Consonant co-occurrence classes and the feature-economy principle*

Dmitry Nikolaev
Stockholm University

Eitan Grossman
Hebrew University of Jerusalem

The feature-economy principle is one of the key theoretical notions which have been postulated to account for the structure of phoneme inventories in the world's languages. In this paper, we test the explanatory power of this principle by conducting a study of the co-occurrence of consonant segments in phonological inventories, based on a sample of 2761 languages. We show that the feature-economy principle is able to account for many important patterns in the structure of the world's phonological inventories; however, there are particular classes of sounds, such as what we term the 'basic consonant inventory' (the core cluster of segments found in the majority of the world's languages), as well as several more peripheral clusters whose organisation follows different principles.

1 Introduction

A central question in phonological typology (and in phonology more generally) is whether there are principles that govern the size, structure and constituent parts of phonological inventories, and if so, what they are. Research in recent decades has proposed numerous factors, often extralinguistic, that predict the composition of phonological inventories. Such proposed factors include demography (Pericliev 2004, Hay & Bauer 2007, Donohue & Nichols 2011, Moran *et al*. 2012, Greenhill 2014), environment and climate (Everett 2013, Everett *et al*. 2015, 2016), genetics (Dediu & Ladd 2007, Creanza *et al*. 2015), geography and population movements (Atkinson 2011), culture (Labov *et al*. 2013) and anatomy (Dediu *et al*. 2017).

Structural, i.e. language-internal or systemic, factors, include the 'size predicts' generalisation: the number of segments in an inventory largely determines its content, such that small systems recruit few (and basic) dimensions, while larger systems entail additional (and secondary) dimensions (Lindblom & Maddieson 1988). In this paper, we focus on another structural factor, namely feature economy. The FEATURE-ECONOMY PRINCIPLE is one of the mainstays of contemporary discussions of phonological segment inventories in the languages of the world. Two different, albeit largely congruent, formulations of this principle were proposed by Lindblom & Maddieson ('small paradigms tend to exhibit 'unmarked' phonetics whereas large systems have 'marked' phonetics'; 1988: 70) and Clements ('languages tend to maximise the ratio of sounds over features'; 2003: 287). This idea goes back at least to early work in structuralist phonology, including Trubetzkoy (1939), Martinet (1952) and Hockett (1955), who were interested in the extent to which phonological inventories are symmetrical with respect to features, or, in other words, how much 'mileage' phonological inventories get out of individual features; see an overview of early developments of this concept in Clements (2003). Similar conclusions were later reached using different formulations and/or different datasets (Marsico *et al.* 2004, Coupé *et al.* 2009, Mackie & Mielke 2011, Moran 2012, Dunbar & Dupoux 2016), and theoretical and experimental investigations of feature economy have become a major line of phonological research: see Pater (2012), Verhoef *et al.* (2016) and Seinhorst (2017).

The aim of the current paper is not to propose another explanation or interpretation of the feature-economy principle, but to take a step back in order to reassess how well it actually fits the structure of phonological segment inventories of the world's languages, focusing on consonants.

Clements (2003: 288–289) hypothesises that the feature-economy principle can only be constrained by functional factors: 'avoided feature combinations can be shown to be inefficient from the point of view of speech communication. That is, their articulation is relatively complex, or their auditory attributes are not distinct enough from those of some other sound in the system'.[1] Marsico *et al.* (2004) and Coupé *et al.* (2009) attempted to quantify the amount of residual variance left unexplained by the feature-economy principle by computing the REDUNDANCY FACTORS and the COHESION of phonological inventories in UPSID (Maddieson & Precoda 1992). Our aim is to provide an exploratory assessment of the structure of this residual variance. Our premise is that if Clements' assessment of the explanatory power of the feature-economy principle were correct, we would be able to explain the majority of exceptions to the feature-economy principle by invoking perception and/or

---

[1] Cf. also: 'The majority of exceptions to feature economy are broadly similar across languages: the great majority of languages exclude nasalised fricatives, voiced ejectives, voiceless vowels and so forth. Only in the case of feature combinations that are only marginally dysfunctional (such as voiced stops or nasalised vowels) do we find much cross-linguistic variation' (Clements 2003: 298).

production factors, and that the variance left unexplained after that would consist of random noise due to the probabilistic nature of sound change. Clements' hypothesis therefore would be falsified (in a non-statistical, observationist way) if we were to discover that there are principles governing inventory structures that do not stem from the abovementioned types of functional factors.

The core of our approach to testing this hypothesis is the notion of a CO-OCCURRENCE CLASS. Co-occurrence classes are groups of sounds that tend to be found together in inventories. We provide a fully algorithmic definition of this notion in §2, but at this stage we would like to explore its implications. We regard co-occurrence classes as a particularly powerful method of phonological analysis, since nearly all principles governing the structure of phonological inventories are plausibly reflected in the structure of these classes. We give two examples in (1).

(1) a. Environment-driven asymmetries between sounds, such as the proposed connection between ejectives and high altitudes (Everett 2013), will lead to sounds having restricted geographical distributions forming tight co-occurrence clusters.
　 b. Markedness asymmetries between sounds[2] should lead to similar outcomes. Unmarked sounds would be lumped together as the dominant class, with more marked segments forming a diffuse and unstructured cloud.

Our primary interest is how the feature-economy principle is reflected in the structure of co-occurrence classes. In order to investigate this, we propose to first reformulate the principle in a more structural fashion. Building on the notion of sound-inventory symmetry explored by Dunbar & Dupoux (2016), we operationalise the feature-economy principle by interpreting it as largely synonymous with the LAYERING PRINCIPLE: new classes of sounds arise by virtue of adding new features to already existing combinations. An empirical confirmation of this is found in Moran (2012: 248) with respect to vowels, such that 'once languages expand their inventories beyond cardinal vowels, they tend to do so by either nasalization or lengthening, and to a lesser extent by adding diphthongs to the inventory'.

This formulation has the advantage of providing a simple way to articulate a structural prediction: if we investigate empirical co-occurrence classes, we should see that they are progressively defined by a succession of additional features. Thus we should see both large classes dominated by basic distinctions (place, manner and VOT) and smaller classes in which these distinctions are augmented by different additional articulations. We call classes that respect the feature-economy/layering principle CONFORMANT CLASSES. Most importantly, the feature-economy/layering principle predicts that

---

[2] With markedness defined on the basis of the dual criterion of 'emergence of the unmarked' in the contexts of neutralisation and 'preservation of the marked' in various assimilation contexts (Rice 2007).

certain constellations should not exist. In particular, it prohibits CROSS-LAYER CONNECTIONS (the close patterning of segments with different numbers of features turned on) and CROSS-FEATURE CONNECTIONS (the close patterning of segments with different privative features turned on). That is, if there exists a class of palatalised segments, we do not expect some of the members of this class to pattern with either labialised segments (which would be a cross-feature connection) or plain segments (which would be a cross-layer connection), as this would imply that languages do not exhaust the usefulness of the [+palatalised] feature. We call classes that do not respect the feature-economy/layering principle NON-CONFORMANT classes. A good example of a conformant class is long voiced stops: /bː dː gː/. They are distinguished by a single distinctive feature value [+long], and exhaust the possible combinations of VOT and manner values for all places of articulation. Short voiced stops /b d g/, however, do not form a conformant class in our data. Instead they are embedded inside a large complex class, the 'first extension set'. We discuss conformant and non-conformant classes in §4.

It is important to stress that this type of analysis is based on bidirectional dependences (the presence of segment A is probabilistically dependent on the presence of segment B, and *vice versa*), not on unidirectional implicational universals (languages with segment A tend to also have segment B, but segment B is also frequently found without segment A). For example, languages with /pʲ/ have a very strong tendency to have /p/ as well. Nevertheless, the bidirectional co-occurrence dependence between these segments is very low: the absence of /pʲ/ is a very weak indicator of the absence of /p/. On the other hand, the absence of /pʲ/ is a strong indicator of the absence of /bʲ/, and *vice versa*.

This paper aims to make the following contributions. First, we propose a statistical method for identifying co-occurrence classes of sounds in the world's languages. Second, using this method, we empirically identify several co-occurrence classes worthy of attention in themselves, one of them being the BASIC CONSONANT INVENTORY. Third, using the structure of the co-occurrence classes identified by this new method, we show the limits of the applicability of the feature-economy principle in its various formulations.

The paper is organised as follows. In §2, the method used to derive co-occurrence classes is described, together with the dataset it is applied to. In §3, the resulting classification of the major types of segments in the languages of the world is presented, and a brief overview of the classes is given. §4 is devoted to the consequences of the structure of the co-occurrence classes for the feature-economy principle, and §5 presents conclusions.

2 Deriving co-occurrence classes

2.1 Dataset

Our data comes from the PHOIBLE database, the most comprehensive openly accessible cross-linguistic repository of phonological inventories

and associated data and metadata. Version 2.0 (Moran & McCloy 2019) includes 3020 phonological inventories, with 3183 segment types found in 2186 distinct languages. One doculect was randomly selected for each language;[3] the resulting dataset contains information about the phonological inventories of 2761 languages. Only segments occurring 30 or more times in the dataset were included in the analysis ($N = 193$), giving a binary 2761×193 matrix with rows corresponding to inventories (with 1's for segments present in the inventory and 0's for segments absent from the inventory) and columns corresponding to segment distributions. The picture presented by the rows of this matrix is incomplete (rare segments are not represented), but this is not an issue, as the analysis is restricted to the columns. In addition to providing more robust inference (frequencies of rare segments are vulnerable to sampling noise), restricting the analysis to relatively frequent segments has empirical support, as the frequency distribution of segments in worldwide samples such as PHOIBLE shows a strong positive skew: it declines exponentially, such that the tenth most frequent segment, /ŋ/, is found in 63% of the languages in the sample, the twentieth most frequent, /ʃ/, is found in 35% and the fortieth most frequent, /ḏ/, in only 21%. Most of PHOIBLE is in fact a very long tail of segments, each of which is documented only in a single language.

