

*"Vers une mesure de la complexité
des espaces phonologiques"*

E. Marsico, C. Coupé, F. Pellegrino & J-M. Hombert
Dynamique Du Langage, Lyon

**III^{èmes} Journées d'Etudes Linguistiques,
"Universaux sonores"
Nantes, Mars 2002**

Objectifs :

- ❖ Evaluer l'architecture des systèmes phonologiques quantitativement et qualitativement à travers une même mesure
- ❖ Proposer une structure de "l'espace phonologique" dans laquelle les dimensions articulatoires sont hiérarchisées
- ❖ Définir une mesure de distance phonologique applicable en synchronie et en diachronie

Problématique :

- ❖ Identifier des indices de mesure exogènes
- ❖ Ne pas considérer la distribution fréquentielle d'un segment comme un indice de sa complexité
- ❖ Proposer une mesure de distance qui soit davantage qu'un simple différentiel du nombre d'unités
- ❖ Identifier les poids respectifs des dimensions articulatoires structurant l'espace de production

Echantillons :

- ❖ UPSID, 451 langues
- ❖ BDPROTO, 120 proto-langues
- ❖ Au total 930 segments différents
- ❖ Chaque segment est défini par une matrice de traits

Modèles théoriques possibles :

"Géométrie des traits" *Clements, 1985*

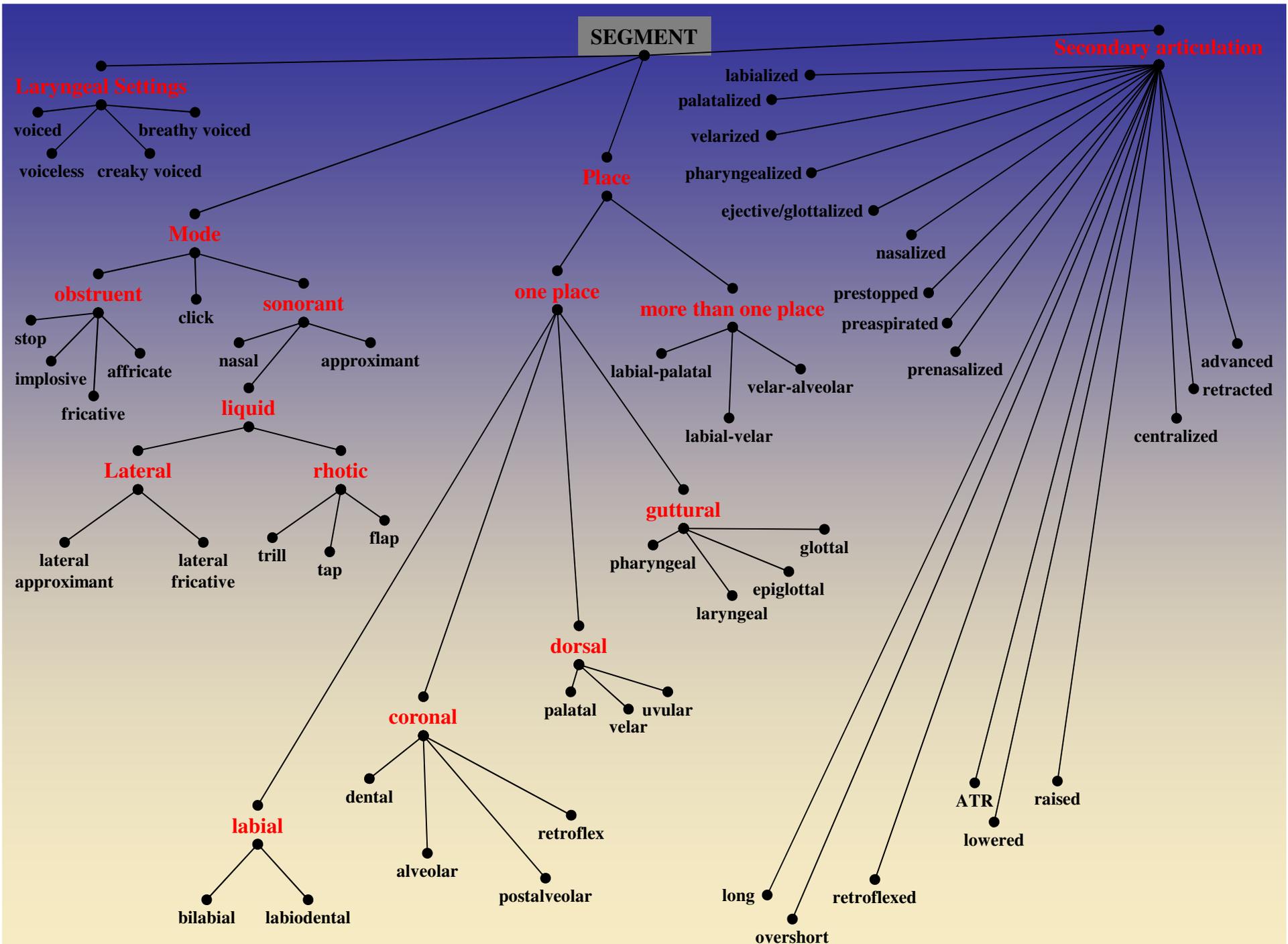
- ❖ Hiérarchie des traits au sein d'une arborescence
- ❖ Calcul de distance envisageable comme une différence d'arbres

