

La typologie des systèmes vocaliques revisitée sous l'angle de la charge fonctionnelle

François Pellegrino, Egidio Marsico, Christophe Coupé

(1) DDL, 14 Avenue Berthelot, 69363 Lyon Cedex 07

francois.pellegrino@univ-lyon2.fr, egidio.marsico@ish-lyon.cnrs.fr, christophe.coupe@ish-lyon.cnrs.fr

RÉSUMÉ

La plupart des études typologiques en phonologie présument que tous les segments d'un inventaire sont d'importance équivalente. Nous nous proposons d'enrichir la représentation classique des systèmes phonologiques en y incluant la notion de charge fonctionnelle. Notre étude, basée sur des corpus, montre que les 12 langues de notre échantillon font un usage très inégal des oppositions disponibles dans leur système vocalique. De plus, aucune tendance à favoriser les contrastes perceptuels maximaux (tels que /a/~i/ et /a/~u/ par exemple) n'apparaît dans les résultats. De fait, même des langues ayant des inventaires similaires présentent des oppositions privilégiées différentes, amenant à réévaluer certaines des tendances universelles bien établies dans le domaine.

ABSTRACT

Vocalic system's typology revisited from the functional load viewpoint

Most studies in phonological typology implicitly assume that all segments in an inventory are equally important. We suggest here that the notion of functional load enriches this usual representation of phonological systems. Focusing on the vowel systems of 12 languages, we developed a corpus-based approach that reveals a very uneven use of the vowel contrasts available from their inventory. Furthermore, the cross-linguistic comparison reveals no uniform tendency to favor maximal contrasts (such as /a~/i/ and /a~/u/). It actually highlights that several languages with similar inventories may exhibit different contrast patterns, questioning some well-established universal tendencies.

MOTS-CLÉS : système vocalique, typologie, opposition distinctive, charge fonctionnelle.

KEYWORDS : vocalic system, typology, phonemic contrast, functional load.

“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.”

Charles F. Hockett (1966)

1 Introduction

Cette citation datant de près d'un demi-siècle contraste avec une tendance marquée des recherches en typologie phonologique : alors que celles-ci visent à comprendre la structure et le fonctionnement des systèmes phonologiques, elles ont pendant longtemps délaissé la notion fondatrice d'opposition pour se concentrer sur l'étude des inventaires. Au début du 20^{ème} siècle, les travaux du Cercle linguistique de Prague et de Trubetzkoi

en particulier (Cercle Linguistique de Prague, 1929) soulignaient pourtant l'importance des oppositions comme éléments de description phonologique synchronique. Aucun rôle structurant ne leur était cependant confié, bien que les linguistes aient eu conscience de l'existence de différences dans le degré d'utilisation des unités phonologiques dans chaque langue. La notion de charge fonctionnelle (CF) est ainsi apparue pour qualifier ces différences. Elle sera popularisée par Martinet (1933) qui lui donnera une dimension diachronique, suivi par Hockett (1955) qui en proposera une formulation mathématique. L'idée de Martinet était que plus la charge fonctionnelle d'une opposition est grande, plus elle offre de résistance au changement phonétique. Cette hypothèse, bien que séduisante, a plutôt été infirmée par de premières études quantitatives (King, 1967, Wang, 1967), menant à une éclipse de la CF durant près de 40 ans. En effet, depuis la fin des années 1960, la typologie phonologique s'est concentrée sur la comparaison des inventaires phonologiques, avec pour objectif d'étudier les tendances universelles de structuration des systèmes phonologiques et de déterminer les contraintes à l'œuvre. Ces recherches ont largement fait progresser notre connaissance des phénomènes en jeu à l'interface phonétique/phonologie, en mettant en évidence un certain nombre de contraintes (« *size principle* », « *ease of articulation* », « *maximal or sufficient perceptual contrast* », « *focalization* » etc.) interagissant dans la structuration des systèmes phonologiques observés (Liljencrants & Lindblom, 1972, Crothers, 1978, Maddieson, 1984, Vallée, 1994).

