Word boundaries in French: Evidence from large speech corpora

Rena Nemoto[⊘⊗], Martine Adda-Decker[⊘], Jacques Durand[♢]

[⊘]LIMSI-CNRS, [⊗]Univ. Paris-Sud 11, Orsay France,
[♢]CLLE-ERSS (UMR5263) CNRS & Univ. Toulouse, France



Outline

- Motivation: acoustic cues for word boundaries?
- Methodology & corpus
- Lexical f_0 profiles
- Lexical duration profiles
- Conclusion

Motivation

- context: French interdisciplinary research projects (*Computer Sciences*, *Linguistics*)
- preliminary question: how do ASR systems locate word boundaries?
  mainly rely on lexical & word n-gram information
- question: are there acoustic cues signaling word boundaries in French?
- make use of large corpora and automatic processing tools
- hypothesis: prosodic cues (f_0, duration)

$\Longrightarrow$ **produce empirical evidence from large corpora**
$\Longrightarrow$ **investigate whether prosodic realisations may contribute to address the word segmentation problem**
$\Longrightarrow$ **increase our knowledge of prosodic realisations in French words**

Hypotheses

- French: f_0 and duration tend to increase on most prosodic word endings (continuation)

Example:

prosodic words	*homophonic*	*French prosody*
(le couple)(est complet)...	/ləkuplɛkõplɛ/	le **coupl**e est com**plet**
(le couplet)(complet)...		le cou**plet** com**plet**

- prosodic word endings are a subset of (content) word endings
- influential factors: word length, word-final schwa, POS...

Corpus

- French TECHNOLANGUE-ESTER1 corpus (Galliano 2005)
- broadcast news shows from French radio stations
- subset of 13 hours of male speakers
- 165k word tokens – 14k word types
- mainly "prepared" journalistic speech style

Methodology: processing steps

audio stream:

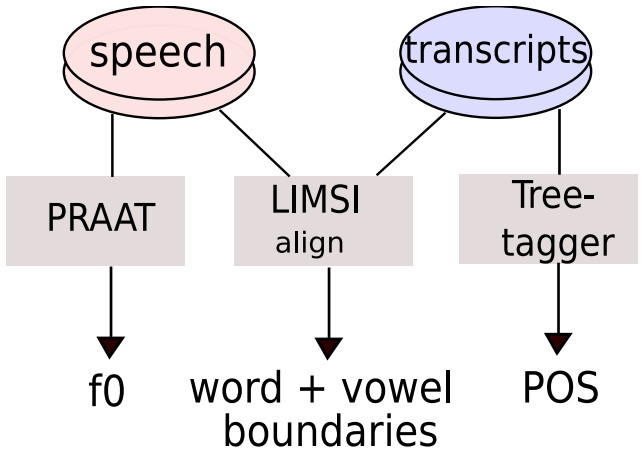
- f_0 measurements each 5 ms (Praat, Boersma 2005)

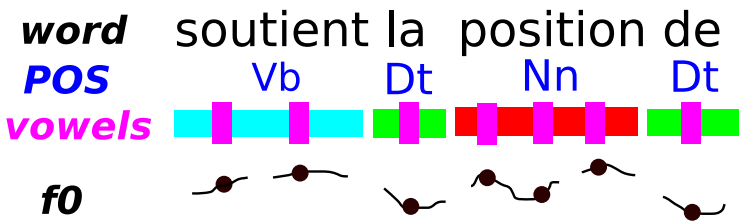
audio + word streams:

- word & vowel boundaries (LIMSI speech alignment system, Gauvain 2005)

word stream:

- POS tags (Treetagger, Schmid 1994)