"Gestes articulatoires" *Browman & Goldstein, 1989*

- ❖ Primitives ancrées dans les contraintes du tractus vocal

Problèmes :

- i) Ces deux modèles ne couvrent pas l'ensemble des traits utilisés dans notre échantillon
- ii) Ils entraînent des pondérations entre catégories de segments dans le calcul des pas entre les arborescences
- iii) Ils ne permettent pas d'aboutir à une seule et unique hiérarchie (eg. pondérations des différentes arborescences entre elles ?)



Choix :

Développer un modèle basé sur des réseaux de neurones dans lequel les pondérations entre traits ne seront pas définies a priori.

* Un premier réseau (réseau Hebbien) permet de pondérer les traits entre eux

* Un deuxième réseau (carte de Kohonen) permettra d'accéder à des prototypes de systèmes

Calcul de la distance :

- 1) Initialisation du réseau de pondération des traits (Etablissement de l'espace des possibles)
- 2) Calcul du réseau pour les langues L1 & L2
- 3) Calcul de la distance entre L1 & L2

1) Initialisation du réseau de pondération des traits :

Features	stop	implosive	nasal	trill	tap	flap	r-sound	fricative	affricate	affricate-lateral	lateral-fricative	lateral	lat.-approx.	approximant	click
bilabial	29	3	13					7						5	
labiodental	2		1			1		10	4					1	
labial-palatal														1	
labial-velar	7	1	2											6	
dental	16	1	5	2	2	1	1	14	26		6		6		14
alveolar	30	2	12	14	1	9	4	20	34	9	8	11	13	2	23
postalveolar	7	1	3			1		17	23				2	1	8
retroflex	12	1	2	3		4		6	10		1	2	2	1	1
palatal	11	1	6	1				6	13	12			1	7	24
velar	34	1	13					12	7		1	2	1	5	
velar-alveolar											1				
uvular	20	2	1	1				14	11					3	
pharyngeal								3							
epiglottal	1														
glottal	4							5							
laryngeal								2							

■ ■ ■

■
■
■

2) Calcul du réseau (graphe) pour les langues L1 & L2 :

Les graphes sont obtenues en procédant comme pour l'initialisation du réseau de départ, on donne en entrées au réseau initialisé, l'ensemble des segments de chaque langue

3) Calcul de la distance entre L1 & L2 :

La distance $D(L1, L2)$ est obtenue en prenant la valeur absolue de la différence des deux graphes

$$D(L1, L2) = \sum_{i,j} |N_{i,j} (L1_{i,j} - L2_{i,j})|$$

Remarques :