There are pitfalls inherent in any cross-linguistic study of phonological segments and segment inventories. The first is the nature of the objects of study. As has been pointed out in earlier critiques of UPSID and PHOIBLE (Lass 1984, Simpson 1999, Vaux 2009, Kiparsky 2018), phonemes are not ideally suitable objects for cross-linguistic study. The primary reason is that phonemic solutions are not unique, as shown long ago by Chao (1934), and suggested even earlier by Sapir (1925). Moreover, in many approaches, phonemes are defined structurally or negatively, i.e. in terms of their oppositions within a system. This relates to the problem of normalisation, especially in descriptions. For example, it is common for grammar writers to note a phoneme /p/, which turns out to be realised as [pʰ] in most environments. However, grammar writers typically do provide usably detailed information about the distribution of allophones. The decision that has informed the collection of data up to now has been to record the allophone with the least limited distribution, rather than the normalised variant. So for example, the /p/ phoneme of Sanandaj Neo-Aramaic (Khan 2009) is recorded in the database as /pʰ/, since it has the least limited distribution of the various allophones. See Moran (2012, 2019) for additional responses to the criticism of segment-based typology.

We would like to address another potential problem with our data, one which is shared with any large cross-linguistic sample and does not have to do with the quality of the data *per se*. The issue, which has been called

---

[3] A doculect ('documented lect') is a documentary record attesting to the existence of some language variety. For additional discussion, see Cysouw & Good (2013).

'Galton's Problem', is that individual languages may not be independent data points. This is due to two main facts. The first is that languages in the sample may have developed from a common ancestor, which means that for a particular feature of interest, two languages, say French and Italian, may have retained this property from Proto-Romance. The second fact is that languages almost universally undergo change as a result of language contact. In particular, languages that are in geographical proximity tend to show the influence of their neighbours. This is often visible in case studies of specific contact situations, but it may also scale up to continent-sized macro-areas, at least with respect to some features. Finally, common descent and geographical proximity can interact: overall, languages are often geographically close to languages of the same family, so it becomes difficult to tease apart the respective contributions of inheritance on the one hand and language contact on the other.

Various methods have been proposed to deal with this problem. One is balanced or stratified sampling, which aims to create samples of languages that are as geographically and genealogically independent as possible. However, since some features may characterise entire macro-areas (Dryer 1989), the *a priori* possibility of constructing a sample of truly independent languages is small. Moreover, carefully balanced samples – if the unit being sampled is individual languages – cannot reach large sizes, so the quantity of data in such samples is by necessity restricted (Piantadosi & Gibson 2014).

Another approach that has been gaining currency is the use of a family of methods that directly test the contribution of genealogy and area by modelling them as predictors. Such approaches, known as 'Distributional Typology' (Bickel & Nichols 2006, Bickel 2013, 2015, 2017), handle large and dense samples well, and in fact require more and denser data than the classical balanced sampling approach. An example illustrating these approaches involves two competing theories regarding the presence of case marking on core arguments in verbal clauses (Bickel 2015). One theory claims that case marking is more prevalent in languages with verb-final basic order, ultimately due to processing constraints (Hawkins 2004); the other theory claims that case marking is especially prevalent in Eurasian languages due to language contact (Jakobson 1931). Bickel (2015) used the 'family bias' method, which investigates diachronic biases *within* families, to test the extent to which both verb-finality and areality predict case marking, and found that the two effects were independent: verb-finality did indeed predict case marking to some extent, but the effect size was much larger in Eurasia than in other parts of the world, suggesting historical contact events as the source of the areal patterning. Similar studies have been conducted in other domains, such as differential object marking (Sinnemäki 2014).

In the present study, we do not attempt to estimate the relative contributions of genealogy, areality and other functional and structural factors. Instead, we point out that areality and common descent are themselves proxies for a potentially large set of factors that shape diachronic pathways

of change. Ultimately, if feature economy – or any other proposed principle – exerts an effect on phonological inventories, it can only do so by influencing the probability of a language to (i) innovate a property A, (ii) lose a property A or (iii) retain a property A. In other words, feature economy is a good candidate for what Bickel (2017: 42) calls 'functional triggers' (or factors), i.e. factors that 'systematically bias the way linguistic structures evolve', the defining property of which 'is that they affect transition probabilities universally, independent of concrete historical events'.

As such, the essentially static picture that we present in this exploratory study is in all likelihood the result of the highly complex interaction of a diverse set of factors that play out in diachrony. Since phonetic variation and sound change are so rampant in human language, we can conceptualise the feature-economy principle as a factor – whether ultimately rooted in the human language faculty or in domain-general cognitive or articulatory and perceptual biases – that works to keep certain sounds grouped together. We would expect this principle to promote the retention of individual sounds in some cases and their loss in others, and perhaps even their borrowing when contact situations permit it.[4] While we do not attempt to directly test the contributions of these factors to the results of our analysis, we do throughout the paper point out cases in which areal factors are suggested by the data.

2.2 The feature set

Clements' original formulation of the feature-economy principle was based on the principle of the economy of representation, 'according to which features are specified in a given language only to the extent that they are needed in order to express generalizations about the phonological system' (2001: 72). This makes it impossible to give a featural description of a segment from an inventory without analysing this inventory as a whole, which would be very time-consuming. Moreover, it aggravates the issue of unstable phonemicisation discussed in the previous section: misleading phonemicisation of a single segment can potentially lead to an incorrect feature specification for the whole inventory. In order to safeguard against this possibility, we adopted a shallow but robust approach to feature specification: all segments from all inventories were analysed in terms of IPA features (*IPA handbook* 1999). The IPA feature set for consonants is articulatory in nature, and is oriented towards describing inventories as assemblages of individual segments. This is different from the feature sets prevalent in the theoretical literature, where the focus is on succinct description of phonological rules or constraints (Hayes 2009).

---

[4] Recent studies have shown that the borrowing of phonological segments is widespread in languages (Grossman *et al.* 2020). Moreover, both cross-linguistically rare and frequent sounds are borrowed. Interestingly, at a featural level, the presence of some features in an inventory, e.g. [labiodental], facilitates the borrowing of phonological segments with this feature, while in other cases an inhibitory effect is observed (Eisen 2019). The point here is that feature economy may have an impact even on contact-induced change.

However, many languages in the sample have not been analysed from the point of view of their rule/constraint systems, and the focus on individual segments seems warranted.

2.3 Methods

2.3.1 *Distance measure*. The focus on bidirectional connections between segment distributions, which we represent as binary vectors with 1's for languages that have a particular segment and 0's for languages that do not have it, makes Pearson's correlation coefficient a natural measure of the strength of the association. The value of the coefficient ranges from −1 (the presence of one segment predicts the absence of another with absolute certainty, and *vice versa*, i.e. they are never found together) to +1 (the presence of one segment predicts the presence of another with absolute certainty, and *vice versa*, i.e. they are always found together). The middle value of 0 indicates the absence of any association (i.e. for any segment of the two, we are as likely to find an inventory that contains only this segment as to find an inventory containing both). Cases of an implicational universal – segment A strongly predicts segment B, but not *vice versa* – give rise to small positive values of the coefficient (which, unfortunately, makes them indistinguishable from chance relationships between segment distributions).

The next step is to divide segments into co-occurrence classes based on their pairwise distributional similarities. Given the high number of segments, it is hardly feasible to do this by manually investigating several thousand correlation coefficients. Methods for dimensionality reduction and clustering offer a convenient shortcut, but in order to use them we need to convert correlational similarity values to a non-negative dissimilarity (more precisely, distance) measure.

The simple step of adding 1 to all the coefficients and subtracting the result from 2 will produce a distance measure satisfying non-negativity, symmetry and zero distance from the object to itself. However, this measure does not obey triangle inequality (a rule that the distance between points A and C cannot be larger than the sum of distances from A to B and from B to C for any point B), which may lead to nonsensical results. Fortunately, there are several methods for converting Pearson's correlation coefficient to a proper distance measure (van Dongen & Enright 2012). One of them is the simple arccosine transformation $D = \arccos(\mathrm{Cor}(x, y))$, and we therefore use it in our analysis.[5]

2.3.2 *Dimensionality reduction and clustering*. Every segment in our dataset is represented as a binary vector of length 2761. The data are very high-dimensional, which makes it impossible to directly visualise them or even to use Principle Components Analysis, which prohibits the

---

[5] It must be noted that such transformations are usually non-linear, and the arccosine function is no exception: differences between neighbouring values close to −1 or 1 are somewhat magnified relative to distances between values close to 0. This can potentially influence the analysis to some extent.

number of variables from being larger than the number of data points (2761 *vs*. 193, in our case).

For the analysis of our dataset, we used 'Uniform Manifold Approximation and Projection' (UMAP; McInnes *et al*. 2018), a novel method for dimensionality reduction which has been adopted in the machine-learning and bioinformatics communities, thanks to its ability to uncover hidden structure in high-dimensional data, both at the level of immediate neighbourhoods of individual points and at the level of larger clusters, if these can be detected.[6]

UMAP takes as input a table with the original data points or a matrix of pairwise distances between them (in our case, arccosine-transformed Pearson's correlation coefficients) and chooses a lower-dimensional representation of the point assemblage in such a way that neighbourhoods of individual points and relative positions of larger clusters are preserved, as far as possible. UMAP has fairly involved mathematical foundations, but in practice its operation boils down to several transparent steps:

(i) A number of neighbouring points, k, is selected, guided by the researcher's guess as to how many points constitute a valid neighbourhood. In this case, this corresponds to the intuition as to how many points constitute a minimal cluster. In the extreme case, k can be taken to be equal to 1, giving rise to two-point clusters. Larger values essentially prohibit small clusters, and make the macro-structure more homogeneous and 'lumpy' (we can also force the algorithm to never lump individual points together by setting a non-negative value for the min_dist parameter, which dictates how far individual points should be from each other in the resulting embedding; this parameter, however, has a weaker influence on the overall structure). $k = 3$ and min_dist = 1 were used for the analysis.

