammar of Kisi: A Southern Atlantic language* (Mouton Grammar Library 16). Berlin: Mouton de Gruyter.

Clahsen, Harald, Claudia Felser, Kathleen Neubauer, Mikako Sato & Renita Silva. 2010. Morphological structure in native and nonnative language processing. *Language Learning* 60(1). 21–43.

Comrie, Bernard. 2013. Alignment of case marking of full noun phrases. 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/chapter/98 (accessed 27 November 2018).

Creissels, Denis. 2000. Typology. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*, 231–258. Cambridge: Cambridge University Press.

Danylenko, Andrii. 2018. The correlation of linguistic patterning and societal structures in systemic typology. Studia Linguistica Universitatis Iagellonicae Cracoviensis 135. 81–96.

Dediu, Dan & Michael Cysouw. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PloS One* 8(1). e55009.

Dixon, Robert M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. Language 68(1). 81–138.

Dryer, Matthew S. 1997. On the six-way word order typology. Studies in Language 21(1). 69–103.

Dryer, Matthew S. 2013a. Position of case affixes. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/51 (accessed 11 November 2019).

Dryer, Matthew S. 2013b. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/81 (accessed 9 October 2018).

- Dryer, Matthew S. 2013c. Determining dominant word order. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/s6 (accessed 27 November 2018).
- Dryer, Matthew S. & Orin D. Gensler. 2013. Order of object, oblique, and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max

Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/84 (accessed 12 February 2019).

- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info (accessed 9 October 2018).
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.
- Farrar, Kimberley & Mari C. Jones. 2002. Introduction. In Mari C. Jones & Edith Esch (eds.), *Language change: The interplay of internal, external and extra-linguistic factors*, 1–16. Berlin: Mouton de Gruyter.
- Fedzechkina, Maryia, Elissa L. Newport & Florian T. Jaeger. 2017. Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science* 41. 416–446.

Fishman, Joshua A. 1991. Reversing language shift. Clevedon: Multilingual Matters.

- Fournier, David A., Hans J. Skaug, Johnoel Ancheta, James Ianelli, Arni Magnusson, Mark N. Maunder, Anders Nielsen & John Sibert. 2012. AD model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* 27. 233–249.
- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 40–70. Cambridge, MA: MIT Press.
- Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In Joseph H.
   Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), Universals of human languages, Volume 1: Method & theory, 61–91. Stanford: Stanford University Press.
- Greenhill, Simon J., Xia Hua, Caela F. Welsh, Hilde Schneemann & Lindell Bromham. 2018. Population size and the rate of language evolution: A test across Indo-European, Austronesian, and Bantu languages. *Frontiers in Psychology* 9. 576.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- Gries, Stefan Th & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: A multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics* 14(3). 293–320.
- Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 112. 13296–13301.
- Halekoh, Ulrich & Søren Højsgaard. 2014. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models the R package pbkrtest. *Journal of Statistical Software* 58(10). 1–30.
- Haspelmath, Martin. 2008. Terminology of case. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 505–517. Oxford: Oxford University Press.

Hawkins, John A. 2004. Efficiency and complexity in grammars. Oxford: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press. Herrmann, Stephanie. 2004. Warao. In Philipp Strazny (ed.), *Encyclopedia of linguistics*, vol. 2,

- 1164–1167. London: Routledge.
- Hewson, John & Vit Bubenik. 2006. From case to adposition: The development of configurational syntax in Indo-European Languages. Amsterdam: John Benjamins.

- Hosmer, David W. & Stanley Lemeshow. 2000[1989]. *Applied logistic regression*, 2nd ed. (Wiley Series in Probability and Statistics). New York: John Wiley & Sons.
- Iggesen, Oliver A. 2013. *Number of cases*. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary

Anthropology. Available at https://wals.info/chapter/49 (accessed 30 November 2018). Jaccard, James. 2001. *Interaction effects in logistic regression*. Thousand Oaks: Sage.

