# Sociolinguistic Typology: Complexification and Simplification

Christian Bentz

*University of Passau, Germany*

Annemarie Verkerk

*Saarland University, Germany*

1 Introduction
2 Data
3 Methods
  3.1 Cross-linguistic statistical analyses
    3.1.1 Linear models (LM)
    3.1.2 Linear mixed effects models (LMM)
    3.1.3 Phylogenetic regression
    3.1.4 Current developments
  3.2 Language learning experiments
    3.2.1 Single generation experiments
    3.2.2 Iterated learning experiments (multiple generations)
    3.2.3 Emerging languages in the laboratory
  3.3 Agent-based computational simulations
4 Outlook
5 Conclusion

## 1 Introduction

Every person uses language in their own unique way. Nuances in pronunciation, choice of words, morphemes, and sentence structures are a 'linguistic fingerprint' for the language user. In twentieth-century linguistics, this variation has often been set aside as superficial noise of language performance, a veil over the universal competence which restricts the space of possible languages.

*The Wiley Blackwell Companion to Diachronic and Historical Linguistics*. Edited by Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin É. Kiss, Joseph Salmons, and Alexandra Simonenko.

DOI: 10.1002/9781119898023.wbcdl046

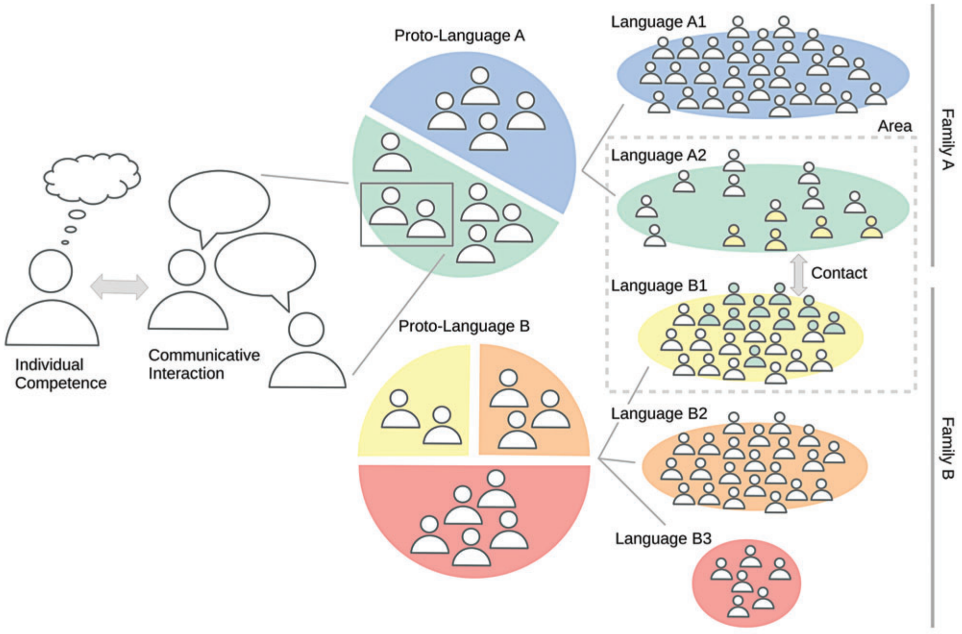

**Figure 1** From individual speakers to language families. Language diversity is the outcome of micro and macro processes of change and interaction. Individual language competence might have a universal, immutable core, but it is certainly shaped in communicative interactions. These are the 'microbiome' of language, a cradle for variance and innovations which might or might not spread to other individuals (local networks), and the entire population (language). Populations, in turn, can split in time, with their subpopulations forming different dialect groups, and finally new languages through geographic and social isolation. Languages can then be grouped into families according to (remaining) similarities. Another level of grouping is geographic: language areas are the outcome of language contact which yields similarity without common descent. Crucially, populations can differ along different dimensions, such as their size, connectedness, proportion of adult learners, etc.

However, inspired by disciplines such as evolutionary biology, genetics, and computer science, language researchers in the twenty-first century have gradually realized that this 'noise' might well function as the cradle for innovations which eventually shape new dialects, languages, and families. Modeling the unfolding of diversity from the individual language user to global networks of communication is a formidable task. Figure 1 unfolds the micro and macro levels of language diversity and change. Without understanding the moving parts of *languages*, it is impossible to delineate the immutable core of *language*.

This change of perspective brings about some fundamental questions: Are random processes sufficient to explain language diversity; or do we need to posit selective pressures which drive its unfolding? Random change and its accumulation are often viewed as a null hypothesis in testing for evidence of selection (Lupyan and Dale 2016). In addition, which parts of language structure are more encapsulated and resistant to external influence, and which are, conversely, frequently shaped by 'extralinguistic' factors?

The latter of these questions in particular has been approached within the framework which became known as *sociolinguistic typology*, synthesized as follows in Trudgill (1992, 195):

> The issue I want to address … is the question to what extent it is possible to link the typology of language varieties to the typology of societies. That is: do different types of society produce different types of language structure; and do they produce different types of linguistic change?

Note that, in this context, the term 'sociolinguistics' is linked specifically to linguistic variation across languages, while variationist sociolinguists is more concerned with variation within single languages. In its broadest interpretation, sociolinguistic typology encompasses all kinds of 'environmental' factors, that is, geography, climate, society, culture, economy, et cetera, and how they are reflected in typological features of languages. It is impossible to cover all of the studies which have proposed some link between the linguistic environment and typological features in this broad sense. In this entry, we more specifically focus on questions relating to complexification and simplification. Most of the time, this relates to morphology, in fewer cases also to the lexicon, phonology, syntax, or 'grammar' more generally. We cast the net even narrower by concentrating on research into the connection between population structure and morphological complexity – not least because this is a nexus which has been scrutinized from different scientific angles, but still remains controversial. However, most research on complexification and simplification deals with morphology; a notable exception is Walkden and Breitbarth (2019).

Early traces of sociolinguistic typology in this narrow sense are found in the second half of the twentieth century, for instance, in the writings of Braunmüller (1984) and Nichols (1992). Braunmüller (1984) discusses morphological phenomena in Germanic languages, in particular, varieties such as Icelandic, Faroese, and Frisian. He notes that the inflectional complexity of these 'small' and 'isolated' varieties seems more pronounced than in their relatives featuring larger populations of speakers with a history of contact. For example, according to Braunmüller (1984, 65), Faroese features seven different forms of the word 'day' according to number and case distinctions: *dagur, dags, degi, dag, dagar, daga, døgum*. In Modern Standard German, this is reduced to four (*Tag, Tages, Tage, Tagen*), and in Modern English to two (*day, days*). Braunmüller (1984, 53) speculates about a (causal?) relationship between the isolated and 'uninterrupted' development of smaller languages and their preservation of more complicated and opaque paradigms.

While Braunmüller (1984) and a few others (see further early examples in Baechler and Seiler 2016, 1–9) based their hypotheses on particular language varieties and genera, Nichols (1992) provided a first birds-eye view. In her more general account, languages of so-called *spread zones* are associated with lower levels of complexity (in different domains of grammar) compared to languages in so-called *residual zones*, that is, mountainous and rather isolated regions. Languages at the fringes of spread zones are in turn less prone to simplifying innovations than languages at the very core of a spread (see also Nichols 2016).

Around the same time as Nichols's account of language in time and space, Trudgill started to popularize the research paradigm of sociolinguistic typology. In a series of publications (Trudgill 1984, 1992, 1995, 2010, 2011), he laid out in more detail the hypothetical characteristics of societies relevant to morphological typology: *population size*, *network density*, *social stability*, degree of *shared information*, and *contact versus isolation*. With regard to the last point, he clarifies that different types of contact are expected to have different consequences on morphology. Contact involving adult L2 learners is expected to engender simplification via imperfect learning, while child bilingualism is hypothesized to lead to complexification via morpheme borrowing. Notably, population size per se and language contact are here seen as two separate variables, while in later accounts they are often seen as two sides of the same coin.

Trudgill's call for a systematic investigation of societal structure and language structure has inspired a whole series of studies of languages and families around the world. Kusters (2003), for instance, analyzes the loss of verbal inflections in Arabic, Scandinavian, Quechua, and Swahili varieties, and shows a positive correlation between transmission stability, that is, relative presence/absence of L2 learners in a community, and preservation of verbal inflections. In a similar vein, McWhorter (2007) investigates the "signs of non-native acquisition" in Standard English, Mandarin Chinese, Persian, Colloquial Arabic, and Malay. Moreover, DeLancey (2014, 2015) takes sociolinguistic typology to languages of India and the Himalayas respectively. It is hard to get an overview of all studies referring to Trudgill's framework for argumentation and analyses in the last three decades. We have named here but a few, for a more extensive sample we refer the reader to volumes such as Miestamo, Sinnemäki, and Karlsson (2008) and Sampson, Gil, and Trudgill (2009).

An important waypoint for the development of sociolinguistic typology came with what we call here the 'quantitative turn' promoted by studies such as Sinnemäki (2009) and Lupyan and Dale (2010). These opened the debate also to other subdisciplines of linguistics and cognitive science, inviting further experimental and computational work. To be sure, research in the tradition of sociolinguistic typology was always prone to some degree of quantification. After all, arguing about simplicity and complexity requires a quantification of said concepts. Fully quantitative accounts, however, go a step further by systematically collecting – or generating – data, and fitting models to answer constrained research questions. Since these studies are currently shaping the debate about sociolinguistic typology, and will hence be relevant to future researchers, this entry is mainly devoted to giving the reader an overview of the current state of the art in terms of data and methods of quantitative studies.

To this end, we first give a primer on the types of data used to test hypotheses of sociolinguistic typology in Section 2. We then divide quantitative studies into three types, to which we devote separate sections: *cross-linguistic statistical analyses* (Section 3.1), *language learning experiments* (Section 3.2), and *agent-based computational models* (Section 3.3). Finally, we touch upon some pertinent issues with the current research, and future research directions.

## 2 Data

In this section, we focus on the types of data which have been used to support sociolinguistic claims. Several key publications (e.g., Trudgill 1992, 2011; McWhorter 2007) are 'qualitative' in the sense of providing theoretical arguments about mechanisms through which complexification and simplification may arise. Of course, these also carefully garner empirical data – mostly in the form of examples from different languages – to support the respective hypotheses.

In the case of cross-linguistic statistical analyses, on the other hand, materials on linguistic and demographic dimensions are gathered as input to statistical models. Many of these studies focus on specific areas or families, for which the data is collected

**Table 1** Selection of databases used in sociolinguistic complexity research.

| Name | Reference | Sample[b] | Description |
|---|---|---|---|
| Ethnologue | Eberhard, Simons, and Fennig (2023) | 7168 languages | language information |
| Glottolog | Hammarström et al. (2022) | 8180 languages | language information |
| AUTOTYP | Bickel et al. (2022) | 3070 languages and dialects | typological database |
| WALS | Dryer and Haspelmath (2013) | 2662 languages and dialects | typological database |
| Grambank | Skirgård et al. (2023) | 2430 languages | typological database |
| GramAdapt | Kashima et al. (2023) | 34 language pairs | sociolinguistic database |
| PHOIBLE | Moran and McCloy (2019) | 2186 languages | phonological inventory database |
| Universal Dependencies (UD) | Zeman et al. (2022) | 135 languages | corpus collection |
| Leipzig Corpus Collection | Goldhahn et al. (2012) | 293 languages | corpus collection |
| Parallel Bible Corpus (PBC) | Mayer and Cysouw (2014) | ca. 1500 languages | parallel corpus |
| UDHR[a] | https://unicode.org/udhr/ | 487 languages | parallel corpus |
| Europarl Parallel Corpus | Koehn (2005) | 21 languages | parallel corpus |

[a] Universal Declaration of Human Rights.