❖ Deux distances peuvent être calculées :

$$D_1 L_1 - L_2 = \sum_{i,j} | 1/N_{i,j} (L_{1 i,j} - L_{2 i,j}) |$$

$$D_2 L_1 - L_2 = \sum_{i,j} | N_{i,j} (L_{1 i,j} - L_{2 i,j}) |$$

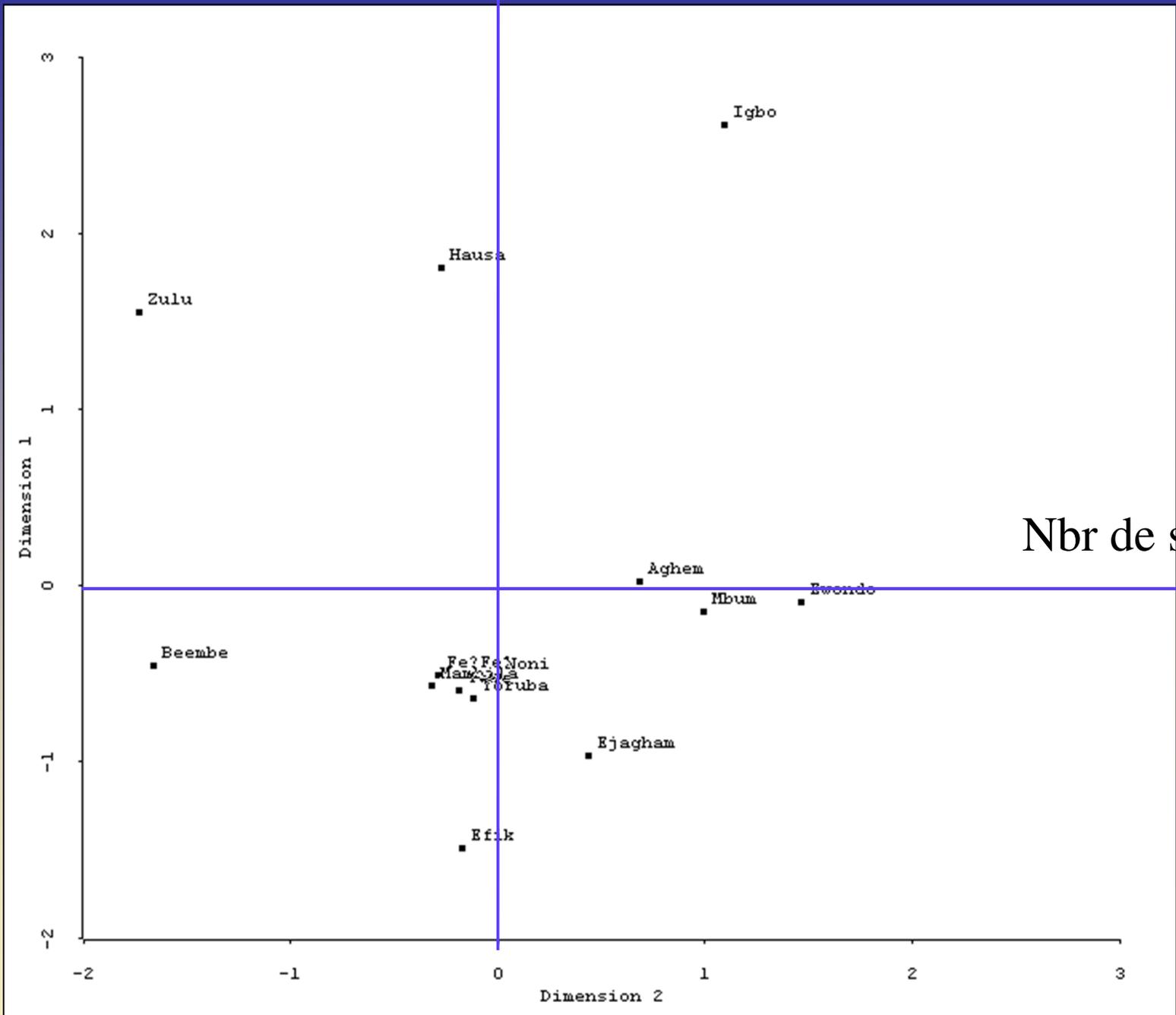
❖ Les distances vocalique et consonantique ont été calculées en plus de la distance globale