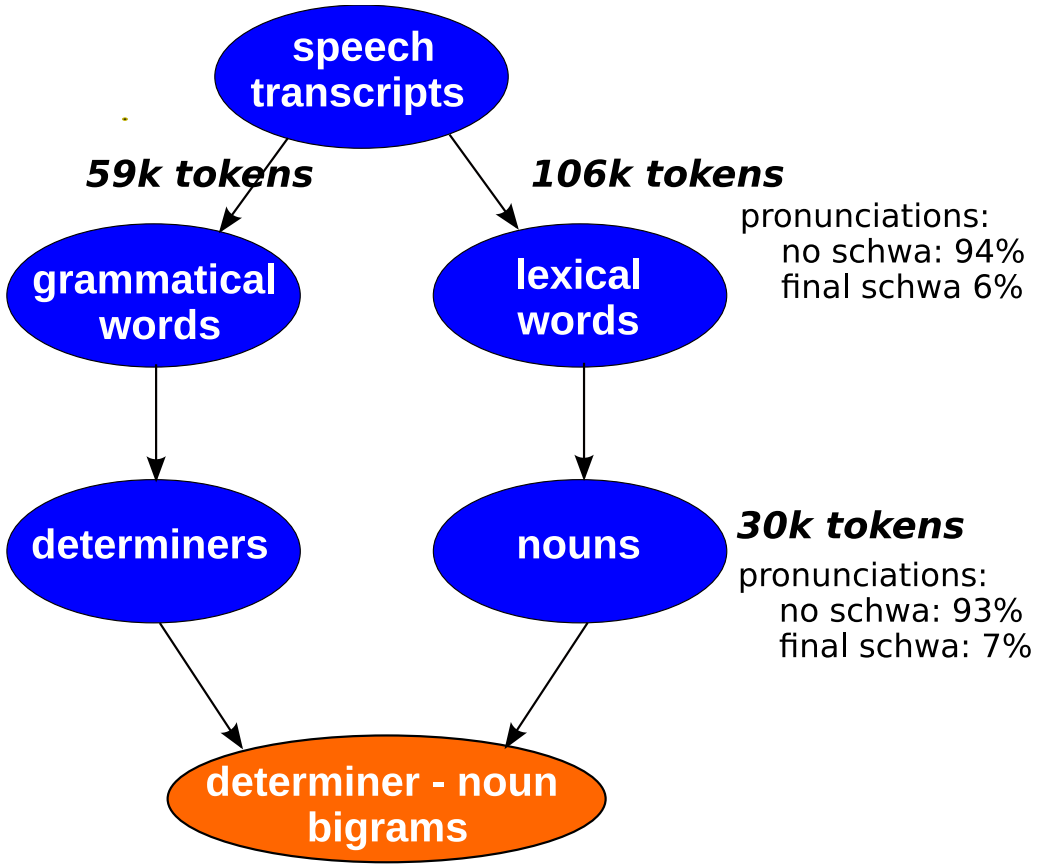


Methodology: syllabic word length classes

n : syllabic word length
word class *n*_**0** : words with *n* syllables and **no final schwa**
word class *n*_**1** : words with *n* syllables and **with final schwa**

n	*n_s*	*#words*	*examples*
0	0_0	13k	l'; d'; de
1	1_0	72k	vingt; reste
2	2_0	36k	beaucoup; journal
3	3_0	16k	notamment; militaire
4	4_0	6k	présidentielle
		#words+ /ə/	
0	0_1	12k	de; le; que
1	1_1	4k	reste; test
2	2_1	2k	ministre
3	3_1	0.7k	véritable
4	4_1	0.2k	nationalistes

Methodology: grammatical vs content word classes



Lexical f_0 profiles

f_0 profiles: computed for each word class (*n_s*,...)
only vowels with *voicing ratio* over 70% were used (rejection rate 10%)

$$(voicing\ ratio = \frac{number\ of\ voiced\ frames}{total\ number\ of\ frames})$$

for each vowel a mean f_0 value was computed (all voiced frames of segment)
values in Hz converted to semitones (st), 120 Hz as reference frequency

example: n_s = 2_0

2_0 : class of bisyllabic words without final schwa:
f_0 profile: (average f_0 of rank 1 vowels) + (average f_0 of rank 2 vowels)

