

Espaces vocaliques dans le français du *PFC*

Analyses préliminaires

Chiara Celata*, Bernard Laks[°], Atanas Tchobanov[°]

*ENS, Pise [°]MoDyCo, Paris

Journées PFC 2011

Paris , 8-10 décembre 2011

Phonologie et variation

- Dans quelle mesure les locuteurs sont sensibles à la perception des différences quantitatives fines engendrées par des règles phonologiques variables ? (e.g. Labov et al. 2011)
- Comment simuler la structuration de catégories phonologiques à partir de faits variables et hétérogènes attestés dans un corpus de langue parlée? (e.g. Laks 2008, 2011)

But général

1. modéliser l'espace vocalique du français selon une analyse multivariée qui permet de prendre en charge *plusieurs dimensions de variation* à la fois (acoustique + distributionnelle)
2. sortir un cadre *psychologiquement plausible* de cette variation
 - quels principes *topologiques* restent stables?
 - quelle *normalisation* est nécessaire pour se débrouiller dans la variation?

Buts

- Nous présentons un premier essai d'analyse multivariée pour un certain nombre de variables **acoustiques** et **lexicales/distributionnelles** qui peuvent définir l'espace de variation des voyelles orales du français.
- Exploitation massive du PFC pour l'analyse acoustique des timbres vocaliques

Dans une étape successive de la recherche:

→ catégorisation perceptive par locuteurs natifs

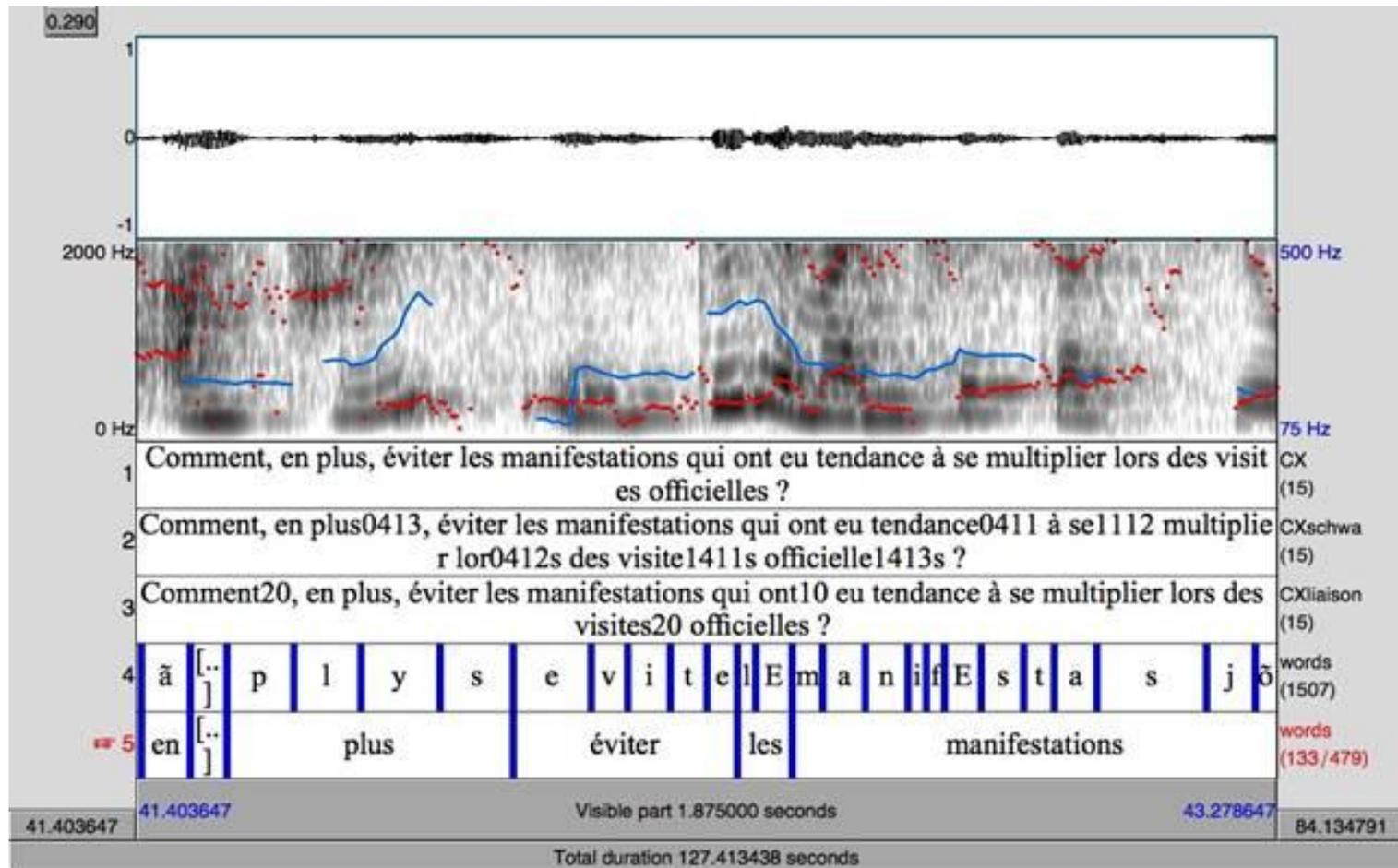
→ coût computationnel nécessaire pour la produire

