CROSS-LINGUISTIC COMPARISON OF COMPLEXITY MEASURES IN PHONOLOGICAL SYSTEMS*

Steven Moran[1,2] and Damián Blasi[3,4]

[1]Department of Linguistics, University of Zurich

[2]Research Unit: Quantitative Language Comparison, University of Marburg

[3]Max Planck Institute for Mathematics in the Sciences

[4]Max Planck Institute for Evolutionary Anthropology

1 Introduction

An assumed truism in linguistics is that if the structure of a language simplifies in one place, it is likely to complicate in another (e.g. Martinet 1955, Hockett 1955, Aitchison 2000).[1] Although the complexity of subsystems may vary within a given language, another assumption is that these differences balance out cross-linguistically, so that all languages tend to be equally complex (e.g. Hockett 1958, Akmajian et al. 1979, Crystal 1987, McMahon 1994, Dixon 1997).[2] This notion is furthered by the long-held view that linguistic structures are not affected by geographic or societal factors, with vocabulary being an exception, e.g. Sapir 1912.

Recently, these assumptions about complexity have been challenged (e.g. McWhorter 2001, Kusters 2003, Dahl 2004, 2008, Hawkins 2004, Shosted 2006, Miestamo et al. 2008, Givón & Shibatani 2009, Sampson et al. 2009, Sinnemäki 2011), and have included findings of correlations between complexity and geographic or sociocultural settings (e.g. Perkins 1992, McWhorter 2001, Kusters 2003, Trudgill 2004, Hay & Bauer 2007, McWhorter 2007, Nichols 2009, Lupyan & Dale 2010, Sinnemäki 2011). These studies are controversial, however, because it is difficult for linguists to come to a consensus on an objective measure for complexity (e.g. Ansaldo & Nordhoff 2009; see also Sinnemäki 2011) and because of the socially sensitive issue of claiming that one language is more complex than another. Nonetheless, at least two common trends for measuring complexity have emerged: measuring relative complexity (the cost or difficulty of using a system) and absolute complexity (the number of parts of a linguistic subsystem).

Assessing the complexity of languages is a non-trivial task and to date there has been no clear methodology to do so at a cross-linguistic aggregate level. Researchers have concentrated on specific aspects of languages, mainly divided along classic distinctions such as syntax, morphology and phonology. In this chapter, we investigate the complexity of phonological systems using a large sample of languages. Under the rubric of absolute measures, researchers have formulated many claims about phonological complexity in terms of univariate or bivariate analyses, i.e. in terms of the shape of statistical distributions or the directions of correlations between two variables (e.g. Hockett 1955, Maddieson 1980, 1984, 2006, 2007, and Justeson & Stephens 1984). The former allows languages to be ranked in regard to a given complexity metric and the latter provides a method for testing compensatory relations between linguistic phenomena. What we find is that claims about distributions and correlations in the literature about phonological systems complexity, seem to be, at best case, dubious.

In our investigation, we first calculate univariate measures that have long been used by linguists, such as the total number of segments in languages.[3] We also apply a novel measure by reducing segment inventories into the minimal set of distinctive features needed to describe their phonemes.[4] The resulting distributions of these measures fail to exhibit any obvious parametric shape and the correlations that serve to defend positions about linguistic complexity do not appear, particularly when we look at the level of language family stock.

Next we look for correlations between different variables, both linguistic and non-linguistic, that have received attention in the literature, including consonants-vowels, obstruents-latitude, mean word length-segments and population-segments. What we find is that, looking at the broadest level (using all languages without any filters) we obtain interesting and significant but weak patterns.[5] However, when we zoom in on the language family stock-specific levels, these correlations turn out to be not significant. This lack of significance may result from either too few data points or may point to "true" insignificance. The first might be true for small families, but it is unlikely to be the case for large families, such as Nilo-Saharan, Afro-Asiatic, Niger-Congo, Trans-New Guinea, Austronesian and Indo-European. Thus when controlling for language family stock, there is only fragmented evidence pointing to universal correlations and there are no general trends.

This chapter is outlined as follows. In Section 2 we provide an overview of how complexity has been defined in the linguistics literature and specifically how absolute measures have been applied to phonological systems. In Section 3 we describe the datasets that we have compiled together for our investigation. In Sections 4 and 5 we present and discuss our results. In Section 6 we put forward our conclusions.

2 Defining complexity

2.1 Linguistic complexity

There are two types of approaches for measuring linguistic complexity, which Miestamo (2006) calls ‘relative’ and ‘absolute’.[6] Relative approaches involve quantifying the cost or difficulty involved in processing or using language, whether as a speaker, listener, L1 acquirer or L2 learner. For example, a measure of relative complexity is proposed by Kusters (2003:6), who defines complexity as the amount of effort a second language learner, or language community ‘outsider’, makes to become acquainted with a target language that he or she is unfamiliar with. Specifically, Kusters measures the effect of sociolinguistic changes on the complexity of verbal morphology by contrasting language learning effort of outsiders of four typologically different language families. One problem with relative complexity measures for typological research is that, as Kusters points out, they invite the question: ‘What is complex to whom?’ Answering this question involves identifying different roles (speaker or listener), different situations (acquirer or learner) and quantifying the numerous factors involved in producing, perceiving, acquiring or learning a language.

< INSERT FIGURE 1 ABOUT HERE >

Figure 1 illustrates, with just four languages, how every spoken language is in a complexity relation with every other language from a language learning standpoint. For example, the native speaker of German learning one of the Chinese languages will have a more difficult time than he or she might have learning English or French. This is due to the genealogical relatedness of German and English (they share many cognates and grammatical features) and the areal proximity of German and French (sustained language contact has led to shared vocabulary). Contrast this situation with the Chinese speaker learning English, French or German. Not only are these three languages typologically very different from Chinese, but the Chinese speaker would have to learn from scratch essentially every word. Kusters (2003) attempts to control for this effect by studying language outsiders like our hypothesized Chinese speaker. Nevertheless, for broad typological studies, measuring the effort of L2 learners is not currently feasible because every pair of languages will have its own unique bidirectional relationship that quantifies the relative complexity of the effort needed to learn the other language, complicated by factors of genealogical relatedness, areal proximity and elements of sheer coincidence, e.g. two languages that share similar sounds but which are not genealogically related or geographically close to each other.

Relative complexity metrics are exemplified by psycholinguistic studies that measure the cost and difficulty to language users in terms of processing or learning. The more difficult or costly it is to learn a linguistic phenomenon, then the more complex it is considered. It is hard, however, to define or quantify 'difficulty' of adult language learning in terms of metrics from psycholinguistic studies (McWhorter 2001, Trudgill 2001). For example, there are different burdens placed on the speaker and listener that require different metrics to capture articulatory- and perceptually-based effort as experienced by the speaker or listener. Therefore a phenomenon that is difficult for one group of language learners may actually be lessening the processing burden of another group. Miestamo (2009:81) gives an example of discontinuous negation, as in French *Je ne chante pas*, which illustrates an extra production burden on the speaker, but eases the comprehension task of the listener. Another example is given by Kusters (2008), who argues that redundant agreement is difficult for L2 learners and speakers, but lessens the burden on L1 learners and hearers. Nevertheless, there is simply not enough data, well-defined methods and psycholinguistic research on all relevant aspects and phenomena associated with different relative complexity measures (Kusters 2008, Miestamo 2009). Moreover, one should also consider non-linguistic factors; Kusters (2003) argues that differences in complexity for outsiders is related to the social and cultural history of their own speech communities.

Given the problems involved in undertaking language acquisition studies or psycholinguistic experiments and quantifying and comparing relative metrics of complexity, Miestamo (2006, 2009) suggests that cross-linguistic studies of complexity should instead define objective measures in absolute terms. Approaches to absolute complexity measure the number of parts of a linguistic system by some formal definition, often based on information theoretic approaches, e.g. McWhorter 2001, Dahl 2004, Miestamo 2006, 2008, among others. Thus complexity is not seen as a metric of the difficulty or cognitive load of a speaker, listener or learner, but as a property of a linguistic (sub)system. In the following section we describe absolute measures that have been proposed to capture the complexity of phonological systems.

2.2 The complexity of phonological systems

Over the years, various absolute complexity measures have been used to describe and compare phonological systems. In this section we provide a historic overview of these measures, focusing on univariate and bivariate measures of complexity and compensatory relations in phonological systems. Univariate measures count the number of pieces in phonological systems, such as the number of contrastive sounds in a language, which can then be compared across languages. Bivariate measures use univariate counts and test for the presence or absence of compensatory relations between different phonological subsystems cross-linguistically; for example do all languages with a great number of consonants have statistically fewer vowels? We also discuss the methodological considerations of these measures and show that the results of the same complexity metric have differed depending on the size and typological coverage of the language sample.

The first step in testing for complexity in phonological systems is to gather a cross-linguistic sample of language data. Once such a sample has been gathered, researchers usually turn their attention to the distribution of variables in the data, such as total segment counts. These discrete variable counts are identical to histograms. Especially interesting are the so-called "parametric" distributions, which can be described with a finite (and typically small) number of parameters such as mean and variance. The parametric shape of distributions is of uttermost important because it allows researchers to make inferences about the diachronic processes that produce them and the constraints that these processes are subject to (Maslova 2000, Cysouw 2007).

Segment inventory size is perhaps the oldest measure used for comparative surveys of cross-linguistic phonological patterns, e.g. Trubetzkoy 1939, Hockett 1955, Greenberg et al. 1978, Maddieson 1984. However, it is questionable whether overall phonological complexity can be captured by simply counting the total number of sounds in languages. Therefore researchers have divided segment inventories into different subsystems, such as consonant, vowel and tone inventories. Each inventory can be further divided into different categorial distinctions, e.g. consonants by type (obstruent, glides, etc.), vowels by quality and non-quality distinctions, and tonal systems by categories such as simple and complex. Segment inventories can also be split into articulation classes (e.g. into sonorants and obstruents) or along phonetic dimensions (e.g. vowels by quality). These finer grained distinctions allow researchers to investigate phonological systems in more detail, but they also allow them to coarsen segment type distinctions into classes of sounds based on their distinctive features.

Apart from distributions, the different aspects of phonological systems are used for investigating hypothesized compensatory relations. One area that has been repeatedly investigated is whether there is a correlation between the size of consonant and vowel inventories. Using a sample of 68 languages, Hockett (1955:138-140) was the first to investigate the "balance" between the number of consonants and vowels. He found a correlation between the number of vowel phonemes and the total number of segmental phonemes. Maddieson (1980) obtained comparable results. With a genealogically stratified sample of 50 languages, Justeson & Stephens (1984) showed that the size of consonant and vowel inventories are statistically independent, concluding that there is no correlation between them. Using an even larger dataset with 317 languages, Maddieson (1984:9) found that vowel ratio is inversely correlated with the number of consonants in an inventory and that the total number of vowels is positively correlated with the consonant total. In contrast, with an even larger dataset of over 600 languages, Maddieson (2007) more recently shows that consonant and vowel inventory sizes show no correlation with each other. And Moran (2012:chp. 5), using a dataset of around 1000 languages, calculates consonant and vowel ratios and shows that for each increase in roughly 13 or 14 consonants, there is an increase in one vowel. Although the p-value suggests that the finding is robust (slope = 0.0738, $R^2 = 0.0143$, $p < .0001$), the correlation is weak. Thus whether consonant and vowel inventories correlate in some way is still an issue of debate. This line of inquiry also shows how the results of one research question may change based on the sample of languages under investigation.

The consonant-vowel correlation has been actively investigated because it has long been suggested that languages compensate for complexity in one subsystem by simplifying another to achieve a roughly constant overall level of complexity. This 'compensation hypothesis' holds that a simplification or complication in one area of an inventory will be counterbalanced by the opposite somewhere else (Martinet 1955). These compensatory relations have been speculated as the outcome of historical processes that drive language change and they have also been investigated in light of languages being equally complex overall with regard to their communicative capacity (cf. Pellegrino et al. 2011). Maddieson (2006, 2007) investigates correlation patterns between different phonological subsystems in an attempt to shed light on compensatory relations in the phonological system, noting that regardless of whether the driving force is historical or communicative (or if it exists at all), if compensatory relationships hold they should present themselves in a survey of phonological properties of languages.

Table 1 illustrates Maddieson's (2006) findings from five phonological system variables taken to represent different metrics of complexity, which were then compared pairwise to see if they correlate positively (both measures increase in complexity), negatively (greater complexity in one measure; less in the other) or if they do not correlate at all. The variables for consonants, vowels, and vowel qualities are numerical, i.e. their values are represented by the total number of each type in the language under investigation. The variables for syllable complexity and tone complexity are categorial; they are based on the range of cross-linguistic types, which are then split into categories as defined by Maddieson (2006, 2007). Syllable complexity is based on the maximally complex syllable type found in the language and tone complexity is defined by the language's number and types of tones.

