

UMR 6057 CNRS



Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

Approche méthodologique et théorique de l'étude des segments en parole spontanée

Christine Meunier

Laboratoire Parole et Langage
CNRS – Université de Provence
Aix-en-Provence

I_PHOM

Analyser l'interaction : de la phonétique à la multimodalité

14 au 18 juin 2010 - Lyon

Les segments en parole spontanée

1. Phonétique et parole "naturelle"

- ▶ L'analyse phonétique
- ▶ Les grands corpus de parole

2. Transcription et alignement du signal

- ▶ Transcription, phonétisation, alignement
- ▶ Annotation manuelle et automatique

3. Quelques résultats sur le français

1. Phonétique et parole naturelle



L'analyse phonétique

- ▶ Objectifs: analyser les sons de la parole dans leurs aspects physiques et physiologiques de façon à éclairer les mécanismes de production et de perception de la parole.
- ▶ Méthodes: acoustique, articulatoire (IRM, radiocinéma, etc.)
- ▶ L'analyse phonétique n'est pas simplement descriptive, mais s'inscrit dans des questionnements généraux sur l'utilisation du langage par les humains et sur les systèmes linguistiques.

Aperçu historique

- ▶ Descriptions phonétiques sur des séquences limitées ou contrôlées (syllabes, mots, phrases)
- ▶ Apparition des outils et techniques expérimentaux:
 - Représentations anatomiques de la parole: radiocinématographie, palatographie, IRM, EMA, etc.
 - Mesures acoustiques et articulatoires: spectrographie, DAB et DAN, PSG, etc.
 - Perception: paradigmes expérimentaux contrôlés (chronométrie mentale)



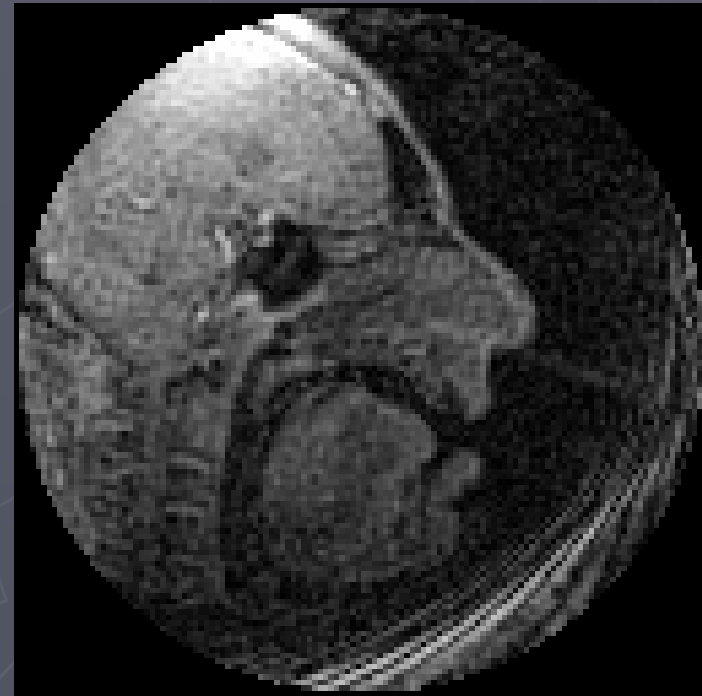
Grandes avancées dans les descriptions articulatoires, acoustiques et perceptives des sons de la parole

Et maintenant...?

- ▶ Que dire de ces connaissances/descriptions dans le mécanisme de production et de perception des sons en situation "naturelle"?
 - Production: cibles prototypiques souvent non atteintes → hypoarticulation, réduction (Duez, 1995, 2003; Johnson, 2004; Pluymaekers et al., 2005)
 - Perception: peu de travaux car méthodologie très complexe

La parole : un phénomène rapide et complexe

- 🔊 200 mots par minute
- 🔊 12 à 15 sons par seconde
- 🔊 Texte



Lindblom (1990): Theory Hypo and Hyper-articulation Theory (H &H)

Parole en situation "naturelle"

- ▶ Parole contrôlée/parole naturelle
 - Informations véhiculées
- ▶ Les conséquences
 - Différences: débit, éloignement des prototypes, occurrences des segments
 - Difficultés: quantité des facteurs de variations, pbs méthodologiques (techniques et statistiques)

Éloignement des prototypes

Détection de la voyelle: isolée (extraite), en contexte syllabique, en contexte lexical (Meunier & Floccia, 1999)

