

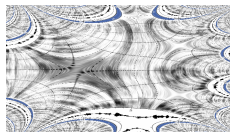
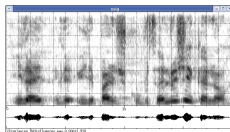
Explorations du corpus PFC à partir d'alignements forcés

Quelques travaux en cours

collaborateurs P. Boula de Mareüil, R. Nemoto, C. Gendrot,
P. Hallé, N. Nguyen, J. Durand, N. Snoeren

Martine Adda-Decker

LIMSI/CNRS



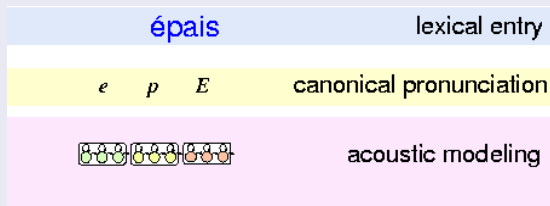
- Motivation:
 - améliorer la modélisation de la parole dans systèmes de transcription automatique
 - tenir compte des variantes de prononciation
 - lesquelles? étude de la variation
- Réduction de parole / Speech reduction
 - réduction vocalique
 - assimilation
 - phénomènes complexes
- Exploration de grands corpus
 - alignement forcé, pron. canonique
 - alignements avec variantes ciblées
- Corpus
 - parole journalistique ESTER
 - conversations PFC



Acoustic word modeling

- via pronunciation dictionaries : $(w; \Phi)$
- few variants
- $\hat{w} = \arg \max_w P(w/x) = \arg \max_w p(x/w, \Phi) P(w)$

- 1 lexical entry w
- 2 pronunciation Φ
- 3 acoustic word model HMM



French ESTER2 (2008) and Quaero (2009)

| | %WER | %sub | %del | % ins | comment |
|---------------|-------------|------|-----------|-------|--|
| French Ester2 | ~ 12 | 7.5 | 3 | 1.5 | news, debates |
| Quaero 2009 | ~ 23 | 11 | 11 | 2 | higher proportion of interactive speech |

Observation:

for Quaero 2009 deletion rates increase more than the other error types

Major error sources:

(near)-homophones (French) and reduced pronunciations (style: spontaneous, repetitive, casual, interactive)



REF

tout le temps
temps de leur installation
chaque année de travail en moins
quai de Seine
c'était le même marasme
confiance appréciable le tandem

HYP

tout _ temps
temps _ leur installation
chaque année **travaillons moins**
quête saine
c'**est elle** même marasme
confiance appréciable _ tandem

deletion of grammatical words : *le, de...*

substitutions (*quai de* ↔ *quête*) (schwa deletion + assimilation)

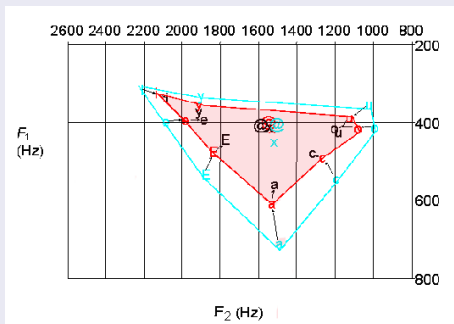


Voyelles orales

collaboration avec Cédric Gendrot, Univ. Paris 3
mesures formantiques en fonction de la **durée**
données journalistiques (extrait du corps ESTER)
(Interspeech, 2005)
résultats similaires sur les données PFC

Triangles vocaliques du français établis automatiquement à partir de corpus journalistiques (coll. Cédric Gendrot)

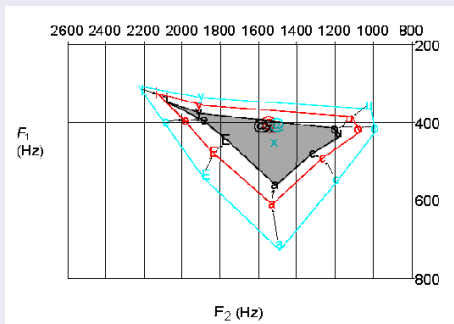
Mesures formantiques en fonction de la durée



$60 \text{ ms} \leq \text{durée} \leq 80 \text{ ms}$

Triangles vocaliques du français établis automatiquement à partir de corpus journalistiques (coll. Cédric Gendrot)

Mesures formantiques en fonction de la durée



$30 \text{ ms} \leq \text{durée} < 60 \text{ ms}$

Assimilation de voisement

collaboration avec Pierre Hallé, Univ. Paris 3 et Natalie Snoeren, post-doc LIMSI

données journalistiques (extrait du corps ESTER)

alignement avec variantes spécifiques

mesure des taux d'alternance de voisement

(ICPhS, 2007)

Voicing assimilation in French

Assimilation is known to be **regressive** (Grammont 1939, Rigault 1967):

Given a consonant sequence (on word boundaries) $C_1\#C_2$ with voice features $f_1\#f_2$.

if $(f_1 = -f_2)$ then $(f_1 \leftarrow f_2)$; i.e. the voice feature is inherited from the following consonant.