Au début des années 2000, la notion de CF a fait une timide réapparition par le biais de quelques études mettant à profit de grands corpus. L'hypothèse diachronique de Martinet a de nouveau été infirmée (Surendran & Niyogi 2003), mais un rôle partiel de la CF dans la séquence d'acquisition des phonèmes par les enfants a été suggéré (Stokes & Surendran 2005). Surendran et Levow (2004) ont également mis en évidence à cette époque que la CF portée par les tons en chinois mandarin est du même ordre de grandeur que celle portée par les timbres vocaliques. Ce dernier résultat met l'accent sur le poids relatif des différents éléments constituant un système phonologique et nous amène à reconsidérer l'approche purement descriptive, qui suppose une équivalence entre le système phonologique et l'inventaire de ses segments. Or, si dans un système phonologique les relations entre segments n'ont pas la même importance pour la langue, comme suggéré dans la citation initiale de Hockett, il est possible que la recherche des tendances universelles et de leurs déterminants ait occulté un pan complet de la structuration des systèmes.

L'objectif de cette étude est donc d'enrichir cette approche classique en intégrant la notion de charge fonctionnelle prise dans son acception première. Ceci consiste à attribuer un poids relatif à chaque opposition – une charge fonctionnelle – déterminé à partir du rôle joué par cette opposition dans la langue. Il s'agit alors i) de tester l'universalité des distributions de charges fonctionnelles à partir d'un échantillon translinguistique et ii) de vérifier si les oppositions les plus fréquemment observées sont en accord avec les principes de structuration des systèmes issus des recherches sur les inventaires phonologiques.

Nous nous limiterons ici à l'étude des systèmes vocaliques pour lesquels les contraintes de production et de perception, ainsi que les tendances universelles, sont les mieux connues.

2 Méthodologie

Dans cette approche, notre calcul de la charge fonctionnelle se base sur la notion d'entropie (d'après Shannon, 1948). Les mesures sont établies à partir de grands corpus disponibles pour plusieurs langues.

2.1 Calcul de la charge fonctionnelle

On considère ici une langue L comme une source de séquences constituées de mots (w) pris dans un ensemble fini (N_L). Les unités dans les séquences sont considérées comme indépendantes les unes des autres. $h(w)$ définit la quantité d'information qu'un mot w apporte en fonction de sa probabilité d'apparition dans une séquence, $p(w)$. Cette quantité d'information est égale à l'opposé du logarithme de cette probabilité. Plus l'apparition est prévisible ($p(w)$ proche de 1), plus $h(w)$ est faible et donc moins le mot est informatif. L'entropie de la langue $H(L)$ est définie comme suit :

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

Nous reprenons la définition de la charge fonctionnelle utilisée dans Surendran (2003), qui s'appuyait sur la formulation de Carter (1987), dérivée de la proposition initiale de Hockett (1955, corrigée en 1966).

La charge fonctionnelle d'une opposition x/y , $CF(x,y)$, se définit ainsi comme l'écart relatif d'entropie entre deux états de langue, l'état observé $H(L)$ et l'état dans lequel l'opposition est neutralisée $H(L^*_{xy})$. La charge fonctionnelle mesure ainsi la perturbation, en termes d'augmentation de l'homophonie et de redistribution des fréquences de mots, engendrée par le fait de remplacer les deux membres de l'opposition x/y par un seul élément ; elle s'exprime en pourcentage.

$$CF(x,y) = \frac{H(L) - H(L^*_{xy})}{H(L)}$$

2.2 Exemple fictif

Considérons une langue fictive avec le système phonologique suivant : /a, i, u, p, b, l/. La Figure 1 (gauche) résume les fréquences d'apparition des mots dans cette « langue ».

Telle que définie en 2.1, l'entropie de cette langue est égale à $H(L) = 2,47$.

On peut calculer la charge fonctionnelle brute de chaque opposition vocalique comme la différence relative entre $H(L)$ et l'entropie calculée après fusion des deux membres de l'opposition dans le lexique. Par exemple, après neutralisation de l'opposition a-i, on obtient $H(L^*) = 1,69$, et pour a-u $H(L^*) = 1,90$. Pour chaque contraste, la Figure 1 (droite) donne la valeur de CF_{BRUTE} obtenue ainsi qu'une valeur normalisée par rapport à l'ensemble des contrastes vocaliques (CF_{NORM}). Dans notre exemple fictif, l'opposition a-i est la plus employée. On peut également étudier la CF du système vocalique considéré dans son ensemble (CF_{SYSVOC}), en fusionnant les trois voyelles en un seul timbre noté \underline{V} (le lexique se résume alors à $p\underline{V}l$ et $b\underline{V}l$). L'entropie $H(L^*) = 0,97$ permet ici d'obtenir $CF_{SYSVOC} = 60,7\%$.

