HOW ECONOMICAL ARE PHONOLOGICAL INVENTORIES?

Christophe Coupé, Egidio Marsico & Gérard Philippson

Laboratoire Dynamique Du Langage, CNRS, Université Lyon 2, France
ccoupe@ish-lyon.cnrs.fr; emarsico@ish-lyon.cnrs.fr; gphilipson@ish-lyon.cnrs.fr

ABSTRACT

Since the work of André Martinet back in the 60's [7], the notion of 'economy' has been accepted as one of the main principles underlying the organization of phonological inventories (hereafter PI). This principle, also named MUAF ('Maximal Use of Available Features') by John Ohala [8], states that PI are organized along a few phonetic dimensions that they tend to maximize in terms of number of segments. In this paper, we propose a method to quantify the extent of this economy. We observe that it is in fact not maximal and better applies to vocalic than consonantal systems. We also study which are the preferred phonetic dimensions organizing this economy.

Keywords: UPSID, phonological inventories, typology, feature economy, underspecification

1. INTRODUCTION

Lindblom and Maddieson [3] describe the 'all-inclusive universal phonetic space' as containing a variety of systems whose structural dimensions vary (both in quantity and quality) according to their size. This variation results from the fact that phonological systems are trade-offs between two opposite constraints. 'Ease of articulation', on the one hand, requires the system to use as few phonetic dimensions as possible to discriminate as much segments as possible. This articulatory economy has the effect of generating similar segments in the system. On the other hand, perceptual salience requires phonetically dissimilar segments to ensure maximum or at least sufficient acoustic distance. Regarding typology, this leads to the prediction that smaller PI will use fewer phonetic dimensions than larger ones since they have fewer segments to contrast. The more segments you add to the phonetic space (i.e., when systems get larger), the more saturated the phonetic dimensions are, thus leading to the recruitment of new dimensions to preserve perceptual contrast.

In this paper, we investigate the role of articulatory economy in the shaping of PI. Following Lindblom and Maddieson [3] and Clements [1, 2], we tackle this issue by considering phonetic features. Lindblom and Maddieson looked in particular at consonantal systems, and showed that when the size of a PI increases, so does the number of phonetically complex consonants in this system. The validity of feature economy was here assessed by showing that the larger a system is, the more phonetic features are needed to describe its segments. As for Clements, he statistically tested the existence of feature economy by looking at the co-occurrences of segments sharing phonetic features in PI. He found that systems tend to be composed of similar segments by showing that the frequency of occurrence of a particular segment is significantly correlated to the number of others in the system bearing the same features.

Our approach constitutes a third alternative. It consists in an attempt to quantify the actual economy of PI and define a tentative hierarchy among phonetic dimensions by identifying their frequency of use in phonological contrasts.

2. DATA

Our data come from an expanded version of the UPSID database ([4, 5]). Aside from being the largest database on PI, UPSID constitutes a genetically and geographically balanced sample of over 7% of the world's languages, thus constituting a very useful tool for phonological typology research. It contains 451 languages, for a total of 833 different segments.

3. APPROACH

3.1. Data coding

Segments are defined by a set of features closely derived from the original UPSID set. The 100 features used are mainly articulatory and are not far from those extracted from the IPA chart. They are phonetic descriptive features but still allow to uniquely and fully specify each segment.

A study by Marsico, et al. [6] based on the UPSID database and relying on a fully specified description of segments, showed that systems are in fact quite redundant. On average, any two segments tend to differ from other segments within a system by two features. This result lessened the impact of economy on the organization of PI. We however argue that using a fully specified description of segments introduces a bias. In order to really capture the role of economy, more parsimonious descriptions are needed, which relate to the notion of underspecification.

We adopt here a 'light' version of under-specification. One of the reasons is that UPSID doesn't contain data on phonological processes, on which choices leading to underspecification are traditionally based. What we actually seek is the minimal description, in terms of features, contrasting all the segments of a PI. This 'contrastive' underspecification is a classical structuralist way to address systemic redundancy.

We developed an algorithm calculating all the possible minimal underspecifications of a system. It relies on massive tests of acceptable descriptions with subsets of features (tens of millions for the largest systems). It turns out that most systems accept several underspecifications (with different subset of features). Since electing the best of them was difficult, we each time averaged our computations - described hereafter - on all possible underspecifications for a system.