Table 1: Direction of correlation between complexity measures (Maddieson 2006:106)

	Syllable complexity	Consonants	Vowel qualities	All vowels
Consonants	Positive			
Vowel qualities	uncorrelated	uncorrelated		
All vowels	uncorrelated	uncorrelated	positive	
Tone complexity	Negative	positive	positive	positive

By comparing different variables, Maddieson (2006, 2007) shows that the compensatory relationship in a sample of phonological systems holds only in the tone and syllable categories. He reports a positive correlation between consonant inventories and syllable complexity, tonal system complexity and consonants, and tonal systems and vowels (for both quality and full systems).[7] Maddieson finds no correlation for the following variable pairs: the number of vowels and syllable complexity, or the size of the inventories of vowels and consonants in languages in the sample. These findings are in line with earlier findings, where Maddieson (1984:21) examines suprasegmentals (tone and stress) in a set of 317 languages and reports that the "overall tendency appears once again to be more that complexity of different kinds goes hand in hand, rather than for complexity of one sort to be balanced by simplicity elsewhere".

The compensation hypothesis has also been tested outside of phonology. Nettle (1995) looks for a correlation between segment inventory size and word length, testing the assumption that as segment inventory size increases, the average word length in the language decreases. This relation might be thought to exemplify the communicative efficiency of a language, since a language with fewer sounds might need longer words to articulate the same number of concepts as a language with more sounds, because the latter has more combinatory possibilities for creating words. For his study, Nettle chose 10 languages and extracted their phonemic inventories from Campbell 1991. For each language Nettle determined an average word length figure by choosing 50 headwords at random from 10 representative dictionaries. Then Nettle compared whether the mean word length is related to segment inventory size and his findings show that correlation holds in the small data sample.

In another and more recent study testing for compensation between grammar and the phonological system, Shosted (2006) tests for compensatory patterning between verbal inflectional markers and potential syllables in a diverse sample of 32 languages. His approach is innovative, not only because he tests for a negative correlation across linguistic subsystems, but also because he develops an algorithm to calculate the combinatory possibilities of all syllable types in each phonological inventory by taking into account phonemic contrasts, syllable types and phonotactic constraints.[8] Shosted uses this calculation as a measure of phonological complexity, reasoning that there must be a discrete maximum syllable count, because elements of a phonological inventory are discrete. Thus Shosted addresses criticism of the approach taken by McWhorter (2001), who does not take syllabic structure into account when evaluating phonological systems complexity, and Shosted provides an empirical approach that can be retested and expanded. He finds no evidence for a negative correlation.

A major problem with many of the studies mentioned so far is their relatively small language samples. As pointed out above, whether the number of consonants and vowels are found to be positively or negatively correlated (or correlated at all) has differed depending on the study and the language sample. Clearly larger sample sizes are more desirous because they are more likely to contain a greater range of variability across languages, including marked and rare phenomena (the problem of course is gathering, analyzing and compiling such resources).

McWhorter (2001:135) discusses complexity of phoneme inventories and defines their complexity through markedness of segment types, hypothesizing that a segment inventory is more complex than another if it has more marked segments. Markedness is calculated by a segment's cross-linguistic distribution, so marked segments are considered those that are less frequent in the world's languages. Again, a problem with this metric of complexity is sample size. For example, in the UCLA Phonological Segment Inventory Database (UPSID) there are 920 segment types that appear in 451 languages (Maddieson 1984, Maddieson & Precoda 1990). Of those 920 segments, 427 (about 46.5%) occur in one and only one language and therefore each is very rare in the sample. In the Phonetic Information Base and Lexicon (PHOIBLE) dataset of 1089 segment inventories, the number of total segment types occurring cross-linguistically increases to 1780. Of those 1780 segments, 909 (51%) are one-off occurrences. Thus in large samples of segment inventories created so far, nearly half of all segment types reportedly occur in just one language. Any study of phonological complexity that uses a small sample of languages is immediately suspect of lacking cross-linguistic representativity. The amount by which languages differ typologically, even within one subsystem, is quite astounding.

An additional methodological concern is the issue of how granular metrics of complexity should be when applied to phonological systems. As has long been noted in the phonological universals literature, phonemes are "not fruitful" universals (Hockett 1963:24), precisely because the analysis of allophones into a set of contrastive phonemes depends on several factors, such as the linguist's theoretical training and his or her ability to perceive sounds in the language being documented and described.[9] In evaluating Trudgill's (2004) thesis that phoneme inventory size correlates with social factors, Kabak (2004) argues that acquiring a phonology is not a process of acquiring an inventory, but of acquiring contrasts. Kusters & Muysken (2001) note that the phonological complexity approach taken by McWhorter (2001) does not take into account distinctive feature contrasts, but merely raw phoneme counts. Clements (2009:29) provides an example of why simple segment counts do not inherently capture complexity by illustrating two historical stages of Zulu, which we replicate in Table 2.

Table 2: Diachronic development of Zulu consonants

Stage 1				Stage 2		
p'	t'	k'		p'	t'	k'
pʰ	tʰ	kʰ		pʰ	tʰ	kʰ
b	d	ɡ		p	t	k
		k		b		ɡ
ɓ						

Stage 1 reflects Zulu nearly a century ago and stage 2 shows a more modern usage of Zulu plosives (excluding fricatives). Both stages have the same number of segments, but through a historical change, the two stops (and sole members of their series) /ɓ/ and /k/ have shifted into a single voiced series. Clements (2009:60) notes the increased feature economy of stage 2, which is due to the elimination of a feature (probably [obstruent]). What is clear is that although both stages of Zulu have the same number of segments, stage 1 requires a greater number of features to encode the phonemic contrasts in the language.

We know of no previous work on complexity that uses the number of distinctive features as a metric for complexity. McCloy et al. (2013) use a multidimensionality reduction algorithm to reduce segment inventories in the PHOIBLE dataset into the minimal set of distinctive features needed to describe the phonemic contrasts in each language. In this chapter we use this dataset as an additional source for evaluating phonological complexity.

To summarize, we have described the absolute measures that have been used to measure the complexity of phonological systems and we have given an overview of some of the ways in which these measures have been used to investigate the compensation hypothesis. We have also raised some of the methodological concerns regarding language samples and how complexity measures have been formulated and tested on both phonological systems and other areas of the grammar. In the next section we describe the large dataset that we have compiled to test metrics of complexity and in Sections 4 and 5 we run statistical tests for absolute complexity measures and discuss our results.

3 Data sample

The goals of our study are to test the absolute complexity measures of phonological systems and to compare previous findings against the results from a much larger and broader sample of the world's languages. To achieve these goals, we first compiled together several datasets from different typological databases to create one large data sample for testing, which covers different units of phonological systems. In this section we describe this dataset and its component parts.

Contrastive segment inventories are the most widely documented aspect of the phonological systems in the world's languages. This is due to the fact that the investigation and analysis of contrastive segments, i.e. phonemes, is the starting point for language documentation and analysis. Thus a great number of segment inventories have been compiled into segment inventory databases of varying genealogical and geographical coverage over the last several decades, among which are: the Stanford Phonology Archive (SPA; Crothers et al. 1979) and the UCLA Phonological Segment Inventory Database (UPSID; Maddieson 1984, Maddieson & Precoda 1990). Phonological typology databases of this sort have provided a rich resource for investigating the segment inventory complexity of languages, allowing for measures such as the number of segments in a language and the ratio of consonants versus vowels.

In this study, segment inventory data is taken from the PHOIBLE database, which is comprised of SPA and UPSID. PHOIBLE contains several hundred additional inventories, bringing the current total number of languages in the sample to 1200 (Moran 2012). Inventories in PHOIBLE are "factual claims" attributed to one or more linguists who either documented a language and used linguistic theory to posit a set of contrastive sounds for that language or consulted various published resources and inferred a set of contrastive sounds based on other linguists' analyses. The first type of factual claim is what one finds in secondary resources such as grammars and phonological descriptions. The second type is exemplified by the inventories in UPSID, in which Maddieson sometimes reinterprets secondary sources' phonemic analyses by a consistent standard. Both claims are influenced by various factors, such as the linguist's training, language background and the linguistic theory used to describe, analyze and posit contrastive sounds in the language under investigation. Thus all measures of complexity using contrastive phonological inventories are at their core based on researchers' assumptions, decisions and skills. In our experience, no two descriptions of the same language by different linguists contain exactly the same description of contrastive segments, so their segment types and counts may differ. A segment inventory is essentially an abstraction over the set of contrastive segments used by a particular language variety's phonological system, as defined by the set of distinctive features employed by that language (Clements 2009:19). Whereas a segment is an abstraction of an articulatory or auditory unit of speech production or perception, features represent the basic phonetic units of a segment as defined by a particular distinctive feature theory. Therefore, in our study we move beyond just segments and on to the level of distinctive features that are needed to describe segment inventories, which allows us to abstract away from individual researcher preferences by using features based on articulation and perception.

The feature system used in this chapter is built on Hayes 2009, which was expanded in Moran 2012 to gain greater typological coverage of the segments in the inventories in the PHOIBLE database. It contains a set of 37 hierarchically organized features. To calculate the number of features needed to encode the contrasts in each segment inventory, McCloy et al. (2013) implemented a multidimensionality reduction algorithm that reduces the feature vectors of all contrastive segments in an inventory into the minimal number of features needed to encode that inventory. These sets of features constitute a data point in the sample of languages that we use in this study.

Additional information in the form of syllable structure data is taken from Maddieson 2011b and available through the World Atlas of Language Structures (WALS; Dryer & Haspelmath 2011) and added to our compiled dataset. The WALS data sample contains 486 languages and syllable canon types are divided into three categorial values: simple, moderately complex and complex. Simple syllable structure is found in 61 languages that only allow (C)V. Another 274 languages with moderately complex syllables have CVC or CCV structures. And complex syllable structures occur in 151 languages that are defined as permitting "freer combinations of two consonants in the position before a vowel, or which allow three or more consonants in this onset position, and/or two or more consonants in the position after the vowel" (Maddieson 2011b:chp. 12).

Other than data about phonological systems, we include information about wordlists, genealogy, geography and demography. Mean word length figures are derived from Swadesh wordlists collected and compiled for the ASJP database (version 15) (Wichmann et al. 2012). Language family stock data come from the Ethnologue 15 (Gordon 2005) via Multitree (LINGUIST List 2009). Geo-coordinates for languages are taken from WALS (Haspelmath et al. 2008). Population figures are from Ethnologue 16 (Lewis 2009).[10]

4 Results of absolute complexity measures

First we examine the total number of segments per language in the PHOIBLE data sample. The distribution of segment inventory sizes by languages is given in Figure 2.

« INSERT FIGURE 2 ABOUT HERE »

At least two proposals for a parametric model underlying this distribution have been put forth: log-normal (Justeson & Stephens 1984) and normal (Nichols 2009). While the normal (or Gaussian) distribution requires little introduction, the log-normal distribution (whose logarithm transform is normal) is characterized by an asymmetric shape around its peak, with a marked right skew. In practical terms this means that, in contrast with the symmetric and concentrated normal distribution, log-normal distributions have a non-negligible number of points far right from their maxima. Fits for normal and log-normal distributions on our dataset are shown in Figure 3.

« INSERT FIGURE 3 ABOUT HERE »

A standard statistical test shows that neither the data (Shapiro-Wilk test, DF = 1299, W = 0.894, $p < 10^{-6}$) nor its logarithm transform (Shapiro-Wilk test, DF = 1299, W = 0.994, $p < 2 \cdot 10^{-4}$) are normally distributed.[9] However, it is important to remember that repertoire size can only be a natural number, whereas normal and log-normal distributions are defined in a continuous domain. Continuity is assumed as well in the normality test, and the violation of that condition may lead to erroneous results. For this reason we test how well the total number of segments can be adjusted by a negative binomial distribution, which is a general model for count data. Once more we find that the data fail to conform to any particular parametric model (Likelihood Ratio Test, χ^2=197.71, DF=74, $p<10^{-6}$).

A closer inspection reveals a structure of sharp local maxima along the distribution, in contrast with the unimodal fit of both normal and log-normal distributions. The reason behind this fine-grained structure becomes clear when stock-specific distributions are analyzed, as is shown in Figure 4.[10,11] The differences in shape and average size of the inventories across families is remarkable, and a first look suggests that areal effects play an important role.

« INSERT FIGURE 4 ABOUT HERE »

Most of what we have pointed out for segment distributions is also true for consonants and vowels when taken separately. The distributions for consonants and vowels in PHOIBLE can be seen in Figure 5. Although Justeson & Stephens (1984) propose a log-normal distribution for consonant and vowel inventories (based on a sample of 50 languages taken from Ruhlen 1975), the histograms show a much more complicated story.

« INSERT FIGURE 5 ABOUT HERE »

Here again we find that, when studied individually, each language family has characteristic distributional features, as is apparent in Figure 6.

« INSERT FIGURE 6 ABOUT HERE »

Together with Figure 4, Figure 6 illustrates why single representative sampling from a language family, as in Shosted 2006, is not appropriate. For instance one can see that the range for consonants (and total segment counts) for Nilo-Saharan, Afro-Asiatic and Niger-Congo is quite large: there is about a fourfold difference between the minimum and the maximum. However, Trans-New Guinea and Austronesian have a more concentrated distribution. A parametric model would force us to collapse many informative quantities into one: for instance, in peak distributions such as normal, Laplacian and Cauchy, the mean, mode and median all coincide. With no obvious parametrization for the families' distributions, then any single scalar (be it the mean, median, mode, any moment or any other statistic) will necessarily lack information about the diversity and the particularities of the distribution it comes from. Thus a single number is inadequate for summarizing the information of a whole stock.