Matériaux

- PFC “Texte lu”
- 69 locuteurs (25 ‘nord’, 44 sud de la France)
- 33 exemplaire vocaliques distribués sur des contextes lexicaux
(33 voyelles x 69 locuteurs = 2277 voyelles orales)

Timbre	Patron phonotactique/prosodique			Timbre	Patron phonotactique/prosodique		
	Fin de mot	Interne	Clitique		Fin de mot	Interne	Clitique
a	étape, hasard	valu	sa	e	lassé	dépêche	ses
i	titres	village		ɛ	fait, appelle, colère	bêtement	
u	jaloux	toujours		œ	jeune, seuls, nonneurs	premier	
y	valu	circuler	du	ø	jeux, jeûne		
o	côte	côté	aux	ə		depuis, indiqueraient	se
ɔ	cote, bords	local					

Analyses



Outil Formants

Recherche de phonèmes dans le texte lu de PFC

Enquête

Age min.

Age max.

Sexe

Age études

Phonème

Finale/Final
 Oui/Yes

Mot unique

Liste de mots

Liste de codes locuteur

Outil Formants

- Filtrer par Enquête, Région, Sexe, Age, Phone, Mot ou liste de mots, Codes locuteurs;
- Ecoute du segment et du mot;
- Visualisation du spectrogramme;
- Exportation des résultats en CSV (tableur);

http://www.projet-pfc/formants

Mot unique

Liste de mots

Liste de codes locuteur

Recherche de phonèmes dans le texte lu de PFC

Sexe

CodeLo...	Transc...	Contex...	Sexe	Age	Point	F0	F1	F2	F3	F4	Duree	Ecoute
11aal1	d	Dépêche	M	75	120.13	101	334	1190	2564	3825	0.04	Ecouter
11aal1	e	Dépêche	M	75	120.185	100	383	1966	2654	3831	0.07	Ecouter
11aal1	p	Dépêche	M	75	120.265	100	376	1319	2366	3566	0.09	Ecouter
11aal1	E	Dépêche	M	75	120.34	95	437	1888	2420	3224	0.06	Ecouter
11aal1	S	Dépêche	M	75	120.43	99	489	1976	2451	3137	0.12	Ecouter
11adp1	d	Dépêche	F	23	110.71...	190	431	1734	3081	4296	0.09	Ecouter
11adp1	e	Dépêche	F	23	110.79...	198	377	2014	2852	4564	0.08	Ecouter
11adp1	p	Dépêche	F	23	110.865	-0	164	1587	2864	4226	0.06	Ecouter
11adp1	E	Dépêche	F	23	110.945	259	670	2093	3118	4240	0.1	Ecouter
11adp1	S	Dépêche	F	23	111.06	497	2343	2829	3964	5120	0.13	Ecouter
11agm1	d	Dépêche	M	76	115.543	153	392	1335	2263	3776	0.05	Ecouter
11agm1	e	Dépêche	M	76	115.598	158	371	2330	2653	4076	0.06	Ecouter

http://www.projet-pfc/formants

Mot unique

Liste de mots

Liste de codes locuteur

Recherche de phonèmes dans le texte lu de PFC

Sexe

CodeLo...	Transc...	Contex...	Sexe
11aal1	d	Dépêche	M
11aal1	e	Dépêche	M
11aal1	p	Dépêche	M
11aal1	E	Dépêche	M
11aal1	S	Dépêche	M
11adp1	d	Dépêche	F
11adp1	e	Dépêche	F
11adp1	p	Dépêche	F
11adp1	E	Dépêche	F
11adp1	S	Dépêche	F
11agm1	d	Dépêche	M
11agm1	e	Dépêche	M

Voir les détails d'un phonème

Code PFC du locuteur: 11aal1

Mot: Dépêche

Phonème: E

Ecoute Mot:

Ecoute Phonème:

Formants en PDF: [Télécharger](#)

	Duree	Ecoute
	0.04	Ecouter
	0.07	Ecouter
	0.09	Ecouter
	0.06	Ecouter
	0.12	Ecouter
	0.09	Ecouter
	0.08	Ecouter
	0.06	Ecouter
	0.1	Ecouter
	0.13	Ecouter
	0.05	Ecouter
	0.06	Ecouter