3.2. Quantifying feature economy

If feature economy (ease of articulation) were to be the only principle acting on the content of PI, we would expect systems to show a maximal use of features. In other words, the ratio of the actual number of segments in a system by the number of possible segments (given the set of features of this system) should be close to 1. For each language we thus calculated the number of possible segments given the set of features of that language (based on the 833 different possible segments contained in UPSID). We did this calculation for both the full and the underspecified descriptions of the systems.

3.3. Describing feature economy

We furthermore looked at the way systems organize their segments around specific dimensions depending on their size. To this end, features were divided into major classes corresponding to the primary dimensions of consonants (laryngeal settings, manner and place) and vowels (height, backness and rounding), and into secondary features (e.g. 'palatalized'). The distinction between primary and secondary features reflects standard articulatory descriptions, and differs for other approaches like [9].

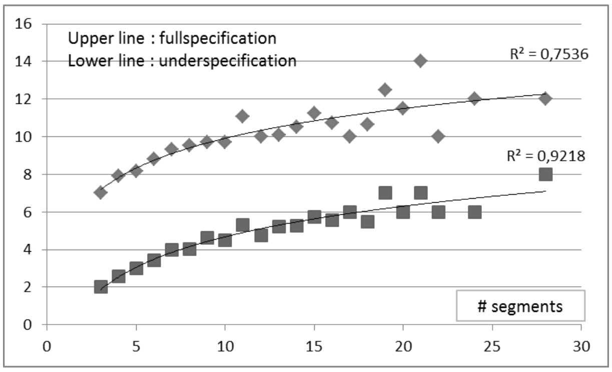
4. RESULTS

Results for individual languages are averaged per systems size. However, large sizes may be represented by a single system, highlighting unwanted local specificities. The higher variability of most right tails of distributions should therefore not be given much significance.

4.1. Quantifying feature economy

Figure 1 shows the average number of features used in vocalic systems as a function of their size. As expected, the more segments in a system, the more features are needed to describe them. Interestingly, the size of the underspecified set of features is on average half the fully specified one. This trend is observed for consonants in comparable proportions. Deriving feature economy from the fact that the number of features increase logarithmically with the number of segments is tempting but wrong, given how any additional feature may create a large number of new possible segments for the system through combinations with other features.

Figure 1: Average number of features of a system against its number of segments for full specification (upper line) and underspecification (lower line). (vowels only).



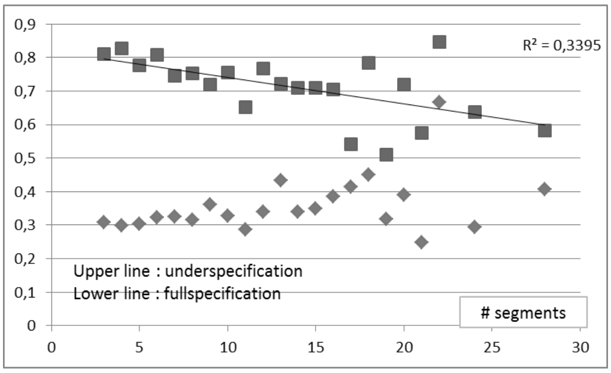
To know the extent of feature economy of a system, we need to look at the rate of use of its features. Table 1 gives the average value and standard deviation of the rate of use of features, for vowels and consonants and full and underspecified descriptions. It should be first noted that when considering full specification, systems are far from being economical, since, they use only a third of the possible segments they could have given their features. More interesting, and probably more surprising are the different effects of underspecification observed for vowels and consonants. Whereas for vowels, economy increases up to 71%, it does not even reach 50% for consonants. The main conclusion here is then that consonantal systems are on average by far less economical than vowel systems when we look at their *effective* use of features.

Table 1: Feature economy calculated as the ratio of the number of actual segments of a system by the number of possible segments given its features.

