

VOWEL INVENTORIES REVISITED: THE FUNCTIONAL LOAD OF VOWEL CONTRASTS

1. Goals

Most studies in phonological typology are based on segment inventories with the hypothesis that observed regularities result from general constraints (articulatory ease, perceptual salience, etc.). This relies on the implicit assumption that all segments in an inventory are equally meaningful. However, segments form a system because they're engaged in phonological contrasts with each other and as noted in [1]: 'some contrasts between the phonemes in a system apparently do more of [the] job than others', thus questioning the supposed equivalence. Recent works have highlighted the importance of frequency of use in characterizing languages' phonology [2]. Building on this idea, we suggest that the notion of functional load (FL, see [1] & [3] among others) would enrich the usual representation of phonological inventories. In this study, we propose a corpus-based and cross-linguistic approach to evaluate the relevance of FL for characterizing vowel systems. Our main objective is to assess whether reanalyzing vowel systems in terms of FL validates the explanations based on inventories or reveals new constraints.

2. Methodology

Data consisted of text corpora (including hundreds of thousands of word tokens) in twelve languages – Amharic, Bulgarian, Spanish, English, Estonian, Finnish, French, German, Swahili, Tagalog, Turkish, and Zulu – mainly from the Leipzig Corpora Collection [4]. FL was computed for each language following the algorithm proposed in [5]: a) Using the frequencies of the wordforms, we evaluated the entropy of their distribution. b) We characterized the FL of a contrast between two vowels as the relative loss of entropy induced by their artificial merging into a single phoneme. c) This procedure was repeated for each vowel pair, and FL was normalized with respect to the sum of the FL borne by vocalic contrasts.

3. Results (overview)

Figure 1a gives the example of Amharic. Edges indicate the FL for the vowel contrasts built upon their endpoints; for instance, the /e~/u/ contrast accounted for 36% of the vocalic FL. Figure 1b displays the distribution of contrasts ranked by decreasing FL. The twelve languages exhibited a similar uneven use of the vowel contrasts available from their inventory. From this paradigmatic viewpoint, they seemed far from an optimal use of their phonological system, even if an optimization in terms of syntagmatic trajectories (involving coarticulation and phonotactics) cannot be ruled out. Furthermore, the cross-linguistic comparison revealed no uniform tendency to favor maximal contrasts (such as /a~/i/ and /a~/u/). It actually highlighted that several languages with similar inventories may exhibit different contrast patterns (see Figure 2). These results and their interpretation in terms of the organization of vowel systems will be discussed during the conference, along with methodological issues.

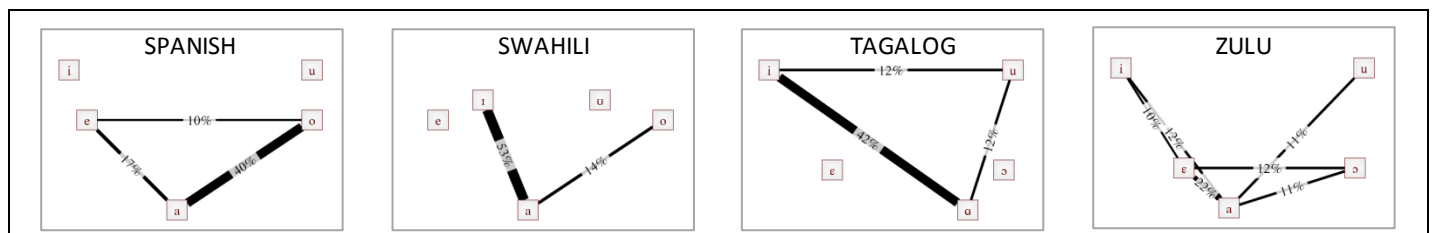
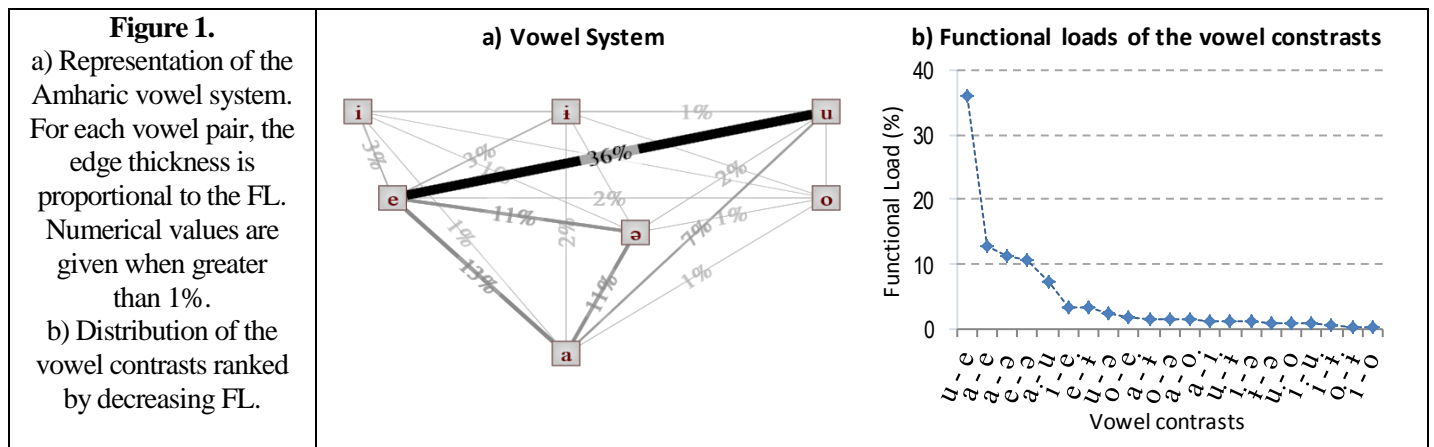


Figure 2. Comparison of four languages with vowel systems akin to /a e i o u/. For legibility sake, only edges corresponding to FL greater than 10% are displayed.

(NUMBER OF WORDS: 500)

4. References

- [1] Hockett, C.F. 1966. *The quantification of functional load: A linguistic problem*, Report Number RM-5168-PR, Rand Corp. Santa Monica.
- [2] Bybee, J. 2001. *Phonology and Language Use*, Cambridge University Press: Cambridge.
- [3] Martinet, A. 1955. *Économie des changements phonétiques. Traité de phonologie diachronique*, Francke: Berne.
- [4] Quasthoff, U., Richter, M., & C. Biemann. 2006. "Corpus Portal for Search in Monolingual Corpora", *proc. of the 5th International Conference on Language Resources and Evaluation LREC*, Genoa, pp. 1799-1802.
- [5] Surendran, D., and P. Niyogi. 2003. *Measuring the Usefulness (Functional Load) of Phonological Contrasts*, Technical Report TR-2003-12, Department of Computer Science, University of Chicago, USA.