MOT	FRÉQUENCE
pal	300
pil	200
bal	150
bil	150
pul	100
bul	100
TOTAL	1000

CONTRASTE	CF _{BRUTE}	CF _{NORM}
a-i	31,6%	42%
a-u	23,1%	30%
i-u	21,0%	28%
TOTAL		100%

FIGURE 1 – Lexique et CF de la langue fictive

La Figure 2 donne deux représentations possibles des résultats, employées dans la suite de l'article. Celle de gauche représente la distribution de la CF_{BRUTE} des contrastes vocaliques ordonnés par ordre décroissant, tandis que celle de droite représente l'organisation du système vocalique sous la forme d'un graphe dont chaque arête porte la CF_{NORM} du contraste défini par ses sommets.

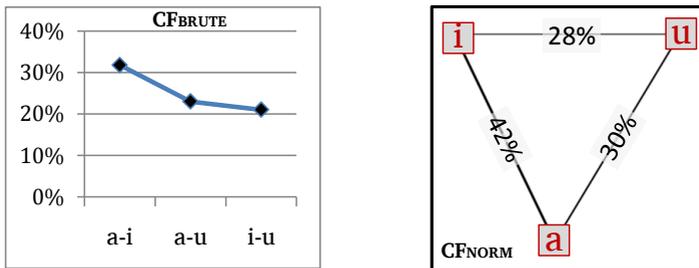


FIGURE 2 – Distribution des CF pour les voyelles de la langue fictive

Enfin, la charge fonctionnelle est calculée pour une opposition, mais il est possible de définir la charge d'une voyelle seule comme la somme des CF_{NORM} des oppositions auxquelles elle participe. Soit par exemple pour i : $(42+28)/2 = 35\%$ (le facteur $\frac{1}{2}$ correspondant à une normalisation du fait qu'une opposition repose sur deux voyelles).

2.3 Données de l'étude

L'étude porte sur 12 langues pour lesquelles nous disposons d'un lexique - assorti des fréquences des mots estimées sur la base de grands corpus écrits - ainsi que d'un système permettant la conversion graphème-phonème. Les corpus proviennent d'Internet ou de livres : (Projet Celex (ENG, GER), Projet Lexique (FRE), Corpus Leipzig (EST, FIN, TGL, TUR), Projet Bulgarian Treebank (BUL), Projet LIFCACH (ChSP), Université de Bristol (ZUL), Université de Grenoble (AMH) et Dynamique Du Langage (SWA)). Afin de limiter l'impact des erreurs de saisie dans les corpus nous n'avons considéré que les 20 000 mots les plus fréquents. Le Tableau 1 donne la liste des langues, leur affiliation génétique, la taille des corpus en nombre de *token*, le pourcentage de couverture du corpus par les

20 000 mots les plus fréquents, la taille du système vocalique ainsi que la charge fonctionnelle du système vocalique CF_{sysvoc} .

Langue	Phylum	Code	Nb Token	Couverture	Nb Voyelles	CF_{sysvoc}
Allemand	Indo-Européen	GER	808k	96,4%	16	2,8%
Amharique	Afro-Asiatique	AMH	1,9M	83,7%	7	4,4%
Anglais britannique	Indo-Européen	ENG	18M	98,6%	16	2,9%
Bulgare	Indo-Européen	BUL	6,2M	90,4%	6	5,1%
Espagnol chilien	Indo-Européen	ChSP	440M	97,7%	5	2,3%
Estonien	Ouralique	EST	3,4M	84,6%	18	4,0%
Finois	Ouralique	FIN	970k	72,2%	16	5,1%
Français	Indo-Européen	FRE	900k	98,6%	15	14,8%
Swahili	Niger-Congo	SWA	27,4M	93,6%	5	4,0%
Tagalog	Austronésien	TGL	180k	98,0%	5	1,7%
Turc	Altaïque	TUR	968k	82,8%	8	3,1%
Zoulou	Niger-Congo	ZUL	217k	75,2%	5	6,3%

TABLEAU 1 – Détails des données de l'étude (voir le texte pour une explication)

3 Résultats

Lorsque l'on considère le système vocalique dans son ensemble, la charge fonctionnelle CF_{sysvoc} varie de 1,7% à 6,3% pour 11 des 12 langues ; le français présente une valeur extrême à 14,8%. Ces valeurs ne sont pas corrélées à la taille du système vocalique (Tableau 1).