Feature Economy	Vowels		Consonants	
	FSpec.	**USpec**	**FSpec.**	**USpec**
Average	0.36	0.71	0.30	0.49
Std. Dev.	0.09	0.09	0.05	0.10

We can observe the variation of this rate as a function of the size of the systems. Figure 2 shows that for full specification, systems do not vary in terms of economy. However, when considering underspecifications, they have a tendency to become less economical as they grow in size.

Figure 2: Average rate of use of features plotted against the size of a system for full specification (lower line) and underspecification (upper line). (vowels only).



4.2. Describing feature economy

We only present here results for underspecified systems. We focus on two primary phonetic dimensions (height for vowels and manner for consonants) and secondary features for both. Figures 3 to 6 give the average number of features of a particular dimension used in the underspecified descriptions of a language.

Figure 3 shows that the more vowels a system has, the more use of height distinctions it makes. No such pattern is observed for rounding and backness. Moreover, we observed that 98% of all underspecifications for a system (100% when size is larger than 5) include a height distinction, whereas this drops down to 91% for backness and 51% for rounding.

Figure 3: Average number of height features in underspecifications as a function of size. (vowels).

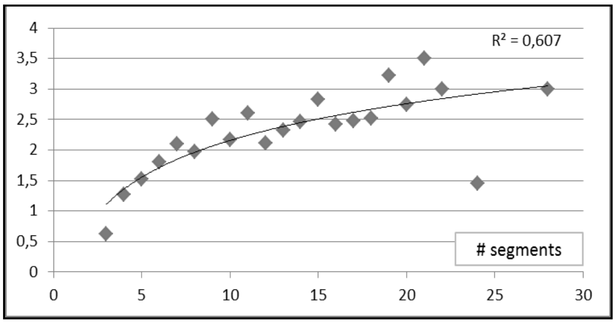


Figure 4 reveals a similar trend for secondary features: the more vowels in the system, the more secondary features are relied on to minimally describe the system. The contrastive use of at least a secondary feature according to size presents an interesting S-type pattern (not shown here). The inflection point is at around 10 vowels, with approximately 60% of underspecifications containing secondary features. This number reaches 100% for sizes larger than 15.

Figure 4: Average number of secondary features in underspecifications as a function of size. (vowels).

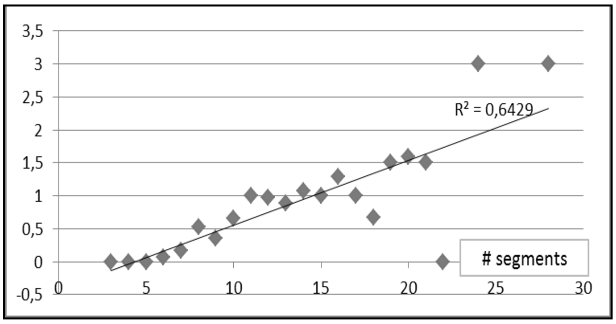
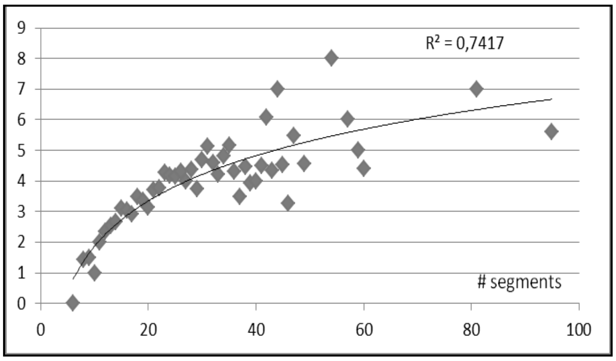


Figure 5: Average number of manner features in underspecifications as a function of size. (consonants).

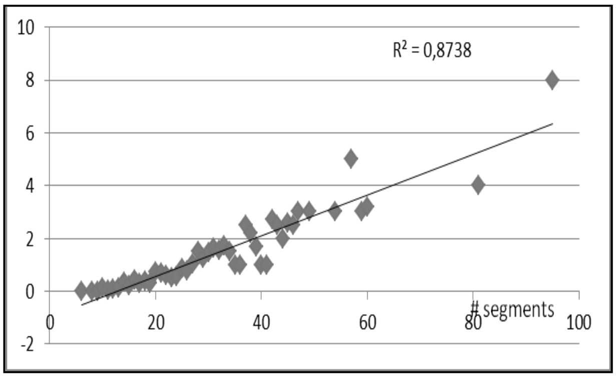


The behavior of consonants is similar. Figure 5 shows the same correlation for size and manner than the one observed for size and height in vowels.