(ii) A multidimensional ball is drawn around each data point, such that it includes its k neighbours. A neighbour graph is then created, where each point is connected to its k neighbours.

(iii) The edges in this graph have varying weights, depending on how far from the target point its neighbours are. Mathematically, the neighbourhood of the points are regarded as fuzzy sets (i.e. sets with probabilistic membership), and the UMAP algorithm defines a way to combine them where the edge lengths are regarded as probabilities of the fact that an edge really exists (if two nodes are not immediate neighbours, the probability that an edge between them exists is set to 0).

(iv) An arrangement of points in a lower-dimensional space is selected in such a way that, when we construct the same union of fuzzy sets based on Euclidean distances, the cross entropy between the probability distributions of the existence of edges in this lower-dimensional neighbour graph and the original high-dimensional neighbour-graph is minimised.

For further details of the mathematical foundations of the method and the implementation details, readers are referred to McInnes *et al*. (2018).

---

[6] For other uses of dimensionality reduction in phonological typology, see Moran (2012), Moran *et al*. (2013) and Macklin-Cordes *et al*. (2016).

UMAP does not propose a ready-made clustering of the points. The latter can be created either (i) on an ad hoc basis, inspecting the scatter plot of the points if the representation is two-dimensional (the most commonly used trade-off between interpretability and the possibility of preserving complex configurations) or (ii) by using a clustering algorithm.

To separate points into clusters, we used the density-based clustering algorithm DBSCAN (Ester *et al.* 1996). It divides the space (in our case, the space of segment distributions interpreted as points on the two-dimensional plane) into regions of high density (those around CORE POINTS of clusters) surrounded by PERIPHERAL POINTS (those reachable from core points) and NOISE POINTS (those not reachable from core points and not belonging to any cluster).

To summarise, in order to derive co-occurrence classes of consonants in the world's languages, we constructed an analytical pipeline from four well-defined mathematical procedures (Pearson's correlation coefficient → arccosine transform → UMAP dimensionality reduction → DBSCAN clustering). The pipeline separates data points into groups without any prior knowledge about their nature or properties. It should be noted that we have a margin of freedom in our choice of the parameters for these procedures. For UMAP, this is the number of nearest neighbours and minimal distance between resulting points; for DBSCAN, it is the minimal number of neighbouring points (again set at 3) and the reachability distance (set at 1.8). The choice of the distance measure for the initial data points is even more important. This freedom, however, is not nearly enough to create an easily interpretable structure from chaos in the initial data: in any case, the pipeline will cluster the points that are closest to each other, and separate them from points lying further afield.[7]

3 Co-occurrence classes in the languages of the world

The distribution of the 193 common segments into co-occurrence classes is shown in Fig. 1, with polygons highlighting the boundaries of clusters identified by DBSCAN. This picture is not particularly informative, since most of the segments are lumped together into a large cluster of relatively frequent segments. We re-ran the analysis on a subset of data consisting only of these segments. The results are presented in Fig. 2, and are discussed in §3.2. First, however, we survey the divergent classes of segments, i.e. those that fall outside the large cluster.

---

[7] In the context of regression modelling or fitting classifiers, it is considered good practice to perform a hyperparameter search to ascertain which values yield best results (Feurer & Hutter 2019). In the context of an exploratory study, however, there is no metric we can optimise against, and the only test we can perform is to check whether the resulting picture is interpretable.

Figure 1
Major consonant co-occurrence classes in the world's languages.

Figure 2
Groups inside the large cluster in Fig. 1.

3.1 Divergent classes

In this section, we survey divergent classes of segments not included in the major cluster in Fig. 1. These classes can be divided into various groups.

3.1.1 *Palatalised segments.* Two classes of palatalised segments were picked out by the pipeline: coronal /tʲ dʲ lʲ nʲ sʲ/ and non-coronal /pʲ bʲ mʲ gʲ kʲ/, together with /rʲ/.

The marginal position of the rhotic with respect to other coronals is relatively easy to account for: it has been observed that palatalised liquids do not form a recurrent pair: /lʲ/ is more typologically frequent and more diachronically stable than /rʲ/ (Kochetov 2005). Similarly, Hall (2000: 1) claims that /rʲ/ is 'far more marked' than other palatalised coronals. Indeed, /rʲ/ ($N = 31$) is more than twice as rare as /lʲ/ ($N = 66$) in our data, as shown in (2).

(2) rʲ hʲ sʲ bʲ nʲ dʲ pʲ mʲ lʲ gʲ tʲ kʲ
    31 34 47 51 56 59 60 62 66 74 87 93

The general coronal *vs.* the rest split, however, is less expected. Palatalised labial segments are considered to be rarer than both palatalised coronals and dorsals, and both coronals and dorsals are evidently more likely than labials to undergo 'full palatalisation', which in most cases means affrication (Bateman 2011).

Inspection of the plot of the correlations of distributions of palatalised segments in Fig. 3, however, shows that there is indeed a noticeable affinity between palatalised dorsals and bilabials on the one hand and between palatalised bilabials and /rʲ/ on the other, suggesting a centre *vs.* periphery articulatory split, with coronals serving as a default locus for palatalisation; this should be explored further.

3.1.2 *Phonologically long consonants.* These are split into the voiced stops, /bː dː gː/, and everything else. As shown in (3), they are largely in the same frequency range as the palatalised segments.

(3) ʧː wː ŋː ɲː dː fː ʃː jː rː gː pː tː bː sː lː kː nː mː
    35 35 35 38 40 41 41 47 47 51 56 56 58 69 77 78 86 89

It is tempting to assume that this split can be explained by the well-known observation that it is hard to maintain voicing in long stops, which makes long voiced plosives susceptible to shortening, loss of voicing or implosion (Ohala 1983), but the fact that /bː/ is more frequent than /pː/ makes this explanation rather weak.

Bilabial long voiced stops are indeed easier to produce, due to the larger oral cavity available for maintaining higher subglottal pressure. Nevertheless, this does not explain why /bː/ does not pattern with /pː tː/.
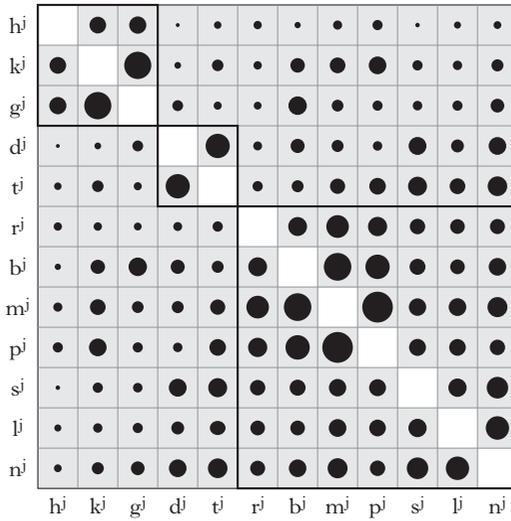
Figure 3

The plot of the correlations between distributions of palatalised segments.
The sizes of the circles are proportional to the values of correlation
coefficients. Thick frames denote groups of segments with high pairwise
mutual correlations established using Ward's clustering algorithm.

Moreover, /gː/ is arguably the most disfavoured voiced long plosive, char-
acterised by the smallest oral cavity, and yet it is more prevalent than /dː/.

A simpler, albeit clearly incomplete, explanation in terms of the layering
principle presents itself: when adding voiced long plosives, languages tend
to exhaust this segment class despite differences in the articulatory
difficulty of different segments. In other words, languages are more
likely to acquire /bː dː gː/ as a group, rather than the individual pairs
/bː pː/, /dː tː/ and /gː kː/.

The strength of connections between segments on the same level of the
layering hierarchy can be illustrated by the fact that there are more lan-
guages with /bː/ but no /pː/ ($N = 29$) than there are with /bː/ but no /dː/
($N = 19$), even though /pː/ is nearly 1.5 times as frequent as /dː/.

3.1.3 *Labialised segments.*   A further divergent cluster is formed by a
subset of labialised segments, together with the palatalised glottal fricative,
as shown in (4).

(4) ʃʷ hʲ tʷ sʷ pʷ fʷ bʷ mʷ hʷ ŋʷ gʷ kʷ
     31 34 38 40 41 43 54 57 67 110 182 342

It is notable that this cluster is coherent, despite substantial frequency
disparities between its elements (/kʷ/ is more than ten times as frequent
as /ʃʷ/, and nine times as frequent as /tʷ/, for example). The visual

structure of this cluster in Fig. 1, further supported by the correlation coefficients, suggests four subgroups, mostly corresponding to frequency bands and exactly corresponding to place of articulation: /hʲ hʷ/, /ŋʷ gʷ kʷ/, /ʃʷ tʷ sʷ/ and /pʷ fʷ bʷ mʷ/.

Three tendencies are discernible. First, different regions of articulation are more productive for labialised stops and fricatives respectively. Second, the use of particular values of the place feature for labialised segments (especially the postvelar articulations, cf. §3.2) is dependent on whether there are other segments with the same value of this feature, rather on whether there are other labialised segments in general. Third, the glottal fricative /h/ either (i) lacks any additional articulation or (ii) tends to have both palatalised and labialised versions.