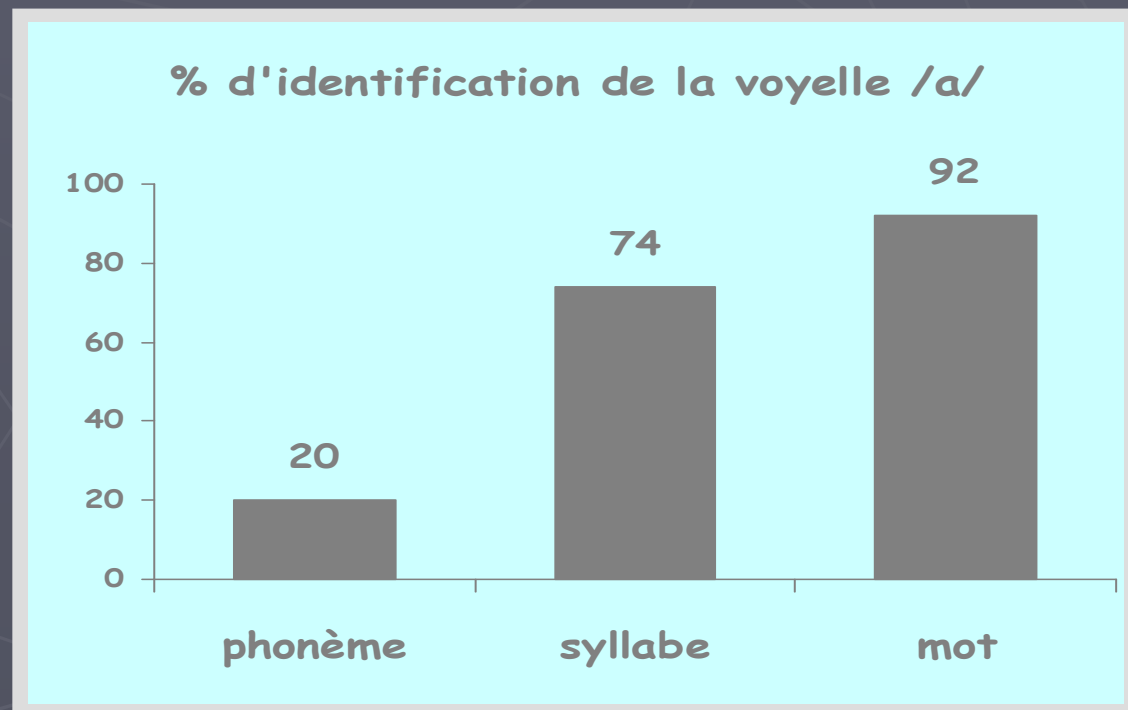
pho



syll



mot



Parole conversationnelle



A après après deux bières là à neuf heures c'est bon je crois que
Y on avait plein d'anecdotes ouais
A tu
A des euh
A tu décoinces
A ouais bon insolite alors insolite qu'est ce qui peut y avoir
Y insolite
Y y avoir d'insolite
A bon puis y a des trucs qui faut qu'on peut quand même pas trop dire quoi insolite euh
A alors
Y euh
A remarque on est pas obligé de parler de trucs
Y non
A complètement insolites hein
Y non non
A je crois que
A c'est pas bien grave
Y non
A alors racontes moi la allez
A ce qui est insolite dans le da
Y non j'ai pas voulu non
A t'avais pas voulu
Y non

Johnson (2002)

- 1 son altéré: 60% des mots
- Au moins 2 sons altérés: 28% des mots
- 1 son omis: 20% des mots
- Au moins 2 sons omis: 5% des mots

/a/

Pourquoi travailler sur des Grands Corpus de Parole?

- ▶ Nécessité d'appréhender la production et la perception dans une dimension plus large
 - Modèles moins limités
 - Compréhension des mécanismes de variation
- ▶ Les phénomènes observés ainsi que la quantité des sources de variation nécessitent de grandes quantités de données
 - Exemple: la durée (phonème, syllabe, prosodie, position lexicale, CMS, discours, etc.)
 - Pertinence des phénomènes observés
 - "Hiérarchie" des indices

Les GCP

- ▶ Multilocuteurs
- ▶ Souvent plusieurs heures de parole
- ▶ Parole "naturelle" ... ou pas?
 - Quel objet linguistique?
 - Le compromis entre « naturalité » et qualité acoustique
- ▶ Français: PFC, ESTER, **CID**, Nijmegen

Annotation multiniveau

- ▶ Phonétique (segmental)
- ▶ Phonétique (supra-segmental)
- ▶ Syllabique
- ▶ Morpho-syntaxique
- ▶ Discursif
- ▶ Gestuel (vidéo)
- ▶ Etc.

2. De la transcription vers l'alignement du signal



Les étapes de conversion

▶ Transcription (manuel)

- Du signal sonore vers une représentation graphique symbolique (via le système perceptif du transcripteur)

▶ Phonétisation (automatique)

- Des symboles orthographiques vers des symboles phonétiques (via un système de règles)

▶ Alignement (automatique)

- Des symboles phonétiques vers des étiquettes localisées sur le signal de parole (via l'identification des segments grâce à des modèles acoustiques)

La transcription

▶ Types

- Phonétique
- Orthographique standard
- Orthographique standard enrichie

▶ Conséquences

▶ Limites

- Le transcripteur doit se limiter à transcrire (pas d'analyse du signal)

Transcribed event	Notation	Representation	Example
Inaudible sequence	Star	*	il a *
Non-writable sequence	Star	[*, sampa]	[*, pf]
Non-linguistic noises	Star	*	*
Several possible transcriptions	Between brackets	<ortho1, ortho2, ...>	<oui, ouais>
Graphic variants	Between braces	{variante 1, ... variante N}	{il chante, ils chantent}
Specific pronunciation	Between square brackets	[ortho, sampa]	[je sais, Se]
Omitted phonemes	Between parentheses	(ortho)	je t(e) le dis
Silent pause	Plus sign	+	demain + ou après-demain
Unusual liaison	Between equal signs	=letter=	Quatre =z= amis
Missing required liaison	Hash mark	#	deux # échecs
Truncated word	Final dash	ortho-	eche-
Laughters	Arobase	@	@
Speech while laughing	Between arobases	@@ ortho @@	@@ demain @@
Proper name	Between dollar symbol	\$ortho, P/\$ \$ortho, T/\$	\$Jean, P/\$ \$Londres, T/\$
Acronym	Between dollar symbols	\$ortho, S/\$	\$RTM, S/\$
Title	Between quotation marks	"ortho"	"rois et reines"
Reported speech	Between paragraph symbols	§ sequence §	il m'a dit § je serai là §

Phonétisation

- ▶ Principe: la transcription orthographique est fournie à un convertisseur graphème-phonème qui s'appuie sur un dictionnaire de formes fléchies et un ensemble de règles modifiables (exception, liaison, etc.). En sortie, le convertisseur fournit une suite de 'tokens' phonétisés (en SAMPA).