So far we have presented observations about the data without considering the possible correlation between variables. First we look for any pattern between consonant and vowel counts. The total counts for consonant and vowel inventories are shown in Figure 7.

« INSERT FIGURE 7 ABOUT HERE »

The most salient trend is a mild, although significant, positive correlation between the number of consonants and the number of vowels (Spearman's ρ: 0.12, $p < 10^{-5}$). When we ask whether this effect holds within the largest language family stocks in the database, we find that only one individual family yields a statistically significant result. A linear correlation for Indo-European turned out to be significant and of moderate effect (Pearson's r: 0.31, $p = 0.018$).

These results should be considered carefully. It might be that the lack of significance is due to the small sample size, but given the fact that we have chosen particularly large language family stocks for our analysis, the reason is likely to lie elsewhere. Within most stocks under investigation, consonants and vowels do not exhibit any special pairing, in the sense that they show remarkably different numbers of consonants and vowels, from what one would expect in a random matching from distinct languages within the stock. Under these circumstances, the original correlation would be a mere by-product of the families' contrasting characteristics instead of denoting a causal connection between consonants and vowels.

Our next analysis concerns the set of features that constitute the set of contrastive segments in each inventory. We first consider the distribution of the number of minimum features required to describe a language's segmental repertoire, shown in Figure 8.

« INSERT FIGURE 8 ABOUT HERE »

The summary statistics of this distribution reveal a much more symmetric and balanced shape than that for segment inventory sizes. As we mentioned above, Justeson & Stephens (1984:531) argue that both consonant and vowel inventories follow a log-normal distribution. They also claim that this distribution "receives straightforward linguistic interpretation in terms of distinctive features". If the number of features were connected with the number of segments through a logarithm, then from the log-normality of the segments we would be able to deduce the normality of the feature distribution.. The distribution of the number of minimum features can be roughly approximated by a binomial distribution (which is the adequate parallel of the normal distribution in our case), although we must conclude that this model is still not satisfactory (Likelihood Ratio Test, χ^2=21.72, DF=12, p=0.04).

However, the biggest issue in Justeson & Stephens' reasoning falls on the conjectured logarithmic relation between features and segments. Such a relation is true, for instance, in the case of a fully productive segment inventory, which makes use of all of the contrasts provided by its feature set. Thus if the feature set is size F, then the segment inventory will be of size $S = 2^F$. However, the relations between segments and features in this dataset exhibit a large degree of variability in the fraction of possible segments attested, as is shown in Figure 9.[14]

« INSERT FIGURE 9 ABOUT HERE »

Besides the form of the distribution of features, another relevant cross-linguistic property of features directly related to complexity is their rarity, i.e. how frequently they are attested in world's languages.

There is a group of six features that are highly represented cross-linguistically in our data sample, namely, dorsal (86%), high (83%), labial (83%), syllable (80%), continuant (80%) and voiced (79%). After that, the frequency of the remaining features decays in a manner that can be reasonably well approximated with an exponential fit (Adjusted $R^2 = 0.988$). This is shown in Figure 10.

« INSERT FIGURE 10 ABOUT HERE »

Next we look at the relation between mean word length from wordlists in the ASJP database (measured in number of distinctive segments) and corresponding languages' segment inventories in the PHOIBLE database. Similarly to what was found by Nettle (1995), mean word length shrinks with larger segment inventories (Spearman's ρ: -0.309, $p < 10^{-5}$), but particularly with the number of vowels (Spearman's ρ: -0.311, $p < 10^{-5}$), instead of with the number of consonants (Spearman's ρ: -0.182, $p < 10^{-5}$). The effect is larger for small repertoires up to ten vowels, and afterwards the mean word length stabilizes. The vowel-to-mean word length correlation is illustrated in Figure 11.

« INSERT FIGURE 11 ABOUT HERE »

When language family stocks are considered in isolation, however, the vowel-mean word length correlations go beyond the threshold of significance in only two families: Indo-European (Spearman's ρ: -0.503, $p = 2 \cdot 10^{-4}$) and Niger-Congo (Spearman's ρ: -0.149, $p = 0.007$). These are illustrated in Figure 12. The other four stocks fail to reach significance, thus showing how widely the patterns differ.

« INSERT FIGURE 12 ABOUT HERE »

Finally, we assess the influence of environmental and societal variables on some of the measures of absolute complexity. First we look for a correlation between latitude and various inventory sizes. Using a larger sample of languages than these previous studies, Maddieson (2011a:32) finds a highly significant correlation between a phonological index (calculated as the sum of consonant inventory size and categorical syllable complexity values) and the distance from the central latitude of the world's temperate zones.[12]

Using the PHOIBLE dataset, we find a positive correlation between latitude and obstruents (Spearman's ρ: 0.451, $p < 10^{-5}$). Interestingly, vowels have a much weaker correlation (Spearman's ρ: 0.162, $p < 10^{-5}$) and sonorants do not even correlate significantly. This information is displayed in Figure 13, along with the best linear fits for illustrative purposes.

« INSERT FIGURE 13 ABOUT HERE »

Latitude elicits the same tendency in syllable structure as well, as can be seen in Figure 14. Far north of the equator, syllabic structure is prominently complex, whereas close to it is mainly of moderate complexity, as defined in Maddieson 2011b (Wilcoxon Rank Sum Test, all Holm-corrected $p < 10^{-6}$).

« INSERT FIGURE 14 ABOUT HERE »

The study of the effect of latitude on segments according to language family can be misleading, since in our database all but six stocks occupy a range larger than 30° latitude, namely Tupi-Guarani, Altaic, Afro-Asiatic, Austronesian, Indo-European and Niger-Congo. Only the last four have enough data to let us infer a significant tendency. The picture that we obtain for the obstruents-latitude relation is mixed: Afro-Asiatic (Spearman's ρ: 0.32, p =0.027) and Austronesian (Spearman's ρ: 0.233, p =0.043) follow the overall trend, whereas for Indo-European (Spearman's ρ: -0.323, p =.024) and Niger-Congo (Spearman's ρ: -0.140 p =0.022) larger latitude is often more related to a decrease in the number of obstruents.

Another important non-linguistic factor that purportedly shapes linguistic complexity, as measured by segment inventory size, is speaker population size (Haudricourt 1961, Trudgill 2004, Hay & Bauer 2007). Our results corroborate the overall correlation between segments, consonants and vowels on the one hand, and population on the other, is positive, with total segments being the correlation with the largest statistic (Spearman's ρ: 0.331, $p < 10^{-5}$). This is shown in Figure 15.

« INSERT FIGURE 15 ABOUT HERE »

Again, a family-specific analysis yields no significant correlations for most of the stocks except two, Afro-Asiatic (Spearman's ρ: 0.341, $p < 0.02$) and Austronesian (Spearman's ρ: 0.381, $p < 0.04$).[13]

5 Discussion

In a nutshell, our results show that claims about either the parametric shape or the correlations among phonological variables appear as unsustained in the light of a simple statistical analysis of the largest phonological database to date. The distributions of the size of segment, consonant and vowel inventories fail to be classified by a standard statistical test as either normal or log-normal distributions (Justeson & Stephens 1984, Nichols 2009). Moreover, when these distributions are inspected in more detail, they reveal a rich structure of local peaks that have their origin in the superposition of the widely variable distributions of individual language family stocks.

We mentioned before that segment counts in the PHOIBLE database are "factual claims", strongly influenced by the knowledge and the praxis of the linguist, and as such, segment counts are not the most reliable indices that can be used for the purpose of measuring complexity. Instead, we claim that counts of minimal features are a more reliable measure, as we discussed in Section 3. This claim turned out to be backed by the analysis of the cross-linguistic frequency of occurrence of each of the 37 features from which minimal sets can be built. From a purely statistical point of view, the fact that these frequencies are well adjusted with an exponential function is relevant, because (in this case) it implies that most of the features will be present in many languages, and only a few will be cross-linguistically rare. This stands in contrast to the so-called "heavy-tailed distributions" that usually have a large number of infrequent events. Conspicuously, segments constitute such a distribution: in both the UPSID and PHOIBLE databases, about half of all segment types are attested in one and only one language. Thus, feature counts are more conservative when it comes to measuring phonological complexity.

Taking together all the results concerning correlations between phonological variables, two commonalities appear. First, all pairwise correlations for the full dataset have a small effect, meaning that the statistical trends emerging from these data were far from revealing dramatic dependencies between variables. In other words, any individual variable does not give us sufficient information about another, e.g. knowing the number of consonants in a language does not give us power to predict the number of vowels in that language.

Second, and perhaps more importantly, for the largest stocks in the data sample used in this chapter, there was no uniform picture emerging from the pairwise correlations. Specifically, while some stocks do coincide in the direction of effect with the overall tendency, most stocks do not reveal any statistically significant trend. Under these circumstances, it is valid to call into question the universality of explanations linking different phonological subsystems. Presently, our data favors a historical (genealogical or areal) explanation behind most of the important correlations found. If there are cognitive and ecological factors molding the system of relations between phonological subsystems, they appear to be slower than the typical dynamics of language change.

6 Conclusions

In the last decade or so, the long-held assumed truism that all languages are equally complex has been challenged by many researchers. These researchers have explored issues of complexity from different perspectives and have proposed various metrics for quantifying complexity in and across linguistic subsystems. However, so far there are no conventionally agreed-upon metrics for quantifying or measuring complexity in the grammar and there has been no comprehensive quantitative method that has proven or disproven the notion of grammatical equality with regard to complexity.

In this chapter we have explored from a typological perspective several absolute measures used to quantify complexity of phonological systems. Despite differences of focus and methods, we have found that most of the research in this area relies on claims about statistical aspects of datasets, and particularly on the parameterization of distributions and the direction of correlations between pairs of variables. After compiling together datasets from some of the largest phonological databases available, we have shown that both strategies are problematic and to date no universal and salient trend can be established. First, distributions for the size of phonological systems are statistically different from the simple parametric models proposed in the literature, i.e. normal and log-normal. We have shown that the counts of minimal sets of features for segment inventories – first introduced in this work – are more robust than those of segments, and as such appear to be more accurate for the purpose of measuring phonological complexity. Second, pairwise analyses of phonological variables yielded significant, yet small correlations. If these correlations were propelled by universal laws of compensation or complexification, we should expect to see them within individual stocks as well – however, this is not the case.

As with most statistically-based research, our conclusions cannot rule out that a more detailed database or the choice of a particular set of variables could yield convincing universally-valid results regarding the structure or the development of phonological systems. However, it might be as well that our results are a consequence of languages being intrinsically flexible, adaptable structures that can allow a broad number of configurations and dynamics that do not fit well within a classification of universal types, cf. Evans & Levinson 2009, Bickel In press.