3.1.4 *Voiceless approximants, liquids and nasals.* The small group in (5) consists of low-frequency and, on average, rather tightly correlated sounds, except for /r̥ ʍ/, which rarely appear together ($r = 0.13$). /m̥ n̥/, on the contrary, are nearly inseparable ($r = 0.847$).

(5) r̥   ʍ   n̥   m̥   l̥
     34 37 65 69 75

3.1.5 *Other clusters.* The remaining groups consist of truly marginal sounds:

(i) Tense and weakly articulated stops, /p̬ p̥ t̬ t̪ t̠ t̥ k̬ k̥/, are very rarely described outside Australia and are almost exclusively found together.[8]

(ii) Phonologically apical stops and liquids, /r̠ ɻ n̠ l̠ t̠/, are also predominantly found in Australia, whose languages are famous for their extended range of place oppositions (Butcher 2006).

(iii) The voiceless prenasalised stops /mp nt ŋk/ are found in Africa, Eurasia and South America, but are rare ($N = 36$, 42, 41 respectively), and are twice as likely as not to be found in inventories that also have corresponding voiced prenasalised stops; see the contingency tables in Table I.

		mb	
		−	+
mp	−	2467	258
	+	10	26

		nd	
		−	+
nt	−	2473	258
	+	10	26

		ŋg	
		−	+
ŋk	−	2485	235
	+	12	29

Table I

Contingency tables for the co-occurrence of voiced and voiceless prenasalised stops. Cells contain counts of inventories with (+) or without (−) a particular segment.

[8] See Austin (1988), Stoakes *et al.* (2007) and Fletcher & Butcher (2012) on the properties of these segments.

3.2 The large cluster

Boundaries between groups inside the large cluster in Fig. 2 identified by DBSCAN are more uncertain, as is evident from the large number of noise points scattered around the figure. In many cases, however, the visual structure of the plot makes it clear to which cluster the unattached points 'really' belong, and we will group them accordingly.

3.2.1 *The basic consonant inventory.* The cluster consisting of /p t k n l r/ and, although it is just outside the core group, /m/ (the most common segment in the dataset) can be considered a serious contender for the title 'basic consonant inventory' (a group of consonantal sounds that we expect to find in a randomly picked language).

The question of which set of segments can be considered basic one has been raised by many scholars, recent notable contributions including Hyman (2008) and Gordon (2016). The main stumbling blocks are the facts that (i) there are no really universal segments (even the most frequent segment, /m/, is found in only 96% of the languages), and (ii) raw frequency counts make it impossible to draw the boundary between quasi-universal segments and merely very frequent ones. Should the boundary be drawn between, say /s/ (frequency 72%, as shown in Fig. 4) and /ŋ/ (63%), or between /g/ (56%) and /d/ (49%)?

The approach based on the identification of co-occurrence classes seems to provide a way of overcoming this impasse. If there is a single co-
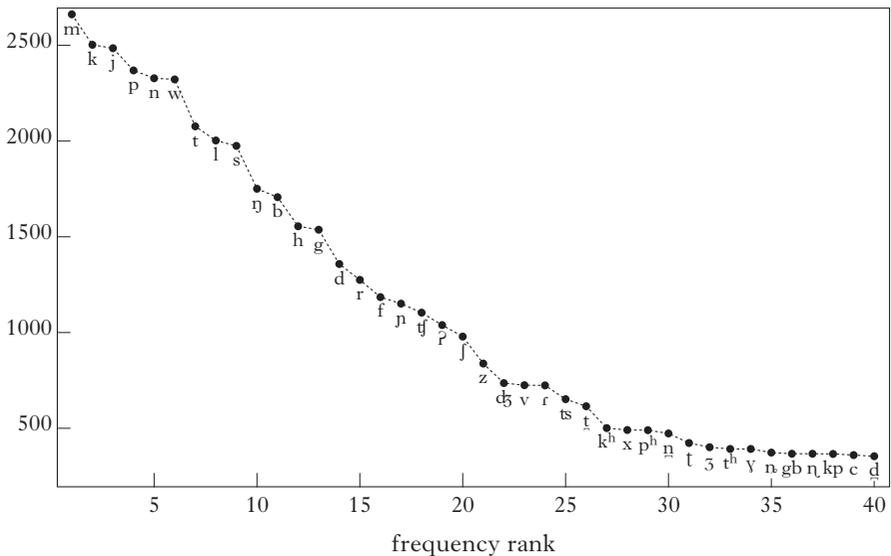
Figure 4

The 40 most frequent segments in the sample.

occurrence cluster consisting of some of the most frequent segments, this means that inventories lacking any of these sounds are somehow statistically anomalous. Such a cluster, therefore, is a good contender for the title of the basic consonant inventory.

A somewhat unexpected feature of the basic inventory cluster in Fig. 2 is the fact that /j/ and /w/, which are also exceedingly common ($N = 2485$, 2322 respectively), do not form part of this cluster, and are grouped together with alveolo-palatal stops, which are found predominantly in Australian languages. What we see here is probably the case of macro-areal influence being strong enough to skew the worldwide distribution of basic segments.[9] In order to test this hypothesis, we first re-ran the analyses on the same dataset, but with Australian languages ($N = 384$) excluded. In the resulting clustering, /j w/ are grouped together with /p t k m n l r/. To further ascertain their status *vis-à-vis* the 'basic set', we then conducted a cluster analysis on the segments from it together with /j w/, based on two versions of the dataset: one with and one without the Australian languages.[10]

The results are shown in Fig. 5. They indicate that in both cases /j w/, /p k/ and /t n/ form basic pairings, and that /p k/ are then merged together with /t n/; this cluster is then augmented by /l r m/ and finally merged with /j w/. The only difference is in the order in which the segments are added: /l/ > /r/ > /m/ in the full dataset, and /m/ > /l/ > /r/ when Australian languages are left out.

Thus it seems that even though /j w/ are very frequent, languages care mostly about having *both* these segments. Losing a segment from the /p t k m n l r/ set, on the other hand, even though clearly possible, leads to a statistically marginal configuration.

3.2.2 *The Australian extension set.* The cluster in Fig. 2 that includes /j w/ otherwise consists of three types of segments: retroflexes, /ɭ ɳ ɽ ʈ/, forming a counterpart to the subset of the basic inventory, alveolo-palatal /ȴ ȵ ȶ/, and /ŋ/. In a worldwide perspective, these sounds fall into three different frequency bands:

(i) /ŋ/ is a very frequent segment ($N = 1751$), but its overall relative frequency (63%) pales in comparison to its relative frequency in Australia (it is present in all Australian languages in the sample). The same applies to /j w/, found in all Australian languages in the sample.

(ii) Retroflex segments are traditionally associated with South Asia and the Himalayas, where a greater variety is indeed found than elsewhere. The relative frequencies of /ɭ ɳ ɽ ʈ/ in Australia, however, are all very

---

[9] On some important methodological issues related to the phonology of Australian languages, as well as explicit discussion of the basis for particular analytical choices relating to their representation in typological databases, see Round (2019).

[10] We again used the arccosine-transformed correlations as distances, and then applied the single-linkage clustering algorithm, which finds two most similar objects, averages them into a single node and continues in this way until there is only a single node left.

with Australian languages   without Australian languages
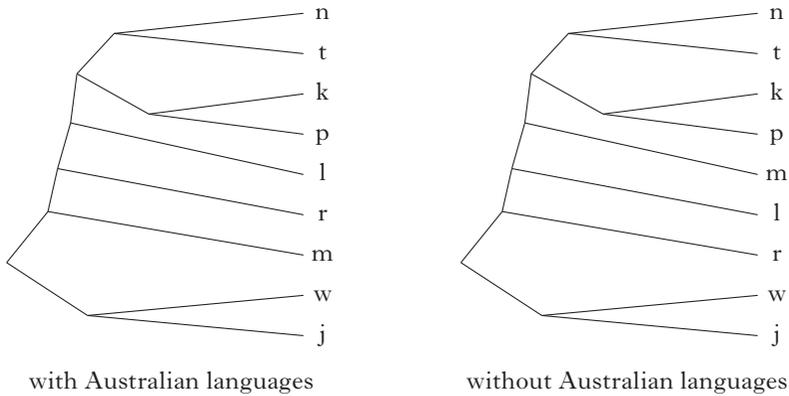
Figure 5
Single-linkage clustering of the most common segments
based on arccosine-transformed correlations.

high (≈ 0.6), while in other regions retroflex inventories display more variation (Nikolaev & Grossman 2018).

(iii) Alveolo-palatal /ȴ ɲ c/ are not reported outside Australia.

Other retroflex segments, /ɖ tʰ ɽ/, are found with similar frequencies in Australian and South Asian languages. As a consequence, they are classified as noise points.

3.2.3 *The non-Australian extension set.* Whereas Australian languages expand the basic inventory by means of an extended set of places of articulation, avoiding both voiced plosives and voiced fricatives, languages from outside Australia almost universally add to the basic set using some or most of the segments in (6). We refer to the resulting set as the FIRST EXTENSION SET.