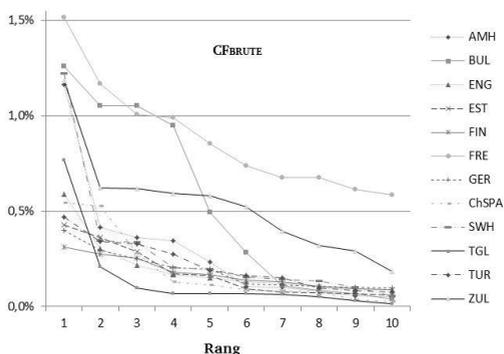


FIGURE 3 – Distribution des 10 CF_{BRUTE} les plus importantes pour chaque langue

La Figure 3 illustre pour chaque langue la distribution des CF_{BRUTE} pour les 10 oppositions vocaliques à plus forte CF (classées par ordre décroissant). Globalement,

ces oppositions se répartissent de manière très inégale, avec très peu d'oppositions à forte CF et beaucoup à faible CF. Là encore le français se distingue par une répartition plus progressive des CF. Le fait que les distributions ne soient pas équilibrées ne s'explique pas facilement au niveau phonologique.

La Figure 4 représente les graphiques des CF_{NORM} pour six langues du corpus.

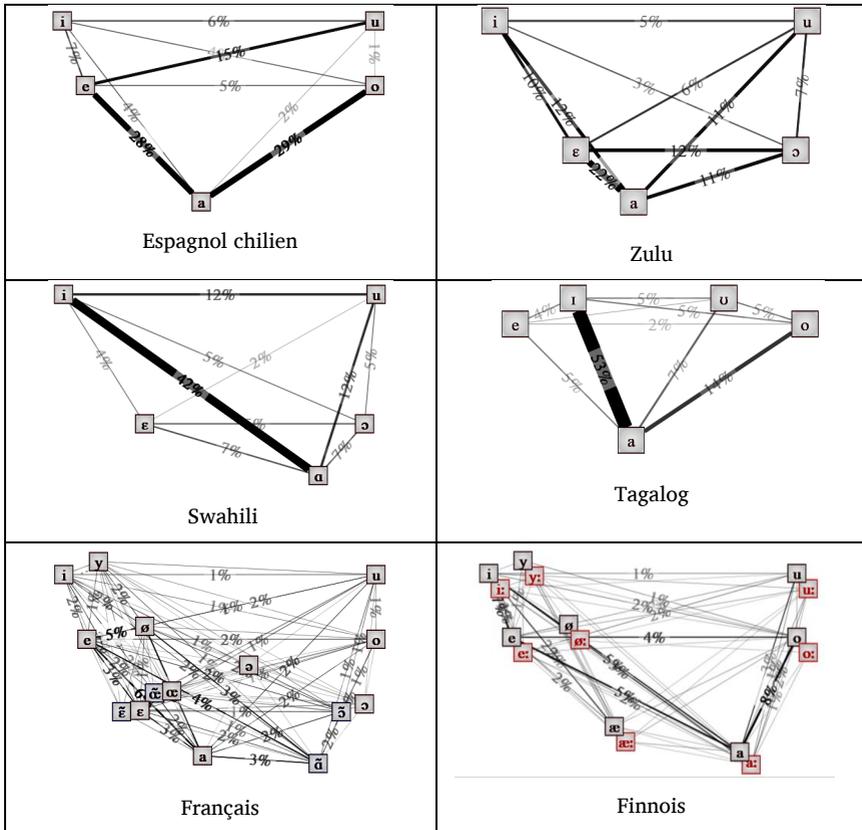


FIGURE 4 –Représentations des distributions de CF_{NORM} pour 6 langues du corpus

On obtient pour certaines langues des CF_{NORM} allant jusqu'à 53% (Tagalog), quand d'autres n'atteignent pas les 10% (8% en finnois et 6% en français). Un autre fait intéressant concerne la variation observée entre les contrastes privilégiés dans les langues de notre corpus. En effet, si ce n'est le fait que les voyelles de grande ouverture semblent privilégiées, les oppositions favorites sont différentes d'une langue à l'autre. Par ailleurs ces oppositions ne respectent pas de manière évidente les contraintes

perceptives mises au jour sur la base des seuls inventaires. Par exemple, ce ne sont pas toujours les oppositions portées par une distance perceptuelle maximale qui présentent les plus grandes CF. Il convient de noter également que les voyelles les plus employées ont tendance à être des voyelles antérieures. Ces résultats, observés sur les 12 langues, se confirment lorsque l'on considère uniquement les systèmes à 5 voyelles : espagnol chilien, swahili, tagalog et zulu, qui partagent une « structure » proche (cf. Figure 5).