Etude de cas :

AGHEM (NK, B.)	ε ι ο υ α β e i o u	ʔ η η̄ m b ɸ v d d̄ ʒ d̄ ʒ f g γ ḡ b j k k̄ p l m n p s t t̄ ʃ t̄ s w z
BEEMBE (NK, B.)	ε ε̄ o ã a ã i i u ũ	f h k' l m n p q̄ f̄ q̄ f̄' p' s t t' t̄ s t̄ s' v
EJAGHAM (NK, B.)	ε o ʌ a i u y	β η b d d̄ f g ḡ b j k k̄ p m n n̄ p r s t t̄ w
EWONDO (NK, B.)	ε ε o a e i o u	η η̄ g η̄ m ḡ b b d d̄ ʒ f g ḡ b j k k̄ p l m m b m̄ v n n d n d̄ ʒ p s t t̄ s v w z
FE?FE? (NK, B.)	a a e i o u u	ʔ η b d d̄ ʒ f g h k m n s t t̄ ʃ v w z
MAMBILA (NK, B.)	o a e i o u	η b ɸ v d d̄ ʒ f g h j k l m n n̄ p s t t̄ ʃ w
NONI (NK, B.)	ε ε ε o o u a a e i i o o u	ʔ η ʃ b d d̄ ʒ f g u j k l m n p s t t̄ ʃ w
TEKE (NK, B.)	ɪ u a e i o u	η b ɸ v d d̄ ʒ f g h j k l m η n n̄ p q̄ f̄ s t t̄ ʃ w
ZULU (NK, B.)	ε o a i u	! s ʃ b d̄ ʒ f h h̄ ɬ j k k' k' k̄ l ʒ m n n̄ p p' p' r s t t' t̄ ʃ t' v w z
MBUM (NK, A.)	a ã e i i o u ũ	η η̄ g η̄ m ḡ b ʒ b b ɸ d d̄ f g ḡ b h j k k̄ p l m m b m̄ v n n ʒ n d n p r s t v w
EFIK (NK, CR.)	ε o a e i o u	η b d f j k k̄ p m n p s t w
IGBO (NK, K.)	ε ε̄ o ã u ũ a ã e ē i i o ã u ũ	d η η̄ m b b̄ b' b' b d d̄ ʒ d̄ ʒ f g γ g* g* g h h* j k k* k* k' l m n n̄ p b̄ p' p' p' t̄ t̄ s t d̄ t̄ ʃ t' v w z
YORUBA (NK, K.)	ε ε̄ o ã a e i i o u ũ	ʃ b d d̄ ʒ f g ḡ b h j k k̄ p l m n r s t w
HAUSA (AA, C.)	a e i o u	ʔ ɸ ʃ b b d d̄ ʒ g g' g' h j j k k' k' k' k* k* l m n r r̄ s s t t̄ ʃ w z