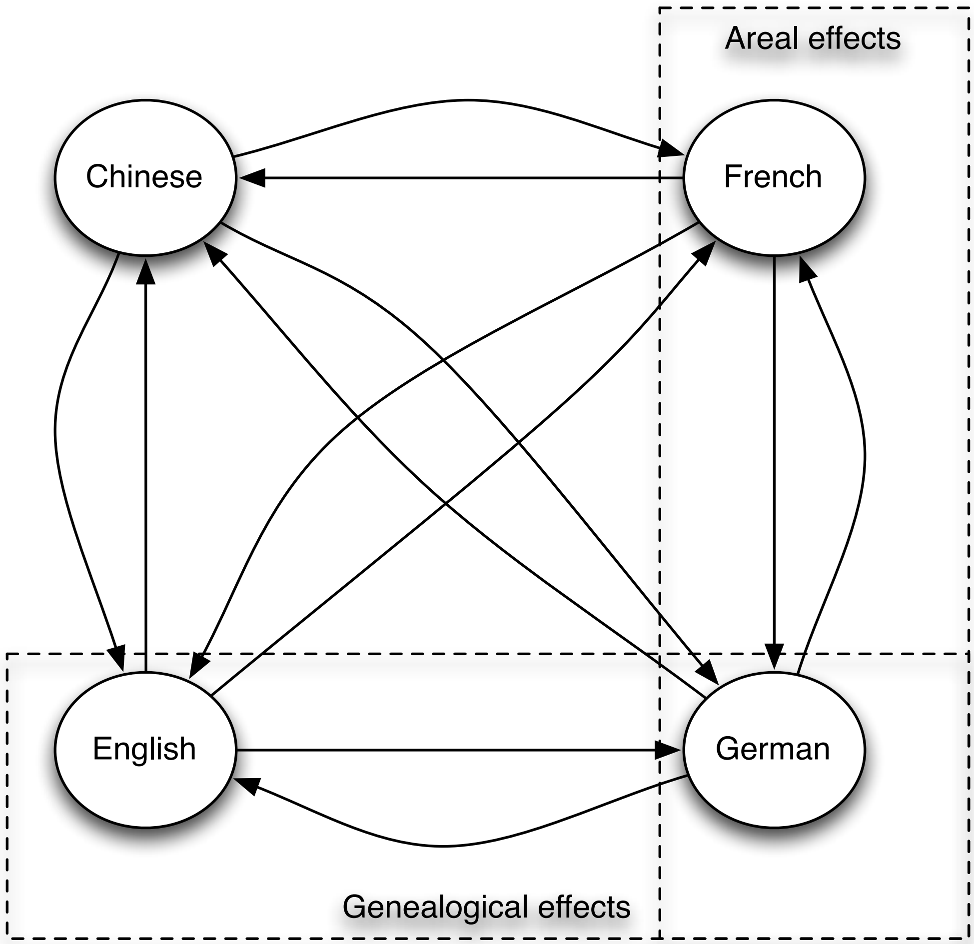

Fig. 1: Relations and effects on L2 learner effort

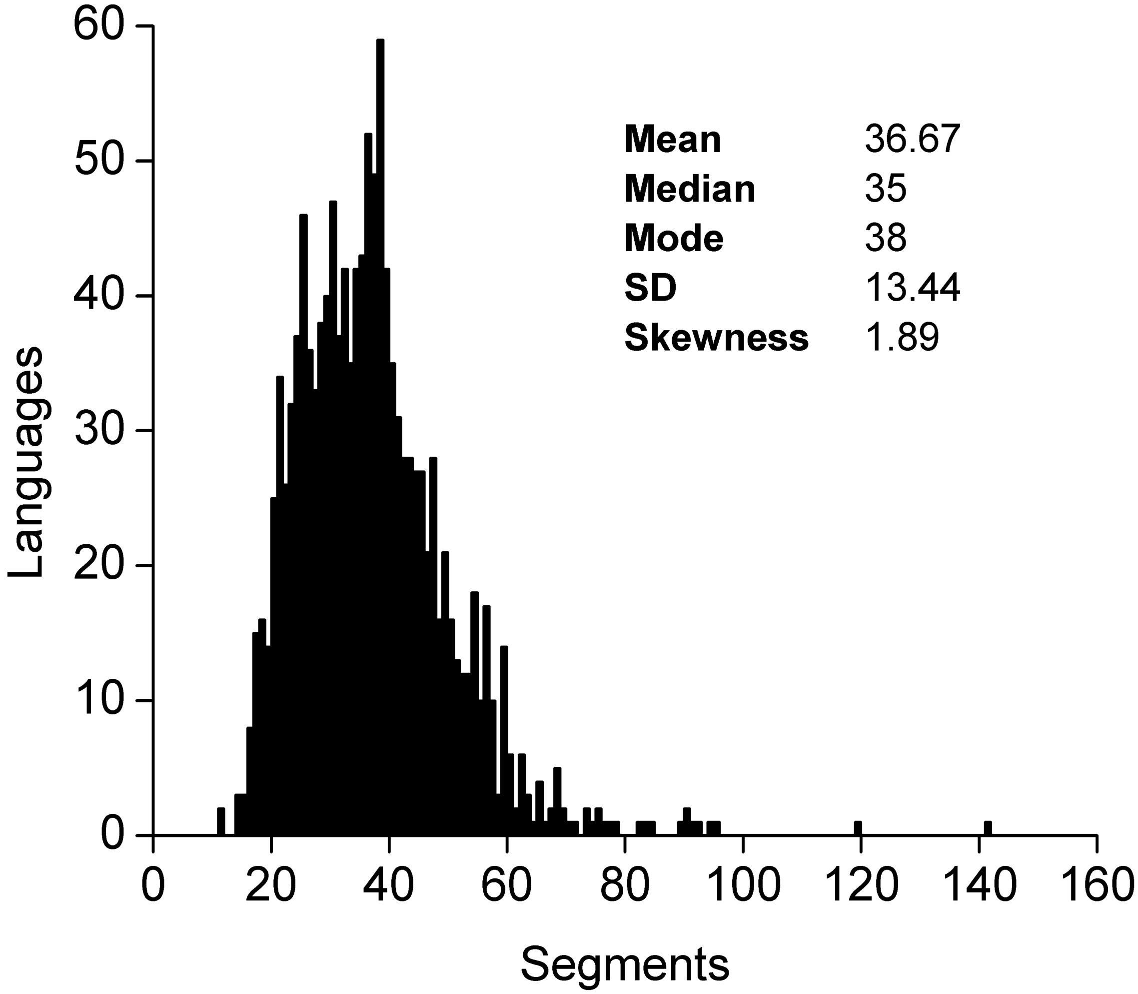

Fig. 2: Distribution of segment inventory size in the PHOIBLE database

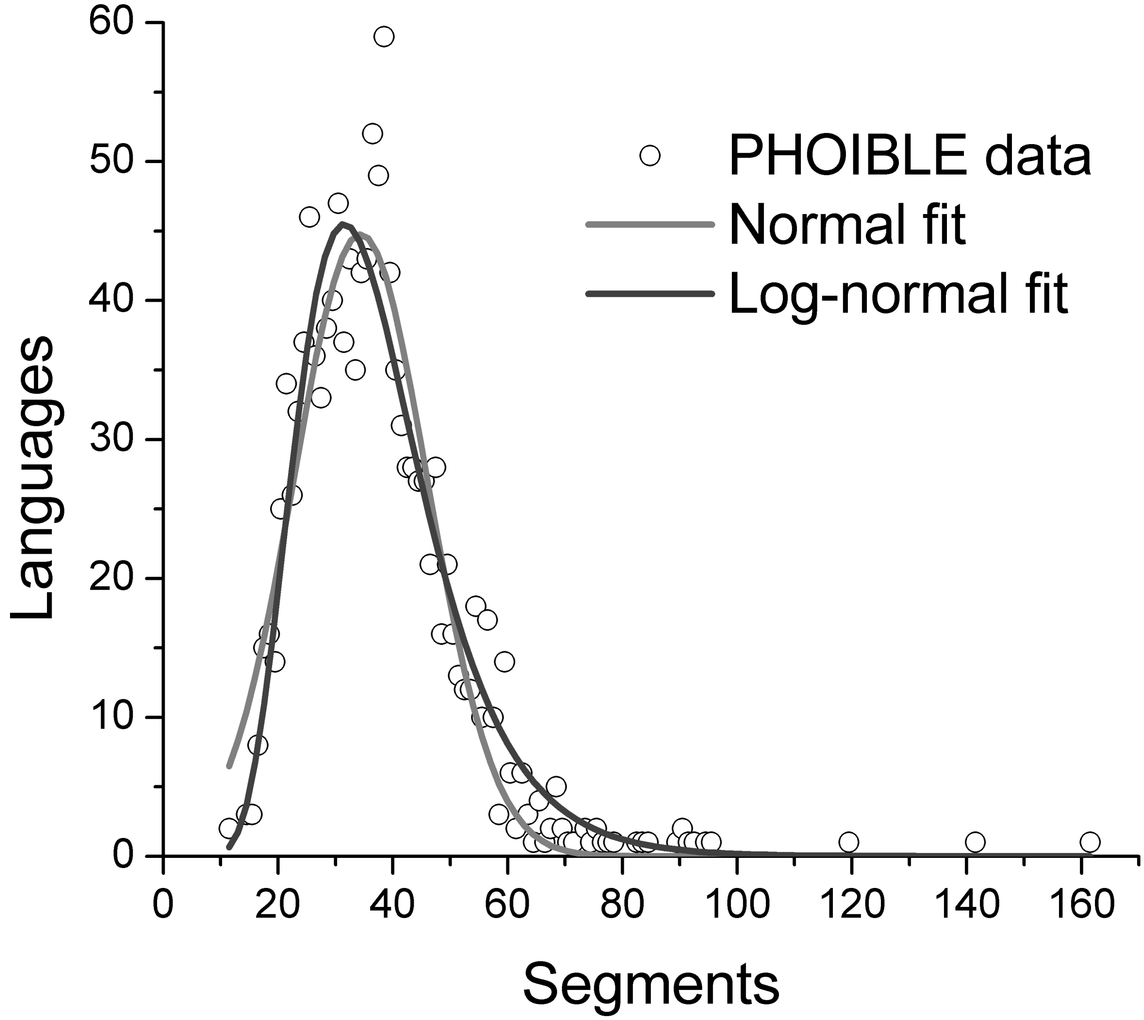

Fig. 3: Normal and log-normal fit of the segments distribution in the PHOIBLE database

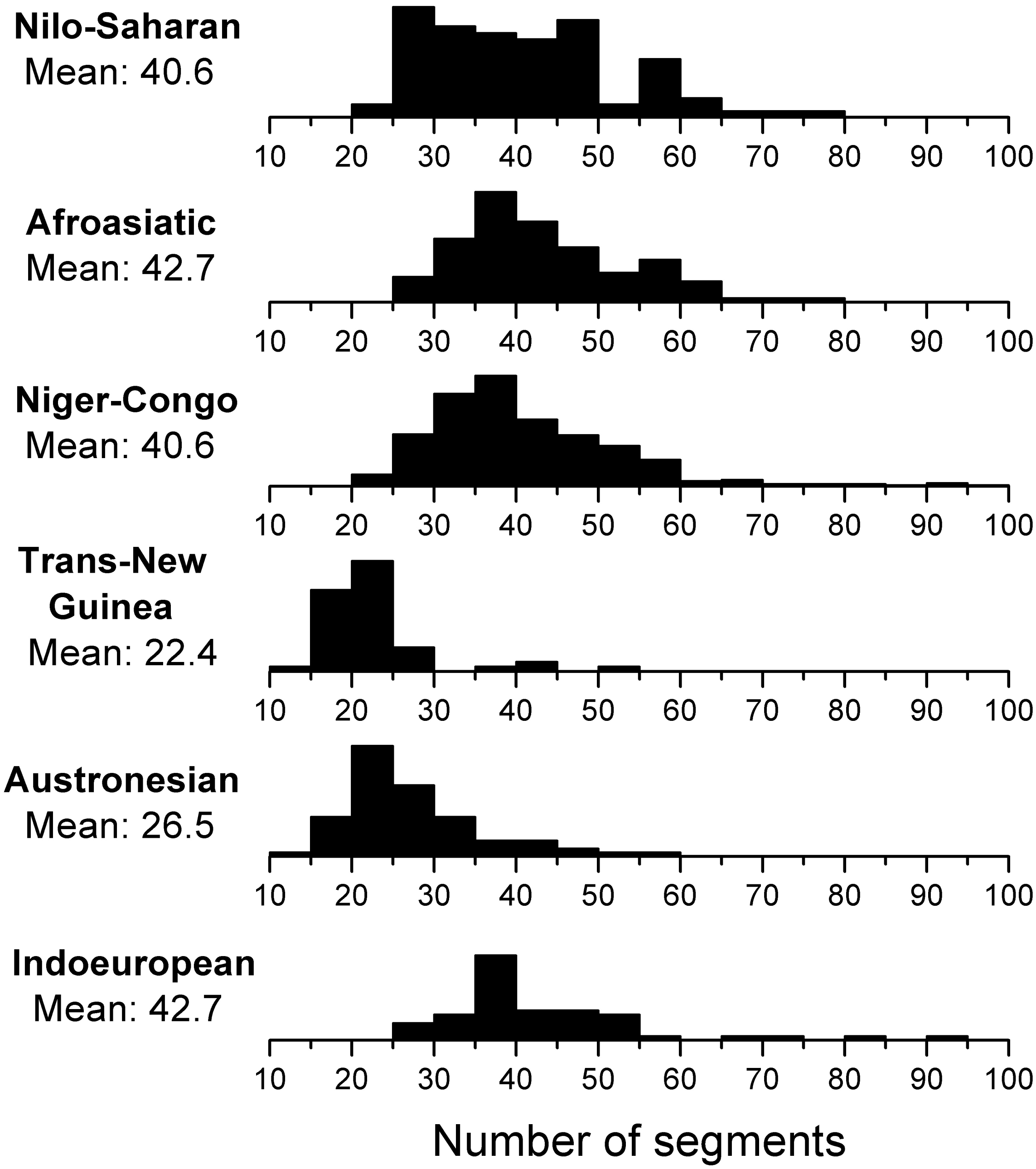

Fig. 4: Distribution of segments for Nilo-Saharan, Afro-Asiatic, Niger-Congo, Trans-New Guinea, Austronesian and Indo-European stocks

