Vowel inventories revisited: the functional load of vowel contrasts

François Pellegrino[1,2], Egidio Marsico[1,2] & Christophe Coupé[1,2]

[1]*Laboratoire Dynamique du Langage, CNRS - Université Lyon 2,* [2]*Institut Rhône-Alpin des systèmes complexes*

Background & Goal

"The function of a phonemic system is to keep the utterances of a language apart. Some contrasts between the phonemes in a system apparently do more of this job than others." (Hockett, 1966)

Yet: differences between contrasts are discarded in phonological analyses. Is this a mistake?

Goal: assess the relevance of a 'Functional Load' approach

The notion of Functional Load (FL)

- Early mentions in Twaddell (1935), Trubetzkoy (1939) & Martinet (1955)
- Is FL a relevant factor of resistance/propensity to sound change (Hoenigswald, 1960)? Negative answer according to (King, 1967)
- Quantitative developments with *Information theory* (e.g. Wang, 1967)

Computing Functional Load

- *Loss of entropy under the hypothesis of a phoneme coalescence* (Hockett, 1966)

$$FL_{Hockett}(x,y) = \frac{H(L) - H(L^*_{xy})}{H(L)}$$ at the word level

- Recent availability of digital corpora allow to measure FL
- Renewal of interest (Surendran & Levow, 2004; Surendran & Niyogi, 2003)

Our approach

- Investigate the distribution of phonemic contrasts
- Compute FL (for vowels) from large corpora in several languages
- Study the organization of phonological systems w.r.t. various constraints

Corpus

Language	Family	Code	# lexical tokens	Corpus co-verage	Word distribution entropy
Amharic	Afro-Asiatic, Semitic	AMH	1.9M	83.7%	12.1
Bulgarian	IE, Slavic	BUL	6.2M	90.4%	10.5
Chilean Spanish	IE, Italic	ChSP	440M	97.7%	9.3
British English	IE, Germanic	ENG	18M	98.6%	9.5
Estonian	Uralic, Finnic	EST	3.4M	84.6%	11.3
Finnish	Uralic, Finnic	FIN	970k	72.2%	11.8
French	IE, Italic	FRE	900k	98.6%	9.6
German	IE, Germanic	GER	808k	96.4%	10.1
Swahili	Niger-Congo, Atlantic-Congo	SWA	27.4M	93.6%	10.2
Tagalog	Austronesian, Malayo Polynesian	TGL	180k	98.0%	9.5
Turkish	Altaic, Turkic	TUR	968k	82.8%	11.8
Zulu	Niger-Congo, Atlantic-Congo	ZUL	217k	75.2%	12.1

Sources: Celex Project (ENG. GER) Lexique Project (FRE). Leipzig Corpora (EST. FIN. TGL, TUR). Bulgarian Treebank project (BUL). LIFCACH project (ChSP). Univ. of Bristol (ZUL). Univ. of Grenoble (AMH). and DDL (SWA)

- Sources: Web-based (mostly newspapers) or book-based corpora
- Only the 20k most frequent words to limit the impact of erroneous entries

Methodology

A toy word corpus and its phonological inventory: /a i u p b l/

Word	Frequency
pal	300
pil	200
bal	150
bil	150
pul	100
bul	100
Overall	**1000**

$$H(L) = \sum_{i=1}^{N_L} p_{w_i} \cdot h(w_i) = -\sum_{i=1}^{N_L} p_{w_i} \log_2(p_{w_i})$$

H(L)=2,47

Contrast	Raw FL	Relative FL
a-i	31.8%	42%
a-u	23%	30%
i-u	21%	28%
Vocalic FL (a-i-u)	**60.7%**	**100%**

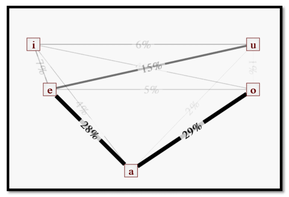
Word	Frequency
p(a,i)l	300+200
b(a,i)l	150+150
pul	100
bul	100
Overall	**1000**

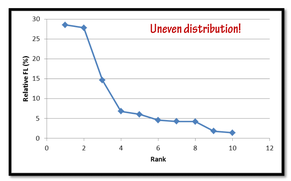
a-i contrast: L*ai :
*H(L*ai)* = 1.69
FL(a-i) = (2.47-1.69)/2.47
= 31.8 %