Voicing distinctive for oral plosives & fricatives:

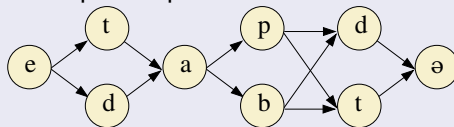
/p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/.

Example: *étape de* /etap#də/ → [etab#də]
NV#V → V#V

Speech alignment

make use of voice alternation specific pronunciation variants

étape de →



vérifier expérimentalement que les variantes de voisement n'arrivent qu'en contexte d'assimilation

Example

Example of alignment result:

sept jours /sɛt#ʒuʁ/
(seven days)

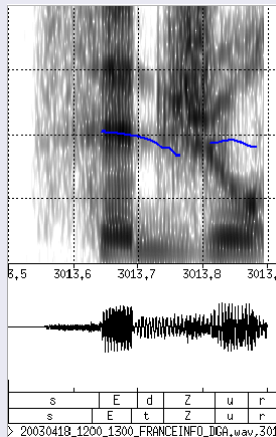
C1#C2 = /t#ʒ/

assimilating NV#V context

ʒ = "Z"

spectrogram labels

surface form labels →
underlying form labels →



Résultats

| %VA | # |
|-----|------|
| 2.7 | 28.7 |
| | 1060 |

%VA : voice alternation rate

: frequency counts in k ((*1000))

nb. of alternations

total nb. of potential alternations;

for different conditions: overall, word-internal, word-boundary.

VA rates as a function of phoneme

| ϕ | overall %VA # | internal %VA# | boundary %VA # | w-start %VA # | w-final %VA # |
|--------|--------------------|--------------------|--------------------|--------------------|-------------------|
| p | 1.7 1.8k 113.4k | 1.2 0.4k 34.7k | 1.8 1.44 78.7k | 1.5 1.1k 76.1k | 10.3 0.2k 2.5k |
| b | 3.2 1.17 36.4k | 2.5 0.4k 18.7k | 4.0 0.7k 17.7k | 3.8 0.6k 16.3k | 6.3 0.00k 1.4k |
| t | 2.6 4.4k 172.7k | 1.3 1.4k 111.4k | 4.8 2.9k 61.3k | 1.6 0.4k 28.2k | 7.5 2.4k 33.1k |
| d | 3.3 5.7k 174.1k | 2.8 0.9k 33.7k | 3.4 4.7k 140.3k | 3.1 4.1k 132.6k | 5.0 1.0k 19.9k |
| k | 1.5 1.9k 131.9k | 0.9 0.44 47.4k | 1.8 1.5k 84.5k | 1.1 0.7k 65.1k | 3.5 0.8k 24.3k |
| g | 4.2 0.8k 20.1k | 3.1 0.3k 11.7k | 5.7 0.48 8.3k | 5.0 0.37 7.4k | 11.4 0.11 0.9k |

| | overall | internal | boundary | w-start | w-final |
|---|--------------------|-------------------|--------------------|-------------------|--------------------|
| f | 1.7 0.7k 43.9k | 1.2 0.1k 16.6k | 2.1 0.5k 27.2k | 1.1 0.2k 24.2k | 9.6 0.29 3.0k |
| v | 2.3 1.3k 57.6k | 1.7 0.59 34.7k | 3.2 0.7k 22.8k | 2.5 0.5k 20.1k | 8.0 0.2k 2.7k |
| s | 2.0 3.9k 202.6k | 1.1 0.9k 93.5k | 2.7 2.9k 109.0k | 1.1 0.9k 88.8k | 7.4 2.27 30.8k |
| z | 9.6 5.2k 54.6k | 5.7 1.3k 23.9k | 12.6 3.8k 30.6k | 6.8 0.00k 1.0k | 12.8 3.8k 29.6k |
| S | 2.5 0.3k 15.9k | 1.6 0.00k 6.0k | 3.1 0.3k 9.9k | 1.4 0.00k 6.7k | 6.8 0.2k 3.1k |
| Z | 2.8 1.0k 36.7k | 2.2 0.3k 16.5k | 3.2 0.6k 20.2k | 2.7 0.4k 15.8k | 5.1 0.2k 5.4k |

Observations on VA rates:

overall: lower for NV cons. than for V ones.

w-internally: lower than on w-boundaries.

w-start: lower than on w-ends.

w-end: tend to be lower for V than for NV.