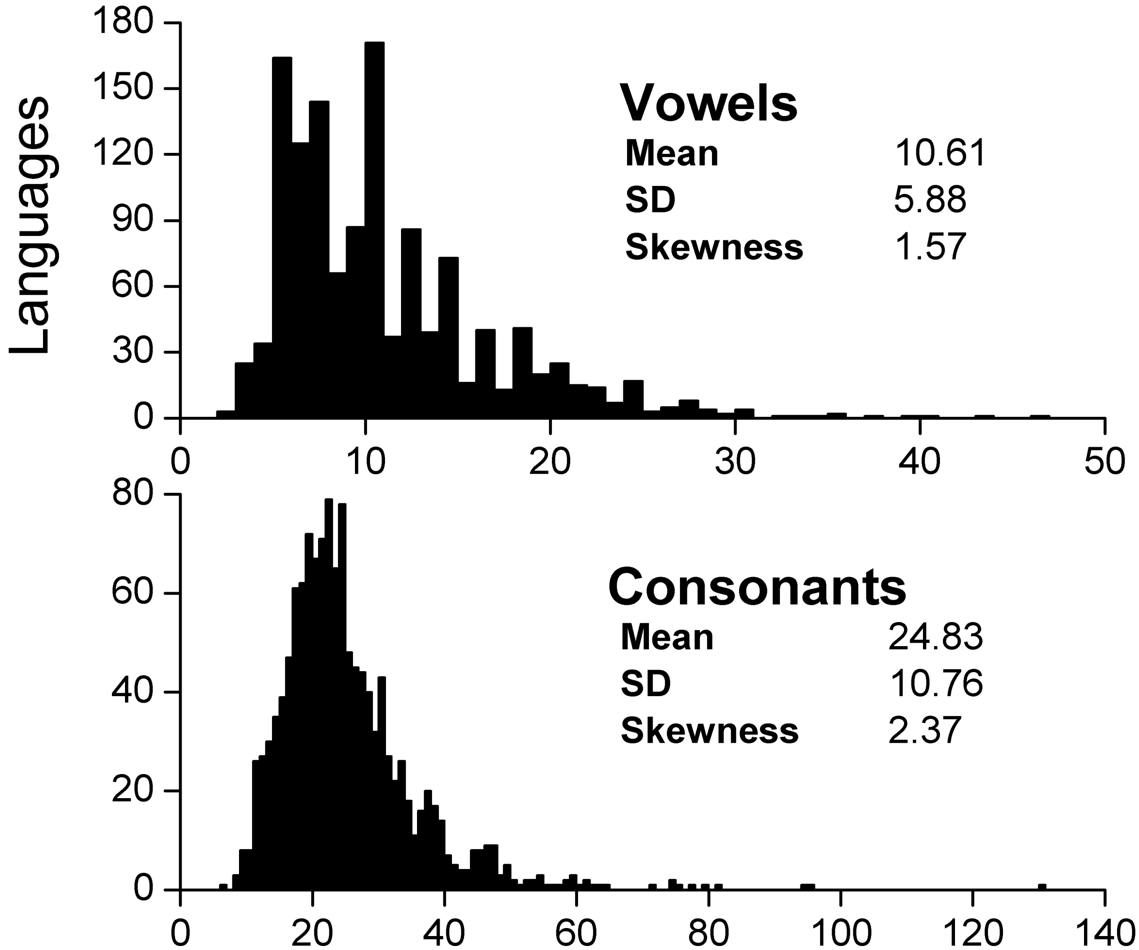

Fig. 5: Distribution of consonants and vowels in the PHOIBLE database

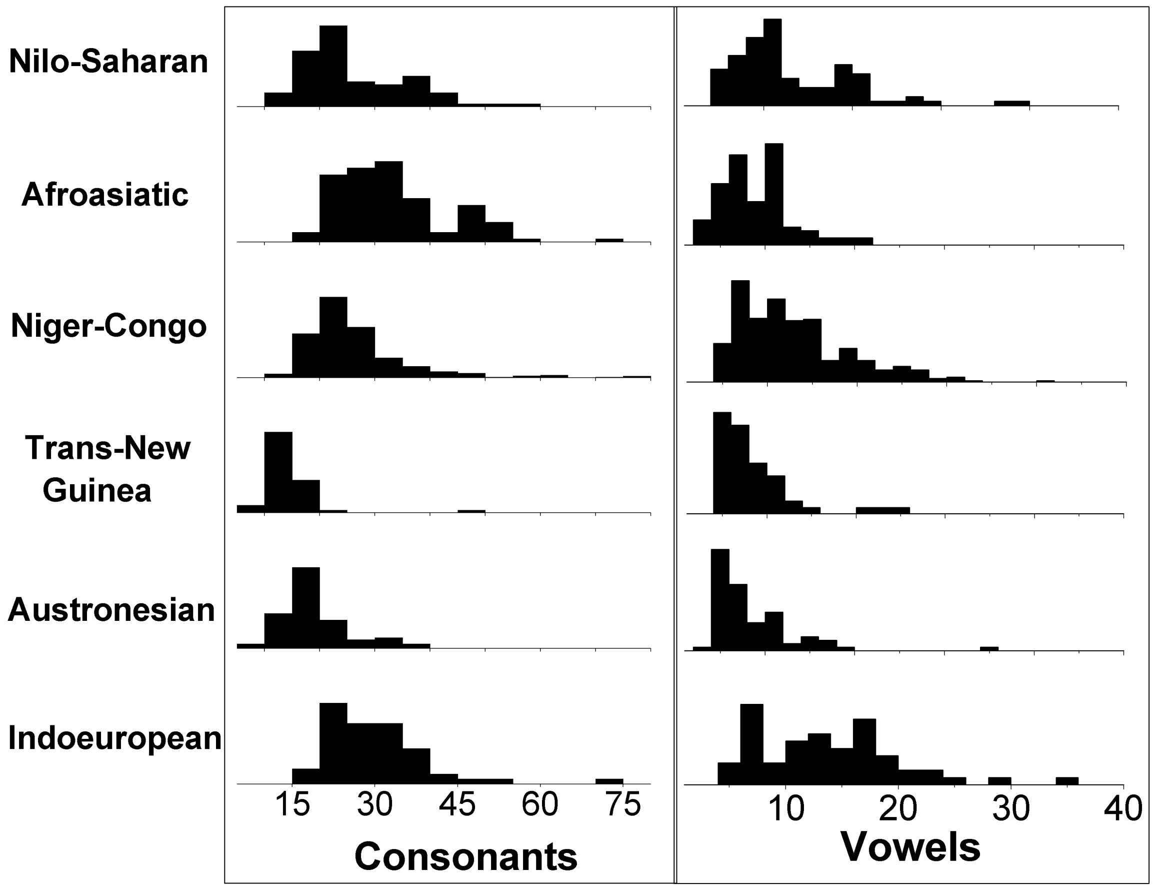

Fig. 6: Distribution of consonants and vowels for Nilo-Saharan, Afro-Asiatic, Niger-Congo, Trans-New Guinea, Austronesian and Indo-European stocks

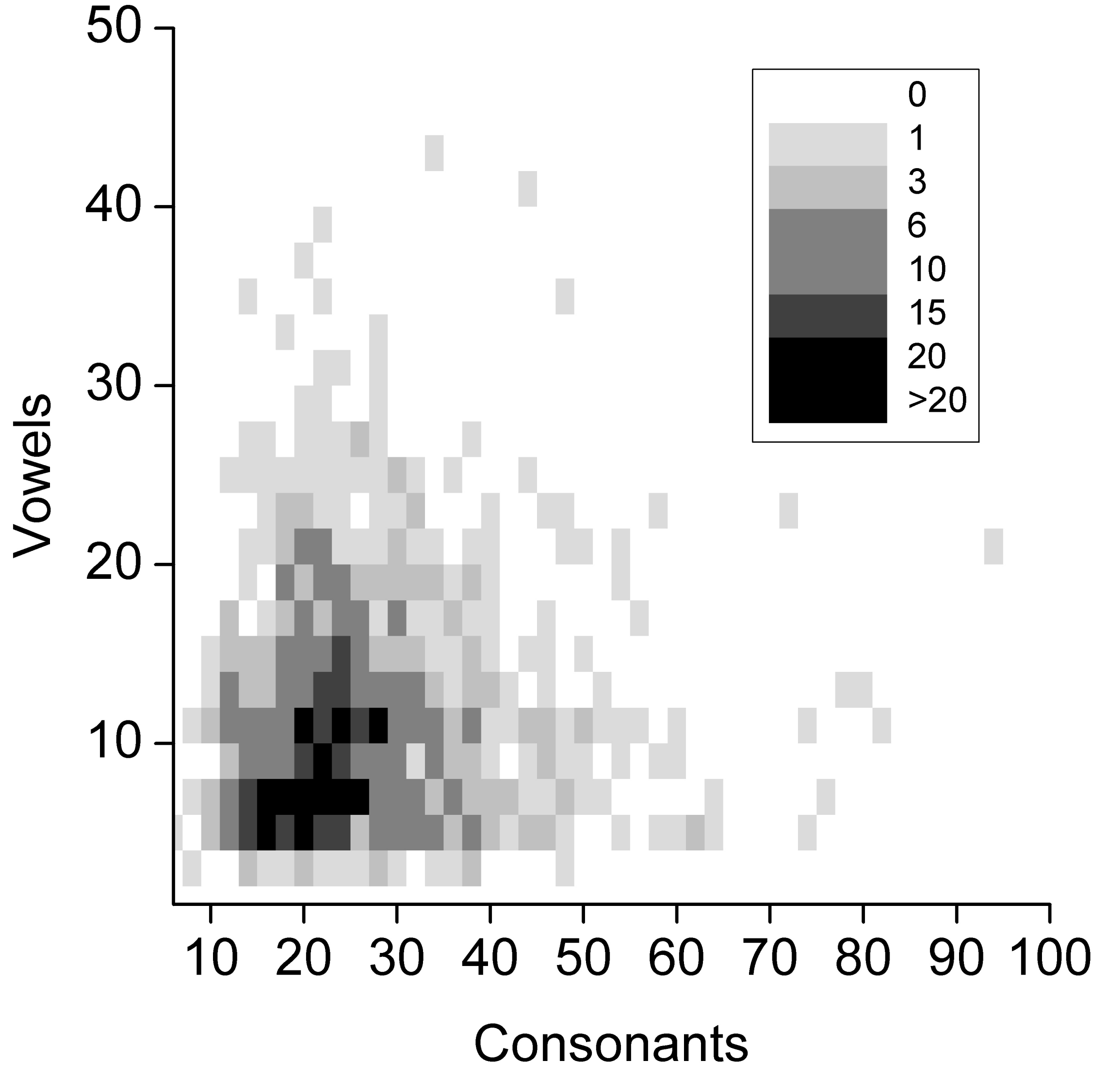

Fig. 7: Vowel and consonant inventories in the PHOIBLE database

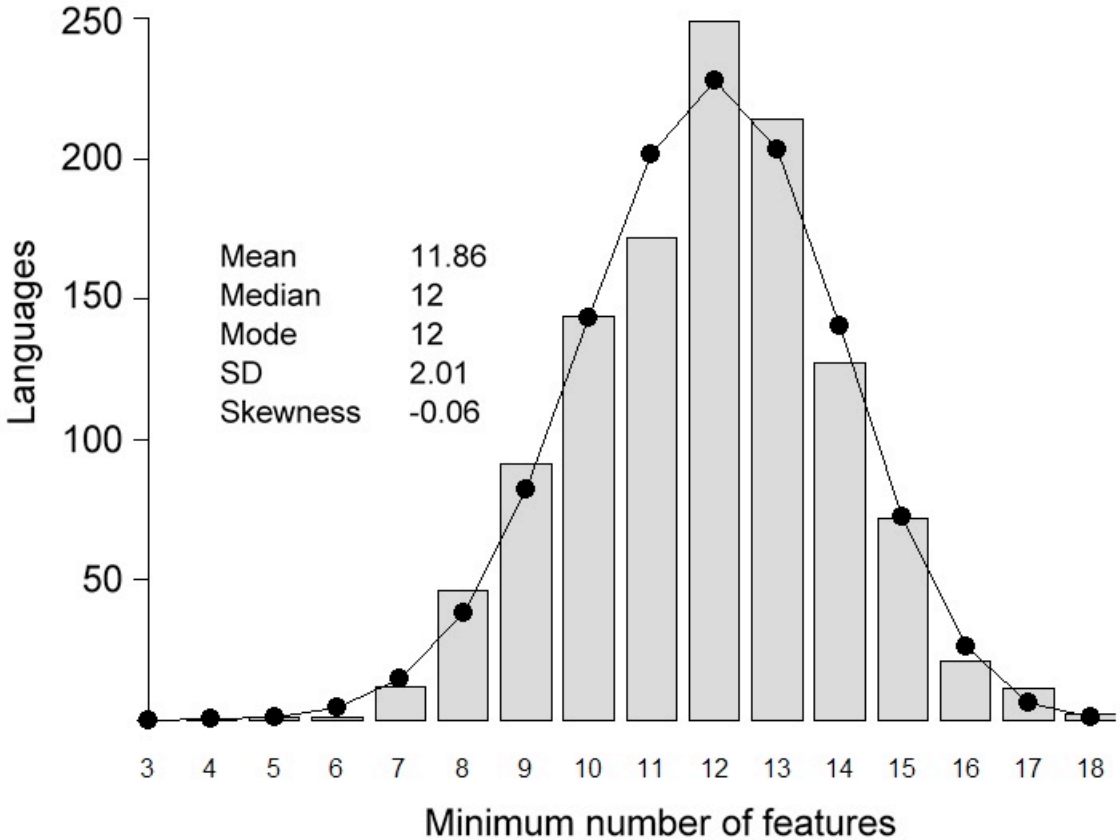

Fig. 8: Distribution of the minimum number of features required to describe each segment inventory in the PHOIBLE database. Black dots correspond to the best fit for a binomial distribution.