[b] The definition of 'language' differs between datasets. Also, the numbers here given are just a snapshot in time, as most of these databases evolve to include more data.

specifically for a particular analysis. Other studies use a range of databases and corpora in order to generate larger samples. A selection of databases and corpora typically used in this context are listed in Table 1. In language learning experiments, on the other hand, the data consist of the input stimuli as well as the output of participants. Finally, in computational models, the data is mostly generated by a predefined process, potentially with parameter settings inspired by empirical observations.

One important distinction we want to draw attention to is between distributional and 'aggregated' data, or between *token* and *type-based* typology (see also discussions in Wälchli 2012; Bentz et al. 2017; Levshina 2022). In a population-based model of language change as outlined in Figure 1, a 'language' comes to be reflected in the communicative interactions at a certain time. In an ideal world, we could measure language structure directly from the entirety of these interactions. In the reality of linguistic research, however, we rather use corpora as a mere reflection of the panoply of linguistic interactions. Observations on corpora are often further aggregated into categorical, and even binary features. We illustrate these aggregation steps for the example of Faroese and German case-marking in Figure 2. It is important to keep in mind that different levels of aggregation will lead to different results when comparing languages. We further discuss this issue in Section 4.

Out of the databases and corpora listed in Table 1, Ethnologue (Eberhard, Simons, and Fennig 2023) and Glottolog (Hammarström et al. 2022) are central. Ethnologue has been vital for cross-linguistic statistical analyses as it contains the most extensive collection of information on speaker numbers of the languages of the world. This is perhaps the most widely used sociolinguistic measure, first correlated with morphosyntactic

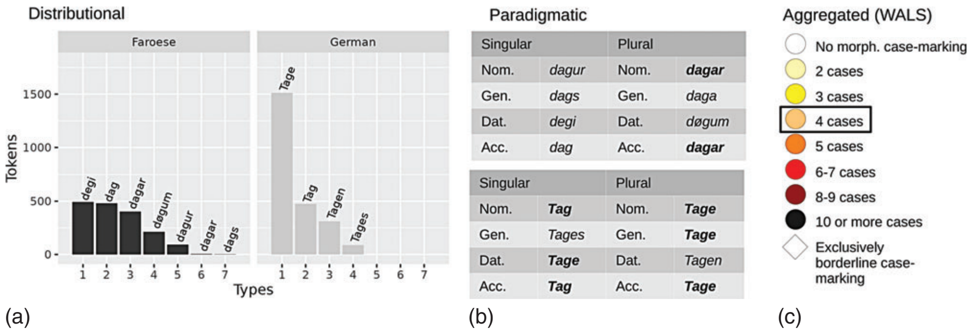

**Figure 2** The granularity of different types of data. Example of case-marking in Faroese (fao) and German (deu) for the word 'day' ('fao' and 'deu' are isocodes, three-letter codes used to identify languages). (a) The actual distributions of token frequencies over word types. These are here taken from texts of the Parallel Bible Corpus (PBC). (b) Paradigms condense the information from corpora into grid cells along dimensions such as NUMBER and CASE, where cells are filled with word types. (c) Typological databases such as WALS further aggregate this to yield discrete categories, which can be multi-factorial (here the example of case-marking categories), or even just binary (case-marking/no case-marking).

features by Sinnemäki (2009) and Lupyan and Dale (2010), but still being used to this day (Shcherbakova et al. 2023, for example). Glottolog is important because it provides a comprehensive view of the world's languages and their varieties, and is being harnessed as a tool to match linguistic and demographic information across datasets. In addition, Gil (2021) uses it to obtain a measure on language family size to explain the prevalence of tense-aspect-mood marking.

Databases collecting typological features have also been widely used in sociolinguistic typology, examples being Lupyan and Dale (2010), Moran, McCloy, and Wright (2012), Sinnemäki and Di Garbo (2018), Sinnemäki (2020), Gil (2021), Shcherbakova et al. (2022), and Shcherbakova et al. (2023). Another major data source is corpora, including parallel corpora; some relevant papers here are Bentz et al. (2015), Koplenig (2019), and Koplenig, Wolfer, and Meyer (in preparation).

As is clear from Table 1, the corpora (except for the Parallel Bible Corpus) feature much smaller language samples than the typological databases. This presents an issue for sociolinguistic typology, as the study of text-based structural features is limited to a fragment of human languages. However, the typological databases only capture part of the world's linguistic diversity, too. This point is illustrated in Figure 3, which illustrates the information on the number of speakers per language from Ethnologue, and coverage in WALS (Dryer and Haspelmath 2013). This serves to visually illustrate that existing typological databases (and the current state of description of the world's languages) clearly miss out on a sizable proportion of the world's linguistic diversity, and rather sample heavily from the same restricted number of languages.

## 3 Methods

### 3.1 Cross-linguistic statistical analyses

Inspired by hypotheses formulated in the sociolinguistic literature discussed in the introduction, cross-linguistic statistical analyses started to emerge in the last two

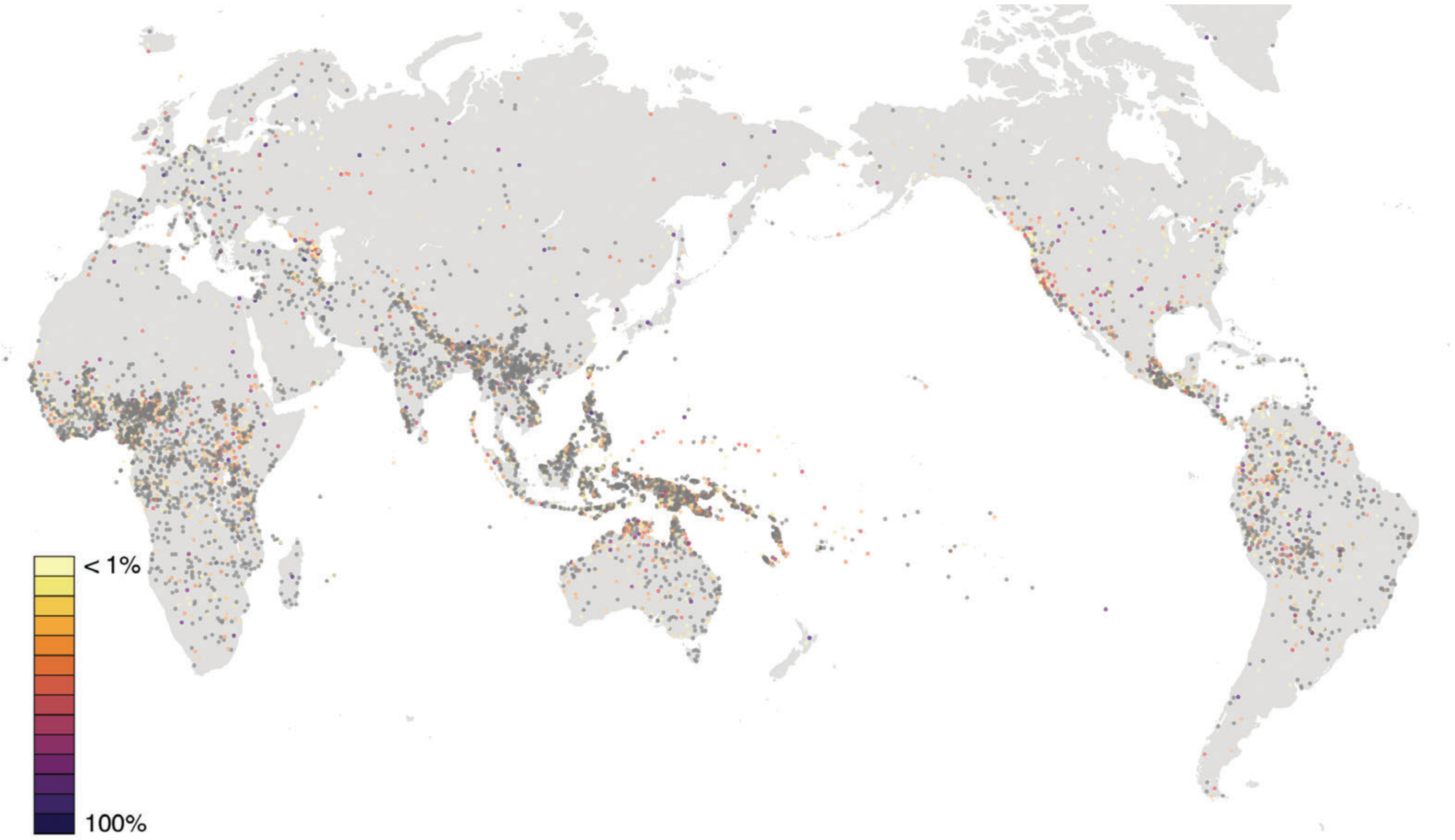

**Figure 3** Languages of the world. Information on the number of speakers is available for all points on the map; the languages indicated by the gray points are not featured in WALS. WALS coverage in terms of how many chapters the language appears in is captured by the color scale, languages in dark purple being featured often, and in light yellow very rarely.

decades. In such analyses, a so-called *dependent variable*, that is, some measure of complexity in our case, is statistically associated with one or more *predictor variables*, for example, population size, percentage of adult learners, endangerment status, et cetera. We here discuss some models which have been used to this effect.

### 3.1.1 Linear models (LM)

Assume we have determined the complexities[1] of $n$ languages according to our measure of choice, such that we have a variable $Y$ of $n$ outcomes, with each outcome $y_i \in \mathbb{R}$, and $i = 1 \ \ldots, n$. In other words, the languages in our sample are assigned real-valued complexities. In statistical analyses, we typically want to predict this so-called *dependent variable* given *predictor variables* $(X_1, X_2, \ldots, X_k)$ which are themselves numerical or categorical (i.e., taking factor levels). An example of a numerical variable would be the size of a speaker population, while the endangerment status of a language would be a categorical variable (e.g., taking factor levels such as *living*, *endangered*, *extinct*). In a first step, we can build a linear model[2] such that (cf. Baayen 2013, 339):

$$\begin{aligned} Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon, \\ \varepsilon \sim N(0, \sigma^2_\varepsilon). \end{aligned} \tag{1}$$

This models the outcomes of $Y$ as the sum of the intercept $\beta_0$ (i.e., the value $Y$ takes when the values of $X_1, X_2, \ldots, X_k$ are zero), the slopes $\beta_1, \beta_2, \ldots, \beta_k$ multiplied with the respective predictor variable values (in a sense, the slopes are the weights given to each predictor variable value), and a random noise term $\varepsilon$. The latter is assumed to be normally distributed around mean zero and variance $\sigma^2_\varepsilon$. Note that the outcomes of $Y$ and $X_1, X_2, \ldots, X_k$ are empirically given – one value for each language (bare any missing data).[3] The parameters to be estimated are the so-called coefficients of the model ($\beta_0$, $\beta_1, \beta_2, \ldots, \beta_k$) as well as the variance of the error term $(\sigma^2_{\varepsilon;})$. A common method in this context is Maximum Likelihood Estimation (MLE). In a nutshell, this includes finding the parameter values which maximize the similarity between the predicted values given our model and the actually observed values for $Y$ (see, for instance, Myung 2003, for an introduction to MLE).