Résultats

| <i>Number of assimilation, control contexts on word boundaries</i> | |
|--|--------------|
| assimilation | #occurrences |
| NV#V | 13k |
| V#NV | 5k |
| control | #occurrences |
| NV#NV | 17k |
| V#V | 6k |

VA rates for C1 (w-final) and for C2 (w-initial):

- 2 assimilating conditions NV#V, V#NV,
- 2 corresponding control conditions V#V, NV#NV

Résultats

| | %VA rate | <i>cond.</i> | %VA rate |
|----------------|----------|--------------|----------|
| | C1 | C1#C2 | C2 |
| <i>control</i> | 9 | NV#NV | 1 |
| <i>assimil</i> | 24 | NV#V | 4.5 |
| <i>assimil</i> | 20 | V#NV | 1 |
| <i>control</i> | 4 | V#V | 3 |

| | %VA rate | cond. | %VA rate |
|----------------|----------|-------|----------|
| | C1 | C1#C2 | C2 |
| <i>control</i> | 9 | NV#NV | 1 |
| <i>assimil</i> | 24 | NV#V | 4.5 |
| <i>assimil</i> | 20 | V#NV | 1 |
| <i>control</i> | 4 | V#V | 3 |

C1 consonant:

Strong tendency to regressive assimilation for both NV#V (24%), V#NV (20%). Slight asymmetry in favour of NV#V: NV becomes more often V than the reverse.

Weak tendency of C1-voicing independently of regressive assimilation: NV easier changes to V (9%) than the opposite (4%).

C2 consonant:

VA rates are very low: **stability of word-start C.**

Cross-condition comparison: two weak tendencies

1) **C2-devoicing** (3% on V#V, 4.5% on NV#V).

2) **progressive** assimilation for NV#V (4.5%)

V-easier changes to NV (3%) than the opposite (1%).



Durées des voyelles

collaboration avec Cédric Gendrot, Univ. Paris 3 et Noël Nguyen, LPP,
Univ. d'Aix-en-Provence

données journalistiques (extraits du corpus ESTER)

lecture et entretiens (extraits du corpus PFC)

accent de mot: allongement de la syllabe finale

alignement forcé

(TAL, 2008)



Durées des voyelles

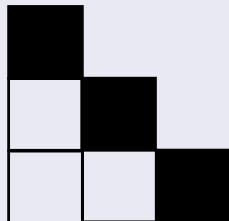
accent de mot: allongement de la syllabe finale

mesure des durées de voyelles en fonction de leur position syllabique dans le mot

mot à 1 syllabe

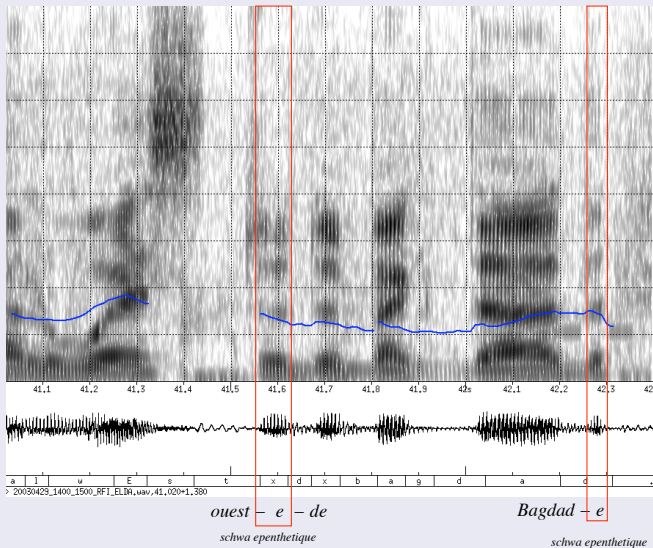
mot à 2 syllabes

mot à 3 syllabes



représentation de Pierre Delattre *Comparing the phonetic features of English, Spanish, German and French*, 1965.