(6) ʎ   ʒ   v   ʤ   z   ʃ   ʔ   ʧ
    121   401   725   736   838   979   1039 1104

     ɲ   f   d   g   h   b   s
    1151 1185 1358 1537 1555 1707 1975

The hierarchical clustering plot for these segments shown in Fig. 6 demonstrates that there is more structure to this extension than mere frequencies: the voiced plosives /b d g/ cluster together, as do /f v z/ and the postalveolar pairs /ʃ ʒ/ and /ʧ ʤ/. /s/ is placed separately, presumably because its very high frequency dilutes correlations with other segments, while other less frequent 'semi-basic' segments do not form pairs. The most notable observation is a relatively weak correlation between the glottal segments /h/ and /ʔ/ ($r = 0.257$). /ʔ/ is more than twice as likely to appear with /h/ than on its own (283 *vs.* 756 cases), but /h/ appears with and without /ʔ/ with nearly equal frequencies, with a slight preference for inventories *without* /ʔ/
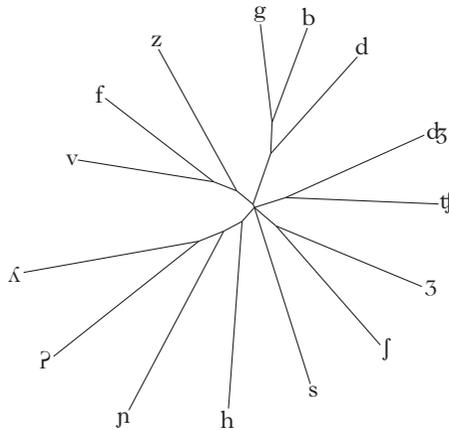
Figure 6
Single-linkage hierarchical clustering of the first major extension to
the basic inventory based on arccosine-transformed correlations.

(756 *vs.* 799 cases). It must be noted, however, that a great number of languages are postulated to have a non-phonemic glottal stop. The influence of this factor on the distribution is unclear.

Close to the extension set but not inside it are the tap /ɾ/ and two groups of what the algorithm deems to be noise points, although they seem to form subclusters of their own:

(i) The affricates /ts dz pf/ have widely different frequencies ($N = 652$, 297, 39 respectively), but /ts dz/ are strongly correlated ($r = 0.52$), and out of 39 occurrences of /pf/, 35 are in languages with /ts/; all the exceptions are found in Africa.

(ii) The palatal plosives /c ɟ/ have comparable frequencies ($N = 360$, 320). Similarly to fricatives and in contrast to basic plosives, they are grouped on the basis of their place of articulation, rather than voice. The position of the much rarer /cʰ/ ($N = 83$) is ambiguous: it is situated between the palatal plosives and the prominent cluster of voiceless aspirates.

Voiceless aspirated plosives and affricates cluster together. The addition of a non-basic VOT feature seems in this case to be more important than place and manner differences. This is not true, however, for the retroflex affricates /tʂ tʂʰ dʐ/, which are separated by their place–manner combination, disregarding VOT. Retroflex affricates form a tight co-occurrence class mostly associated with the affricate-rich languages of Eurasia (see Nikolaev & Grossman 2018).

The retroflex fricatives /ʂ ʐ/, treated as noise points by the clustering algorithm, enjoy a much wider and geographically unstructured distribution (languages with retroflex affricates tend to also have corresponding fricatives, but this connection is one-directional).

The alveolo-palatal fricatives and affricates /ɕ ʑ tɕ tɕʰ dʑ/ are medium-frequency segments forming the cluster characterised exclusively by place of articulation in (7).

(7) tɕʰ   ʑ   dʑ   ɕ   tɕ
   52   64   66   99   101

They are found in all major areas, and seem to be especially prevalent in the Circum-Tibetan area. The frequencies, however, may be unreliable, as the frequency of reports of alveolo-palatal segments in descriptions has increased in recent years, replacing earlier reports of postalveolar /ʃ ʒ tʃ dʒ/.

Languages rarely have both alveolo-palatals and postalveolars, and it has even been suggested that this is impossible, due to their acoustic and articulatory similarity (Hall 1997). However, there are 16 languages – nearly all of them Eurasian, and mostly Circum-Tibetan – that have both /ʃ/ and /ɕ/, and 13, mostly from the same set, with both /tʃ/ and /tɕ/.

The breathy voiced/murmured stops /b̤ d̤ ḏ̤ g̈/ are found predominantly in South Asia and the surrounding regions. They were evidently inherited in the Indo-Aryan languages of the area, and later spread, due to phonological borrowing and convergent processes of sound change. A similar scenario has been proposed to account for the distribution of retroflex segments, which presumably spread from their homeland in the Dravidian southern periphery of South Asia (Bomhard 1986, Bohnert *et al*. 2018).

Another group of segments augmenting the extension set are the bilabial and interdental fricatives together with the voiced velar fricative and, surprisingly, the retracted dental nasal stop /ṉ/, as in (8).

(8) ṉ   θ   ɸ   ð   β   ɣ
   46   108   144   147   296   392

The appearance of /ṉ/ here is probably a statistical artefact: it is very weakly correlated with other segments (no $r \geq 0.06$), which, it must be said, are generally rather weakly interconnected, except for /θ ð/ ($r = 0.44$).

Interestingly, voiced bilabial and interdental fricatives seem to be more common than their voiceless counterparts. /x/ ($N = 491$) does trump /ɣ/, but it also tends to co-occur with the 'guttural segments' (cf. below), which separates it from this cluster. Another possible peripheral member of this cluster is the alveolar approximant /ɹ/, a surprisingly rare segment ($N = 54$). It is known, however, that the description of rhotic segments often suffers from overnormalisation, with taps and approximants being inaccurately described as trills (Catford 2001), so this datum is unreliable.

Postvelar (uvular and pharyngeal segments) have low to mid frequencies, as in (9).

(9) ɢ ʕ qʰ ħ ʁ χ q
   41 53 67 70 119 181 228

Uvulars are generally more common than pharyngeals, with the notable exception of the voiced uvular stop /ɢ/, which suffers from the same kind of articulatory disadvantage as described above for /g/ (an extremely small cavity for maintaining high subglottal pressure for voicing). The interleaving of frequencies, however, does not influence the internal structure of the cluster, where the two pharyngeals are opposed to all the uvulars (cf. Fig. 7); uvular fricatives form a tight unit, but uvular stops, due to a wide discrepancy in frequencies, do not.
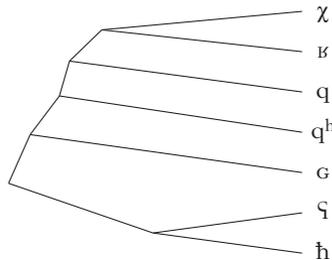
Figure 7
Single-linkage clustering of guttural segments
based on arccosine-transformed correlations.

Ejective segments, whose geographic distribution has been debated since the publication of Everett (2013) (see Hammarström 2013 and Roberts 2013, for example), undoubtedly form a clearly delineated group subsuming different places and manners of articulation, as in (10) ('R' is a symbol for an underspecified rhotic used in UPSID, one of the sources for PHOIBLE).

(10) t̪' R tɬ' q' ts' p' tʃ' t' k'
     37 46 38 71 126 145 156 171 212

Interestingly, /R/ tends to co-occur with ejective segments. It is also notable that the non-ejective voiceless lateral affricate occurs more rarely than its ejective counterpart, and does not pass the 30-language threshold.

The labialised velar and uvular segments in (11) form a cluster of their own, midway between ejectives and plain (i.e. non-labialised) postvelars. /xʷ/, a nearby noise point, can also probably be considered a part of this cluster.

(11) kʰʷ qʷ χʷ kʷ' xʷ
     45 39 48 67 72

These segments tend to appear in large inventories, which, in addition to sporting a wide range of places of articulation in the postvelar region, also employ labialisation, in accordance with the layering principle.

The appearance of the ejective /kʷ'/ in this cluster can be explained by the fact that labialised velar and uvular segments tend to appear in large inventories, especially those with ejectives, i.e. in the Caucasus and parts of North America. It should be pointed out that some languages with extremely large inventories (Saamic languages and some Tibetan dialects) lack both labialised plosives and ejectives.

Lateral fricatives and creaky voiced approximants form a surprising grouping. /j̰/ and /w̰/ are both quite rare (appearing in 30 languages each), and are very highly correlated ($r = 0.9$). /ɬ/ appears in half of the languages in which they appear in (14 for /j̰/ and 15 for /w̰/); these languages are all from North America, making this connection a localised one.

3.2.4 *Other clusters.*   The remaining subgroups of the large cluster are:

(i) The implosives /ɓ ɗ ʄ/ ($N = 293, 254, 44$), which form a natural group. /ʄ/ is mostly found in Africa, but /ɓ ɗ/ have an intriguing distribution: they are found in Africa, Eurasia, both Americas and in Oceania, but never above the 41st northern parallel.

(ii) The labial-velar segments /kp gb ŋm/, together with /ɥ/, the former almost exclusive to Africa.

(iii) The prenasalised segments /mb nd ŋg ndʒ ŋmgb ŋgʷ ndʑ nz ɱv ɲɟ/ ($N = 284, 278, 264, 127, 93, 57, 53, 46, 46, 36$), which are also confined to the southern hemisphere, with /mb nd ŋg/ enjoying a rather wide distribution, and others found predominantly in Africa.

(iv) The palatal affricates /cç cçʰ ɟʝ/ ($N = 112, 43, 97$), which occur primarily in languages of the Indian subcontinent, together with the palatal fricatives /ç ʝ/ ($N = 116, 54$), which have a much wider unstructured distribution.

(v) The dental segments /t̪ n̪ l̪/, together with nearby /d̪/ ($N = 615, 473, 191, 354$). Common descriptive practices may suppress the real frequency of these segments, which are often described as alveolar or 'denti-alveolar', but there is little doubt about the status of this cluster.

4 Co-occurrence classes and the feature-economy principle

In this section, we review the co-occurrence clusters identified by the method described in §2 in light of their conformity with the layering principle, which we consider a reformulation of the feature-economy principle.

In order to compare clusters in a principled manner, we first explain and illustrate the notion of CONFORMANT CLASS of segments. A conformant class

(i) uses all of the feature-value combinations available after pooling feature values from all its members, and (ii) if it includes stops and fricatives, also includes bilabials/labiodentals, denti-alveolars and velars.