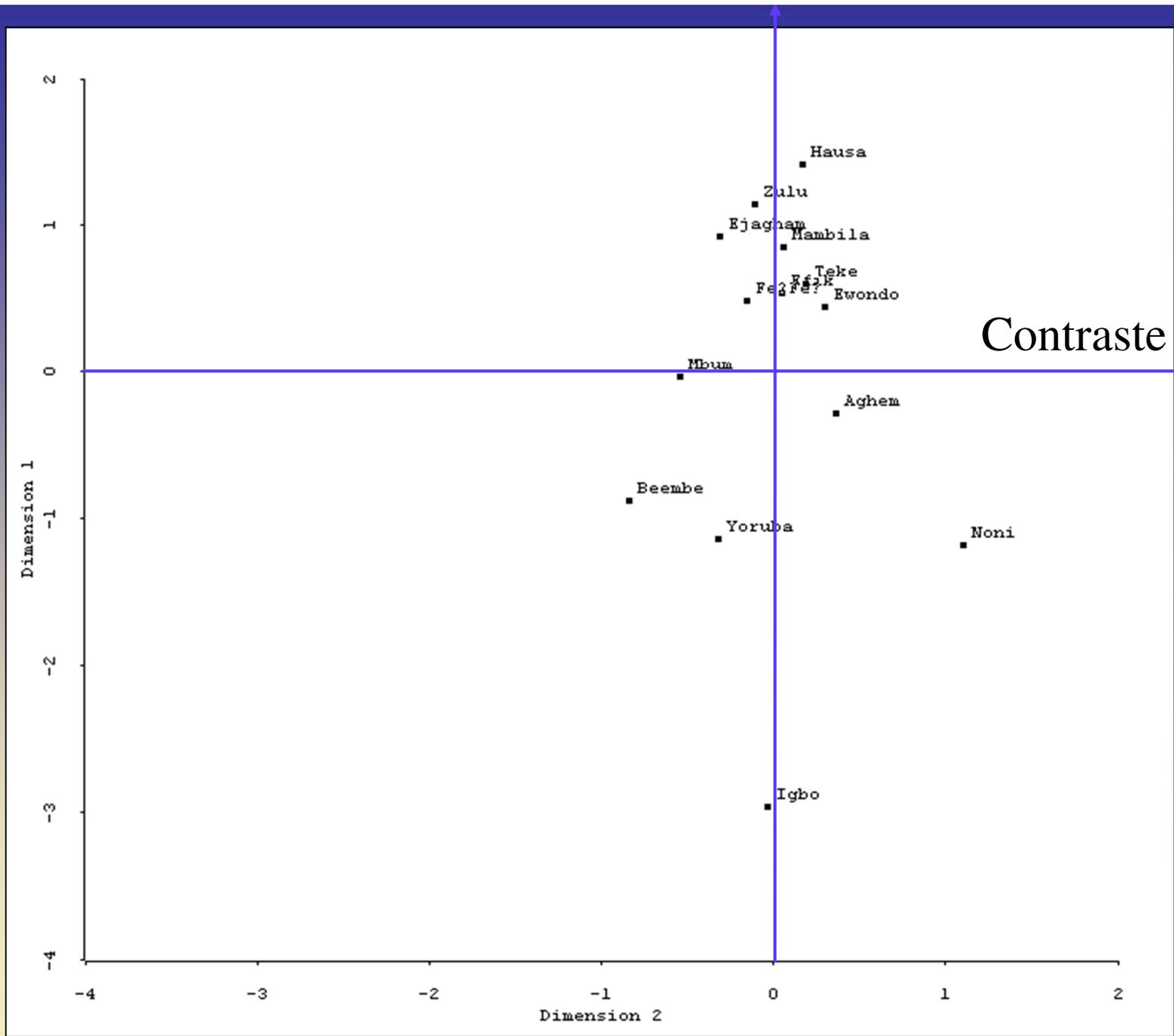
Résultats :