Ortho. "ouais, mais en fait là tu sais pas"

Phonétique [wɛmɛãfɛtlatysɛpa]

SAMPA wE mE a~ fEt la ty sE pa

<http://www.phon.ucl.ac.uk/home/sampa/>

Alignement

- ▶ À partir de la séquence de phonèmes et du signal audio, l'aligneur fournit en sortie la localisation temporelle de chaque phonème sur le signal
- ▶ Modèles acoustiques de phonèmes du français standard
- ▶ En parole conversationnelle, l'aligneur fonctionne dans des conditions difficiles

Automatique versus manuel

- ▶ Principes de l'alignement automatique
 - Modèles acoustiques des phonèmes
 - Régularité
 - + grande quantité de données
 - imprécisions, erreurs, "cécité"
- ▶ Principes de la segmentation manuelle
 - Représentations des réalisations acoustiques
 - Adaptabilité
 - + précisions, adaptation, découverte de nouveaux phénomènes (conglomérats, etc)
 - temps, coût, faible quantité de données, irrégularité (?)
- ▶ Conserver les avantages de chacun...

Phénomènes spécifiques

- ▶ Métaplasme (Adda-Decker et al, 2008)

"Je crois que c'est quelque chose"

[ʃrwaksekʃoz]

- ▶ Conglomérats:

- La réduction n'est pas un phénomène uniquement paradigmatique

- Fusion des segments

- Préservation d'un "socle" articulatoire?

- ▶ → top-down et bottom-up

3. Quelques travaux sur le français



Domaines d'interaction

▶ Suprasegmental (prosodie)

- Phonologie
- lexique
- Syntaxe
- Discours
- Pragmatique

▶ Segmental (phonèmes)

- phonologie
- Morphologie
- lexique

Questions

► Questions linguistiques

- Comment fonctionne la communication orale?
- Quelles interactions?
- Comment fonctionne l'équilibre.

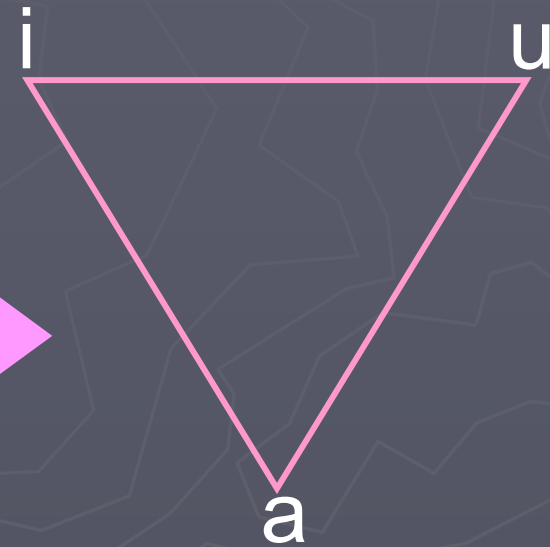
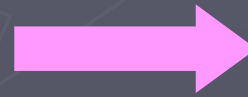
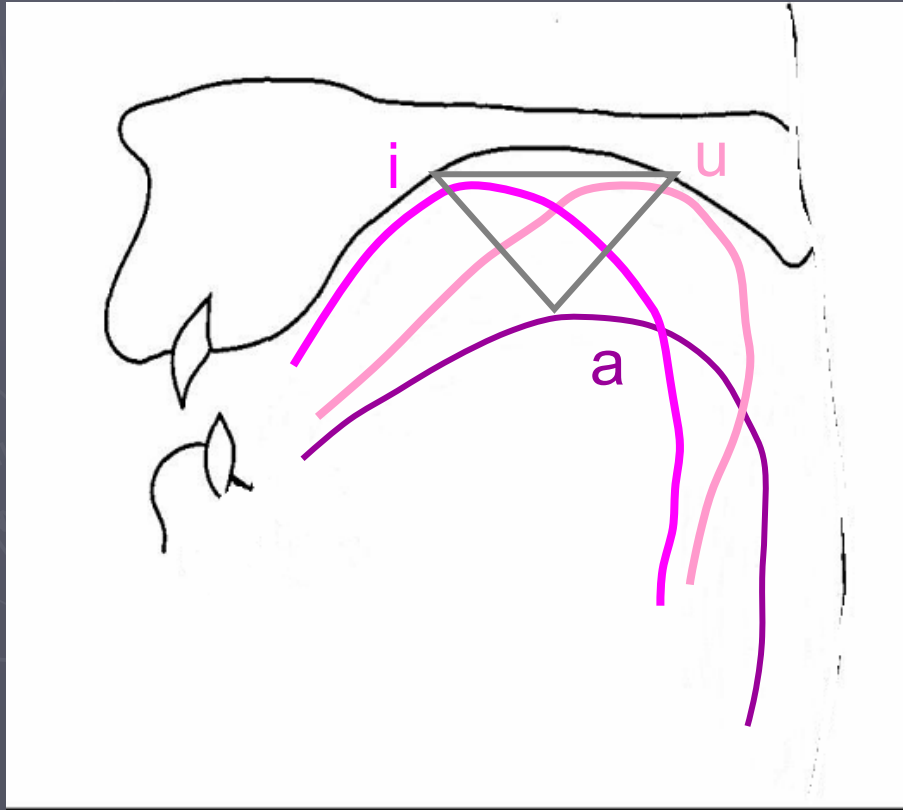
► Questions phonétiques