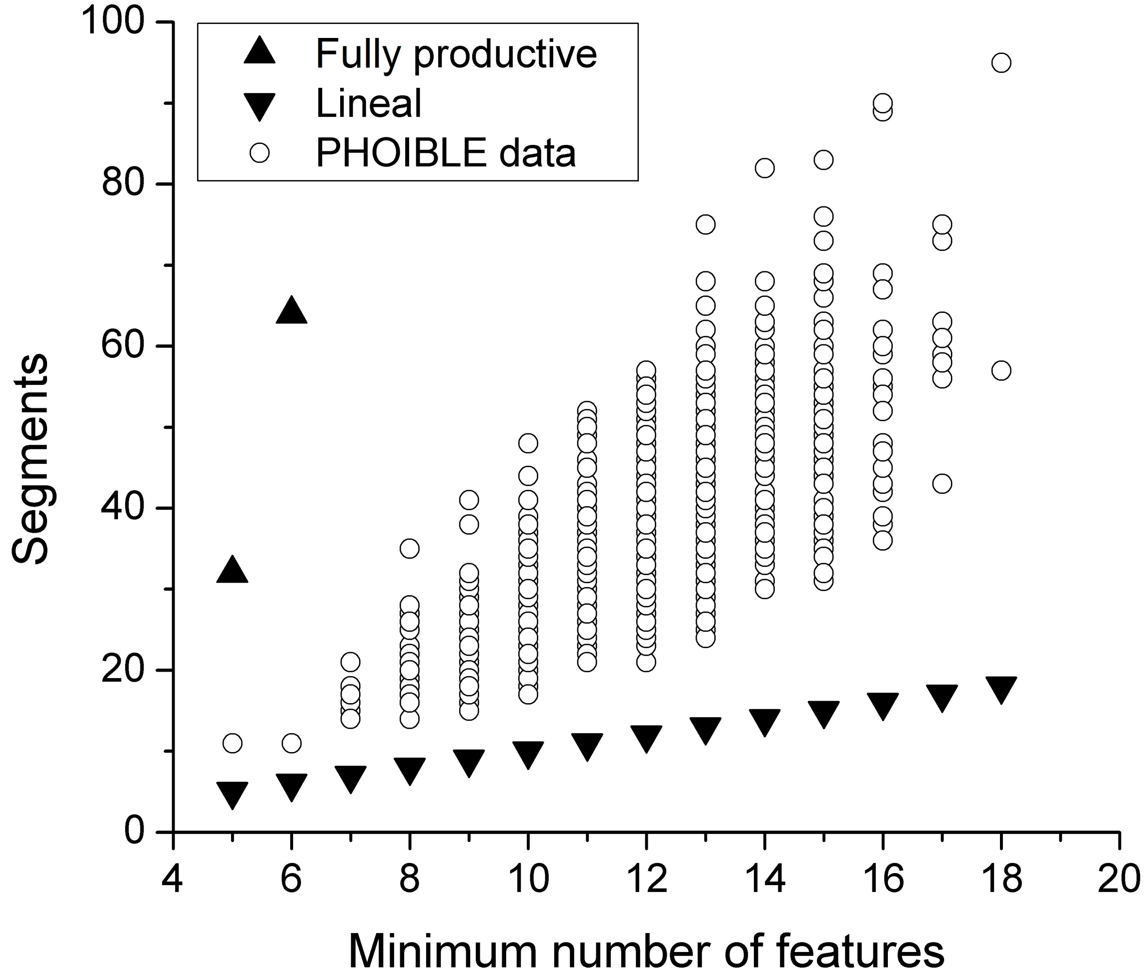

Fig. 9: Segment inventory size vs minimum number of features needed to encode each inventory

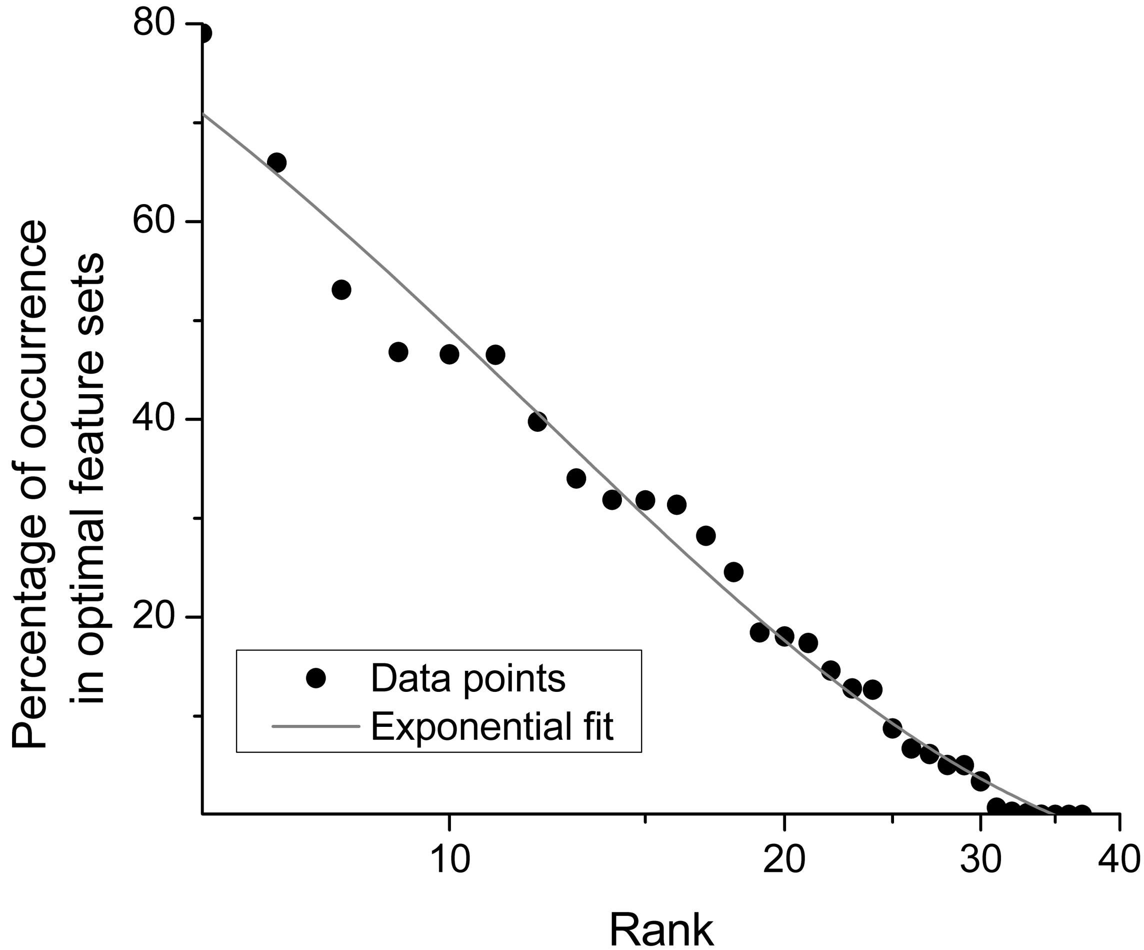

Fig. 10: Ranking of features according to their cross-linguistic frequency and exponential fit (first 6 values not shown)

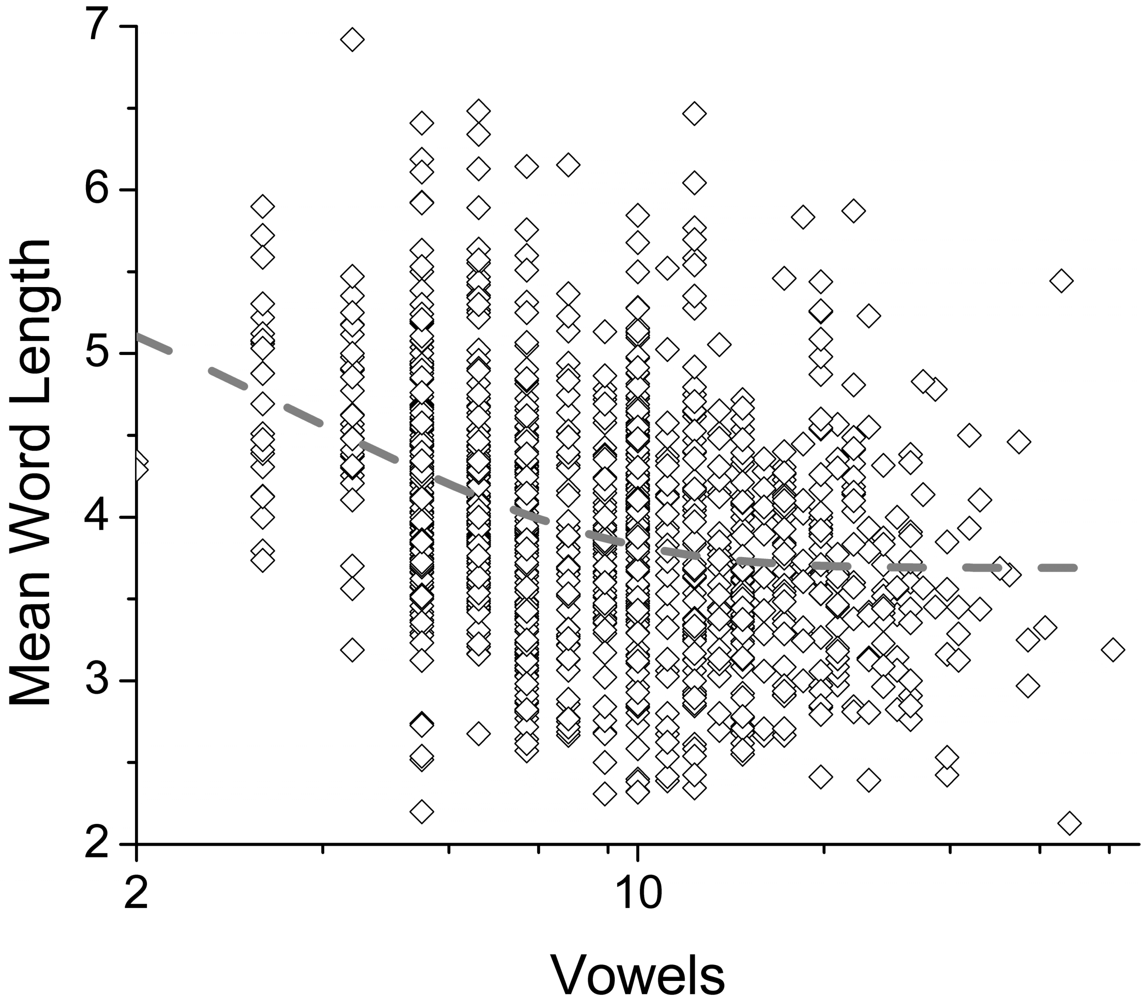

Fig. 11: Mean word length (from ASJP) vs vowel inventories (from PHOIBLE). The dotted gray curve is an exponential approximation of the moving average of data.

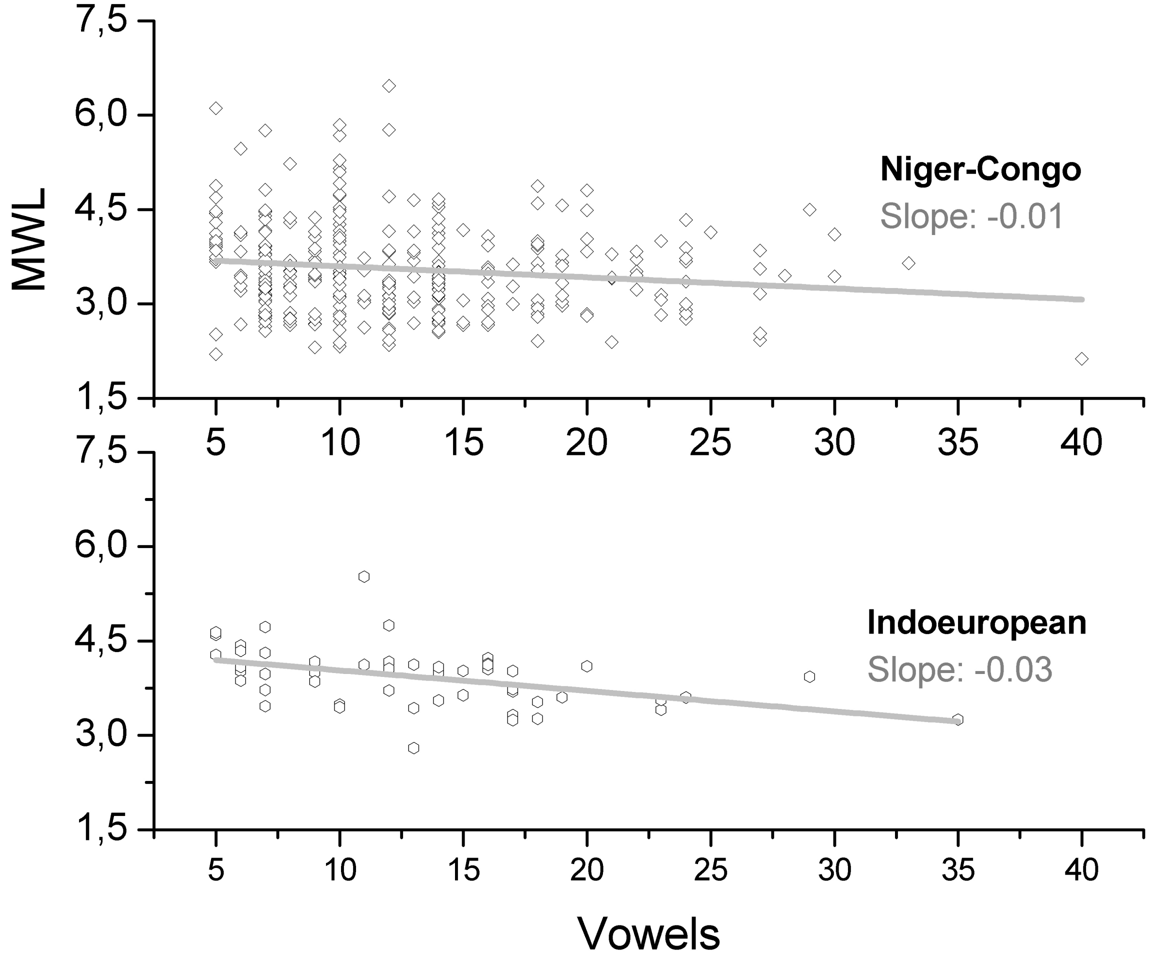

Fig. 12: Mean word length versus number of vowels for Niger-Congo and Indo-European families. Straight lines correspond to the best linear fits.