Analyses

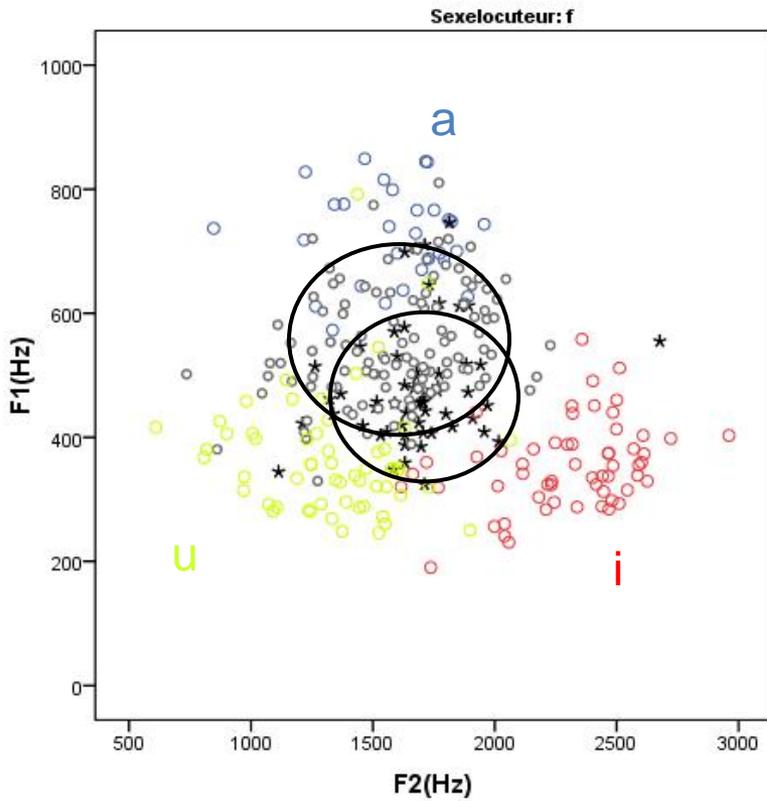
- Procédure d'extraction automatique de F1, F2, F3, f0, durée
 - coordonnées standard en *PRAAT* 5.1.32
- Inspection et filtrage manuel des valeurs formantiques
- Procédure de normalisation formantique en Bark (Traunmüller 1997)
- Analyse statistique en *SPSS* 16.0
 - analyse des dispersions
 - modèle de classification par analyse discriminante