- Les variations phonétiques comme "marqueurs" des propriétés de plus haut niveau?
- Interactions entre contraintes physiologiques et contraintes linguistiques
- Universalité des phénomènes de réduction

La réduction

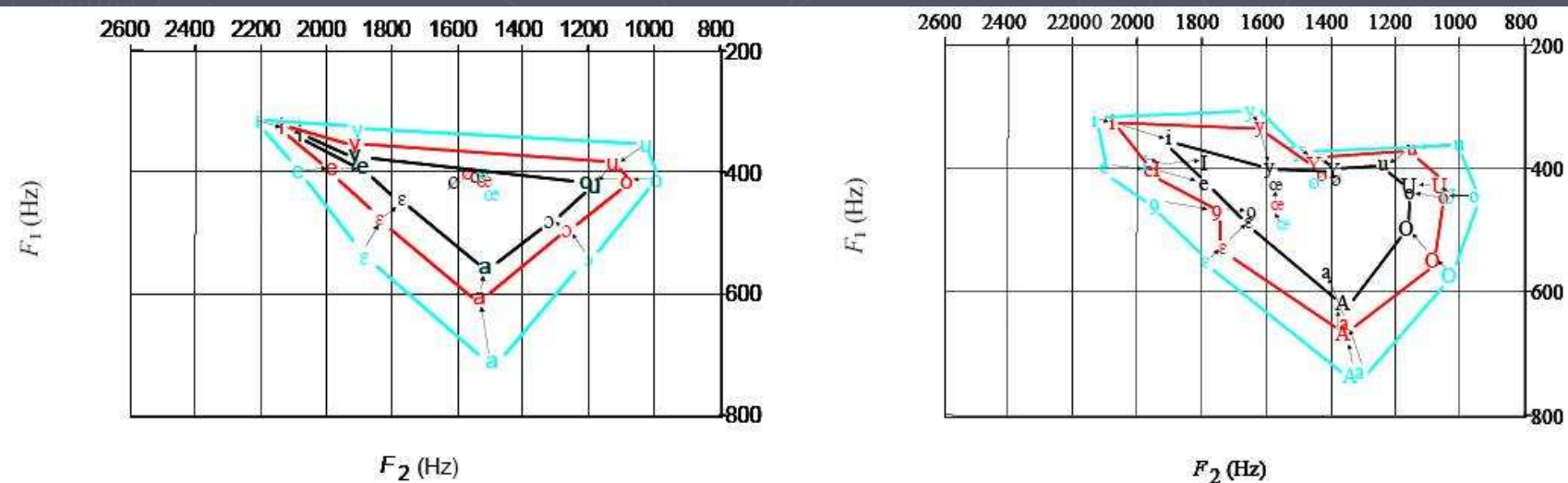
- ▶ Phénomène "phare" de la phonétique segmentale en parole spontanée
 - Réduction et assimilation des consonnes occlusives du français (Duez, 1995,2001)
 - Réduction vocalique:
 - ▶ Gendrot & Adda-Decker (2005, 2007,2008)
 - ▶ Meunier & Espesser (2006,2008)
 - Torreira, Adda-Decker & Ernestus (The Nijmegen Corpus of Casual Speech, 2010)

Triangle vocalique



Réduction vocalique

- ▶ Lindblöm (1963): réduction (centralisation) des voyelles inaccentuées (donc brèves) du suédois lorsque le débit augmente
- ▶ Gendrot & Adda-Decker (2005, 2007): centralisation des voyelles françaises et allemandes corrélée avec la diminution des durées



Figures 2- 3: Measures mean average values of F1 and F2 for French (left) and German (right) vowels according to their duration. By ascending order (black [30 – 50ms], red [60 - 80], blue [90 - 110]). Male and female results are merged.

Objectifs

- ▶ Réduction vocalique:
 - outil d'observation des variations
 - Lié à l'hypoarticulation
 - Variations graduelles
- ▶ Corrélé avec:
 - Facteurs lexicaux?
 - Structure prosodique?
 - Organisation discursives?

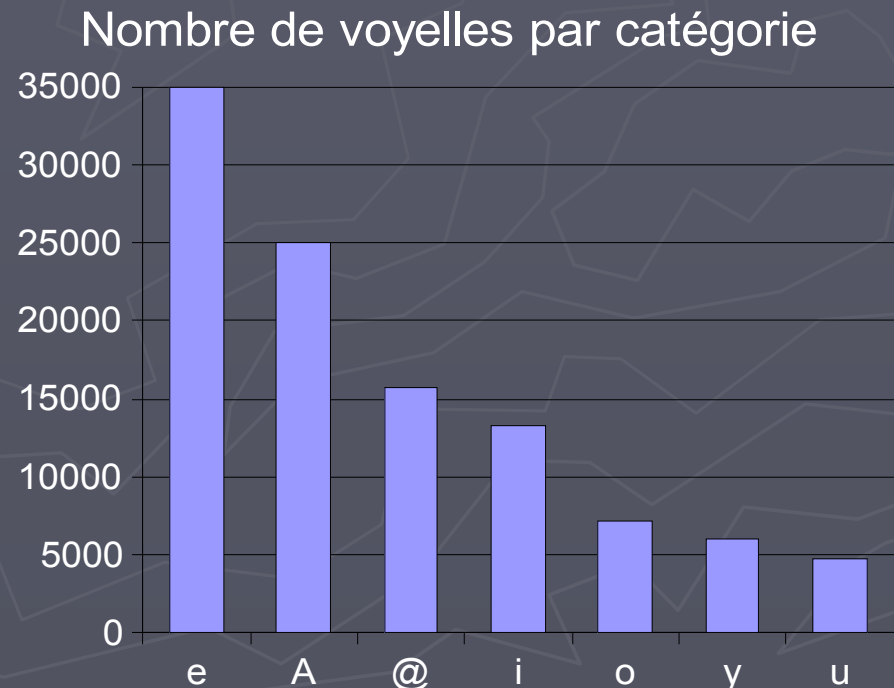
Corpus

- ▶ CID: **C**orpus of **I**nteractional **D**ata (Bertrand & al., 2007)
- ▶ 16 dialogues d'1h chacun (6 hommes, 10 femmes)
- ▶ conversation familière
- ▶ Annotation multi-niveau
 - Orthographique, phonétique, morpho-syntaxique, prosodique, gestuel