For example, /p t k/ would be a conformant class, along with both /b d g/ and /p t k b d g/, because (i) it is impossible to define a new phoneme which would include a combination of the feature values found in the segments of these classes, and (ii) no requisite places of articulation are missing.[11] We regard place, manner and VOT as categorical features, and all other features as privative, i.e. features that can be either present or absent in a given segment's specification (e.g. a voiced implosive segment is taken as having a feature [+implosive] and non-modal voicing is modelled as [+breathy voice] or [+creaky voice]). A typical example of a non-conformant class will be /t k b d/, because available feature combinations {voiceless bilabial plosive} and {voiced velar plosive} are not represented. A class consisting of /t k d g/ is also non-conformant, because it does not include bilabials.

In the basic inventory, the cluster /p t k m n l r/ does not display any additional articulation or non-modal voicing, and utilises a comprehensive set of place features, but it is still hardly minimal. Even if we contend that voiceless stops cannot be in opposition with voiced continuants on the basis of a difference in their VOT values, this array of segments uses two privative features ([+lateral], [+rhotic]), which would not be required if, for example, /l r/ were replaced with /m̥ n̥/. It is therefore evident that, for currently unknown reasons, languages are remarkably reluctant to skip items in a minimal set of combinations of place and manner of articulation values with some 'default' voicing settings for these combinations.

In Table II, the basic consonant inventory augmented with /j w/ is compared to the minimal inventories discussed by Hyman (2008). We see that the basic consonant inventory can be considered a good predictor for what a minimal consonantal inventory will look like. The two noticeable discrepancies are that minimal inventories tend not to have both a rhotic and a liquid, and that the basic consonant inventory does not include /h/.

Perhaps more interesting is the comparison of our basic consonant inventory with that proposed by Clements (2009: 46), which is based on the 15 most common segment types in UPSID. Clements' proposed basic consonant inventory is phrased in terms of feature combinations, given as capital letters, with representative sounds given in IPA. The first five sounds, /t k n m p/, converge, as do /l r/, which are less frequent in UPSID. However, other sound types in Clements' basic consonant inventory, in particular all fricatives, affricates and glottals, as well as voiced stops, are not part of our basic consonant inventory proper, but

---

[11] Here we do not take into account the glottal stop, whose phonological status is problematic across a wide range of languages: it is often automatically inserted word-initially in languages that prohibit open syllables, and when found only in this position, it is hard to be certain whether it is contrastive.

[12] A revised account of the Maxakalí segment inventory is given in Silva (2020); cf. the short overview in English in Silva *et al.* (2020).

	P	T	TŠ	K	ʔ	B	D,R,L	ndʒ	G	M	N	f	S	H	W	J
BCI	p	t		k			r,l			m	n				w	j
Rotokas	p	t		k		β	ɾ		g							
Hawaiian	p			k	ʔ		l			m	n			h	w	
Nasioi	p	t		k	ʔ	b	ɾ			m	n					
Pirahã	p	t		k	ʔ	b			g				s	h		
Taoripi	p	t		k			l			m		f	s	h		
Gadsup	p	t		k	ʔ	β	d			m	n					j
Roro	p	t		k	ʔ	b	ɾ̥			m	n̥			h		
Ekari	p	t		k		b	d		gl	m	n				w	j
Maxakali[12]	p	t	tʃ	k	ʔ	mb	nd	ndʒ	ŋg					h		
Sentani	p	t		k			d			m	n	f		h	w	j
BCI (Clements)	P	T	TŠ	K		B	D,R,L		G	M	N		S	H	W	J

Table II

The basic consonant inventory compared with the minimal
inventories discussed by Hyman (2008) and the basic
consonant inventory proposed by Clements (2009: 46).

rather belong to extension sets. Furthermore, as we have shown, the glides /j w/ are part of the basic consonant inventory only if Australian languages are excluded. In other words, our approach shows that even within the set of basic sounds proposed in prior work on the basis of frequency, there is additional structure to be uncovered.

In the first extension set, the cluster /ʎ ʒ v dʒ ʐ z ʃ ʔ tʃ ɲ f d g h b s/ enlarges the minimal inventory by adding a series of voiced stops, but instead of exhausting the usefulness of the VOT opposition (by introducing voiceless sonorants and liquids), it adds two new major classes of sounds: fricatives and affricates, both voiced and voiceless, at two places of articulation (denti-alveolar and postalveolar; affricates are only postalveolar), together with the glottal stop (the status of /ʎ/ in this cluster is clearly peripheral, due to its lower frequency). This can be regarded as the 'second best' set of sounds that languages tend to have when the inventory is large enough to go beyond the basic set. It has been shown that some members of this inventory may be rather recent historically (Blasi *et al*. 2019), demonstrating that even the structure of comparatively basic parts of phonological inventories can be shaped by the forces of language contact and historical contingency. Furthermore, some of these sounds, in particular the fricatives and affricates, are among the most frequently 'borrowed' sounds, i.e. sounds that become contrastive as the result of lexical borrowing (Grossman *et al*. 2020). Table III shows the basic consonant inventory and the first extension set together.

		bilabial/ labiodental/ labial-velar	denti- alveolar	post- alveolar	palatal	velar	glottal
stop	voiceless	p	t			k	ʔ
	voiced	b	d			g	
fricative	voiceless	f	s	ʃ			h
	voiced	v	z	ʒ			
affricate	voiceless			tʃ			
	voiced			dʒ			
nasal	voiced	m	n			ŋ	
lateral	voiced		l		ʎ		
rhotic	voiced		r				
glide	voiced	w			j		

Table III
The basic consonant inventory (white cells) and the non-Australian extension set (light grey cells). Dark grey cells fall into the basic consonant inventory if Australian languages are excluded from the clusterisation.

4.1 Conformant classes

Voiceless aspirated segments are a conformant class: voiceless stops and affricates from different places of articulation have an additional distinctive feature. The problem with this class, however, is that it can be argued that it is incomplete: as will be shown below, voiceless affricates at several places of articulation form non-conformant classes.

The algorithm divided palatalised segments into coronals and non-coronals, but both classes are strictly conformant with the layering principle: they consist of voiced and voiceless consonants at different places of articulation with no other additional articulations.

Phonologically long segments form another conformant class: this cluster consists of voiced and voiceless consonants of different places of articulation characterised by no other features, such as additional articulations or non-modal phonation.

The treatment of the class of voiceless approximants and liquids depends on the problematic setting of the feature VOT for this class of segments. If voicing of approximants and liquids is regarded as a basic opposition, this class is not conformant, because it should also include voiced approximants and liquids. However, voicing may be treated as an integral feature of these segments, enforced by a fill-in rule of sorts (Keating 1988), in which case their voiceless counterparts are united by the presence of an additional feature value [−voice].

Breathy voiced stops form a conformant class defined by the addition of non-modal voicing to basic voiced stops. This class has a tight areal distribution, being a characteristic feature of South Asia.

Ejective segments as selected by the algorithm also form a well-behaved cluster, defined by an additional articulation superimposed on voiceless stops and affricates. Ejective fricatives are not frequent, and were not included in the data analysed.

The phonological analysis and feature specification of implosives are a matter of debate (Clements & Osu 2002). However, their articulation, acoustic properties and historical origins (they usually derive from geminated voiced stops; Xi 2009) point to their being an augmented version of voiced plosives, which is appropriately reflected by their place in the layered system.

4.2 Non-conformant classes

4.2.1 *Classes defined by a particular value for place.*   Several classes violate the layering principle by being defined not by an additional feature or a combination of features, but by a particular value for place. Most place values do not determine their own co-occurrence classes, and are subsumed under broader conformant classes.

The alveolo-palatal segments /z ɕ dʑ tɕ tɕʰ/, in addition to being defined by a particular place value, contain only a single element with an additional articulation (aspiration).

The uvular segments /qʰ ɢ q χ/ are similar to alveolo-palatals in consisting of voiced and voiceless segments with different manners of articulation and including one aspirated segment.

4.2.2 *Classes defined by a particular value of the manner feature.*   The group /θ ɸ ð β ɣ/, together with /n̪/, whose presence may be a statistical artefact due to its relatively low frequency, constitutes an intriguing extension to the basic inventory. Most notably, three groups of fricatives defined by different places of articulation – interdental, bilabial and voiced velar – form a group, even though they are all comparatively rare, and do not have obvious features in common which distinguish them from other modally voiced fricatives. A further violation of the layering principle is the fact that even though bilabial fricatives are in nearly complementary distribution with labiodental fricatives, it seems that precisely those languages which have bilabial fricatives tend to also have interdental fricatives and the voiced velar fricative.

The voiceless prenasalised stops /mp nt ŋk/ form a separate class by virtue of occurring in a small subset of languages with voiced prenasalised stops. This may be regarded as a prominent violation of the layering principle, since most languages with prenasalised stops do not have voiced ones, and therefore underutilise the available feature combinations.

A prenasalised stop, however, may be also regarded as voiced by default in the same way as a sonorant, as one of its constituent elements is

sonorant. In this scenario, voiceless prenasalised stops are characterised by an additional feature [−voice], making them a conformant class.

4.2.3 *Classes defined by a combination of place and manner features.* The two small classes of palatal and retroflex affricates are defined by their combination of place and manner features and their ability to incorporate voiceless aspirated segments (under the layering principle, voiceless aspirated fricatives are expected to cluster with voiceless aspirated stops, as we see at other places of articulation). These segments tend to appear as a group in large inventories (cf. the analysis of retroflex affricates by Nikolaev & Grossman 2018), and are largely geographically concentrated in and around South Asia. In the dimensionality-reduction plot in Fig. 2, both classes are also connected with respective fricative segments, which enjoy much wider and less structured distributions.