Résultats

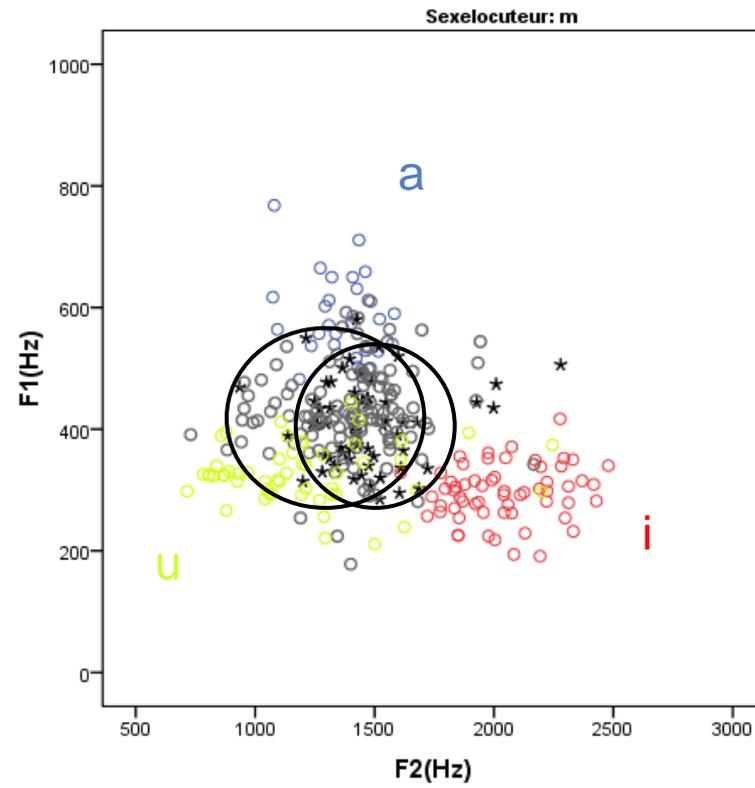
1) Aires d'existence de /ə ø œ/

/a i u ø œ/ en position finale de mot

femmes



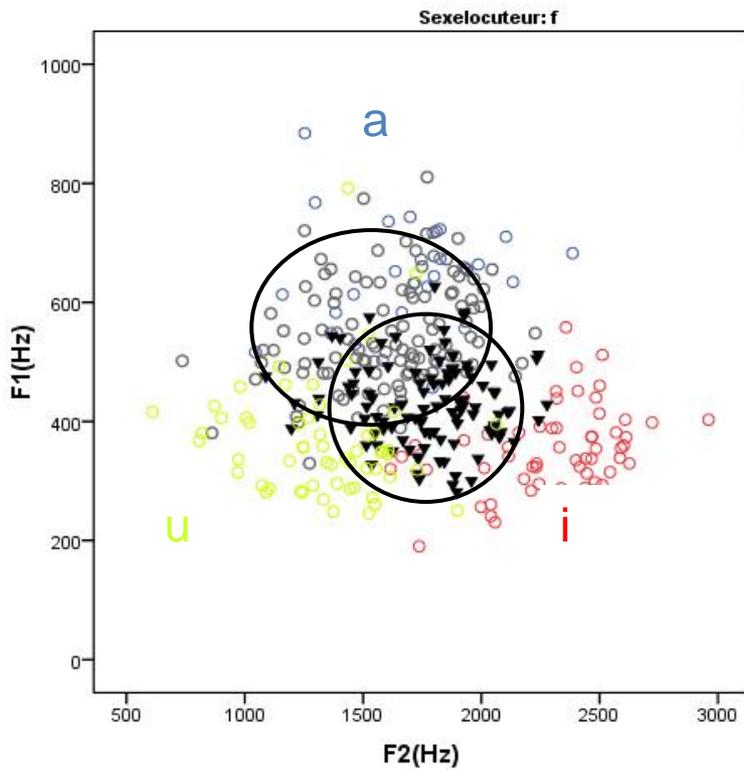
hommes



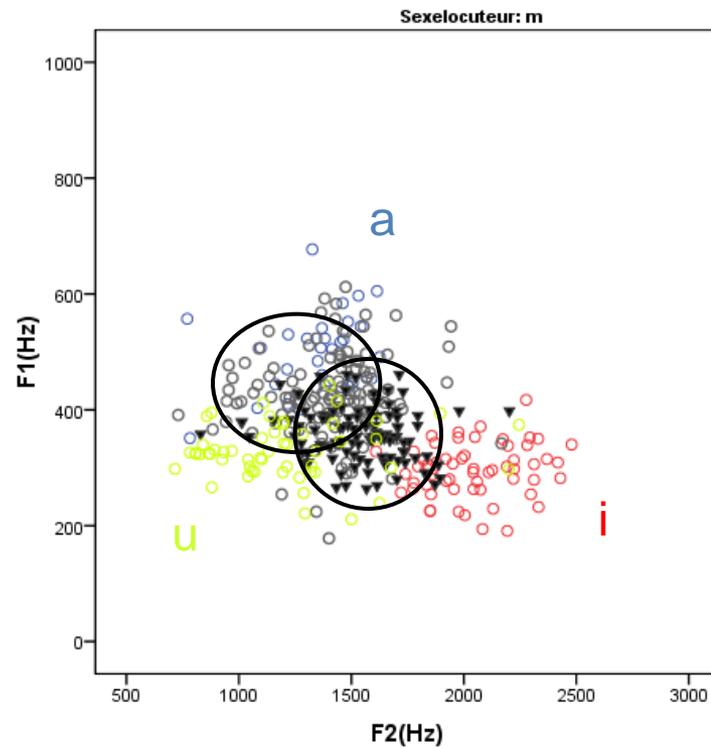
* /ø/
○ /œ/

/a i u œ ə/ en position interne de mot

femmes



hommes

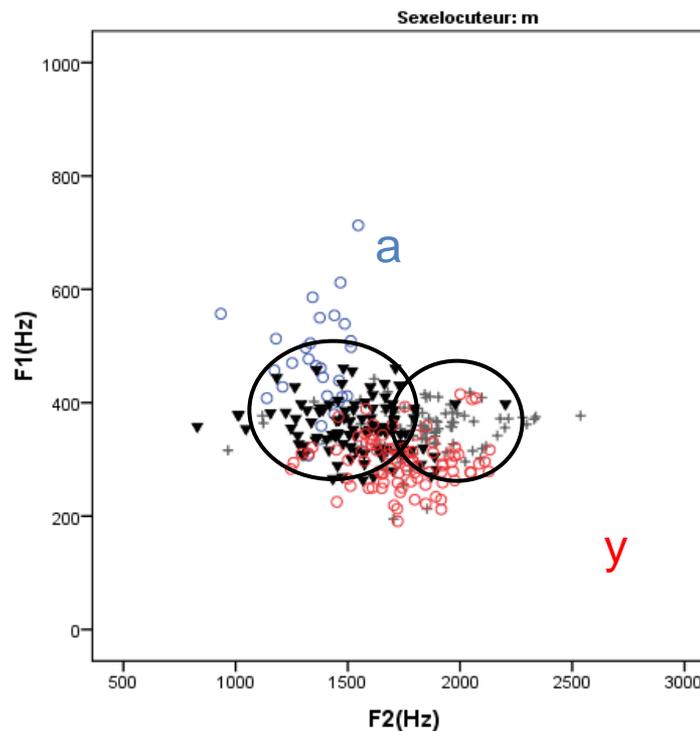
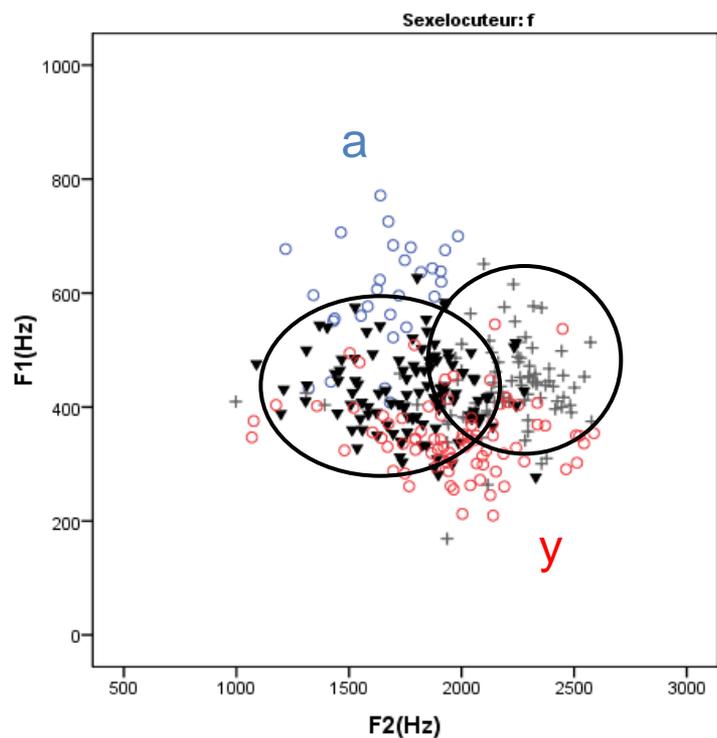


▼ /ə/
○ /œ/

/a y e ə/ dans les mots clitiques

femmes

hommes



▼ /ə/
+ /e/

Les contextes sont empiriquement différents

Les catégories se distribuent d'une façon
non-uniforme sur les différents contextes

La forme (ou *topologie*) des relations peut être conservée
(là où le contraste est maintenu)

Résultats

2) Classification par analyse discriminante

L'analyse discriminante permet de modéliser la valeur d'une variable catégorielle dépendante sur la base de sa relation avec un ou plus prédictors.

Les fonctions sont générées à partir d'un échantillon de données pour lesquelles l'appartenance catégorielle est connue;

les fonctions peuvent ensuite être appliquées aux cas nouveaux qui ont des mesures pour les prédictors mais dont l'appartenance catégorielle est inconnue.