As a concrete example relating to sociolinguistic typology, Lupyan and Dale (2010) use generalized[4] linear models (GLM) to establish statistical links between overall 28 morphological features of the WALS, on one hand, and three numerical predictors relating to social structure of a given language: population size, language area in km$^2$, and the number of neighboring languages. Additionally, they factor in the effect of language family and geographic location (i.e., latitudes and longitudes) by using them as further predictors (i.e., covariates) in the model. We thus have models like the one specified in (1), with the respective morphological feature as dependent variable, four numerical predictors, and one categorical predictor (language family). Among other findings, Lupyan and Dale (2010) show that for most of the 28 morphological features population size is a significant predictor: more complex morphological coding strategies are associated with smaller population sizes.

Similar statistical models have been used in a range of further studies over the past two decades. For example, Sinnemäki (2009) establishes a statistical link between the complexity of argument marking (head marking, dependent marking, word order) and population size (i.e., 'small' or 'large' with different thresholds).[5] Larger populations tend to adhere to the one-meaning-one-form principle in argument marking, while small populations tend to deviate from it.

### *3.1.2 Linear mixed effects models (LMM)*

A pertinent issue for statistical models involving data from different languages is *non-independence*. Languages are often similar because they are related and have inherited characteristics from a common ancestor, or because they are spoken in close geographic proximity and have borrowed from one another – or both (remember Figure 1).

A possible remedy is to sample languages from different families (or genera) and different geographic areas (cf. Dryer 1989; Sinnemäki 2009). However, as information on a linguistic feature might be available for hundreds or thousands of languages, only including independent languages will drastically shrink the sample. An alternative approach is introduced in Cysouw (2010) and Jaeger et al. (2011). In the context of typological analyses, they propose to include fixed effects (predictors) and random effects (levels of grouping) in a so-called *mixed effects* model.[6] For instance, we can include language family (f) and language area (a) as random effects in our model from above (with just a single predictor $X_1$ for simplicity), such that (cf. Baayen 2013; Jaeger et al. 2011; Barr et al. 2013):

$$
\begin{aligned}
Y_{f\,a} &= \beta_0 + b_{0f} + b_{0a} + (\beta_1 + b_{1f} + b_{1a})X_1 + \varepsilon_{fa}, \\
&\qquad\qquad\qquad\qquad\qquad \varepsilon_{fa} \sim N(0, \sigma_\varepsilon^2), \\
(b_{0a}, b_{1a}) &\sim N\left(0, \begin{bmatrix} \sigma_{b0a}^2 & \rho\sigma_{b0a}\sigma_{b1a} \\ \rho\sigma_{b0a}\sigma_{b1a} & \sigma_{b0a}^2 \end{bmatrix}\right), \\
(b_{0f}, b_{1f}) &\sim N\left(0, \begin{bmatrix} \sigma_{b0f}^2 & \rho\sigma_{b0f}\sigma_{b1f} \\ \rho\sigma_{b0f}\sigma_{b1f} & \sigma_{b0f}^2 \end{bmatrix}\right),
\end{aligned}
\tag{2}
$$

In this example, we introduce *random intercepts* ($b_{0f}$, $b_{0a}$) and *random slopes* ($b_{1f}$, $b_{1a}$) for family and area.[7] Random intercepts are family and area-wise adjustments to the general intercept ($\beta_0$). Imagine that $Y$ is a variable reflecting measurements of morphological complexity. It is possible that languages in our sample which belong to the Sino-Tibetan family have systematically lower scores than the global average, while languages belonging to the Nakh-Daghestanian family have systematically higher scores. The introduction of $b_{0f}$ will adjust the intercept downwards for all Sino-Tibetan languages, and upwards for all Nakh-Daghestanian languages. Random slopes, on the other hand, adjust the effect ($\beta_1$) of our predictor variable ($X_1$) according to family and area. Imagine that a negative relationship between population size and morphological complexity holds within the Sino-Tibetan family, but not within the Nakh-Daghestanian family. This is taken into account by the addition of a random slope ($b_{1f}$) per family.

Introducing these adjustments to intercepts and slopes considerably increases the complexity of the model compared to a simple linear model. This becomes apparent also in the model specification in (2). While normality is still assumed for the error term ($\varepsilon$) as before, the random intercepts and slopes now also come with their own normality assumptions, and, on top of this, the variance–covariance matrices for random intercepts and slopes need to be taken into account (lines 3 and 4 in (2)). For instance, it is possible that the intercept adjustments for families are correlated with the family-wise slopes. This is represented by the covariance term $\rho\sigma_{b0f}\,\sigma_{b1f}$, where $\rho$ is the correlation parameter, see also Barr et al. (2013, 260). In our hypothetical example, it is possible that Sino-Tibetan languages have lower morphological complexity on average *exactly because* they have large populations which are connected to simplification, while for Nakh-Daghestanian languages the inverse effect occurs. If this is generally the case across families, then we will find correlations between the intercept and slope adjustments. In extreme cases of almost perfect correlations, it might not be justified to include both random intercepts and slopes, since they unnecessarily inflate the model complexity.

In addition, adding random intercepts and slopes is only possible if there are enough languages within families and areas. For instance, Jaeger et al. (2011, 298) only include families with at least four members in their analyses. However, this automatically excludes many language isolates from the analyses. For further problems, pitfalls, and solutions with mixed-effects model fitting see also Baayen, Davidson, and Bates (2008), Jaeger et al. (2011), Barr et al. (2013), Bates et al. (2015), and Coupé (2018).

Having said this, the general purpose of linear mixed-effects models (LMM) is still the same as in the simple linear model, namely, estimating the coefficients of the predictor variable(s) which we are mainly interested in – but taking levels of grouping into account. As a concrete example, Sinnemäki and Di Garbo (2018) use LMMs to test the link between population sizes and adult speaker proportions (as predictors), on one hand, and verbal synthesis and number of genders (as dependent variables), on the other. Their samples comprise several hundred languages. They introduce random intercepts and slopes for AUTOTYP stocks and geographic areas (Bickel et al. 2022). Among other results, they argue that population size is a significant predictor for verbal synthesis, but not for the number of genders.

### 3.1.3 Phylogenetic regression

In evolutionary biology, it has long been noted that – since species share common ancestries – standard regression models are not appropriate for modeling species data. To address this issue, *phylogenetic generalized least squares* (PGLS) models have been developed (Grafen 1989; Pagel 1997). Remember that in the linear regression of (1), the residual errors are assumed to be normally distributed, that is, $\varepsilon \sim N\,(0, \sigma_\varepsilon^2)$. However, this is only the case if all species – or languages in our case – have evolved independently from one another. In a scenario where, on the contrary, species or languages are dependent due to common descent, we need to take this into account. Namely, a variance–covariance matrix **C** is multiplied with the variance term, such that (cf. Revell 2010, 319):

(3) $\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{C}).$

The variance–covariance matrix can be derived from a phylogeny. For example, assume the phylogeny with branch lengths as given in Figure 4. The matrix of expected[8] variances and covariances for this phylogeny is:

(4) $$\mathrm{C} = \begin{bmatrix} \mathbf{5} & 3 & 0 & 0 \\ 3 & \mathbf{6} & 0 & 0 \\ 0 & 0 & \mathbf{5} & 2 \\ 0 & 0 & 2 & \mathbf{4} \end{bmatrix}.$$

The diagonal elements in this matrix give the summed lengths of the branches from root to tip, for example, $3 + 2 = 5$ for Italian (row 1, column 1), and $3 + 3 = 6$ for French

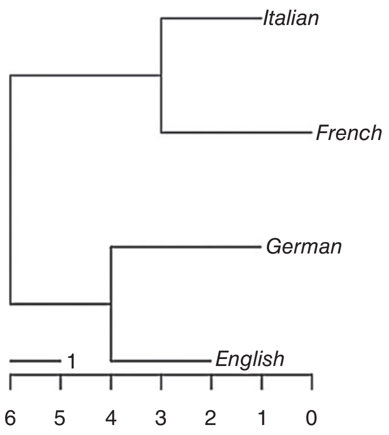

**Figure 4** Hypothetical phylogeny for four languages.

(row 2, column 2), et cetera. This corresponds to the *expected variance* in the tip value of a given language. The off-diagonal elements, on the other hand, give the branch lengths of the shared ancestry of any two tips, for example, 3 for Italian and French (row 1, column 2), 0 for Italian and German (row 1, column 3), et cetera. These correspond to the *expected covariances* between languages. In other words, we expect Italian and French to display more covariance in their tip values than Italian and German, given this particular phylogeny. In a phylogenetic regression, this variance–covariance structure is taken into account when estimating the other regression parameters, such as the coefficient(s) and intercept(s). For an introduction with a worked example see also Symonds and Blomberg (2014).

The general idea behind phylogenetic regression is similar to mixed-effects regression with random intercepts and slopes per language family, namely, to take non-independence due to phylogeny into account. However, note that phylogenetic regression is 'finer-grained' in the sense that the variance–covariance matrix reflects even minor differences in the tree structure, while random effects are typically restricted to the language family or genus level. As a downside, however, phylogenetic regression can only be applied to languages within a given family – since the available phylogenetic trees in linguistics are typically restricted in depth. Also, it is not straightforward how to include further factors such as areal patterns into this type of regression (but see Guzmán Naranjo and Becker 2022; Guzmán Naranjo and Mertner 2022; Shcherbakova et al. 2023).

An application of PGLS is found in Bentz et al. (2015). In this study, the regression coefficient of the ratio of adult learners (L2) over first-language learners (L1) on measures of lexical diversity is estimated for 26 Indo-European languages. The phylogenetic relatedness of these languages is factored in using a tree sample for the Indo-European family. It is shown that the effect of L2/L1 ratio on lexical diversity is significant even after controlling for phylogeny. However, this study also illustrates one of the major drawbacks of PGLS, namely, the issue of data sparsity. Given that PGLS requires not only the data for the dependent and independent variable(s), but also specifically formatted family trees, the respective samples for such analyses are likely small. This, in turn, strongly limits our ability to draw more general conclusions about languages across the world.

On a more general note, when using phylogenetic regression models for language data, we should be aware that we buy into some of the fundamental assumptions about species' evolution also for languages. For instance, if the variance−covariance matrix **C** is derived as in the working example above, we implicitly assume that languages have evolved via Brownian motion (i.e., a random walk) along the branches of the phylogeny. To what extent such assumptions are justified for language data is an open research question (see for example Bowern 2018).