Results

Example: Chilean Spanish: /a i u e o/

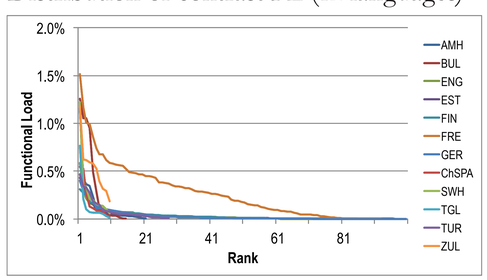
Contrast	Raw FL	Relative FL
a i	0.08%	4.3%
a u	0.03%	1.8%
a o	0.54%	28.5%
a e	0.53%	27.8%
i u	0.11%	6.0%
i o	0.08%	4.2%
i e	0.13%	6.8%
u o	0.03%	1.4%
u e	0.28%	14.6%
o e	0.09%	4.6%
Vowel FL	1.90%	100%





Cross-linguistic comparisons

Distribution of contrast FL (12 languages)

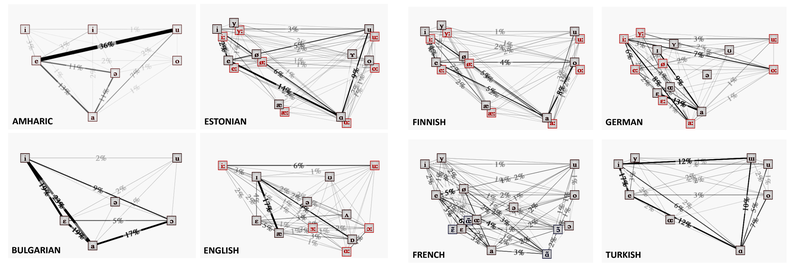


Whole vowel system FL

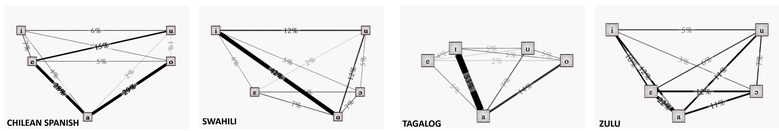
Language	Vocalic FL	Size of the vowel system
AMH	4.4.%	7
BUL	5.1%	6
ENG	2.9%	16
EST	4.0%	18
FIN	5.1%	16
FRE	14.8%	15
GER	2.8%	16
ChSPA	2.3%	5
SWH	4.0%	5
TGL	1.7%	5
TUR	3.1%	8
ZUL	6.3%	5

- Uneven distribution of FL
- Low FL of the vowel system on average (from 1.9% to 5.5%)
- French as an outlier (14.8%)
- No direct relationship with vowel system size
- Distributions too irregular to be power-law

Vowel inventories revisited



- /a/-like vowel intensively utilized, Front dimension more than Back
- Tendency to favor redundant contrasts
- Longest perceptual distances not always favored



- "Symmetric" vowel systems (e.g. 5 vowels) may be irregular in light of FL
- Identification of well-known constraints not straightforward
- Uneven distribution of FL: not optimal in a strict information-theoretical framework, but guarantees an efficient level of redundancy

Conclusions & Perspectives

- Existence of a kernel phonemic network distinguishing between i) the most exploited phonemes & contrasts and ii) the other phonemes & contrasts
- In this view, the latter are not useless because they guarantee the adaptive capacity of the language during its evolution (*Complex Adaptive Systems).*

→ Broaden the spectrum of languages considered

→ Potential bias due to limited lexical coverage

→ Compute feature-based functional load

→ Explore the relationship between FL and phonological and morpho-syntactic patterns (small-world network, preferential attachment)

→ Impact of the corpus: most frequent words, Swadesh list, small texts

References

1. Hockett, C.F. (1966). The quantification of functional load: A linguistic problem, Report Number RM-5168-PR, Rand Corp. Santa Monica.
2. Hoenigswald, H. (1960). Language change and linguistic reconstruction. Chicago: University of Chicago Press
3. King, R.D. (1967). Functional Load and Sound Change. Language, 43(4): 831-852.
4. Martinet, A. (1955). Économie des changements phonétiques. Traité de phonologie diachronique. Francke.
5. Stokes, S. & Surendran, D. (2005). Articulatory complexity, ambient frequency and functional load as predictors of consonant development in children, J. of Speech and Hearing Research 48(3).
6. Surendran, D. & Levow, G.-A. (2004). The Functional Load of Tone in Mandarin is as High as that of Vowels. Proceedings of Speech Prosody 2004, Japan.
7. Surendran, D. & Niyogi, P. (2003). Measuring the Usefulness (Functional Load) of Phonological Contrasts. Tech. Report TR-2003-12., Department of Computer Science, Univ. of Chicago.
8. Twaddell, W.F. (1935). On Defining the Phoneme, Language 11(1): 5-62.

Financial support: CNRS – Grant PICS n°05766 'Dartmouth-DDL'; ASLAN Laboratory of Excellence

Contact: François PELLEGRINO, francois.pellegrino@univ-lyon2.fr