Dans notre modèle,

les **catégories** → les 33 contextes lexicaux

les **variables** prédictives → F1-F2-F3 (Bark), max-f0 (Bark), durée (msec)

On évalue si les paramètres acoustiques disponibles permettent au système de récupérer une partition catégorielle des différents timbres vocaliques dans leur contexte spécifique (phonotactique et prosodique)

Le modèle est forcé à catégoriser des exemplaires dans un espace qui est très parcellisé

L'analyse des erreurs de catégorisation révèle les forces de l'attraction exercée sur les différents exemplaires, en fournissant au même temps des informations sur les prédicteurs les plus efficaces

→ Deux modèles, locuteurs du 'nord' vs locuteurs du Midi

Sommaire des fonctions discriminantes canoniques

Locuteurs du 'nord'

- F1
- F2
- durée

Function	% of Variance	Cumulative %	Canonical Correlation
1	64.1	64.1	0.869
2	20	84.1	0.700
3	15.9	100.0	0.650

- F1
- F2
- durée
- F3

Function	% of Variance	Cumulative %	Canonical Correlation
1	61.7	61.7	0.896
2	25.8	87.6	0.794
3	6.7	94.3	0.555
4	5.7	100.0	0.523

Locuteurs du Midi

Prédictions d'appartenance catégorielle, timbres moyens antérieurs/centraux

Locuteurs du 'nord'