4.2.4 *Mixed classes.* Labialised segments are a complex case. The clearly separated class of non-guttural labialised segments in Fig. 1 is largely conformant. At the same time, the analysis shows that /hʲ/ forms a part of this cluster, due to its tight distributional connection with /hʷ/. This can be considered a minor violation of the layering principle. The predictive power of the principle is severely damaged, however, by the existence of a separate cluster of labialised segments consisting of /kʰʷ kʷ' χʷ qʷ/. These segments do not have a common distinguishing feature (one of them is aspirated, one is ejective and the other two are plain voiceless labialised). It is evident that this particular region of feature space tends to be adopted in its entirety by languages with rich inventories, including ejective sounds.

The lateral fricatives and creaky voiced approximants, /ɬ ʐ j̰ w̰/, form a rather surprising grouping, as they have nothing in common. Languages underlying this connection are North American, and it may have arisen due to geographically and phylogenetically restricted sound-change processes.

The featural analysis of the labial-velar segments, /kp gb ŋm/, presents difficulties, due to their dual nature. However, Cahill (1999) presents evidence in favour treating them as underlyingly labial segments with an additional velar coarticulation. In our data they strongly pattern with /ɥ/, which can also be analysed as a labial sound with an additional coarticulation. However, as this coarticulation is palatal, the class is non-conformant.

5 Conclusion

The contributions of this study are largely empirical, although they have a bearing on theoretical concerns. The primary contribution is the articulation of an empirical method for identifying co-occurrence clusters of segments in inventories. These are essentially groups of sounds between which bilateral dependences obtain. This in itself can contribute to future studies of the structure of sound inventories, but also to the analysis

of the geographical distribution of rule sets, grammatical categories or indeed any phenomenon that is governed by a point process (i.e. a process where each observation is a subset of some base set; see the analysis of vowel inventories from this perspective in Cotterell & Eisner 2017).

From a theoretical perspective, we have investigated the extent to which the feature-economy principle accounts for sound systems. The interpretation of our findings depends somewhat on the reader's outlook. On the one hand, we found that a large number of clusters conform to the feature-economy principle. To this extent, the feature-economy principle is validated by these classes. However, we also identified a number of non-conformant classes. Rather than reflecting mere noise, these classes tend to reflect fairly clear organisational principles, such as particular values for place or manner features, or a combination thereof.

Nonetheless, the feature-economy principle cannot really explain the structure of the most basic parts of inventories, in particular the classes we termed the basic consonant inventory and the first extension set, which always span a 'non-minimal' set of place and manner feature combinations with preferred values for VOT for some of these combinations. An intriguing finding of our study is that some earlier proposals for a basic consonant inventory, which were based purely on frequency in cross-linguistic databases, conceal additional structure. For example, Clements' proposal includes the basic consonant inventory proper, as well as parts of the first extension set and classes whose inclusion or exclusion depends on areal factors.

Secondly, several well-defined classes identified by the algorithm are carved out of the wider space of feature combinations with no regard for the feature-economy principle. Among the important exceptions to the principle are alveolo-palatal and postvelar segments, and retroflex and palatal affricates. These segments tend to be found mostly in complex inventories that utilise several additional articulations, and are usually found together. Consequently, these co-occurrence classes include members with and without additional articulations, going against the expected sequence of co-occurrence classes defined by adding extra features: plain segments appearing before palatalised, aspirated and labialised segments. It therefore seems that particular combinations of values of the basic place and manner features, very frequently found individually, are underrepresented in small and medium-sized inventories, to the detriment of feature economy.

Moreover, some combinations of feature values are highly correlated, especially labialisation of velar and postvelar segments.[13] This is unexpected on the basis of the feature-economy principle, which predicts that labialisation, when introduced to an inventory, should not be preferentially associated with any particular place of articulation.

[13] This can be treated as an example of feature enhancement, since labialisation is congruent with the 'grave' acoustic nature of velars and postvelars (Backley & Nasukawa 2009).

Additionally, it is apparent that there are strong macro-areal effects in the number and structure of co-occurrence classes. Australian inventories are a clear outlier, demonstrating deviating tendencies even in the structure of the most basic parts of the system, but Africa, North America and some large sections of Eurasia also have their own co-occurrence classes. We also have to account for the highly localised worldwide distribution of ejectives and for the skewed worldwide distribution of prenasalised segments and implosives, which are primarily found in the southern hemisphere.

These findings suggest that, in order to more fully account for the structure of actual phonological inventories of the world's languages, the feature-economy principle, which is structural in nature, should probably be augmented with other principles. The most important of these seems to be the tendency of consonant inventories to be MODULAR. Instead of progressive multiplicative expansion by means of adding new features to some base set of segments, predicted by the feature-economy principle and related hypotheses, we see that inventories grow by exploring particular regions in the segment space. Some of these regions and co-occurrence clusters can be construed as following the feature-economy principle, but many cannot, and this leads to the conclusion that there is perhaps no universal principle of inventory growth. And if there is one, it may be something along the lines of Stevens' quantal theory of speech (1989), with particular regions of feature-combination space having outsized prominence. The ECONOMY of inventories then can be construed as consisting of a minimisation not of the number of features used, but of the number of explored co-occurrence classes. Interestingly, apart from the fact that languages do not skip the basic consonant inventory and, for the most part, the first extension set, all other attractor regions are essentially unordered.

After presenting our conclusions, we would like to end the paper by repeating that this study is first and foremost an exploratory one. The analytical pipeline we used consists of a set of well-proven components, but as a whole it is very much a novel method that has not been extensively tested on other datasets in order to evaluate proposed phonological principles like feature economy. We therefore acknowledge that our results, despite being in line with reasonable expectations, may turn out to be somewhat brittle, and that other pipelines may identify other patterns of co-occurrence classes, although we would expect the structure as a whole to be relatively robust. We hope that our effort will spur a search for new, data-driven approaches in the field of consonant-inventory typology and that, as a result, the field will be able to break new ground, and at some point converge on a set of commonly accepted procedures for testing hypotheses.

REFERENCES

Atkinson, Quentin D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* **332**. 346–349.
Austin, Peter K. (1988). Phonological voicing contrasts in Australian Aboriginal languages. *La Trobe Working Papers in Linguistics* **1**. 17–42.

Backley, Phillip & Kuniya Nasukawa (2009). Representing labials and velars: a single 'dark' element. *Phonological Studies* **12**. 3–10.

Bateman, Nicoleta (2011). On the typology of palatalization. *Language and Linguistics Compass* **5**. 588–602. https://doi.org/10.1111/j.1749-818X.2011.00294.x.

Bickel, Balthasar (2013). Distributional biases in language families. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.) *Language typology and historical contingency*. Amsterdam & Philadelphia: Benjamins. 415–444.

Bickel, Balthasar (2015). Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.) *The Oxford handbook of linguistic analysis*. 2nd edn. Oxford: Oxford University Press. 901–923.

Bickel, Balthasar (2017). Areas and universals. In Raymond Hickey (ed.) *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press. 40–54.

Bickel, Balthasar & Johanna Nichols (2006). Oceania, the Pacific Rim, and the theory of linguistic areas. *BLS* **32:2**. 3–15.

Blasi, D. E., S. Moran, S. R. Moisik, P. Widmer, D. Dediu & B. Bickel (2019). Human sound systems are shaped by post-neolithic changes in bite configuration. *Science* **363**. https://doi.org/10.1126/science.aav3218.

Bohnert, Amanda, Samantha Myers, James Smith & Kelly Berkson (2018). Indic and Dravidian retroflexes: an ultrasound investigation. *JASA* **143**. 1755.

Bomhard, Allan R. (1986). The aspirated stops of Proto-Indo-European. *Diachronica* **3**. 67–79.

Butcher, Andrew (2006). Australian aboriginal languages: consonant-salient phonologies and the 'place-of-articulation imperative'. In Jonathan Harrington & Marija Tabain (eds.) *Speech production: models, phonetic processes, and techniques*. New York & Hove: Psychology Press. 187–210.

Cahill, Michael (1999). Aspects of the phonology of labial-velar stops. *Studies in African Linguistics* **28**. 155–184.

Catford, J. C. (2001). On Rs, rhotacism and paleophony. *Journal of the International Phonetic Association* **31**. 171–185.

Chao, Yuen-Ren (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of Institute of History and Philology, Academia Sinica* **4**. 363–397.

Clements, G. N. (2001). Representational economy in constraint-based phonology. In T. A. Hall (ed.) *Distinctive feature theory*. Berlin & New York: Mouton de Gruyter. 71–146.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology* **20**. 287–333.

Clements, G. N. (2009). The role of features in phonological inventories. In Raimy & Cairns (2009). 19–68.

Clements, G. N. & Sylvester Osu (2002). Explosives, implosives, and nonexplosives: the phonological function of air pressure differences in stops. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7. Berlin: Mouton de Gruyter. 299–350.

Cotterell, Ryan & Jason Eisner (2017). Probabilistic typology: deep generative models of vowel inventories. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Vol. 1: *Long papers*. 1182–1192. Vancouver: Association for Computational Linguistics. https://doi.org/10.18653/v1/P17-1109.

Coupé, Christophe, Egidio Marsico & François Pellegrino (2009). Structural complexity of phonological systems. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.) *Approaches to phonological complexity*. Berlin & New York: Mouton de Gruyter. 141–169.

Creanza, Nicole, Merritt Ruhlen, Trevor J. Pemberton, Noah A. Rosenberg, Marcus W. Feldman & Sohini Ramachandran (2015). A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences* **112**. 1265–1272.

Cysouw, Michael & Jeff Good (2013). Languoid, doculect and glossonym: formalizing the notion 'language'. *Language Documentation and Conservation* **7**. 331–359.

Dediu, Dan, Rick Janssen & Scott R. Moisik (2017). Language is not isolated from its wider environment: vocal tract influences on the evolution of speech and language. *Language and Communication* **54**. 9–20.

Dediu, Dan & D. Robert Ladd (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proceedings of the National Academy of Sciences* **104**. 10944–10949.