### *3.1.4 Current developments*

In recent years, several proposals have emerged to counter some of the shortcomings of the statistical models discussed above. For instance, Jäger and Wahle (2021) extend phylogenetic analyses of typological characters beyond single family trees by deriving a global tree of relatedness from word lists. Another avenue is to develop models which take both phylogenetic and geographic information into account (Cathcart et al. 2018; Ranacher et al. 2021), for instance, by including a Gaussian process term to model areal diffusion in a phylogenetic generalized linear mixed model (PGLMM) (Guzmán Naranjo and Becker 2022; Guzmán Naranjo and Mertner 2022).

Finally, besides controlling for potential confounds, another pertinent issue in statistical modeling is causality: *correlation does not imply causation* is an often repeated mantra (see Roberts 2018, for a discussion relating to language data). Applying methods such as Granger-causality to historical data (Moscoso del Prado Martin and Brendel 2016), as well as synchronic data (Levshina 2021) will help to disentangle cause and effect in the change of complexity. To a similar effect, the Causal Hypotheses in Evolutionary Linguistics Database (CHIELD) is mapping out the network of alleged causal connections between sociolinguistic variables and language structures – also beyond research into language complexity (Roberts et al. 2020).

We have here given an overview of the statistical methods currently used to test sociolinguistic hypotheses. We picked certain studies as examples for applications of the respective models. For the interested reader, we give a condensed overview of further studies in Table 2.

## 3.2 Language learning experiments

The cross-linguistic statistical studies summarized here have helped spread the topic of sociolinguistic typology to cognitive science. Researchers in this field often work with experiments bringing human participants into the laboratory. In a sense, this is like holding a 'looking glass' to individual language learners to understand the micro processes fueling language change. This also serves to uncover the causal mechanisms underlying simplification and complexification.

Language learning experiments revolve around simulating language change 'in the lab' by teaching a real or artificial language to participants of an experiment, and measuring how well they fare at learning the respective structures. There is a large number of studies in the first and second language learning literature which at least indirectly speak to the question of how simple and complex language structures are learned. However, we here focus on experimental studies which explicitly design task setups to directly shed light on the link between language learning and language simplification/complexification.

**Table 2** Overview of further statistical studies about hypotheses derived from sociolinguistic typology.

| Study | Language sample[a] | Dependent variable(s) | Predictor variable(s) | Model(s)[b] |
|---|---|---|---|---|
| Sinnemäki (2009) | 50 languages | core argument marking | population size | Chi-squared Test |
| Szmrecsanyi and Kortmann (2009) | 46 varieties of English | 76 morphosyntactic features | type of variety (LI, L2) | Correlation, MDS, CA, PCA |
| Lupyan and Dale (2010) | 2236 languages | 28 WALS features (morph.) | population size, spread, neighbors | GLM |
| Bentz and Winter (2013) | 66 languages | number of nominal cases | L2 perc., population size | GLMM |
| Nichols (2013) | 26 languages (Caucasus) | inventory sizes, opacity | altitude, longitude, population size | Correlation |
| Maitz and Németh (2014) | 4 varieties of German | syntheticity, analyticity, etc. | high versus low contact varieties | ANOVA |
| Bentz et al. (2015) | 647 languages | lexical diversity | L2/L1 ratio | LM, LMM, PGLS |
| Nichols and Bentz (2018) | 193 languages | unigram entropy, inventory size | altitude, longitude | Correlation |
| Sinnemäki and Di Garbo (2018) | 309/345 languages | inflection synthesis, gender | LI speakers, L2 prop. | GLMM |
| Bentz (2018) | 115/1193/1197 languages | unigram entropy | population size, L2 perc., status | Correlation, LM, LMM. |
| Cathcart et al. (2018) | 117 languages | 118 binary morph, features | areality (language contact) | Phylogeogr. mode, GLMM |
| Koplenig (2019) | 1088/1581 languages | entropy rate, morph. complexity | population size, vehicularity | Correlation, LMM |
| Meakins et al. (2019) | Gurundji Kriol | 120 features | low, medium, high complexity | Write-Fisher model |
| Sinnemäki (2020) | 66 languages | number of nominal cases | population size, L2 perc., word order | GLMM |
| Gil (2021) | 868 languages | TMA marking | language family size | chi-square test |
| Widmer et al. (2021) | 4 languages | 5 variables (e.g., morph. synthesis) | low versus high contact | GLMM |
| Morozova, Escher, and Rusakov (2022) | 17 South Slavic dialects | 29 features | altitude, distance to Albanian | LM |
| Kauhanen, Einhaus, and Walkden (2023) | 1499/717/171 languages | entropy rate, morph. complexity | population size, vehicularity, L2 prop. | LMM |
| Shcherbakova et al. (2023) | 1314 languages | fusion, informativity | population size, L2 prop. | spatiophylogenetic |

[a] We here give the overall number of languages/varieties encompassed in a given study. Note that this can differ substantially from the number actually included in any given analysis. In cases where these different numbers are explicitly given in the study, we include these delimited by slashes.

[b] ANOVA: Analysis of Variance; CA: Cluster Analysis; GLM: Generalized Linear Model; GLMM: Generalized Linear Mixed-Effects Model; LM: linear model; LMM: Linear Mixed-Effects Model, MDS: Multidimensional Scaling; PCA: Principal Component Analysis; PGLS: Phylogenetic Generalized Least Squares.

### 3.2.1 Single generation experiments

For example, Cuskley et al. (2015) ask native and non-native speakers of English to give past tense forms for English-sounding non-words in a so-called 'Wug' test.[9] They find that native speakers are actually more prone to regularize (i.e., using an *-ed* form) than non-native speakers. De Smet, Rosseel, and Van de Velde (2022) reproduce these findings for native and non-native speakers of Dutch. In a similar experimental setup, they too find that regularization of past tense markers is rather associated with native speakers than non-native speakers. If this behavior applies to more natural settings of language usage as well as a wider variety of languages, then we would expect that higher numbers of native speakers (rather than non-native speakers) lead to the regularization of verb forms. However, an important question then is how regularization relates to other measures of morphological simplicity/complexity.

Taking a different angle, Atkinson, Kirby, and Smith (2015) investigate the impact that input variability might have on learning outcomes. They hypothesize that an inverse relation between population size and morphological complexity could be explained by input variability increasing with the growth of a speaker population. They test this with English speakers learning Hungarian sentences featuring locative case markers in two conditions: the input coming from a single Hungarian speaker or multiple speakers. However, they do not find a significant difference between the conditions.

### 3.2.2 Iterated learning experiments (multiple generations)

The studies discussed above all feature a single generation of learners, that is, participants in the lab are exposed to stimuli, and their learning performance is measured and compared between different conditions. However, language change often happens over multiple generations. To model this, a particular flavor of language learning experiments uses so-called *iterated* setups (Kirby, Cornish, and Smith 2008): The output of a given participant when tested on the respective language structure serves as the input for the participant(s) of the next generation. The data which is generated this way is thus unique for each experimental chain of iterated learning, consisting of several generations. See example chains of an actual experiment in Figure 6.

Atkinson, Smith, and Kirby (2018) use this setup with two generations of learners. In their first experiment, they have participants learn number-marking morphology of a miniature language. For example, *-o* and *-op* mark singular and plural on nouns denoting different animals: 'crocodile' is *snap-o*, and 'crocodiles' is *snap-op*. Learners indeed simplify the inflectional system, especially in early rounds of learning. Simplification here occurs mainly by overgeneralization (i.e., using *-o* for both singular and plural) rather than dropping of markers (i.e., zero marking). Interestingly, when the output of learners with higher and lower degrees of simplification is mixed, and given as input to the second generation, this does not lead to general simplification in the output of the second generation. This begs the question if and how simplifications of inflectional systems due to imperfect learning spread throughout a population, and are handed down to subsequent generations in language change. Based on a third experiment, Atkinson, Smith, and Kirby (2018) argue that it might be the accommodation of advanced learners to beginners which magnifies the effect of simplification.

In two subsequent studies, Berdicevskis and Semenuks (2020, 2022) create an artificial language where nouns take plural suffixes and verbs agreement suffixes (see Figures 5 and 6). This is a miniature language reflecting extremely simplified Russian morphosyntax, and the participants in the experiment are Russian speakers. This

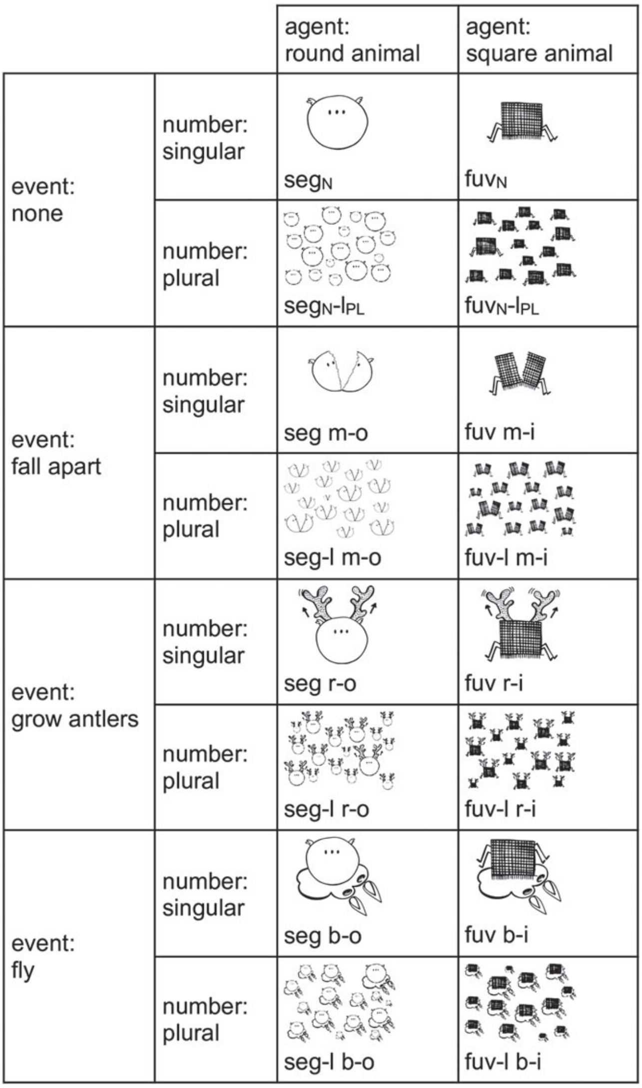

**Figure 5** Artificial language created for iterated learning experiments in Berdicevskis and Semenuks (2020, 2022). Plural is here marked with a regular *-l* suffix on nouns, and agreement with the noun is marked with suffixes on the verb (*-o* and *-i* ). Drawings were produced by Tanja Russita.

setup thus minimizes the morphosyntactic distance between the artificial language to be learned and the native language of participants, such that imperfect learning effects cannot be attributed to native language influence. Simple verb phrases (noun–verb combinations) in this language are paired with pictures, and participants are required to learn the picture–phrase mappings (see Figure 5). Their output when tested on the pictures is handed to the next generation for learning. Crucially, in these studies, there are chains of overall ten generations of iterated learners. Additionally, 'native' versus 'non-native' learning is taken into account by varying the amount of input a given generation of learners receives. Berdicevskis and Semenuks (2020, 2022) find that especially the verbal agreement system is prone to disappear over several generations of learning.