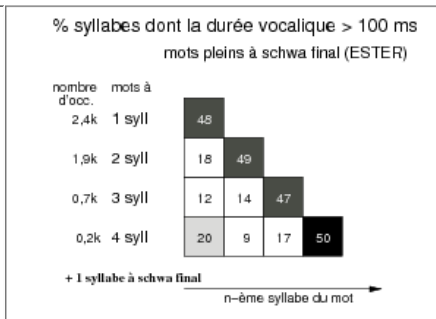
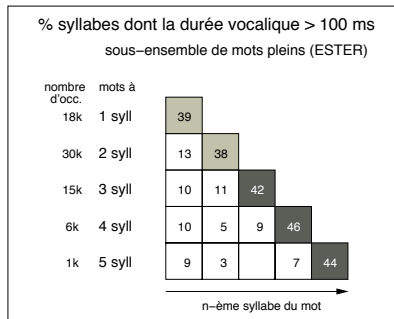
Exemple de segmentation : à l'ouest de Bagdad



corrélat acoustique : durée syllabique (vocalique)

Mots lexicaux: sans/avec schwa final

Taux de segments vocaliques dont la durée est $>$ à 100 ms

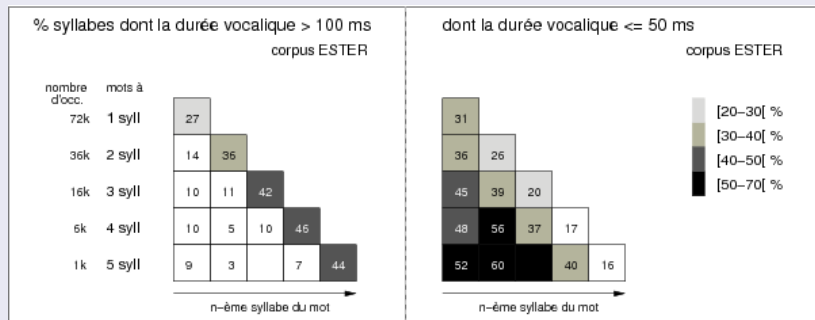


Représentation similaire à celle de Delattre.

Taux de syllabes longues en fin de mot augmente avec la longueur du mot
Taux de syllabes longues plus élevé si schwa final.



Ensemble des mots: voyelles longues / courtes

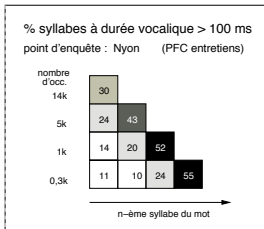
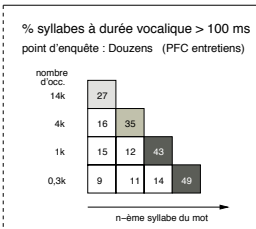
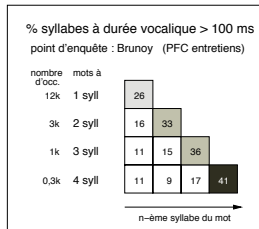


Taux de segments à durée > à 100 ms (gauche) et ≤ 50 ms (droite).
Les syllabes internes présentent les taux les plus élevés de durées faibles: réduction temporelle, prononciations réduites...

Comparaison Nord - Sud - Est

3 points d'enquête Nord (Brunoy, Ile-de-France), Sud (Douzens, Languedoc) et Est (Nyon, Suisse romande)

Taux de segments vocaliques dont la durée est $>$ à 100 ms.



Similarités entre les points du Nord et du Sud.

Particularité de Nyon: plus d'allongements sur la pénultième.



Frontières de mot - profils de f_0

avec Rena Nemoto, doctorante, LIMSI

collaboration avec Jacques Durand, Univ. Toulouse

question: indices de frontière de mot?

hypothèse: prosodie peut informer sur la position des frontières:

(allongements), intonation

données journalistiques (extraits du corpus ESTER)

entretiens (extraits du corpus PFC)

alignement forcé

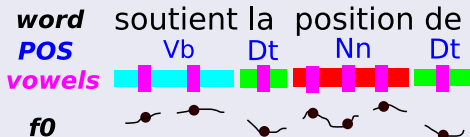
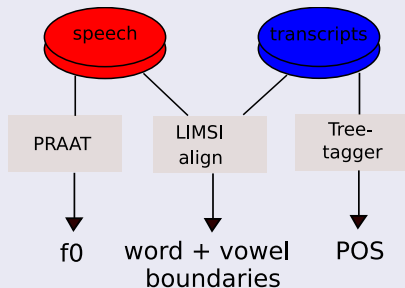
mesure de f_0

calculs de f_0 moyen par rang syllabique

(JEL, 2009)