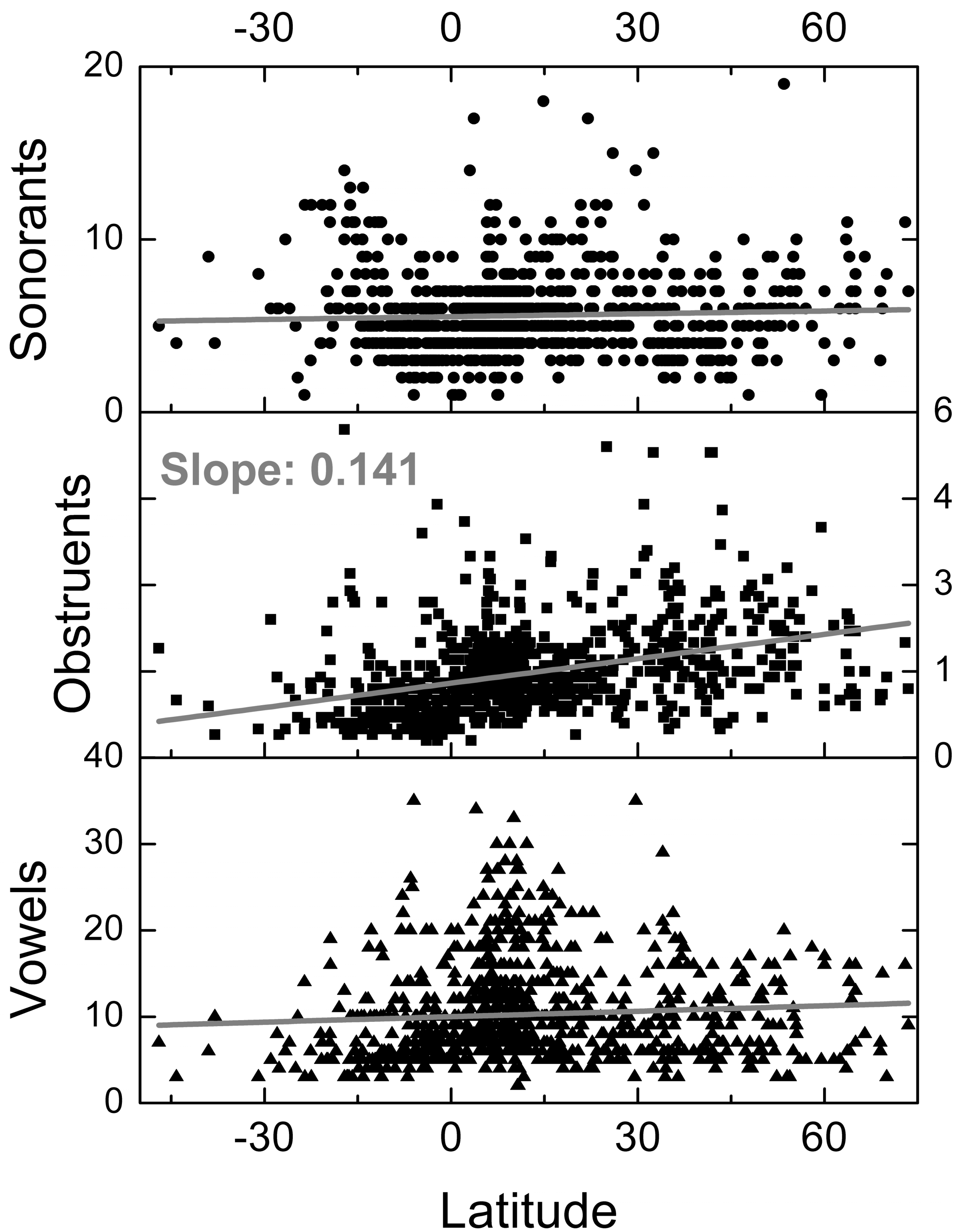

Fig. 13: Sonorants, obstruents and vowels repertory size versus latitude. Best linear fits are displayed for reference. Slope values for vowels and sonorants are not displayed because they were not significantly different from zero.

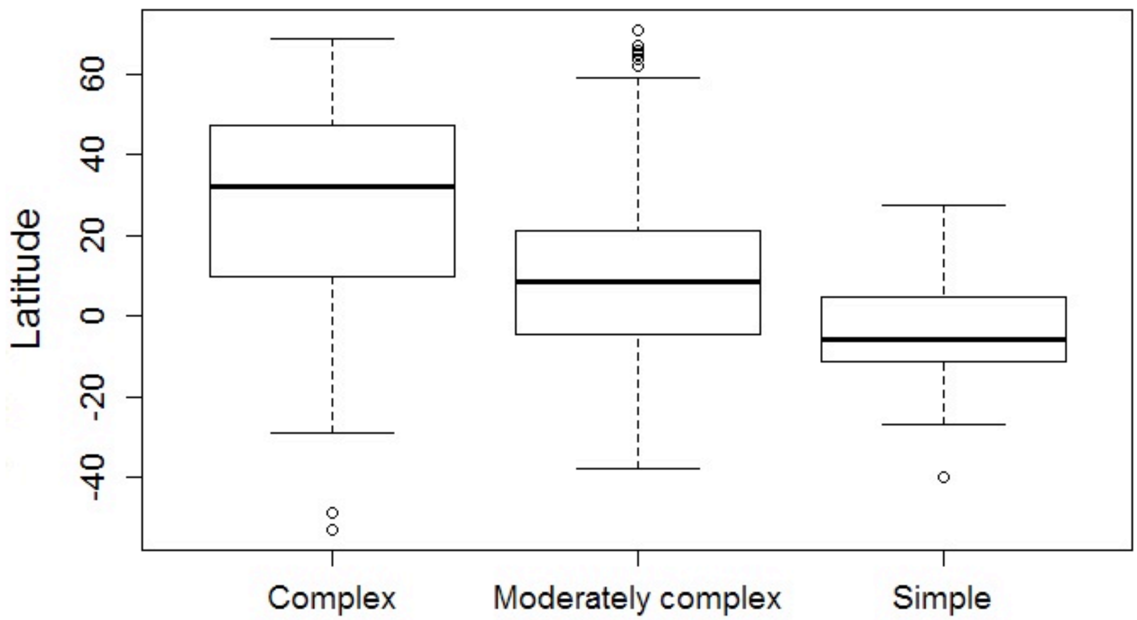

Fig. 14: Complexity of syllable structure versus latitude

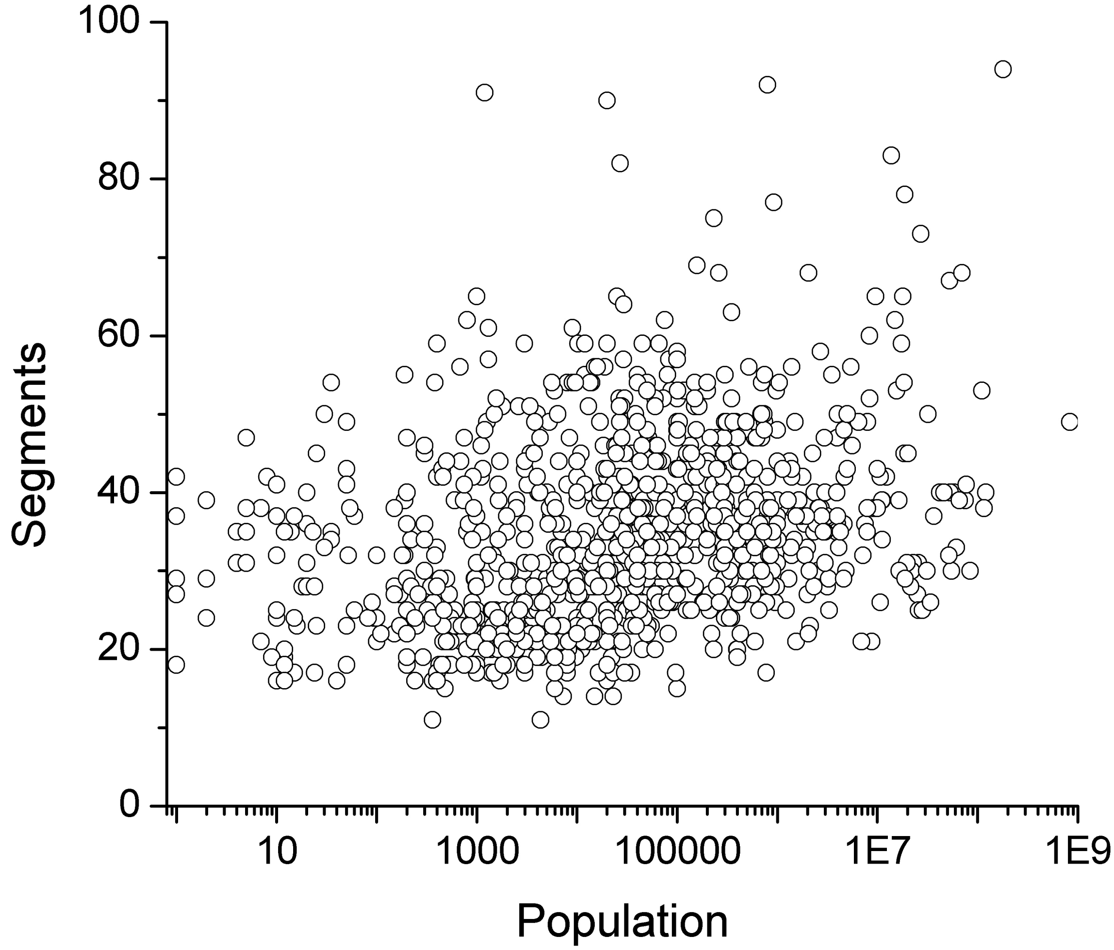

Fig. 15: Segment repertoire size vs population

NOTES

[1] The idea that languages compensate for complexity in different subsystems has been termed, among other things, the *compensation hypothesis* and the *negative correlation hypothesis* (Shosted 2006).

[2] The idea that all languages are equally complex has also been given several labels, e.g. the *equi-complexity hypothesis* (Sinnemäki 2011), the *invariance of language complexity* (Sampson 2009), or simply "ALEC" (Deutscher 2009).

[3] More recently, the number of segments has been used as a proxy for a language's "phonemic diversity" and has been correlated to factors such as increases in population size (Nichols 2009) and serial founder effects (Atkinson 2011). However, see the rebuttals in *Linguistic Typology* 15(2) and in Cysouw et al. 2012.

[4] Acquiring a phonology is a process of acquiring contrasts rather than inventories; thus changes in phoneme inventories are better understood in terms of contrastive features and phonological contexts (Kabak 2004).

[5] In this chapter the significance level was taken as $\alpha=0.05$.

[6] Kolmogorov complexity and its variants have also been invoked as an ideal measure for complexity in some of the linguistic complexity literature (e.g. Juola 1998, Benedetto et al. 2002, Dahl 2009), but in practical terms, with its many problems of interpretation and implementation, "it is quite useless as a measure of the complexity of natural systems" (Shalizi 2006:53). Furthermore, although a full argument against Kolmogorov complexity is beyond the scope of this chapter, let it suffice to say that it is not computable, meaning that one cannot devise a computer program to determine its value.

[7] As Maddieson points out, there is an obvious positive correlation between vowel quality-only distinctions and full vowel systems.