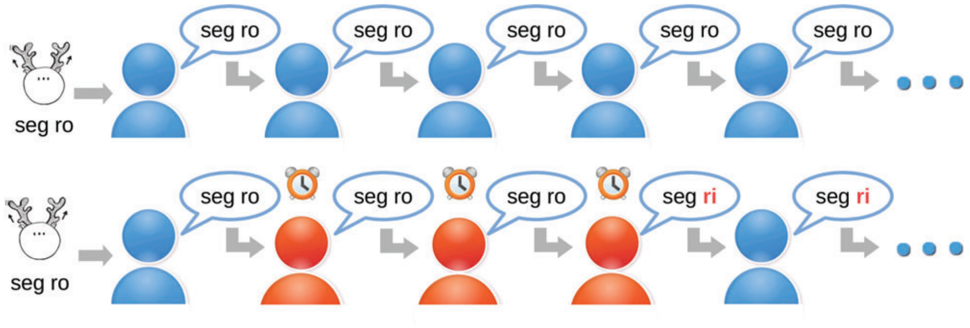

**Figure 6** Example of iterated artificial language learning experiment adapted from Berdicevskis and Semenuks (2022). A picture displaying an intransitive action is paired with a noun–verb combination (*seg* for round object, *r-* for growing antlers, and *-o* is an agreement marker to be used with this type of object). The output of one participant when naming the object (after a learning phase) is given to the next participant as label when learning the mappings. The upper chain consists only of learners with long exposure to the mapping rules, while the lower chain is 'interrupted' with three low exposure learners. Note that one of them uses an incorrect agreement marker, which is then handed on to the subsequent generations of learners. These example chains are taken from the actual experimental chains reported in the appendix of Berdicevskis and Semenuks (2022).

This effect is more pronounced when 'non-native' learners with less exposure to the system are introduced in the chains. Interestingly, however, the inflectional systems – if preserved – also become more irregular, rather than more regular. Berdicevskis and Semenuks (2020) hence argue that imperfect learning causes different trajectories on overspecification (i.e., loss of agreement), on one hand, and irregularity of inflectional markers, on the other.

### 3.2.3 *Emerging languages in the laboratory*

Despite different experimental setups, the studies falling under the rubrics above share a common feature: the target language – whether real or artificial – is given at the outset. Raviv, Meyer, and Lev-Ari (2019) take a different route: they have miniature language(s) emerge in the lab by letting participants label objects of a certain shape and movement direction, and subsequently use these labels in communicative tasks. They show that larger groups (8 participants) converge onto more structured labels than smaller groups (4 participants). Language structure is here measured as the correlation between distances of labels and distances of meanings (shapes and movement directions). In other words, language structure is measured as the degree to which similar labels are used to denote similar meanings. As an example, one of the participants in a large group developed a set of labels where the string *wowo* denotes a particular shape, the character *i* denotes an upward movement, and the character *k* denotes rightward movement, such that *wowoik* could be read as 'shape x moving right upwards'. Notably, as pointed out by Raviv, Meyer, and Lev-Ari (2019), there is no clearcut distinction between syntax and morphology in this context, as the resulting labels could be interpreted as either whole sentences or complex word forms. Having said this, they argue that population size per se contributes to differences in the structural regularity of the evolving languages, independent of other sociolinguistic variables such as proportions of adult learners or connectedness.

In a follow up study, Raviv, Heer Kloots, and Meyer (2021) illustrate that the highly structured languages evolved in their earlier experiment are indeed easier to learn for adults, while the partly structured languages – those where some irregularity remains – are not.

## 3.3 Agent-based computational simulations

Both cross-linguistic statistical analyses and language learning experiments come with their own limitations. Statistical analyses are often 'high level', that is, they collapse the many facets of real-world complexity into a few variables. Needless to say that there is a danger of losing important information in this process. Language learning experiments hold the magnifying glass to individual interactions in the lab, but are also highly simplified sketches of real-world interactions. It is simply not feasible to build the intricate complexities of real languages into stimuli, and have these learned by participants.

Given this state of affairs, some researchers turn to agent-based computational simulations to study the effects of micro interactions onto macro outcomes. They are a versatile tool to explore the space of possible evolutionary scenarios, especially when many factors can play a role for a given outcome. This is certainly the case for the link between socioecological factors and language structure. On both sides of the equation, there is a multitude of variables which could be connected. Also, the development of linguistic structure from individual speakers (personalized language), to pairs of speakers (language in interaction), and finally groups of speakers (language under conventionalization) is hard to disentangle for real-world data. In the controlled setting of computational simulations, different parameter values and their interactions can be explored. On the downside, the parameter settings need to be rail-guarded in some way in order to not become completely disconnected from the real-world questions we are asking.

Computational simulations have been used to better understand language change from the behavior of individuals to whole populations, for instance, in the case of evolving agreement systems (Beuls and Steels 2013), the rise of weak preterite forms (Pijpops, Beuls, and Van de Velde 2015), and the evolution of definite and indefinite articles (Blythe and Croft 2021). A crucial question in this context is whether *neutral* change, that is, replication of variants without systematic bias, suffices to explain an observed replacement of one linguistic variant for another, or whether *selective* pressure has to be invoked (cf. Blythe and Croft 2012; Kauhanen 2017; Blythe and Croft 2021).

To date, there are relatively few studies which directly relate their agent-based models to hypotheses on simplification and complexification from sociolinguistic typology. One example is Dale and Lupyan (2012). In their model, they represent agents simply as binary vectors, where '1' could be seen to stand for using a morphological marker in a message, while '0' could stand for using a different strategy. Agents have randomized bit vectors to start with, and interact with one another to adjust their vectors over time. Given this basic setup, Dale and Lupyan (2012) run four different simulations. They include a bias toward morphological marking for 'infant' agents, and a bias against morphological marking in 'adult' agents. With their last simulation they illustrate how a small population (5 agents), in which adults are replaced by infants in a birth–death process, can feature a wide variety of vectors with low or high proportions of ones (i.e., 'morphological markers'). However, when a large number of adult learners is introduced (15 adult agents), then the five 'native' agents gradually reduce the number of ‘1’s in their vectors. This is seen as a proof of concept for the causal role which adult learning might play for the link between population size and morphological simplification.

Cuskley, Loreto, and Kirby (2018) elaborate on the agent-based model architecture of Dale and Lupyan (2012) by more realistically crafting the message space which agents use to communicate: there is a distinction between lemmas and inflections, with lemmas being distributed according to a Zipfian distribution, and inflections can follow overall 12 different rules. Regularization is conceptualized as the spread of the usage of one inflectional rule to the majority of lemmas. After 100 runs of interactions, the larger populations of agents ($N = 100$) feature systematically more regularization behavior than smaller populations ($N = 20$). Cuskley, Loreto, and Kirby (2018) emphasize the relevance of population characteristics such as the introduction of new agents (growth) or replacement of agents (turnover), which possibly outweigh the contribution of relative population size.

Reali, Chater, and Christiansen (2018) are not only interested in grammatical simplification but also in vocabulary growth. They discuss how both phenomena might be associated with larger population sizes, and subsequently model this effect with agents who share ‘conventions’. The crucial mechanism in their model is that certain conventions, once invented by an agent spontaneously, are easily spread throughout the population due to being more ‘learnable’, that is, being picked up by other agents in a single encounter. Other conventions are harder to learn, that is, require more encounters for being transmitted. Reali, Chater, and Christiansen (2018) argue that solely the difference in ‘learnability’ could be enough to explain complexity differences in small and large populations: as populations grow, the number of innovations grows, but only the easily learnable ones spread. Thus, smaller populations might maintain higher proportions of difficult-to-learn conventions, while in larger populations, there is a bias against them.

The models discussed above are relatively unconstrained in their setting of parameters. Some models are emerging which are more explicitly checked against data collected for actual languages and contact situations. For instance, Jon-And and Aguilar (2019) model the loss of verbal inflectional markers in Mozambican Portuguese driven by repeated introductions of reduced variants by L2 speakers. They use census data of Mozambique spanning several decades as a backbone to their model. In a similar vein, Dekker, Klamer, and De Boer (2022) model the impact of L2 learning on verbal inflections in Alorese, an Austronesian language of the Alor and Pantar islands in Indonesia, which has been in contact with Papuan languages, while some of its sister languages have not.

Agent-based computational modeling is an under-used but effective tool to model the interaction between sociolinguistic variables and language structure from micro to macro levels. An avenue for future research in this direction is models which allow researchers to assess the interdependence and relevance of multiple variables, and models which are explicitly linked to cross-linguistic and historical data.

## 4 Outlook

In the introduction, we laid out the complex interplay of micro and macro levels in the development of language diversity (remember in particular Figure 1). To fully understand language change, it is necessary to trace micro variation from individual speakers to macro variation at the level of families and areas. This is a lot to ask. However, within the research program of sociolinguistic typology, some advances have been made in this direction.

A consensus seems to emerge between cross-linguistic statistical studies, language learning experiments, and computational models. We here formulate this tentatively as: *the structure of a speaker population is indeed systematically linked with the morphological complexity of the respective language*. It is worth pausing for a moment to realize how radical a statement this is. A few decades ago it was the general consensus of linguists that linguistic structure is genuinely independent of the socioecological dimensions which languages are used in. Researchers working within the framework of sociolinguistic typology have accrued several lines of evidence to the contrary.

However, this is also where the consensus ends. The statement above is relatively vague, and there are still contradictions between studies. Both 'population structure' and 'morphological complexity' are terms which are defined and interpreted in many different ways. Depending on the interpretation, and the quantitative models used, the outcomes change. This has also become apparent in the studies discussed in this entry. In the following, we want to outline some of the major issues which hamper a broader consensus.

- *Linguistic measures: granularity and type*. As discussed in Section 3, there are many different types of data which can be used to assess claims of sociolinguistic typology. Remember the example of nominal case-marking in Faroese and German in Figure 2. There is clearly a difference in the distributions of case-marked word forms – the type–token ratio is around three times higher for Faroese. This difference is already concealed in the paradigmatic view, since both languages feature eight cells along the dimensions of NUMBER and CASE. If we factor in syncretism of forms by counting unique word types, then we get to seven for Faroese versus four for German. The difference between the languages has decreased in line with the granularity of the measure. The type count is now less than two times higher for Faroese. Typological databases like WALS further collapse such counts into discrete categories. Note that in this case, Faroese and German now fall into the same category for 'number of cases' (4 cases), and are hence indistinguishable in terms of 'case complexity'. Thus, the confirmation or rejection of the link between population structure and nominal morphology – as proposed by Braunmüller (1984) for Germanic varieties – crucially hinges upon the level of aggregation. It is important to keep this issue of granularity in mind in statistical studies and experiments.