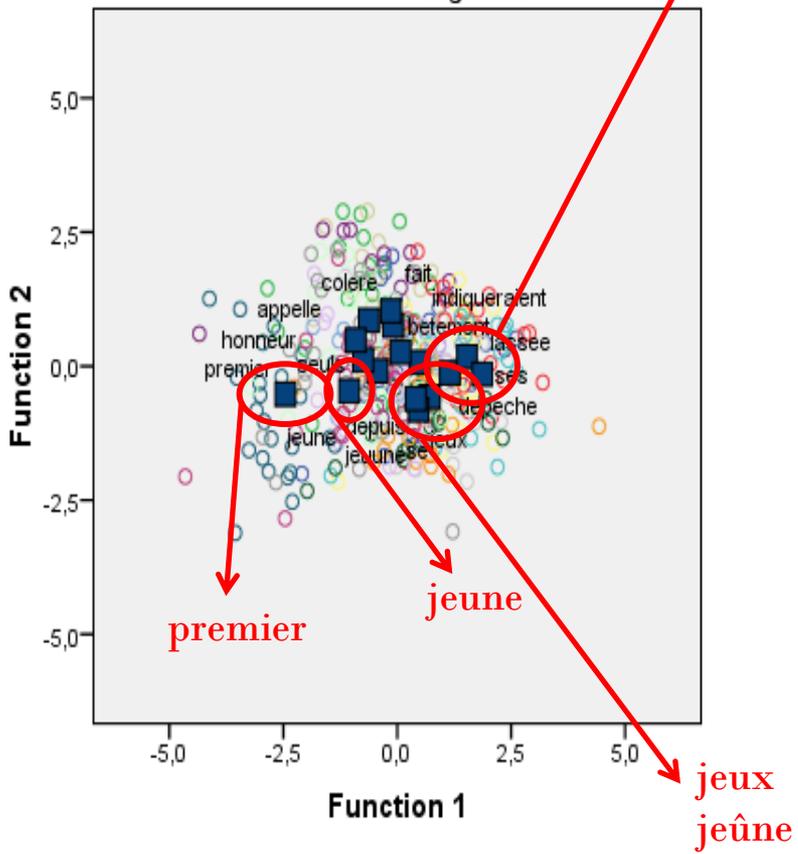
Classification Results																	
Predicted Group Membership																	
MotN	betement	appelle	fait	colere	depeche	lassee	ses	depuis	indiqueraient	se	jeux	jeune	premier	seuls	jeune	honneur	Total
%	betement	33.3	11.1	11.1	0	0	0	0	33.3	0	0	0	11.1	0	0	0	100
	appelle	0	33.3	0	16.7	16.7	0	0	16.7	0	0	0	0	0	16.7	0	100
	fait	22.2	0	33.3	11.1	22.2	0	0	0	0	0	0	0	0	0	11.1	100
	colere	0	0	25	50	0	0	0	0	0	0	0	0	0	0	25	100
	depeche	0	0	11.1	0	22.2	0	33.3	0	11.1	11.1	0	0	11.1	0	0	100
	lassee	0	0	0	0	0	37.5	25	0	12.5	0	25	0	0	0	0	100
	ses	0	0	0	0	16.7	33.3	33.3	0	0	0	0	0	0	16.7	0	100
	depuis	0	0	0	0	0	14.3	57.1	0	0	28.6	0	0	0	0	0	100
	indiqueraient	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	100
	se	0	0	0	0	37.5	0	0	12.5	0	12.5	12.5	0	12.5	0	12.5	100
	jeux	0	11.1	0	0	0	11.1	0	33.3	0	22.2	22.2	0	0	0	0	100
	jeune	0	12.5	12.5	12.5	0	0	0	12.5	12.5	0	12.5	25	0	0	0	100
	premier	0	0	0	0	0	0	0	10	0	0	0	80	0	0	10	100
	seuls	14.3	0	0	0	0	14.3	0	0	28.6	0	0	0	0	14.3	28.6	100
	jeune	0	25	0	0	0	0	0	25	0	12.5	0	12.5	0	25	0	100
	honneur	16.7	0	16.7	0	0	0	0	0	0	0	0	16.7	0	0	50	100

Classification Results																		
Predicted Group Membership																		
MotN	betement	appelle	fait	colere	depeche	lassee	ses	depuis	indiqueraient	se	jeux	jeune	premier	seuls	jeune	honneur	Total	
%	betement	15.4	7.7	7.7	7.7	0	0	7.7	0	23.1	0	0	0	0	15.4	15.4	0	100
	appelle	18.8	37.5	12.5	18.8	0	0	0	0	0	0	0	0	0	6.2	6.2	0	100
	fait	7.7	0	46.2	0	15.4	23.1	0	0	0	0	0	0	0	0	0	7.7	100
	colere	7.1	14.3	14.3	28.6	7.1	7.1	0	0	7.1	0	0	7.1	0	0	7.1	0	100
	depeche	0	0	15.4	0	23.1	30.8	7.7	15.4	0	0	0	0	7.7	0	0	0	100
	lassee	0	0	0	0	16.7	66.7	0	8.3	0	0	0	0	0	0	0	8.3	100
	ses	0	0	0	0	0	45.5	36.4	9.1	0	0	0	0	0	9.1	0	0	100
	depuis	0	0	0	0	0	0	23.1	61.5	0	7.7	0	0	7.7	0	0	0	100
	indiqueraient	0	0	0	0	0	0	0	7.7	23.1	7.7	7.7	7.7	23.1	7.7	7.7	7.7	100
	se	0	0	0	0	0	0	0	23.1	7.7	38.5	15.4	0	7.7	0	0	7.7	100
	jeux	0	0	0	0	0	0	0	30	10	10	30	0	20	0	0	0	100
	jeune	0	7.1	0	0	0	0	0	7.1	7.1	0	28.6	7.1	7.1	21.4	7.1	0	100
	premier	0	0	0	0	0	7.7	0	0	7.7	7.7	15.4	53.8	0	0	0	7.7	100
	seuls	6.2	0	6.2	6.2	0	0	0	0	0	6.2	6.2	6.2	25	25	12.5	0	100
	jeune	0	0	0	12.5	0	0	0	0	0	6.2	0	18.8	12.5	43.8	6.2	0	100
	honneur	0	9.1	0	0	0	0	0	0	18.2	9.1	0	0	0	36.4	27.3	0	100