Les voyelles dans le CID: occurrences et représentativité

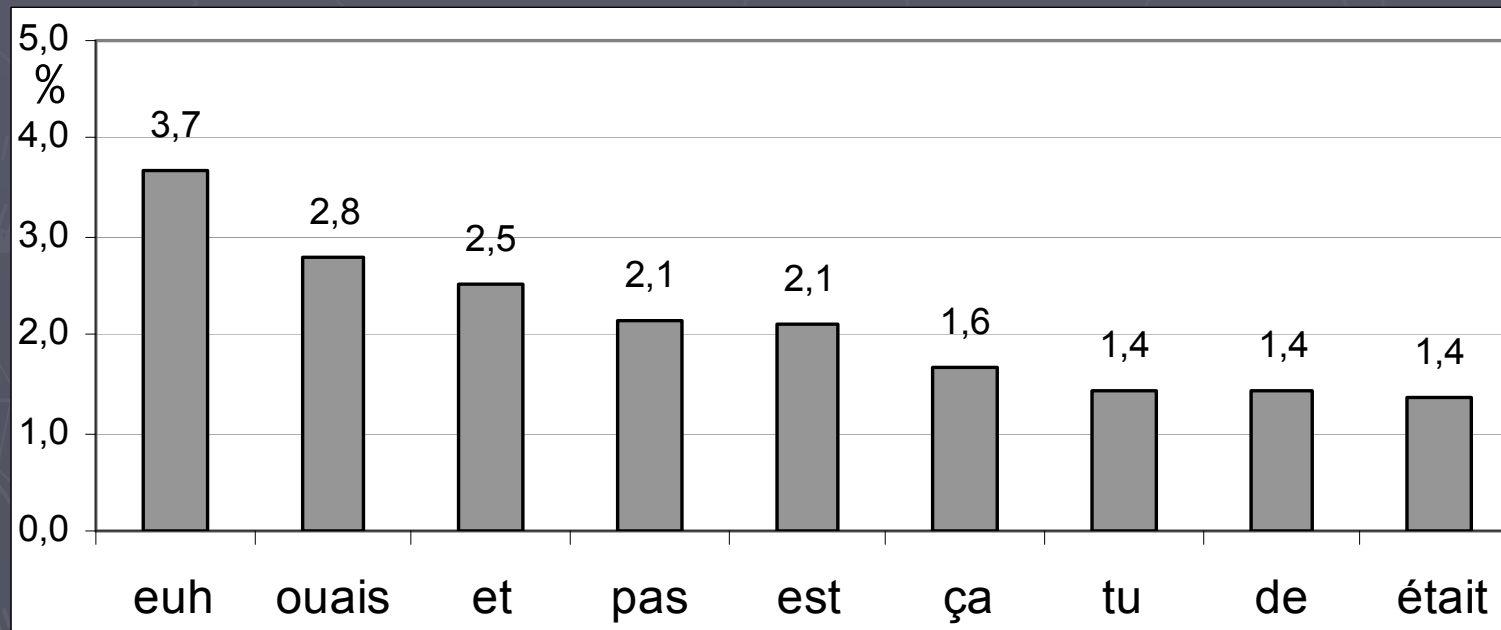
- ▶ Aligneur (LORIA): 7 macro-classes de voyelles (A,e,@,o,i,y,u)
- ▶ 107 000 voyelles orales (272 000 phonèmes; 39%)
- ▶ e + A = 56% des voyelles du corpus (+@ = 71%)

Étant donné le regroupement en macro-classes, cette répartition est globalement conforme aux données de la littérature