Méthodologie



Corpus journalistique : classes de mots

Description of the corpus according to word tokens of syllable length n ($n = 0-4$). Separate counts are given for words w/wo realized final schwa(right/left)). *syll.class* counts 1) the number of full syllables and 2) presence/absence of final schwa.

| n | <i>syll.class</i> | #words | examples | <i>syll.class</i> | #words + /ə/ | examples |
|-----|-------------------|--------|----------------------|-------------------|-----------------|---------------|
| 0 | 0_0 | 12578 | l'; d'; de | 0_1 | 12295 | de; le; que |
| 1 | 1_0 | 72249 | vingt; reste | 1_1 | 3918 | reste; test |
| 2 | 2_0 | 36027 | beaucoup; journal | 2_1 | 2087 | ministre |
| 3 | 3_0 | 15994 | notamment; militaire | 3_1 | 698 | véritable |
| 4 | 4_0 | 6053 | présidentielle | 4_1 | 174 | nationalistes |

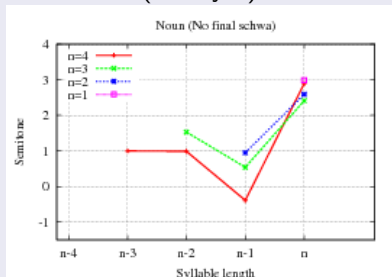
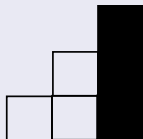
Quelques résultats : parole journalistique

Mean f_0 profiles of n -syllabic lexical words.
Words without final schwa (1-4 syll.)

mot à 1 syllabe

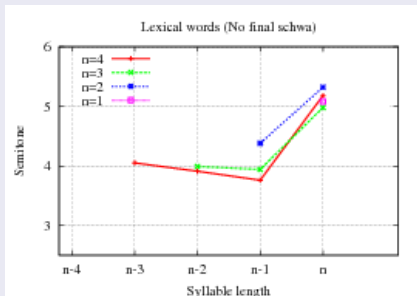
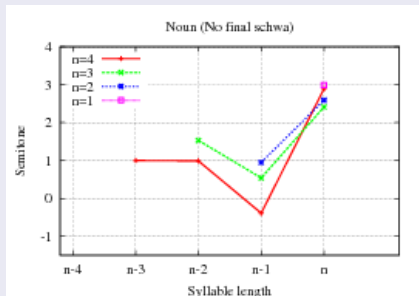
mot à 2 syllabes

mot à 3 syllabes



- (i) Mean f_0 is much higher for final syll. n than for preceding ones.
- (ii) For trisyllables or more, f_0 delta between two consecutive vowels is maximal between penultimate and final vowels. The corresponding delta increases with word syllable length.
- (iii) Initial word accent remains weak on mean f_0 contours.

Quelques résultats : ESTER vs entretiens PFC



PFC profiles similar, although flatter than journalistic profiles.
... ongoing work ...

Variation diatopique en français

avec Philippe Boula de Mareüil

le cas du /O/

alignements à variantes spécifiques

corpus PFC

- Variation & reconnaissance automatique de la parole
 - parole spontanée/interactive:
défi pour la reconnaissance automatique de la parole
 - erreurs de transcription automatique:
taux d'omission de mots augmente pour la parole spontanée/interactive par rapport à la parole préparée (lecture, journalistique)
 - hypothèse: plus de réduction temporelle en parole spont./interactive
 - → étudier la variation en lien avec la durée segmentale
 - quelles variantes de prononciation (plus courtes)?

- Exemples d'étude de la variation à l'aide d'alignement forcé & résultats
 - quantification de la variation:
 - réduction des voyelles orales (+résultats similaires sur un grand nombre de corpus (y compris PFC) et langues différentes)
 - allongements des durées sur les syllabes finales de mot (plus d'allongement si schwa final, cas des pénultièmes pour Nyon)
 - réduction des durées sur les syllabes internes de mot
 - assimilation de voisement régressif pour plosives et fricatives (plus de voisement que de dévoisement)
 - taux d'assimilation plus élevés pour segments courts (cf. ICPHS 2007)
 - profils moyens de f_0 : montée de f_0 sur les syllabes finales (descente sur la pénultième)
 - variation diatopique du /O/ dans PFC (cf. présentation de P. BdM)
- Perspectives
 - compléter les études de la variation à travers les différents styles de PFC
 - plus de points d'enquête de PFC

***Merci
de votre
attention***

10