Locuteurs du Midi

Canonical Discriminant Functions

Region: oil

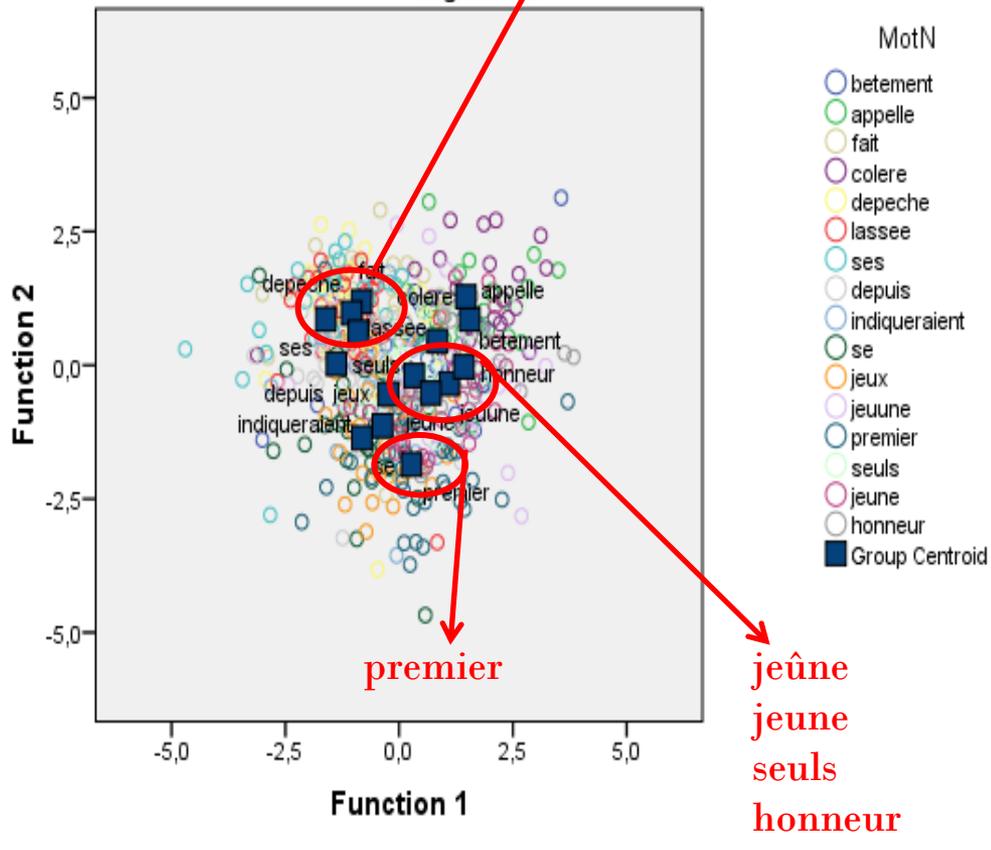


Locuteurs du 'nord'

lassé
dépêche
ses

Canonical Discriminant Functions

Region: sud



Locuteurs du Midi

lassé
dépêche
ses
+ fait

- MotN
- betement
 - appelle
 - fait
 - colere
 - depeche
 - lassee
 - ses
 - depuis
 - indiqueraient
 - se
 - jeux
 - jeune
 - premier
 - seuls
 - jeune
 - honneur
 - Group Centroid

Procédures pour la ‘normalisation cognitive’ de l’espace vocalique

- Réductions de dimensionnalité
- Quantification des différences:
cibles différentes pour accents différents dans
le même espace de relations

Discussion et conclusion

Notre but est celui de parvenir à une représentation de l'espace vocalique du français

à partir de **variables lexicales et patrons prosodiques/phonotactiques** qui définissent certains contextes d'occurrence.