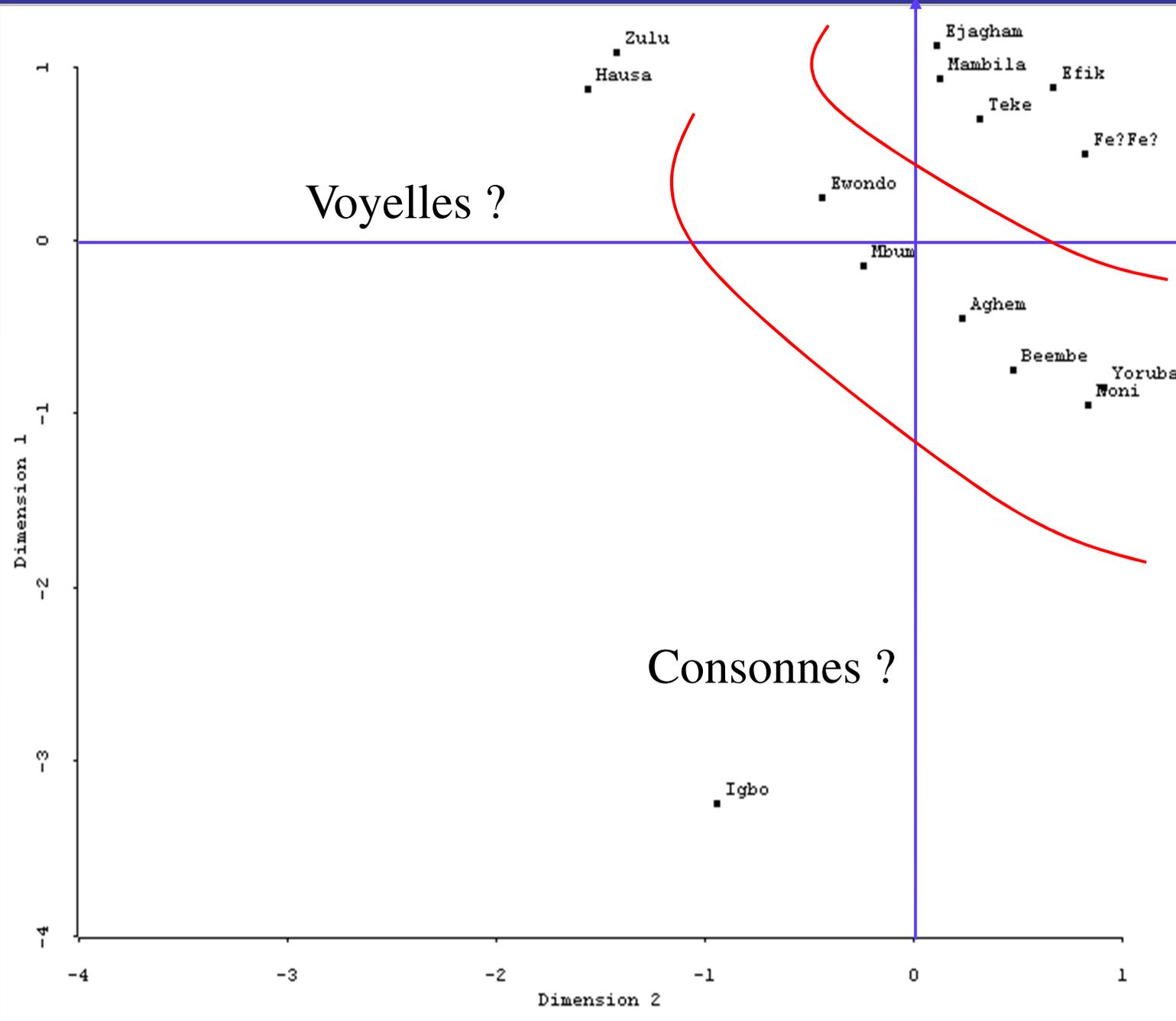
Après avoir obtenu des matrices de distance pour D1 et D2 (systèmes, voyelles & consonnes), nous avons appliqué la méthode du Multidimensional Scaling, afin d'obtenir des représentations topologiques



Nbr de segments ?

D2V





Interprétations :

Les différentes topologies observées notamment avec la mesure D2, permettent de montrer :

i) / Voyelles :

un regroupement des langues à caractéristique secondaire &/ou ayant beaucoup de timbre vs. langues sans trait secondaire et avec peu de timbres

ii) / Consonnes :

interprétation difficile du fait de la grande variabilité des systèmes consonantiques, le facteur nombre d'éléments semble important

iii) / Systèmes :

un regroupement croisant "complexité consonantique" et "complexité vocalique"

Conclusion :

- ❖ Les résultats préliminaires sont intéressants
- ❖ Nécessité d'affiner l'interprétation en considérant plus de dimensions

Perspectives :

- ❖ Développer les réseaux de niveau supérieur afin d'accéder à une hiérarchisation des segments et des systèmes et à une définition de la complexité
- ❖ Intégrer des contraintes plus "naturelles" (gestes articulatoires ?)
- ❖ Appliquer ces mesures à une analyse qualitative de l'évolution des systèmes phonologiques : complexification, simplification ou stabilité ?