[8] Maddieson (2009:94) points out some problems with Shosted's method.

[9]Moreover, phonologists simply do not all agree on how to do phonemic analysis, as is apparent when comparing two or more segment inventories of the same language variety as described by different linguists (Moran 2012:chp. 2).

[10]As this study goes to press, new versions of WALS and Ethnologue (Lewis 2013) have recently been published.

[9] The results from all normality and log-normality tests conducted here hold as well when the three most extreme segment inventories are discarded. These are: !Xóõ (161), Ju|’hoan (141) and Khwe (119).

[10] We restricted our analysis to the six largest language families in PHOIBLE, which account for about 60% of the total languages.

[11] When no explicit scales are displayed, the area of the distributions are modified for expositive purposes.

[14]As a comparison, we include points for a binary productive segment inventory (see text) and a linear one (where each feature grants a single segment in the repertoire).

[12] The sample size is reported as between 450-650 languages.

[13] In fact, a detailed analysis of this phenomenon was carried out by Moran et al. (2012), who introduced mixed effect models that can distinguish the linear contribution of the stock and population to the number of segments. Their findings show that, at least under a linear model, the contribution of population, independently of language stock, is not significant.

References

Aitchison, Jean. (2000). *Language Change: Progress or Decay?* Cambridge, UK: Cambridge University Press.

Akmajian, Andrew, Demers, Richard A. and Hamish, Robert M. (1979). *Linguistics: An Introduction to Language and Communication*. Cambridge, Mass: MIT Press.

Ansaldo, Umberto and Nordhoff, Sebastian. (2009). 'Complexity and the Age of Languages', in Enoch O. Aboh and Norval Smith (eds.), *Complex Processes in New Languages*. Amsterdam: John Benjamins, 345–63.

Atkinson, Quentin D. (2011). 'Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion From Africa', *Science* 332: 346–59.

Benedetto, Dario, Caglioti, Emanuele and Loreto, Vittorio. (2002). 'Language Trees and Zipping', *Physical Review Letters* 88(4): 48702.

Bickel, Balthasar. (In press). 'Linguistic Diversity and Universals', in Nick Enfield, Paul Kockelman Enfield and Jack Sidnell (eds.), *The Cambridge Handbook of Linguistic Anthropology*. Cambridge: Cambridge University Press.

Campbell, George. (1991). *Compendium of the World's Languages*. London: Routledge.

Clements, G. N. (2009). 'The Role of Features in Phonological Inventories', in Eric Raimy and Charles E. Cairns (eds.), *Contemporary Views on Architecture and Representations in Phonology*. Cambridge, Mass: MIT Press, 19–68.

Crothers, John H., Lorentz, James P., Sherman, Donald A., and Vihman, Marilyn M. (1979). *Handbook of Phonological Data From a Sample of the World's Languages: A Report of the Stanford Phonology Archive*. Publisher?

Crystal, David. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Cysouw, Michael (2007). 'On the probability distribution of typological frequencies', in *Proceedings of the 10th and 11th Biennial conference on the mathematics of language*, 29–35.

Cysouw, Michael, Dediu, Dan and Moran, Steven. (2012). 'Still No Evidence for an Ancient Language Expansion From Africa', *Science* 335: 657–b.

Dahl, Östen. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins Publishing Company.

Dahl, Östen. (2008). ‘Grammatical Resources and Linguistic Complexity’, in Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 3–22.

Dahl, Östen. (2009). ‘Testing the Assumption of Complexity Invariance: the Case of Elfdalian and Swedish’, in Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 50–63.

Deutscher, Guy. (2009). Overall Complexity – a Wild Goose Chase?, in Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 243–251.

Dixon, Robert M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Dryer, Matthew S. and Haspelmath, Martin (eds.). (2011). *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/.

Ember, Carol R. and Ember, Melvin. (2007). ‘Climate, Econiche, and Sexuality: Influences on Sonority in Language’, *American Anthropologist* 109(1): 180–185.

Evans, Nicholas and Levinson, Stephen C. (2009). ‘The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science’, *Behavioral and Brain Sciences* 32(5): 429–448.

Fought, John G., Munroe, Robert L., Fought, Carmen R. and Good, Erin M. (2004). ‘Sonority and Climate in a World Sample of Languages: Findings and Prospects’, *Cross-Cultural Research* 38: 27–51.

Givón, T. and Shibatani, M. (2009). *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution*. Amsterdam: John Benjamins Publishing Company.

Gordon, Raymond G. (ed.). 2005. *Ethnologue: Languages of the World, Fifteenth Edition*. Dallas, TX: Summer Institute of Linguistics.

Greenberg, Joseph H., Ferguson, Charles A. and Moravcsi, Edith (eds.). (1978).

Universals of Human Language, Volume 2: Phonology. Stanford, CA: Stanford University Press.

Haspelmath, Martin, Dryer, Matthew S., Gil, David, and Comrie, Bernard. (2008). *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http: //wals.info/.

Haudricourt, André G. (1961). 'Richesse en phonèmes et richesse en locuteurs', *L'Homme* 1:5–10.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hay, Jennifer and Bauer, Laurie. (2007). 'Phoneme Inventory Size and Population Size', *Language* 83:388–400.

Hayes, Bruce. (2009). *Introductory Phonology*. Oxford: Blackwell.

Hockett, Charles F. (1955). *A Manual of Phonology*. Baltimore, MD: Waverly Press.

Hockett, Charles F. (1958). *A Course in Modern Linguistics*. Macmillan Company.

Hockett, Charles F. (1963). 'The Problem of Universals in Language', in Joseph H. Greenberg (ed.), *Universals of Language*. Cambridge, MA: MIT Press. 1–29.

Juola, Patrick. 1998. 'Measuring Linguistic Complexity: The Morphological Tier', *Journal of Quantitative Linguistics* 5(3): 206–213.

Justeson, John S., and Stephens, Laurence D. (1984). 'On the Relationship Between the Numbers of Vowels and Consonants in Phonological Systems', *Linguistics* 22(4): 531–545.

Kabak, B. (2004). 'Acquiring Phonology Is Not Acquiring Inventories but Contrasts: The Loss of Turkic and Korean Primary Long Vowels', *Linguistic Typology* 8(3): 351–368.

Kusters, Wouter. (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.

Kusters, Wouter. (2008). 'Complexity in Linguistic Theory, Language Learning and Language Change', in Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 3–22.

Kusters, Wouter, and Muysken, Pieter. (2001). 'The Complexities of Arguing About Complexity', *Linguistic Typology* 5:182–185.

Lewis, M. Paul (ed.). (2009). *Ethnologue: Languages of the World, Sixteenth Edition*. Dallas, TX: SIL International.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2013. *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

LINGUIST List. (2009). *Multitree: A Digital Library of Language Relationships*. Institute for Language Information and Technology (LINGUIST List), Eastern Michigan University. Ypsilanti, MI. http://multitree.org/.

Lupyan, Gary, and Dale, Rick. (2010). 'Language Structure Is Partly Determined by Social Structure', *PLoS ONE* 5(1): 1–10.

Maddieson, Ian. (1980). 'Phonological Generalizations from the UCLA Phonological Segment Inventory Database (UPSID)', in *UCLA Working Papers in Phonetics* 50: 57–68.

Maddieson, Ian. (1984). *Pattern of Sounds*. Cambridge, UK: Cambridge University Press.

Maddieson, Ian. (2006). 'Correlating Phonological Complexity: Data and Validation', *Linguistic Typology* 10(1): 106–123.

Maddieson, Ian. (2007). 'Issues of Phonological Complexity: Statistical Analysis of the Relationship Between Syllable Structures, Segment Inventories, and Tone Contrasts', in Maria-Josep Solé, Patrice Speeter Beddor and Manjari Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press.

Maddieson, Ian. (2009). 'Calculating Phonological Complexity', in Iona Chitoran, Christophe Coupe and Egidio Marsico (eds.), *Approaches to Phonological Complexity*. Berlin: Mouton De Gruyter, 85–109.

Maddieson, Ian. (2011a). 'Phonological Complexity in Linguistic Patterning', in *ICPhS XVII*, 28–34.

Maddieson, Ian. (2011b). 'Syllable Structure', in Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/feature/12A.

Maddieson, Ian, and Precoda, Kristin. (1990). 'Updating UPSID', in *UCLA Working Papers in Phonetics* 74: 104–111.

Martinet, André. (1955). 'Économie des changements phonétiques'. A. Francke.

Maslova, Elena. (2000). 'A Dynamic Approach to the Verification of Distributional Universals', *Linguistic Typology* 4: 307–333.

McCloy, Daniel, Moran, Steven, and Wright, Richard. (2013). 'Revisiting "The role of Features in Phonological Inventories"', Paper Presented at the CUNY Conference on the Feature in Phonology and Phonetics. New York, NY.

McMahon, April M. S. (1994). *Understanding language change*. Cambridge: Cambridge University Press.

McWhorter, John H. (2001). 'The World's Simplest Grammars Are Creole Grammars', *Linguistic Typology* 5: 125–166.

McWhorter, John H. (2007). *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. New York, NY: Oxford University Press.

Miestamo, Matti. (2006). 'On the Feasibility of Complexity Metrics', in K. Kerge and M.-M. Sepper (eds.), *FinEst Linguistics, Proceedings of the Annual Finnish and Estonian Conference of Linguistics*. Tallinn, Estonia, 11–26.

Miestamo, Matti. (2008). 'Grammatical complexity in a cross-linguistic perspective', in Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 23–42.

Miestamo, Matti. (2009). 'Implicational Hierarchies and Grammatical Complexity', in Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 80–97.

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

Moran, Steven. (2012). *Phonetics Information Base and Lexicon*. University of Washington, PhD thesis.

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

Morton, Eugene S. (1975). ‘Ecological sources of selection on avian sounds’, *American Naturalist* 109: 17–34.

Munroe, Robert. L., Munroe, R. H., and Winters, S. (1996). ‘Cross-cultural Correlates of the Consonant-vowel (CV) Syllable’, *Cross-Cultural Research* 30: 60–83.

Munroe, Robert L., and Silander, Megan. (1999). ‘Climate and the Consonant-Vowel (CV) Syllable: A Replication Within Language Families’, *Cross-Cultural Research* 33: 43–62.

Munroe, Robert L., Fought, John G., and Macaulay, Ronald K. S. (2009). ‘Warm Climates and Sonority Classes: Not Simply More Vowels and Fewer Consonants’, *Cross-Cultural Research* 43(2): 123–133.

Nettle, Daniel. (1995). ‘Segmental Inventory Size, Word Length, and Communicative Efficiency’, *Linguistics* 33: 359–367.

Nichols, Johanna. (2009). ‘Linguistic Complexity: A Comprehensive Definition and Survey’, in Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 64–79.

Pellegrino, François, Coupé, Christophe and Marsico, Egidio. (2011). A Cross-language Perspective on Speech Information Rate, *Language* 87(3): 539–558.

Perkins, Revere D. (1992). *Deixis, Grammar, and Culture*. Amsterdam: Benjamins.

Ruhlen, Merritt. (1975). *A Guide to the Languages of the World*. Language Universals Project, Stanford University.

Wilkins, M. R., Seddon, N., & Safran, R. J. (2012). Evolutionary divergence in acoustic signals: causes and consequences. Trends in ecology & evolution. Is this entry complete?

Sampson, Geoffrey. (2009). ‘A Linguistic Axiom Challenged’, in Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 1–18.

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

Sapir, Edward. (1912). ‘Language and Environment’, *American Anthropologist* 14: 226–242.

Shalizi, Cosma Rohilla. (2006). ‘Methods and Techniques of Complex Systems Science: An Overview’, *Complex Systems Science in Biomedicine* 33–114.

Shosted, Ryan K. (2006). ‘Correlating Complexity: A Typological Approach’, *Linguistic Typology* 10: 1–40.

Sinnemäki, Kaius. (2011). *Language Universals and Linguistic Complexity: Three Case Studies in Core Argument Marking*. University of Helsinki, PhD thesis.

Trubetzkoy, Nikolai. (1939). ‘Grundzüge der Phonologie’, *Travaux du cercle linguistique de Prague* 7.

Trudgill, Peter. (2001). ‘Contact and Simplification: Historical Baggage and Directionality in Linguistic Change’, *Linguistic Typology* 5: 371–374.

Trudgill, Peter. (2004). ‘Linguistic and Social Typology: The Austronesian Migrations and Phoneme Inventories’, *Linguistic Typology* 8(3): 305–320.

Wichmann, Søren, André Müller, Viveka Velupillai, Annkathrin Wett, Cecil H. Brown, Zarina Molochieva, Julia Bishoffberger, Eric W. Holman, Sebastian Sauppe, Pamela Brown, Dik Bakker, Johann-Mattis List, Dmitry Egorov, Oleg Belyaev, Matthias Urban, Harald Hammarström, Agustina Carrizo, Robert Mailhammer, Helen Geyer, David Beck, Evgenia Korovina, Pattie Epps, Pilar Valenzuela & Anthony Grant. (2012). *The ASJP Database (version 15)*. http://email.eva.mpg.de/~wichmann/listss15.zip.