  The type of measurement matters beyond granularity. Where Sinnemäki and Di Garbo (2018) find no correlation between the number of grammatical genders and population size, Verkerk and Di Garbo (2022) find evidence for a link between population size and gender systems when focusing on the agreement system, rather than 'simply' the number of genders.
- *Measures of population structure*. A similar rationale applies on the socioecological side. There are many different measures which reflect 'population structure' in some way. Trudgill (2011) mentions five: population size, network density, social stability, shared information, and contact versus isolation. He also teases apart different contact scenarios involving adult second-language learners, on

one hand, and child bilingualism, on the other. Subsequent studies have added further measures, for instance, the number of neighbors and geographic spread (Lupyan and Dale 2010), altitude as a proxy of isolation (Nichols 2013; Nichols and Bentz 2018), or ‘vehicularity’ (Koplenig 2019). If and how all these measures correlate, and whether they constitute genuinely different dimensions of ‘population structure’ is a research question in itself. There is currently work underway to ‘typologize’ different aspects of population structure, especially for areas of small-scale multilingualism such as the Caucasus (Dobrushina and Moroz 2021), Sub-Saharan Africa (Lüpke 2016), Northern Australia (Singer and Harris 2016), and across the world (Pakendorf, Dobrushina, and Khanina 2021).

A better understanding of small-scale multilingualism is directly relevant to sociolinguistic typology. Firstly, Trudgill (2011) himself discusses “additive complexification” of morphological paradigms due to long-term bilingualism. Along similar lines, Evans (2019) gives concrete examples of “summative complexification” of noun class systems in the bilingual setting of Central Arnhem Land. In fact, Evans (2018, 22) argues that differences in morphological complexity “may result from ‘esterogeny effects’ ..., in which hard to learn forms (e.g. irregular or suppletive paradigm cells) are selected for, so as to stymy the language-learning efforts of outsiders.” In this case, morphology might complexify – rather than simplify – under certain kinds of close contact between language varieties.

- *Measuring morphological complexity*. Likewise, measures of morphological complexity, let alone ‘language complexity’ in general are plenty. Many articles and whole volumes have been written on the topic (e.g., Newmeyer and Preston 2014; Baerman, Brown, and Corbett 2015; Berdicevskis et al. 2018; Ehret et al. 2021). It seems likely that as the number of measures for a given concept increases, the disagreement between the measures will increase, too. For the domain of morphology, Çöltekin and Rama (2022) have shown that different complexity measures are significantly correlated across the board, and hence ‘agree’ reasonably well. However, when complexity measures from the domains of morphology and syntax are mixed, languages become statistically indistinguishable (Bentz et al. 2022). This suggests that very general statements about differences in overall ‘grammatical complexity’ are unlikely to hold – simply because of the different ideas about complexity, and the resulting variance in measurements. Even within the domain of morphology, there are a multitude of different dimensions in which complexity could be measured (e.g., verbal inflection, gender marking, case marking, etc.). In fact, Sinnemäki and Di Garbo (2018) argue that the statistical associations with population size and adult learner percentages differ between verbal synthesis and gender marking. Sinnemäki (2020), on the other hand, replicates the findings in Bentz and Winter (2013) on the statistical link between the loss of case marking and L2 percentages. These results suggest that also within the domain of morphology there are systematic differences, and that general statements about ‘all of morphology’ are possibly not warranted.
- *Statistical models: usage and interpretation*. Once we have agreed on a set of variables to investigate, another fundamental issue awaits us: the choice of a statistical model. As outlined in Section 3.1, the trend over the last two decades has been to increase the complexity of models by adding more and more fine-grained methods to include phylogenetic and geographic information. This seems to follow the mantra: *the finer grained the better*. However, this is to some extend misguided. From

a theoretical point of view, there is always a trade-off between Type I and Type II errors, that is, between being anti-conservative and over-conservative when aiming to detect significant effects. For this reason, there is an ongoing discussion of how to strike the right balance in mixed-effects models used in psycholinguistics (cf. Barr et al. 2013; Bates et al. 2015; Matuschek et al. 2017). Namely, including all possible controls will safeguard against Type I errors, that is, detecting spurious effects, but it will also increase the Type II error rate, that is, reduce the power to detect an actual effect given a certain amount of data. The same discussion needs to happen for cross-linguistic statistical models in typology.

More generally, applying models from psycholinguistics, evolutionary biology, and geoscience to typological data requires us to not only think about the practical issues of data processing and model fitting, but, even more importantly, the issue of *interpretation*. For example, in psycholinguistics, if an effect ceases to be significant after controlling for differences between subjects and items, then it is typically seen as 'not interesting', and hence not in need of further explanation. Is this true also for typological data when controlling for language families and areas? What if a link between population size and the usage of morphological marking exists in one language family but not in another? A mixed-effects model will tell us that after adjusting family intercepts and slopes the effect is gone. But, arguably, these family differences are not just a nuisance which we want to get rid of, but in fact, part of what we came to investigate.

- *The connection between statistical models and experiments*. Once we have established a statistical association between some socioecological variable and a linguistic variable, we would like to see this confirmed in experiments. This will give us the link between micro processes of learning and usage, and macro patterns of feature distributions across populations. However, the devil lies in the detail here, and there is always a temptation to gloss over details for the sake of a more general interpretation.

For example, the experiments reported in Cuskley et al. (2015) and De Smet, Rosseel, and Van de Velde (2022) might seem to contradict the statistical results and argumentation in Lupyan and Dale (2010). Namely, the experiments show that native speakers regularize past tense verb morphology in English and Dutch more than non-native speakers. Which in turn might suggest that the link between population size and morphological simplification is not imperfect adult language learning. However, at the face of it, Lupyan and Dale (2010) use 28 WALS features relevant to 'morphology' in a broad sense, but none of them directly speak to the difference between *regular and irregular marking of past tense*. The features closest to this topic are probably 'the past tense' (Dahl and Velupillai 2013), and 'inflectional synthesis of the verb' (Bickel and Nichols 2013). Here, languages are put into discrete bins according to the number of remoteness distinctions, and the number of inflectional categories potentially expressed on a verb respectively. For example, English verbs inflect for tense, and for number agreement. English is thus put into the bin of 'present, no remoteness distinctions', and '2–3 categories per word' alongside languages such as German, Finnish, Lezgian, Kannada, Tagalog, and Martuthunira. Needless to say that the exact realization of inflectional markers on verbs in these languages can look very different. The bottom line is that the question of regularity of past tense markers is simply irrelevant to these WALS chapters, since both irregular and regular markers count as inflections of past tense. Hence, the connection between the statistical and experimental results in this case is extremely vague. They simply operate at different levels of granularity when it comes to measuring 'morphological complexity'. Of course, all statistical and experimental studies we have discussed here are valuable by themselves, as they shed light on different aspects of the problem. However, when relating them to one another, we need to look at the details of *what* is modeled *how*.

- *Language learning literature*. Apart from experiments specifically designed to test sociolinguistic hypotheses, there is a wealth of studies in the first and second language learning literature which directly or indirectly speak to the question of complexity. As a starting point, Kuiken (2022) gives an overview of complexity measures developed in studies on second language acquisition (SLA) since the 1990s. Some first thoughts about insights from language learning to typology are found in Schepens, Slik, and Hout (2013) and Kempe and Brooks (2018). But what is currently lacking is a systematic overview of the language learning literature with sociolinguistic typology in mind. In particular, given that experiments in the laboratory necessarily feature extremely simplified stimuli, and that there are many different morphological dimensions which could be linked to socioecological factors, it would be informative to derive hypotheses from the language learning literature on 'real languages'. Morphological patterns which emerge as being hard to learn for children and/or adults in these studies are predicted to be dispreferred in language varieties shaped by the respective learning scenarios.

## 5 Conclusion

Since the inception of sociolinguistic typology in the 1980s and 1990s, qualitative and quantitative evidence has been collected to probe the fundamental link between population structure and language structure. We have here discussed the data and methods of the quantitative accounts which have emerged in the past two decades. We have also outlined some issues and caveats to be considered in the next decades. While we have here focused on the specific topic of morphological simplification and complexification, the conceptual and methodological challenges tackled in this domain are also relevant to research into language diversity more generally.

**SEE ALSO:** Abruptness in Language Change; Acquisition and Learnability; Autonomy of Morphology: Paradigmatic Structure in Diachrony; Contact and Borrowing; Convergence and Linguistic Areas; (De)Grammaticalization; Diachrony of Morphological Strategies; Endogeny and Exogeny; Inflectional Change and Morphological Theory; Markedness, Naturalness, and Complexity; Non-Morphological Sources and Triggers of Morphological Change; Rise and Fall of Case: When Syntax Meets Morphology; Syntactic Methods for Language Phylogenies.

## Acknowledgments

We would like to thank Kaius Sinnemäki, Matías Guzmán Naranjo, Aleksandrs Berdicevskis, Nicholas Evans, and an anonymous reviewer for their comments and criticisms. Remaining errors and shortcomings are our own.

## Notes

1. Here we used the example of complexity measurements, but the statistical analyses discussed are relevant for any measurable dimension in languages. Also, we here do not delve into the issue of measuring language complexity. There are many articles and whole volumes dealing with this issue, including Miestamo, Sinnemäki and Karlsson (2008), Sampson, Gil, and Trudgill (2009), Pallotti (2015), and Hale (2016).
2. The term *linear* refers to the predicted values falling on a straight line for a single predictor variable (Baayen 2013, 339). The term *regression* model is then used when all the predictors are numerical, while exclusively categorical predictors constitute an analysis of variance (ANOVA), and a mix of numerical and categorical predictors is referred to as analysis of covariance (ANCOVA).
3. In case of categorical predictor variables, languages would take numerical values according to a so-called dummy coding, e.g., 0 and 1 if we have binary factor levels. See Baayen (2013, 339–341) for further details.
4. *Generalized* here means that the dependent variables are not necessarily numerical, but can also be binary (e.g., presence or absence of a morphological marker), or categorical with multiple factor levels. These types of models are made possible by link functions, which transform response variables with non-linear distributions in such a way that they can vary linearly with the predictors (see McCullagh and Nelder 1989).
5. In this case, there is a categorical dependent variable, and one categorical predictor variable. Chi-squared tests and Fisher's exact tests are used in this setup (see Sinnemäki 2009, 136). Additionally, Sinnemäki (2011, 143–146) provides a logistic regression for the same data.
6. While in psycholinguistics and typology, random effects are typically seen as 'controls' in the face of various types of non-independence, there are different ideas about fixed and random effects. Gelman (2005, 21) proposes a more general terminology defining "effects (or coefficients) in a multilevel model as constant [i.e., fixed] if they are identical for all groups in a population and varying [i.e., random] if they are allowed to differ from group to group." Thanks to Matías Guzmán Naranjo for pointing this out to us.
7. These adjustments are called Best Linear Unbiased Predictors (BLUPs). They are strictly speaking not part of the parameters to be estimated. This is the reason they are written in Latin rather than Greek script here (cf. Baayen, Davidson, and Bates 2008, 394).
8. Expected under a Brownian motion model of evolution, i.e., a random walk with a constant rate. This implies that, for a continuous character, e.g., morphological complexity, change is not directed but can happen in either direction (i.e., languages may become more or less morphologically complex) with the expected mean over time being zero. To alleviate this restriction, an off-diagonal multiplier, e.g., Pagel's $\lambda$ can be introduced. See Revell (2010, 320) for further details. Whether these assumptions (non-directed change, constant rate) hold within and across language families is an empirical question.
9. In a Wug test, a participant is required to respond to a non-word with an inflected form according to the morphological rules of a given language, i.e., extrapolate rules to unseen items. In the original setup, English children were asked to give the plural form of a 'wug', hence the name.