Jaeger, T. Florian, Peter Graff, William Croft & Daniel Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15. 281–319.

- Jake, Janice Lynn. 1983. *Grammatical relations in Imbabura Quechua*. Urbana-Champaign, IL: University of Illinois dissertation.
- Jespersen, Otto. 1922. Language: Its nature, development and origin. London: Allen & Unwin.
- Kempe, Vera & Patricia J. Brooks. 2018. Linking adult second language learning and diachronic change: A cautionary note. *Frontiers in Psychology* 9. 480.
- Klein, Wolfgang & Clive Perdue. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13(4). 301–347.

Koplenig, Alexander. 2019. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science* 6. 181274.

Koplenig, Alexander, Peter Meyer, Sascha Wolfer & Carolin Müller-Spitzer. 2017. The statistical trade-off between word order and word structure – large-scale evidence for the principle of least effort. *PLoS ONE* 12(3). e0173614.

Kortmann, Bernd (ed.). 2008. *Dialectology meets typology: Dialect grammar from a crosslinguistic perspective* (Trends in Linguistics. Studies and Monographs 153). Berlin: De Gruyter Mouton.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT Publications.

- Ladd, Robert D., Seán G. Roberts & Dan Dediu. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1. 221–241.
- Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.). 2016a. Ethnologue: Languages of the world, 19th ed. Dallas, Texas: SIL International. Available at https://www.ethnologue.com/ 19/ (accessed 9 November 2018).
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2016b. *Ethnologue global dataset*, 19th ed. Available at https://www.ethnologue.com/sites/default/files/Ethnologue-19-Global% 20Dataset%20Doc.pdf (accessed 30 January 2019).
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PloS One* 5(1). e8559.

Lüdecke, Daniel. 2018. sjPlot: Data visualization for statistics in social science, R package version 2.6.2. Available at https://CRAN.R-project.org/package=sjPlot (accessed 30 January 2019).

- Lüpke, Friederike. 2016. Uncovering small-scale multilingualism. *Critical Multilingualism Studies* 4(2). 35–74.
- Macneill, Bryan N. & Marc J. Lajeunesse. 2019. Effects of river hydrology and physicochemistry on Anchovy abundance and Cymothoid Isopod parasitism. *Journal of Parasitology* 105(5). 760–768. (accessed 15 November 2019).
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4(3). 307–333. https://doi.org/10.1515/lity.2000.4.3.307.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald R. Baayen & Douglas M. Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94(2). 305–315.