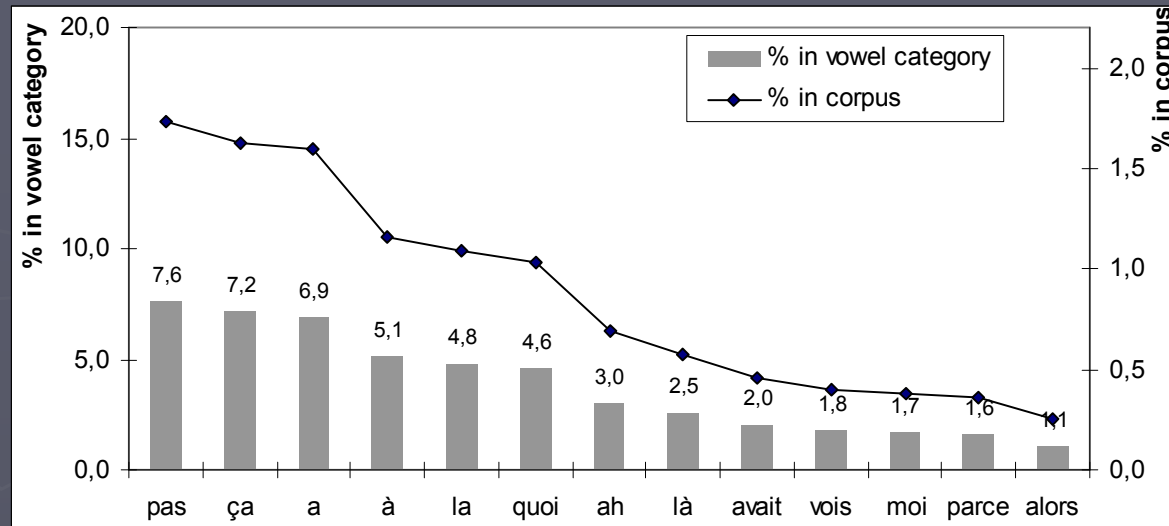
Les voyelles dans les mots

- ▶ La plupart des voyelles sont produites dans un nombre très restreint de mots (50% sont produites dans moins de 70 mots différents, et 10% dans 9 mots différents).