This is valid for place features too, but not as obvious for features of laryngeal settings.

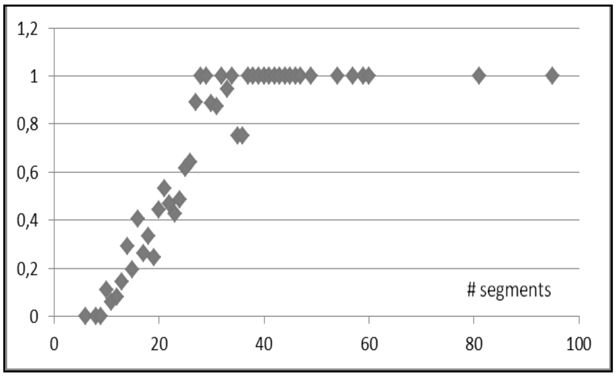
The relation between the size of the system and the contrastive use of secondary features is even higher than for vowels: R^2=0.87 (cf. Figure 6).

Figure 6: Average number of secondary features in underspecifications as a function of size. (consonants).



When we consider the 'need' for a specific dimension in underspecifications, we see that all languages have at least a manner and a place distinction, but have to have more than 30 segments for secondary features to systematically be part of underspecifications (cf. Figure 7).

Figure 7: Average number of underspecifications containing a secondary feature as a function of size. (consonants).



5. SUMMARY

In this study, we investigated the degree of feature economy in phonological inventories. We found that PI do not show any economy if fully specified segments are considered. Economy is only revealed when using a more parsimonious 'contrastive underspecification'. Furthermore, this principle is mostly active in vowel systems where the rate of use of features reaches 71%, whereas it remains around 50% for consonants. It also seems like the economy decreases with the size of the system (for vowels and consonants).

It also appears that some phonetic dimensions are primarily used to generate contrasts in systems (height and backness for vowels, manner and place for consonants). These dimensions are the only ones used in small inventories. Secondary dimensions appear in small systems, but only become indispensable - that is not at least partially redundant with primary features when it comes to contrasting segments - when the systems get larger.

6. DISCUSSION

Our results reveal an interesting asymmetry between vowels and consonants when it comes to feature economy. This might be due to the fact that vowels constitute a continuous phonetic space, contrary to consonants. John Ohala's MUAF may also need to be lessened, as the economy is never maximized in languages. This could be due to the need for a sufficient perceptual contrast, which can affect systems in two complementary ways: either by ensuring a certain amount of redundancy or by not saturating particular dimensions in order to make contrasts easier. Further studies will be carried out to answer these questions.

7. REFERENCES

[1] Clements, G.N. 2003a. Testing feature economy. *15th International Congress of Phonetic Sciences* Barcelona, Spain.

[2] Clements, G.N. 2003b. Feature economy as a phonological universal. *15th International Congress of Phonetic Sciences* Barcelona, Spain.

[3] Lindblom, B., Maddieson, I. 1988. Phonetic universals in consonant systems. In Li, C.N., Hyman, L.H. (eds.), *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*. Beckenham: Croom Helm, 62-78.

[4] Maddieson, I. 1984. *Patterns of Sounds*. Cambridge, MA: Cambridge University Press.

[5] Maddieson, I., Precoda, K. 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74, 104-111.

[6] Marsico, E., Maddieson, I., Coupé, C., Pellegrino, F. 2003. Investigating the "hidden" structure of phonological systems. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, 256-267.

[7] Martinet, A. 1955. *L'économie des changements phonétiques*. Berne: A. Francke.

[8] Ohala, J.J. 1980. Chairman's introduction to symposium on phonetic universals in phonological systems and their explanation. *Proceedings of the Ninth International Congress of Phonetic Sciences* Copenhagen, 184-185.

[9] Stevens, K.N., Keyser, S.J. 1989. Primary features and their enhancement in consonants. *Language* 65(1), 81-106.