## References and Suggested Readings

Atkinson, Mark, Simon Kirby, and Kenny Smith. 2015. "Speaker Input Variability Does Not Explain Why Larger Populations Have Simpler Languages." *PLOS ONE* 10 (6): e0129463.

Atkinson, Mark, Kenny Smith, and Simon Kirby. 2018. "Adult Learning and Language Simplification." *Cognitive Science* 42 (8): 2818–2854.

Baayen, R. Harald. 2013. "Multivariate Statistics." In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 337–372. Cambridge: Cambridge University Press.

Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. 2008. “Mixed Effects Modeling with Crossed Random Effects for Subjects and Items.” *Journal of Memory and Language* 59 (4): 390–412.

Baechler, Raffaela, and Guido Seiler, eds. 2016. *Complexity, Isolation, and Variation*. Berlin and Boston: Walter de Gruyter.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2015. *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. “Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal.” *Journal of Memory and Language* 68 (3): 255–278.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. “Parsimonious Mixed Models.” *arXiv* preprint, arXiv:1506.04967.

Bentz, Christian. 2018. *Adaptive Languages: An Information-Theoretic Account of Linguistic Diversity*. Berlin: De Gruyter Mouton.

Bentz, Christian, Dimitrios Alikaniotis, Tanja Samardžić, and Paula Buttery. 2017. “Variation in Word Frequency Distributions: Definitions, Measures and Implications for a Corpus-Based Language Typology.” *Journal of Quantitative Linguistics* 24 (2–3): 128–162.

Bentz, Christian, Ximena Gutierrez-Vasques, Olga Sozinova, and Tanja Samardžić. 2022. “Complexity Trade-Offs and Equi-Complexity in Natural Languages: A Meta-Analysis.” *Linguistics Vanguard.* DOI: 10.1515/lingvan-2021-0054.

Bentz, Christian, Annemarie Verkerk, Douwe Kiela, Felix Hill, and Paula Buttery. 2015. “Adaptive Communication: Languages with More Non-Native Speakers Tend to Have Fewer Word Forms.” *PLOS ONE* 10 (6): e0128254.

Bentz, Christian, and Bodo Winter. 2013. “Languages with More Second Language Learners Tend to Lose Nominal Case.” *Language Dynamics and Change* 3 (1): 1–27.

Berdicevskis, Aleksandrs, Çağri Çöltekin, Katharina Ehret, Kilu von Prince, Daniel Ross et al. 2018. “Using Universal Dependencies in Cross-Linguistic Complexity Research.” In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, 8–17.

Berdicevskis, Aleksandrs, and Arturs Semenuks. 2020. “Different Trajectories of Morphological Overspecification and Irregularity under Imperfect Language Learning,” In *The Complexities of Morphology*, edited by Peter Arkadiev and Francesco Gardani, 283–305. Oxford: Oxford University Press.

Berdicevskis, Aleksandrs, and Arturs Semenuks. 2022. “Imperfect Language Learning Reduces Morphological Overspecification: Experimental Evidence.” *PLOS ONE* 17 (1): e0262876.

Beuls, Katrien, and Luc Steels. 2013. “Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement.” *PLOS ONE* 8 (3): e58960.

Bickel, Balthasar, and Johanna Nichols. 2013. “Inflectional Synthesis of the Verb (v2020.3).” In *The World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. DOI: 10.5281/zenodo.7385533.

Bickel, Balthasar, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt et al. 2022. *The AUTOTYP database (v1.1.0)*. DOI: 10.5281/zenodo.6793367.

Blythe, Richard A, and William Croft. 2012. “S-Curves and the Mechanisms of Propagation in Language Change.” *Language* 88 (2): 269–304.

Blythe, Richard A., and William Croft. 2021. “How Individuals Change Language.” *PLOS ONE* 16 (6): e0252582.

Bowern, Claire. 2018. “Computational Phylogenetics.” *Annual Review of Linguistics* 4 (1): 281–296.

Braunmüller, Kurt. 1984. “Morphologische Undurchsichtigkeit: ein Charakteristikum kleiner Sprachen.” *Kopenhagener Beiträge zur Germanistischen Linguistik* (Kopenhagen (Dänemark)) 22: 48–68.

Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson, and Erich Round. 2018. “Areal Pressure in Grammatical Evolution: An Indo-European Case Study.” *Diachronica* 35 (1): 1–34.

Çöltekin, Çağri, and Taraka Rama. 2022. "What Do Complexity Measures Measure? Correlating and Validating Corpus-Based Measures of Morphological Complexity." *Linguistics Vanguard* 9: 27–43.

Coupé, Christophe. 2018. "Modeling Linguistic Variables with Regression Models: Addressing Non-Gaussian Distributions, Non-Independent Observations, and Non-Linear Predictors with Random Effects and Generalized Additive Models for Location, Scale, and Shape." *Frontiers in Psychology* 9: 513.

Cuskley, Christine, Francesca Colaiori, Claudio Castellano, Vittorio Loreto, Martina Pugliese, and Francesca Tria. 2015. "The Adoption of Linguistic Rules in Native and Non-Native Speakers: Evidence from a Wug Task." *Journal of Memory and Language* 84: 205–223.

Cuskley, Christine, Vittorio Loreto, and Simon Kirby. 2018. "A Social Approach to Rule Dynamics Using an Agent-Based Model." *Topics in Cognitive Science* 10 (4): 745–758.

Cysouw, Michael. 2010. "Dealing with Diversity: Towards an Explanation of NP-Internal Word Order Frequencies." *Linguistic Typology* 14: 253–286.

Dahl, Östen, and Viveka Velupillai. 2013. "The Past Tense (v2020.3)." In *The World Atlas of Language Structures Online,* edited by Matthew S. Dryer and Martin Haspelmath. DOI: 10.5281/zenodo.7385533.

Dale, Rick, and Gary Lupyan. 2012. "Understanding the Origins of Morphological Diversity: The Linguistic Niche Hypothesis." *Advances in Complex Systems* 15 (03n04): 1150017.

De Smet, Isabeau, Laura Rosseel, and Freek Van de Velde. 2022. "Are Non-Native Speakers the Drivers of Morphological Simplification? A Wug Experiment on the Dutch Past Tense System." *Journal of Language Evolution* (August). DOI: 10.1093/jole/lzac008.

Dekker, Peter, Marian Klamer, and Bart De Boer. 2022. "Language-Specific and Universal Factors behind Morphological Simplification: An Agent-Based Modelling Study of Alorese." In *Proceedings of the Joint Conference on Language Evolution.* DOI: 10.17617/2.3398549.

DeLancey, Scott. 2014. "Sociolinguistic Typology in North East India: A Tale of Two Branches." *Journal of South Asian Languages and Linguistics* 1: 59–82.

DeLancey, Scott. 2015. "The Historical Dynamics of Morphological Complexity in Trans-Himalayan." *Linguistic Discovery* 13 (2): 60–79.

Dobrushina, Nina, and George Moroz. 2021. "The Speakers of Minority Languages Are More Multilingual." *International Journal of Bilingualism* 25 (4): 921–938. DOI: 10.1177/13670069211023150.

Dryer, Matthew S. 1989. "Large Linguistic Areas and Language Sampling." *Studies in Language: International Journal Sponsored by the Foundation "Foundations of Language"* 13 (2): 257–292.

Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS Online (v2020.3)*. DOI: 10.5281/zenodo.7385533.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig. 2023. *Ethnologue: Languages of the World*, 26th ed. https://www.ethnologue.com.

Ehret, Katharina, Alice Blumenthal-Dramé, Christian Bentz, and Aleksandrs Berdicevskis. 2021. "Meaning and Measures: Interpreting and Evaluating Complexity Metrics." *Frontiers in Communication* 6: 640510.

Evans, Nicholas. 2018. "The Dynamics of Language Diversity." In *The Dynamics of Language: Plenary and Focus Lectures from the 20th International Congress of Linguists,* edited by Rajend Mesthrie and David Bradley, 12–41. Cape Town, South Africa: UCT Press.

Evans, Nicholas. 2019. "Linguistic Divergence under Contact." In *Historical Linguistics 2015: Selected Papers from the 22nd International Conference on Historical Linguistics, Naples, 2731 July 2015,* edited by Michela Cennamo and Claudia Fabrizio, 564–591. Amsterdam: John Benjamins.

Gelman, Andrew. 2005. "Analysis of Variance: Why It Is More Important than Ever." *The Annals of Statistics* 33 (1): 1–53.

Gil, David. 2021. "Tense-Aspect-Mood Marking, Language-Family Size and the Evolution of Predication." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1824): 20200194. DOI: 10.1098/rstb.2020.0194.

Goldhahn, Dirk, Thomas Eckart, and Uwe Quasthoff. 2012. "Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, 2012*, 31–43.

Grafen, Alan. 1989. "The Phylogenetic Regression." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326 (1233): 119–157.

Guzmán Naranjo, Matias, and Laura Becker. 2022. "Statistical Bias Control in Typology." *Linguistic Typology* 26 (3): 605–670.

Guzmán Naranjo, Matias, and Miri Mertner. 2022. "Estimating Areal Effects in Typology: A Case Study of African Phoneme Inventories." *Linguistic Typology* 27 (2): 455–480.

Hale, J. 2016. "Information-Theoretical Complexity Metrics." *Language and Linguistics Compass* 10 (9): 397–412.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2022. *Glottolog* 4.7. https://glottolog.org.

Jaeger, T. Florian, Peter Graff, William Croft, and Daniel Pontillo. 2011. "Mixed Effect Models for Genetic and Areal Dependencies in Linguistic Typology." *Linguistic Typology* 15: 281–320.

Jäger, Gerhard, and Johannes Wahle. 2021. "Phylogenetic Typology." *Frontiers in Psychology* 12. DOI: 10.3389/fpsyg.2021.682132.

Jon-And, Anna, and Elliot Aguilar. 2019. "A Model of Contact-Induced Language Change: Testing the Role of Second Language Speakers in the Evolution of Mozambican Portuguese." *PLOS ONE* 14 (4): e0212303. DOI: 10.1371/journal.pone.0212303.

Kashima, Eri, Francesca Di Garbo, Oona Raatikainen, Rosnátaly Avelino, Sasha Beck et al. 2023. "GramAdapt Social Contact Dataset." https://hal.science/hal-03929516.

Kauhanen, Henri. 2017. "Neutral Change." *Journal of Linguistics* 53 (2): 327–358.

Kauhanen, Henri, Sarah Einhaus, and George Walkden. 2023. "Language Structure Is Influenced by the Proportion of Non-Native Speakers: A Reply to Koplenig (2019)." *Journal of Language Evolution* 8 (1): 90–101. DOI: 10.1093/jole/lzad005.