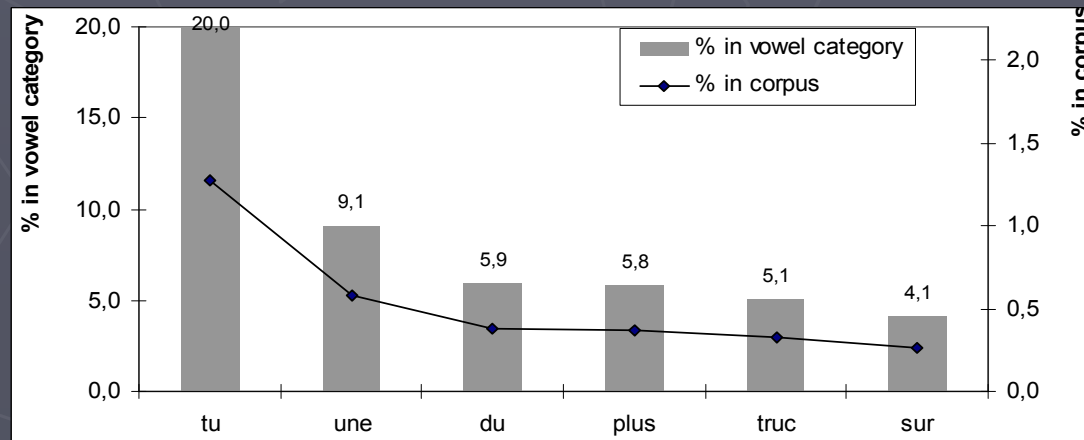


Occurrence des voyelles en parole conversationnelle

50% des /a/
sont produits
dans 13 mots



50% des /y/
sont produits
dans 6 mots



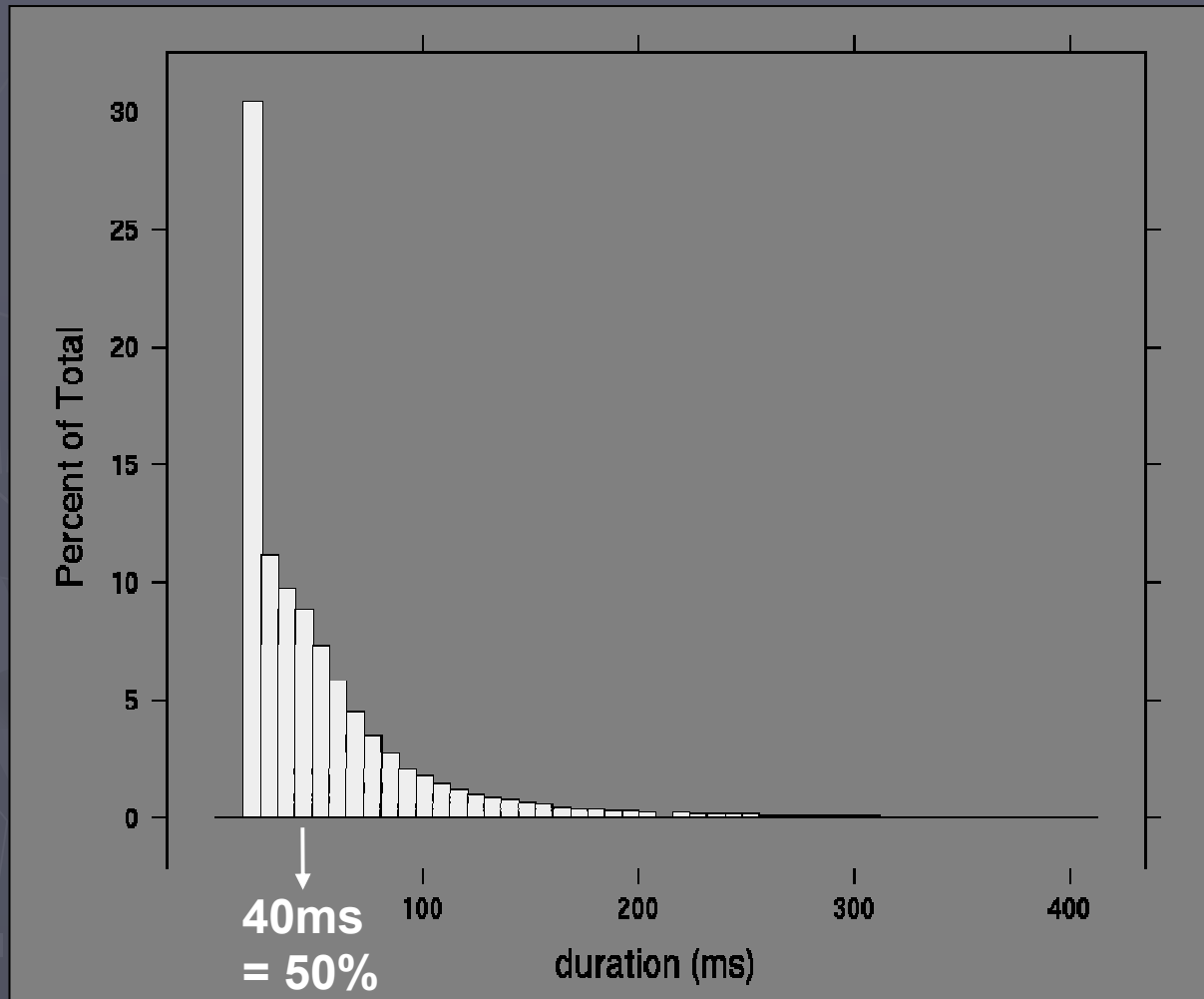
→ Mots fonction monosyllabiques

Catégories morpho-syntaxiques

- ▶ Content words (CW)
 - Nouns, verbs, adverbs, adjectives
- ▶ Function words (FW)
 - Conjunctions, auxiliaries, pronouns, prepositions
- ▶ Other words ("euh", "ouais") are extracted from the analysis

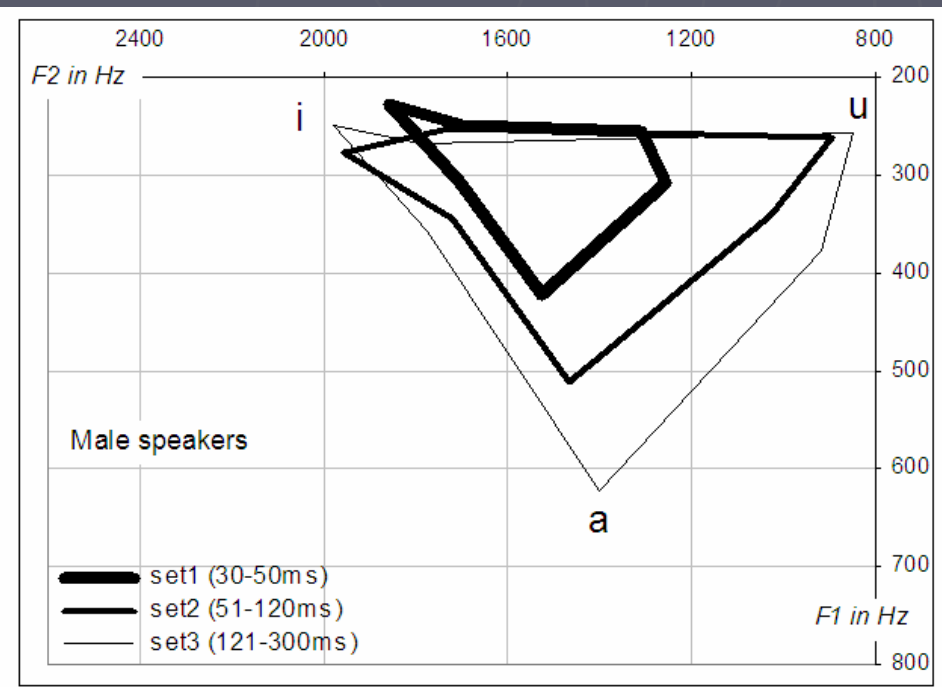
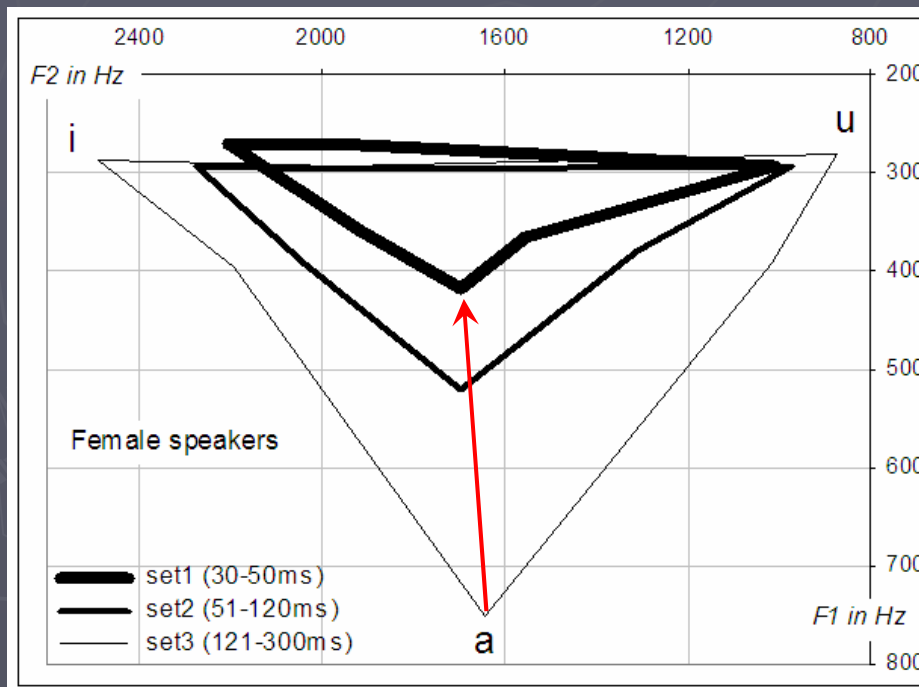
Mesures acoustiques

DISTRIBUTION DES VOYELLES EN FONCTION DE LEUR DUREE



→ timbre?