- Mosonyi, Esteban E. & Jorge C. Mosonyi. 2000. *Manual de lenguas indígenas de Venezuela*, vol. 2. Caracas: Fundación Bigott.
- Nettle, Daniel. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B* 367. 1829–1836.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.
- Nichols, Johanna. 2003. Diversity and stability in language. In Richard D. Janda & Brian D. Joseph (eds), *Handbook of historical linguistics*, 283–310. London: Blackwell.
- Olthof, Marieke. 2017. Transparency in Norwegian and Icelandic: Language contact vs. language isolation. *Nordic Journal of Linguistics* 40(1). 73–115.
- Operstein, Natalie. 2015. Contact-genetic linguistics: Toward a contact-based theory of language change. *Language Sciences* 48. 1–15.
- R Core Team. 2018. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available at https://www.R-project.org (accessed 30 January 2019).
- Reichle, Verena. 1981. *Bawm language and lore: Tibeto-burman area* (Europäische Hochschulschriften Reihe, 21). Bern: Peter Lang.
- Richerson, Peter J., Robert Boyd & Robert L. Bettinger. 2009. Cultural innovations and demographic change. *Human Biology* 81(3). 211–235.
- Roberts, Sarah J. & Joan Bresnan. 2008. Retained inflectional morphology in pidgins: A typological study. *Linguistic Typology* 12. 269–302.
- Roberts, Seán G., James Winters & Keith Chen. 2015. Future tense and economic decisions: Controlling for cultural evolution. *PloS One* 10. e0132145.
- Roberts, Seán, Anton Killin, Angarika Deb, Catherine Sheard, Simon J. Greenhill, Kaius Sinnemäki, José Segovia Martín, Jonas Nölle, Aleksandrs Berdicevskis, Archie Humphreys-Balkwill, Hannah Little, Kit Opie, Guillaume Jacques, Lindell Bromham, Peeter Tinits, Robert Ross, Sean Lee, Emily Gasser, Jasmine Calladine, Matthew Spike, Stephen Mann, Olena Shcherbakova, Ruth Singer, Shuya Zhang, Antonio Benítez-Burraco, Christian Kliesch, Ewan Thomas-Colquhoun, Hedvig Skirgård, Monica Tamariz, Sam Passmore, Thomas Pellard & Fiona Jordan. 2020. Chield: The causal hypotheses in evolutionary linguistics database. *Journal of Language Evolution* 5(2). 101–120.
- Romero-Figeroa, Andrés. 1985. OSV as the basic order in Warao. *Lingua* 23(1). 105–121.
- Schepens, Job, Frans van der Slik & Roeland van Hout. 2013. Learning complex features: A morphological account of L2 learnability. *Language Dynamics and Change* 3(2). 218–244.
- Seiler, Walter. 1985. *Imonda, a Papuan language* (Pacific Linguistics B 93). Canberra: Australian National University.
- Siewierska, Anna & Dik Bakker. 2008. Case and alternative strategies: Word order and agreement marking. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 290–304. Oxford: Oxford University Press.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), Language complexity: Typology, contact, change, 67–88. Amsterdam: John Benjamins.
- Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), Language complexity as an evolving variable (Oxford Studies in the Evolution of Language 13), 125–140. Oxford: Oxford University Press.
- Sinnemäki, Kaius. 2010. Word order in zero-marking languages. *Studies in Language* 34(4). 869–912.

- Sinnemäki, Kaius. 2014. Cognitive processing, language typology, and variation. *WIREs Cognitive Science* 5(4). 477–487.
- Sinnemäki, Kaius & Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9. 1141. https://doi.org/10.3389/ fpsyg.2018.01141.
- Skaug, Hans J. & David A. Fournier. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics & Data Analysis* 51. 699–709.
- Skaug, Hans J., David A. Fournier, Ben Bolker, Arni Magnusson & Anders Nielsen. 2016. Generalized linear mixed models using 'AD model builder'. R package version 0.8.3.3. Available at https://CRAN.R-project.org/package=glmmADMB (accessed 14 May 2020).
- Song, Jae Jung. 2012. Word order. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt. 2016. An analytic-synthetic spiral in the history of English. In Elly van Gelderen (ed.), *Cyclical change continued* (Linguistics Today 227), 93–112. Amsterdam: John Benjamins.
- Thomason, Sarah Grey. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & R. Harald Baayen. 2019. Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 1–39. https://doi.org/10.1017/ S0022226719000203.
- Torres Cacoullos, Rena & Catherine E. Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692.
- Trudgill, Peter. 1998. Typology and sociolinguistics: Linguistic structure, social structure and explanatory comparative dialectology. *Folia Linguistica* 31. 349–360.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford: Oxford University Press.

UCLA Statistical Consulting Group. no date. Introduction to SAS. UCLA: Statistical Consulting group. Available at https://stats.idre.ucla.edu/sas/modules/sas-learningmoduleintroduction-to-the-features-of-sas/ (accessed June 3, 2020).

- Wichmann, Søren. 2015. Diachronic stability and typology. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 212–224. London: Routledge.
- Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Witzlack-Makarevich, Alena. 2019. Argument selectors: A new perspective on grammatical

relations. An introduction. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), Argument selectors: A new perspective on grammatical relations, 1–38. Amsterdam: John Benjamins.

**Supplementary material:** The online version of this article offers supplementary material (https://doi.org/10.1515/jhsl-2019-1010).