Kempe, Vera, and Patricia J Brooks. 2018. "Linking Adult Second Language Learning and Diachronic Change: A Cautionary Note." *Frontiers in Psychology* 9: 480.

Kirby, Simon, Hannah Cornish, and Kenny Smith. 2008. "Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language." *Proceedings of the National Academy of Sciences* 105 (31): 10681–10686.

Koehn, Philipp. 2005. "Europarl: A Parallel Corpus for Statistical Machine Translation." In *Proceedings of Machine Translation Summit x: Papers*, 79–86.

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 (2): 181274.

Koplenig, Alexander, Sascha Wolfer, and Peter Meyer. In preparation. "Human Languages Trade off Complexity against Efficiency."

Kuiken, Folkert. 2022. "Linguistic Complexity in Second Language Acquisition." *Linguistics Vanguard* 9 (s1): 83–93.

Kusters, Wouter. 2003. "Linguistic Complexity." Diss., Netherlands Graduate School of Linguistics.

Levshina, Natalia. 2021. "Cross-Linguistic Trade-Offs and Causal Relationships between Cues to Grammatical Subject and Object, and the Problem of Efficiency-Related Explanations." *Frontiers in Psychology* 12: 648200.

Levshina, Natalia. 2022. "Corpus-Based Typology: Applications, Challenges and Some Solutions." *Linguistic Typology* 26 (1): 129–160.

Lüpke, Friederike. 2016. "Uncovering Small-Scale Multilingualism." *Critical Multilingualism Studies* 4 (2): 35–74.

Lupyan, Gary, and Rick Dale. 2010. "Language Structure Is Partly Determined by Social Structure." *PLOS ONE* 5 (1): e8559.

Lupyan, Gary, and Rick Dale. 2016. "Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity." *Trends in Cognitive Sciences* 20 (9): 649–660.

Maitz, Péter, and Attila Németh. 2014. "Language Contact and Morphosyntactic Complexity: Evidence from German." *Journal of Germanic Linguistics* 26: 1–29.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. "Balancing Type I Error and Power in Linear Mixed Models." *Journal of Memory and Language* 94: 305–315.

Mayer, Thomas, and Michael Cysouw. 2014. "Creating a Massively Parallel Bible Corpus." *Oceania* 135 (273): 40.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*. London: Chapman Hall.

McWhorter, John. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford: Oxford University Press.

Meakins, Felicity, Xia Hua, Cassandra Algy, and Lindell Bromham. 2019. "Birth of a Contact Language Did Not Favor Simplification." *Language* 95 (2): 294–332.

Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson, eds. 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins.

Moran, Steven, and Daniel McCloy, eds. 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. https://phoible.org.

Moran, Steven, Daniel McCloy, and Richard Wright. 2012. "Revisiting Population Size vs. Phoneme Inventory Size." *Language* 88 (4): 877–893.

Morozova, Maria, Anastasia Escher, and Alexander Rusakov. 2022. "Linguistic Complexity of South Slavic Dialects: A New Perspective on Old Data." *Linguistics Vanguard* 9 (s1): 133–154.

Moscoso del Prado Martin, Fermin, and Christian Brendel. 2016. "Case and Cause in Icelandic: Reconstructing Causal Networks of Cascaded Language Changes." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1: *Long Papers*, 2421–2430.

Myung, In Jae. 2003. "Tutorial on Maximum Likelihood Estimation." *Journal of Mathematical Psychology* 47 (1): 90–100.

Nettle, Daniel. 2012. "Social Scale and Structural Complexity in Human Languages." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1597): 1829–1836.

Newmeyer, Frederick J., and Laurel B. Preston. 2014. *Measuring Grammatical Complexity*. Oxford: Oxford University Press.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Nichols, Johanna. 2013. "The Vertical Archipelago: Adding the Third Dimension to Linguistic Geography." In *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives,* edited by Peter Auer, Martin Hilpert, Anja Stukenbrock, and Benedikt Szmrecsanyi, 38–60. Berlin and Boston: De Gruyter.

Nichols, Johanna. 2016. "Complex Edges, Transparent Frontiers: Grammatical Complexity and Language Spreads." In *Complexity, Isolation, and Variation*, edited by Raffaela Baechler and Guido Seiler, 117–137. Berlin and Boston: De Gruyter.

Nichols, Johanna, and Christian Bentz. 2018. "Morphological Complexity of Languages Reflects the Settlement History of the Americas." In *Perspectives on the Peopling of the Americas*, edited by Katerina Harvati, Gerhard Jäger, and Hugo Reyes-Centeno, 13–26. Tübingen: Kerns Verlag.

Pagel, Mark. 1997. "Inferring Evolutionary Processes from Phylogenies." *Zoologica Scripta* 26 (4): 331–348.

Pakendorf, Brigitte, Nina Dobrushina, and Olesya Khanina. 2021. "A Typology of Small-Scale Multilingualism." *International Journal of Bilingualism* 25 (4): 835–859.

Pallotti, G. 2015. "A Simple View of Linguistic Complexity." *Second Language Research* 31 (1): 117–134. DOI: 10.1177/0267658314536435.

Pijpops, Dirk, Katrien Beuls, and Freek Van de Velde. 2015. "The Rise of the Verbal Weak Inflection in Germanic: An Agent-Based Model." *Computational Linguistics in the Netherlands Journal* 5: 81–102.

Ranacher, Peter, Nico Neureiter, Rik van Gijn, Barbara Sonnenhauser, Anastasia Escher et al. 2021. “Contact-Tracing in Cultural Evolution: A Bayesian Mixture Model to Detect Geographic Areas of Language Contact.” *Journal of the Royal Society Interface* 18 (181): 20201031.

Raviv, Limor, Marianne de Heer Kloots, and Antje Meyer. 2021. “What Makes a Language Easy to Learn? A Preregistered Study on How Systematic Structure and Community Size Affect Language Learnability.” *Cognition* 210: 104620.

Raviv, Limor, Antje Meyer, and Shiri Lev-Ari. 2019. “Larger Communities Create More Systematic Languages.” *Proceedings of the Royal Society B* 286 (1907): 20191262.

Reali, Florencia, Nick Chater, and Morten H. Christiansen. 2018. “Simpler Grammar, Larger Vocabulary: How Population Size Affects Language.” *Proceedings of the Royal Society B: Biological Sciences* 285 (1871): 20172586.

Revell, Liam J. 2010. “Phylogenetic Signal and Linear Regression on Species Data.” *Methods in Ecology and Evolution* 1 (4): 319–329.

Roberts, Seán G. 2018. “Robust, Causal, and Incremental Approaches to Investigating Linguistic Adaptation.” *Frontiers in Psychology* 9: 166.

Roberts, Seán G., Anton Killin, Angarika Deb, Catherine Sheard, Simon J. Greenhill et al. 2020. “CHIELD: The Causal Hypotheses in Evolutionary Linguistics Database.” *Journal of Language Evolution* 5 (2): 101–120.

Sampson, Geoffrey, David Gil, and Peter Trudgill, eds. 2009. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Schepens, Job, Frans van der Slik, and Roeland van Hout. 2013. “Learning Complex Features: A Morphological Account of L2 Learnability.” *Language Dynamics and Change* 3 (2): 218–244.

Shcherbakova, Olena, Volker Gast, Damián E. Blasi, Hedvig Skirgård, Russell D. Gray, and Simon J. Greenhill. 2022. *Linguistics Vanguard*. DOI: 10.1515/lingvan-2021-0011.

Shcherbakova, Olena, Susanne Maria Michaelis, Hannah J. Haynie, Sam Passmore, Volker Gast et al. 2023. “Societies of Strangers Do Not Speak Grammatically Simpler Languages.” *Science Advances* 9: eadf7704.

Singer, Ruth, and Salome Harris. 2016. “What Practices and Ideologies Support Small-Scale Multilingualism? A Case Study of Warruwi Community, Northern Australia.” *International Journal of the Sociology of Language* 2016 (241): 163–208.

Sinnemäki, Kaius. 2009. “Complexity in Core Argument Marking and Population Size.” In *Language Complexity as an Evolving Variable*, edited by Geoffrey Sampson, David Gil, and Peter Trudgill, 126–140. Oxford: Oxford University Press.

Sinnemäki, Kaius. 2011. “Language Universals and Linguistic Complexity: Three Case Studies in Core Argument Marking.” PhD diss., University of Helsinki.

Sinnemäki, Kaius. 2014. “Cognitive Processing, Language Typology, and Variation.” *Wiley Interdisciplinary Reviews: Cognitive Science* 5 (4): 477–487.

Sinnemäki, Kaius. 2020. “Linguistic System and Sociolinguistic Environment as Competing Factors in Linguistic Variation: A Typological Approach.” *Journal of Historical Sociolinguistics* 6 (2). DOI: 10.1515/jhsl-2019-1010.

Sinnemäki, Kaius, and 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.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins et al. 2023. “Grambank Reveals the Importance of Genealogical Constraints on Linguistic Diversity and Highlights the Impact of Language Loss.” *Science Advances* 9: eadg6175.

Symonds, Matthew R.E., and Simon P. Blomberg. 2014. “A Primer on Phylogenetic Generalised Least Squares.” In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology,* edited by Laszlo Z. Garamszegi, 105–130. Berlin and Heidelberg: Springer.

Szmrecsanyi, Benedikt, and Bernd Kortmann. 2009. “The Morphosyntax of Varieties of English Worldwide: A Quantitative Perspective.” *Lingua* 119 (1643–1663). DOI: 10.1515/jhsl-2019-1010.

Trudgill, Peter. 1984. *On Dialect: Social and Geographical Perspectives*. New York: New York University Press.

Trudgill, Peter. 1992. *Dialect Typology and Social Structure*. Berlin and New York: Mouton de Gruyter.

Trudgill, Peter. 1995. "Dialect Typology: Isolation, Social Network and Phonological Structure." In *Towards a Social Science of Language*, edited by R. Gregory Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh, 3–22. Amsterdam: John Benjamins.

Trudgill, Peter. 2010. "Contact and Sociolinguistic Typology." In *The Handbook of Language Contact*, edited by Raymond Hickey, 299–319. Oxford: Wiley Blackwell.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Verkerk, Annemarie, and Francesca Di Garbo. 2022. "Sociogeographic Correlates of Typological Variation in Northwestern Bantu Gender Systems." *Language Dynamics and Change* 12: 155–223.

Wälchli, Bernhard. 2012. "Indirect Measurement in Morphological Typology." In *Methods in Contemporary Linguistics*, edited by Andrea Ender, Adrian Leemann, and Bernhard Wälchli, 69–92. Berlin: De Gruyter Mouton.

Walkden, George, and Anne Breitbarth. 2019. "Complexity as L2-Difficulty: Implications for Syntactic Change." *Theoretical Linguistics* 45 (3–4): 183–209.

Widmer, Manuel, Mathias Jenny, Wolfgang Behr, and Balthasar Bickel. 2021. "Morphological Structure Can Escape Reduction Effects from Mass Admixture of Second Language Speakers." *Studies in Language* 45 (4): 707–752.

Zeman, Daniel, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli et al. 2022. *Universal Dependencies* 2.11. http://hdl.handle.net/11234/1-4923.