Dongen, Stijn van & Anton J. Enright (2012). Metric distances derived from cosine similarity and Pearson and Spearman correlations. Available (August 2020) at http://arxiv.org/abs/1208.3145.

Donohue, Mark & Johanna Nichols (2011). Does phoneme inventory size correlate with population size? *Linguistic Typology* **15**. 161–170.

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* **13**. 257–292.

Dunbar, Ewan & Emmanuel Dupoux (2016). Geometric constraints on human speech sound inventories. *Frontiers in Psychology* **7**:1061. https://doi.org/10.3389/fpsyg.2016.01061.

Eisen, Elad (2019). *The typology of phonological segment borrowing*. MA thesis, Hebrew University of Jerusalem.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander & Xiaowei Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han & Usama Fayyad (eds.) *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. Menlo Park, Ca.: AAAI Press. 226–231.

Everett, Caleb (2013). Evidence for direct geographic influences on linguistic sounds: the case of ejectives. *PLoS One* **8**. https://doi.org/10.1371/journal.pone.0065275.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts (2015). Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* **112**. 1322–1327.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts (2016). Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution* **1**. 33–46.

Feurer, Matthias & Frank Hutter (2019). Hyperparameter optimization. In Frank Hutter, Lars Kotthoff & Joaquin Vanschoren (eds.) *Automated machine learning: methods, systems, challenges*. Springer. 3–33.

Fletcher, Janet & Andrew Butcher (2012). Sound patterns of Australian languages. In Harold Koch & Rachel Nordlinger (eds.) *The languages and linguistics of Australia: a comprehensive guide*. Berlin & Boston: De Gruyter Mouton. 91–138.

Gordon, Matthew K. (2016). *Phonological typology*. Oxford: Oxford University Press.

Greenhill, Simon J. (2014). Demographic correlates of language diversity. In Claire Bowern & Bethwyn Evans (eds.) *The Routledge handbook of historical linguistics*. London & New York: Routledge. 557–578.

Grossman, Eitan, Elad Eisen, Dmitry Nikolaev & Steven Moran (2020). SegBo: a database of borrowed sounds in the world's languages. In *Proceedings of the 12th Conference on Language Resources and Evaluation*. Marseille: European Language Resources Association. 5316–5322. https://www.aclweb.org/anthology/2020.lrec-1.654.

Hall, T. A. (1997). The historical development of retroflex consonants in Indo-Aryan. *Lingua* **102**. 203–221.

Hall, T. A. (2000). Typological generalizations concerning secondary palatalization. *Lingua* **110**. 1–25.

Hammarström, Harald (2013). C. Everett's ejectives/altitude correlation is not significant. https://dlc.hypotheses.org/491.

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

Hay, Jennifer & Laurie Bauer (2007). Phoneme inventory size and population size. *Lg* **83**. 388–400.

Hayes, Bruce (2009). *Introductory phonology*. Malden, Mass. & Oxford: Wiley-Blackwell.

Hockett, Charles F. (1955). *A manual of phonology*. Baltimore: Waverly Press.

Hyman, Larry M. (2008). Universals in phonology. *The Linguistic Review* **25**. 83–137.

IPA handbook (1999). *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jakobson, Roman (1931). K xarakteristike èvrazijskogo jazykovogo sojuza. In Roman Jakobson. *Selected writings*. Vol. 1: *Phonological studies*. The Hague & Paris: Mouton. 144–201.

Keating, Patricia A. (1988). Underspecification in phonetics. *Phonology* **5**. 275–292.

Khan, Geoffrey (2009). *The Jewish Neo-Aramaic dialect of Sanandaj*. Piscataway, N.J.: Gorgias Press.

Kiparsky, Paul (2018). Formal and empirical issues in phonological typology. In Larry M. Hyman & Frans Plank (eds.) *Phonological typology*. Berlin: de Gruyter Mouton. 54–106.

Kochetov, Alexei (2005). Phonetic sources of phonological asymmetries: Russian laterals and rhotics. In Claire Gurski (ed.) *Proceedings of the 2005 Annual Conference of the Canadian Linguistic Association*. Available (August 2020) at https://cla-acl. artsci.utoronto.ca/actes-2005-proceedings.

Labov, William, Ingrid Rosenfelder & Josef Fruehwald (2013). One hundred years of sound change in Philadelphia: linear incrementation, reversal, and reanalysis. *Lg* **89**. 30–65.

Lass, Roger (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.

Lindblom, Björn & Ian Maddieson (1988). Phonetic universals in consonant systems. In Larry M. Hyman & Charles N. Li (eds.) *Language, speech and mind: studies in honour of Victoria A. Fromkin*. London: Routledge. 62–78.

McInnes, Leland, John Healy, Nathaniel Saul & Lukas Großberger (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**(**29**). https://doi.org/10.21105/joss.00861.

Mackie, Scott & Jeff Mielke (2011). Feature economy in natural, random, and synthetic inventories. In G. N. Clements & Rachid Ridouane (eds.) *Where do phonological contrasts come from? Cognitive, physical and developmental bases of distinctive speech patterns*. Amsterdam & Philadelphia: Benjamins. 43–63.

Macklin-Cordes, Jayden, Steven Moran & Erich Round (2016). Evaluating information loss from phonological dimensionality reduction. Paper presented at the Symposium on the Role of Predictability in Shaping Human Language Sound Patterns, 16th Speech Science and Technology Conference (SST2016), Western Sydney University. Slides available (August 2020) at https://doi.org/https://dx.doi. org/10.6084/m9.figshare.4465976.

Maddieson, Ian & Kristin Precoda (1992). UPSID and phoneme. Version 1.1. Available (August 2020) at http://phonetics.linguistics.ucla.edu/sales/UPSID-guide.pdf.

Marsico, Egidio, Ian Maddieson, Christophe Coupé & François Pellegrino (2004). Investigating the 'hidden' structure of phonological systems. *BLS* **30**. 256–267.

Martinet, André (1952). Function, structure, and sound change. *Word* **8**. 1–32.

Moran, Steven (2012). *Phonetics information base and lexicon*. PhD dissertation, University of Washington.

Moran, Steven (2019). Phonological inventories. In Mark Aronoff (ed.) *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi. org/10.1093/acrefore/9780199384655.013.112.

Moran, Steven & Daniel McCloy (eds.) (2019). PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. Available (August 2020) at https:// phoible.org.

Moran, Steven, Daniel McCloy & Richard Wright (2012). Revisiting population size vs. phoneme inventory size. *Lg* **88**. 877–893.

Moran, Steven, Taras Zakharko & Balthasar Bickel (2013). Estimating diachronic preferences of phonological features cross-linguistically. Paper presented at the 10th Biennial Conference of the Association for Linguistic Typology, Leipzig. Abstract available (August 2020) at https://www.eva.mpg.de/lingua/conference/2013_ALT10/pdf/abstracts/abstract_112_DiachroPreferencesPhonologFeatures.pdf.

Nikolaev, Dmitry & Eitan Grossman (2018). Areal sound change and the distributional typology of affricate richness in Eurasia. *Studies in Language* **42**. 562–599.

Ohala, John J. (1983). The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*. New York: Springer. 189–216.

Pater, Joe (2012). Emergent systemic simplicity (and complexity). *McGill Working Papers in Linguistics* **22**. Available (August 2020) at https://scholarworks.umass.edu/linguist_faculty_pubs/165.

Pericliev, Vladimir (2004). There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. *Linguistic Typology* **8**. 376–383.

Piantadosi, Steven T. & Edward Gibson (2014). Quantitative standards for absolute linguistic universals. *Cognitive Science* **38**. 736–756.

Raimy, Eric & Charles E. Cairns (eds.) (2009). *Contemporary views on architecture and representations in phonology*. Cambridge, Mass.: MIT Press.

Rice, Keren (2007). Markedness in phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. 79–97.

Roberts, Sean (2013). Altitude and ejectives: hypotheses up in the air. http://www.replicatedtypo.com/altitude-and-ejectives-hypotheses-up-in-the-air/6324.html.

Round, Erich (2019). Australian phonemic inventories contributed to PHOIBLE 2.0: essential explanatory notes. https://doi.org/10.5281/zenodo.3464333.

Sapir, Edward (1925). Sound patterns in language. *Lg* **1**. 37–51.

Seinhorst, Klaas T. (2017). Feature economy vs. logical complexity in phonological pattern learning. *Language Sciences* **60**. 69–79.

Silva, Mário Coelho da (2020). *Sons e pedaços da língua Maxakalí: descrição da fonologia e morfologia de uma língua Macro-Jê*. PhD dissertation, Federal University of Minas Gerais.

Silva, Mário Coelho da, Andrew Nevins & James White (2020). Domains and prominence in nasal harmonization of Maxakalí loanwords. *IJAL* **86**. 285–321.

Simpson, Adrian P. (1999). Fundamental problems in comparative phonetics and phonology: does UPSID help to solve them? In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*. Vol. 1. Berkeley: University of California. 349–352.

Sinnemäki, Kaius (2014). A typological perspective on Differential Object Marking. *Linguistics* **52**. 281–313.

Stevens, Kenneth N. (1989). On the quantal nature of speech. *JPh* **17**. 3–45.

Stoakes, Hywel, Janet Fletcher & Andy Butcher (2007). An acoustic study of Bininj Gun-Wok medial stop consonants. In Jürgen Trouvain & William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Saarland University. 869–872.

Trubetzkoy, Nikolai S. (1939). *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Vaux, Bert (2009). The role of features in a symbolic theory of phonology. In Raimy & Cairns (2009). 75–97.

Verhoef, Tessa, Simon Kirby & Bart de Boer (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science* **40**. 1969–1994.

Xi, Cun (2009). *A phonetic study on implosives in China*. PhD dissertation: Hong Kong University of Science and Technology.