ChSPA		SWH		TGL		ZUL	
Voyelle	Poids	Voyelle	Poids	Voyelle	Poids	Voyelle	Poids
a	31%	ɑ	34%	a	40%	a	28%
e	27%	i	31%	ɪ	33%	ɛ	25%
o	19%	u	15%	o	13%	ɔ	17%
u	12%	ɔ	11%	ʊ	9%	i	15%
i	11%	ɛ	9%	ɛ	6%	u	15%

FIGURE 5 – CF rapportée aux voyelles pour les systèmes à 5 voyelles

4 Conclusions et perspectives

Une première conclusion est que la prise en compte du poids relatif des oppositions vocaliques apporte des éléments nouveaux quant aux systèmes vocaliques. Là où les recherches basées sur les inventaires ont montré que les systèmes vocaliques étaient plutôt réguliers et répondaient à des contraintes claires, l'étude des charges fonctionnelles montre une grande diversité inter-linguistique dans l'utilisation des oppositions distinctives des voyelles. Ceci est vrai aussi bien dans le choix des « meilleures » oppositions que dans la répartition des CF au sein du système. Le lien avec des contraintes phonético-phonologiques bien connues semble difficile à établir.

Ce travail souligne l'importance des charges fonctionnelles (au niveau d'une opposition ou rapportées à chaque voyelle) dans la compréhension de l'organisation des systèmes phonologiques. S'il n'invalide pas que des contraintes propres au niveau phonologique puissent expliquer les motifs paradigmatiques observés, il suggère également que ces motifs peuvent être expliqués par des facteurs syntagmatiques (exprimés partiellement dans la structure des mots). Les contraintes phonotactiques au niveau des structures syllabiques peuvent ainsi se révéler pertinentes, tout comme celles observables aux niveaux morphologique et syntaxique. La forte CF d'une opposition entre 2 voyelles peut ainsi dans certains cas refléter une opposition entre 2 morphèmes grammaticaux. Plus globalement, l'éclairage apporté par l'étude de la CF permet de reconnecter le système vocalique aux principes à l'œuvre dans la composition des mots d'une langue.

Références

- CERCLE LINGUISTIQUE DE PRAGUE. (1929). « Thèses présentées au Premier congrès de philologues slaves ». *Travaux du Cercle linguistique de Prague 1*, 5-29, consulté sur Internet le 01/05/2008 à l'adresse <http://www2.unil.ch/slav/ling/textes/theses29.html>
- CROTHERS, J. (1978). Typology and universals of vowel systems in phonology. In J. H. Greenberg, Ed., *Universals of human language*, vol. 2. Stanford, CA: Stanford University Press, pages 95-152.
- HOCKETT, C.F. (1955). *A manual of phonology*. Waverly Press: Baltimore.
- HOCKETT, C.F. (1966). *The quantification of functional load: A linguistic problem*. Report Number RM-5168-PR, Rand Corp. Santa Monica.
- KING, R.D. (1967). Functional Load and Sound Change. *Language*, 43(4), 831-852.
- LILJENCRAFTS, J. et LINDBLOM, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48, 839-862
- MADDIESON, I. (1984). *Patterns of sounds*. Cambridge, MA: Cambridge University Press.
- MARTINET, A. (1933). « Remarques sur le système phonologique du français », *Bulletin de la Société Linguistique de Paris*, 34, 192-202.
- MARTINET, A. (1955). *Économie des changements phonétiques. Traité de phonologie diachronique*. Francke: Berne.
- SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 & 623-656, July and October, 1948.
- STOKES, S. et SURENDRAN. D. (2005). Articulatory complexity, ambient frequency and functional load as predictors of consonant development in children. *Journal of Speech and Hearing Research*, 48(3), 577-591.
- SURENDRAN, D. et LEVOW, G.-A. (2004). The Functional Load of Tone in Mandarin is as High as that of Vowels. In *Proc. of Speech Prosody 2004*, Japan.
- SURENDRAN, D. et NIYOGI, P. (2003). *Measuring the Usefulness (Functional Load) of Phonological Contrasts*. Technical Report TR-2003-12. Department of Computer Science, University of Chicago.
- TWADDELL, W.F. (1935). "On Defining the Phoneme", *Language*, 11(1), 5-62.
- VALLÉE, N. (1994). *Systèmes vocaliques : de la typologie aux prédictions*. PhD dissertation, Université Stendhal, Grenoble, France.