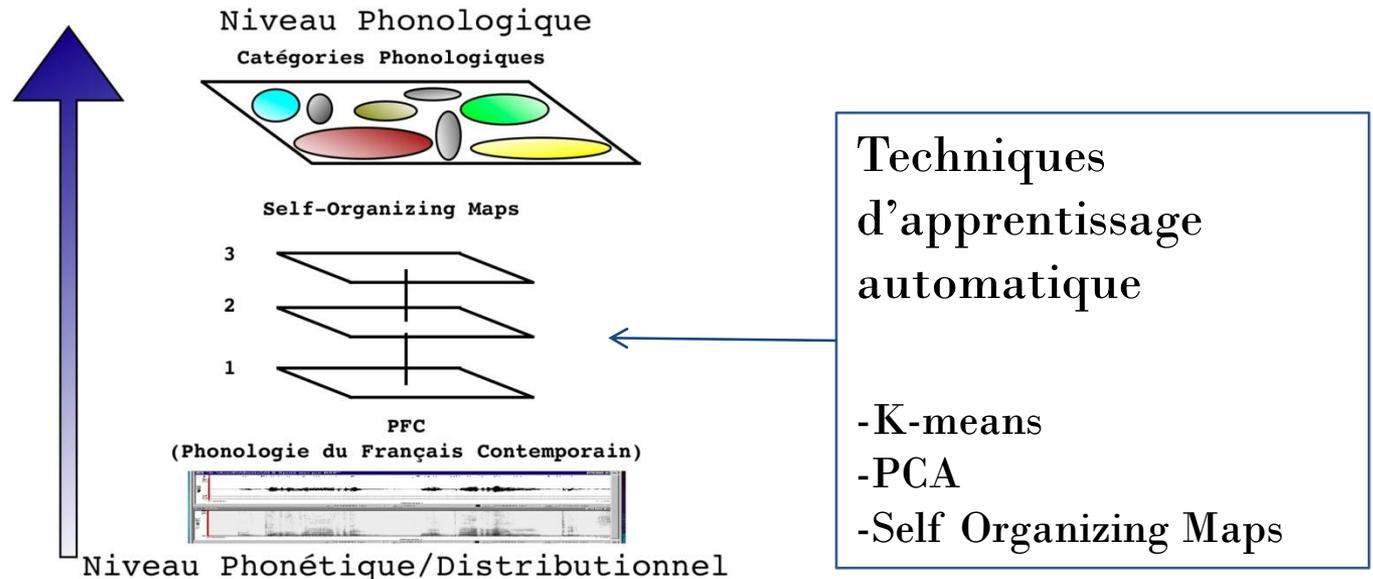
Les contextes, nécessairement en nombre limité pour pouvoir être traités dans une analyse discriminante, simulent sur une échelle extrêmement réduite l'espace vocalique de la langue comme il est implémenté dans la variété lexicale.

Le traitement de la variabilité du timbre de schwa (par rapport à /œ ø/) montre d'une façon optimale que toute question portant sur la variation doit se mesurer avec le problème de la *normalisation d'un espace multidimensionnel*

Le *nombre* et la *nature* des variables qui entrent dans un espace de variation ne peuvent pas être décidées a priori

Ce travail est en cours de développement selon plusieurs pistes:

- Extension à un plus grand nombre de locuteurs, variétés régionales/accents, styles de parole, contextes etc.
- Jugements perceptifs subjectifs, pour tester la validité de la catégorisation issue des données de corpus
- Aspects computationnels



**Projet financé par PICS2011/13-CNRS
et GR2011 n°23, SNS, Pise**

Nos remerciements à

B. Calderone (CLLE-ERSS, CNRS/ Université de Toulouse)

C. Bertini (SNS, Pise)

Bibliographie

- Labov W., S. Ash, M. Ravindranath, T. Weldon, M. Baranowski & N. Nagy (2011) “Properties of the sociolinguistic monitor”. *Journal of Sociolinguistics* 15/4: 431-463.
- Laks B. (2008) “Pour une phonologie de corpus”. *Journal of French Language Studies* 18/1: 3-32.
- Laks B. (2011) “Pourquoi y-a-t-il de la variation plutôt que rien?” Ms., Université de Paris Ouest – Nanterre La Défense.
- Fougeron C., C. Gendrot & A. Bürki (2007). On the phonetic identity of French schwa compared to /ø/ and /œ/. In *Schwa(s)*, 5èmes Journées d'études linguistique, Nantes, 2007. 191-197.
- Nguyen, N. & Z. Fagyal (2008) “Acoustic aspects of vowel harmony in French”. *Journal of Phonetics* 36/1: 1-27.
- Traunmüller, H. (1997) “Auditory scales of frequency representation”. [Online: <http://www.ling.su.se/staff/hartmut/bark.htm>]