Hypo-articulation des voyelles en fonction de leur durée



Femmes (CID) 10 locutrices

Hommes (CID) 6 locuteurs

Réduction vocalique et facteurs lexicaux

- ▶ **Fréquence lexicale:** cette réduction (hypo-articulation) est-elle corrélée avec la fréquence des mots produits?
- ▶ **Catégorie lexicale:** cette réduction est –elle dépendante du type de mot produit?

Fréquence et catégorie lexicale

- ▶ Sélection des mots monosyllabiques (CV et CVC)
- ▶ Fréquence des mots
- ▶ Deux catégories: CW (Content Words) et FW (Function words)
- ▶ Hypothèses
 - les mots fréquents devraient être plus réduits (Pluymakers et al., 2005; Ernestus et al., 2006)
 - Les FW devraient être plus réduits (van Bergem, 1993)
- ▶ Durée des 7 voyelles et timbre de /a/

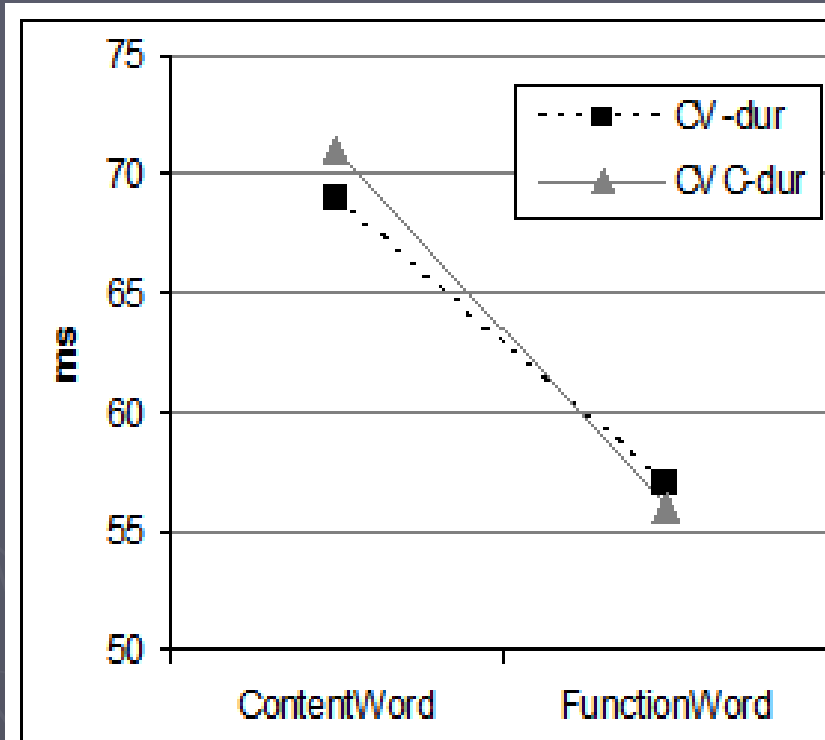
Fréquence

- ▶ Effet de la fréquence sur la durée des voyelle → seulement pour les CVC
 - mots de haute fréquence=64ms
 - mots de basse fréquence=71ms
 - ▶ Différence=7ms
- (résolution de l'aligneur=8ms)

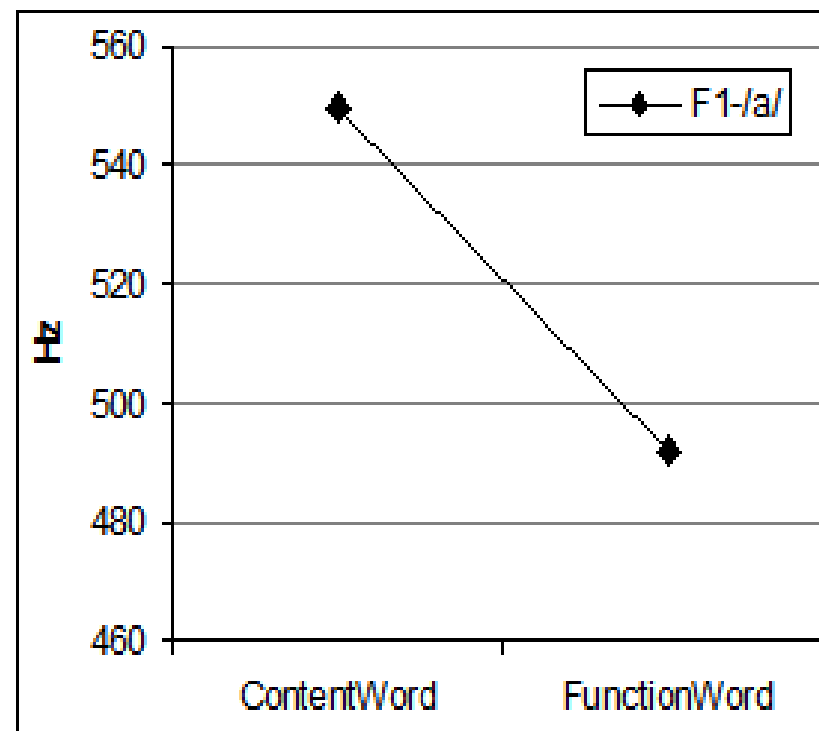


Peu pertinent pour l'interprétation phonétique

Catégories lexicales



Durées



Timbre du /a/

→ les voyelles produites dans des mots fonction sont plus hypo-articulées...

... Mais

Perspectives

- ▶ Plus que les catégories morpho-syntaxiques, les fonctions discursives/pragmatique du lexique pourraient jouer un rôle déterminant dans les phénomènes de réduction