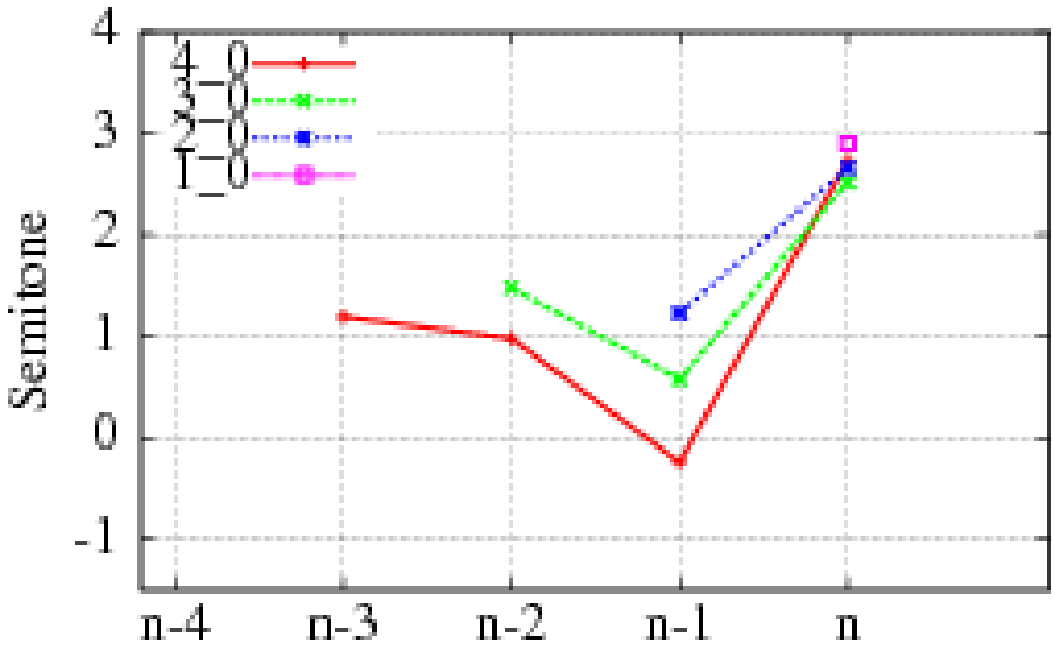
Mean f_0 profiles of *n*-syllabic lexical words

lexical words without final schwa (1-4 syll.)

word classes:

1_0 monosyllabic words without final schwa
2_0 bisyllabic words without final schwa
3_0 trisyllabic words without final schwa
4_0 4-syllabic words without final schwa

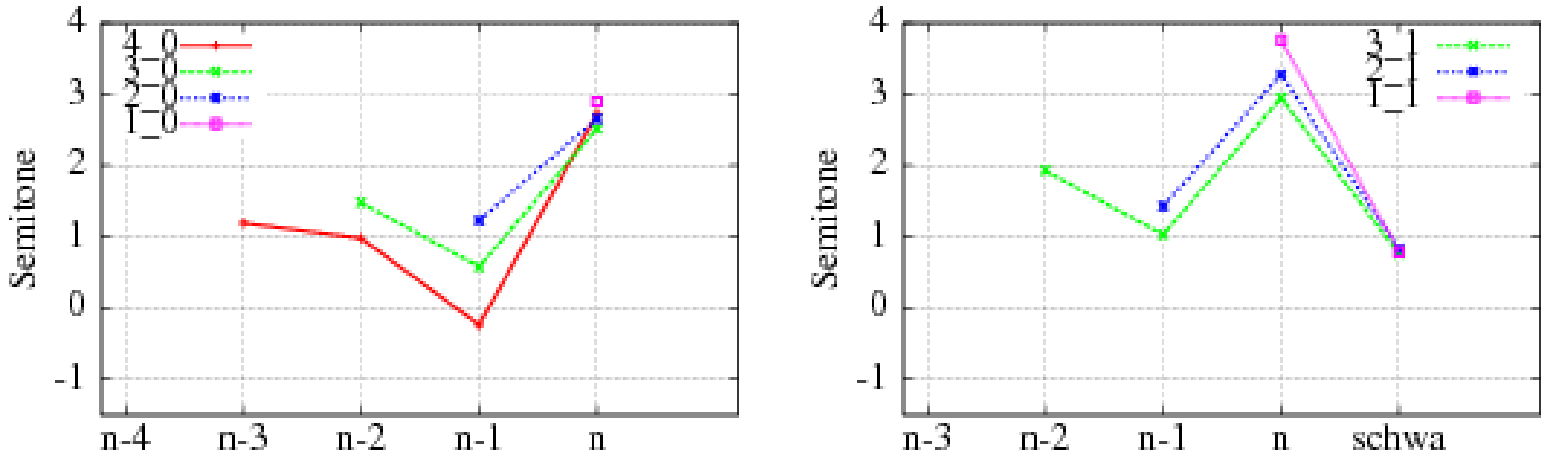
profiles are aligned w.r.t. to the final syllable *n*



x-axis: vowel rank (w.r.t. final syllable vowel) - y-axis: f_0 (in semitones)

Mean f_0 profiles of *n*-syllabic lexical words

left: words without final schwa (1-4 syll.) **right:** with final schwa (1-3 syll.)



x-axis: vowel rank (w.r.t. final syllable vowel) - y-axis: f_0 (in semitones)

(i) f_0 much higher for the final syllable *n* than for the preceding ones.

(ii) for trisyllables+, f_0 delta maximal between final & penultimate vowels difference tends to increase with word syllabic length.

(iii) monosyllabic f_0 as high as that of the final syllable of longer words.

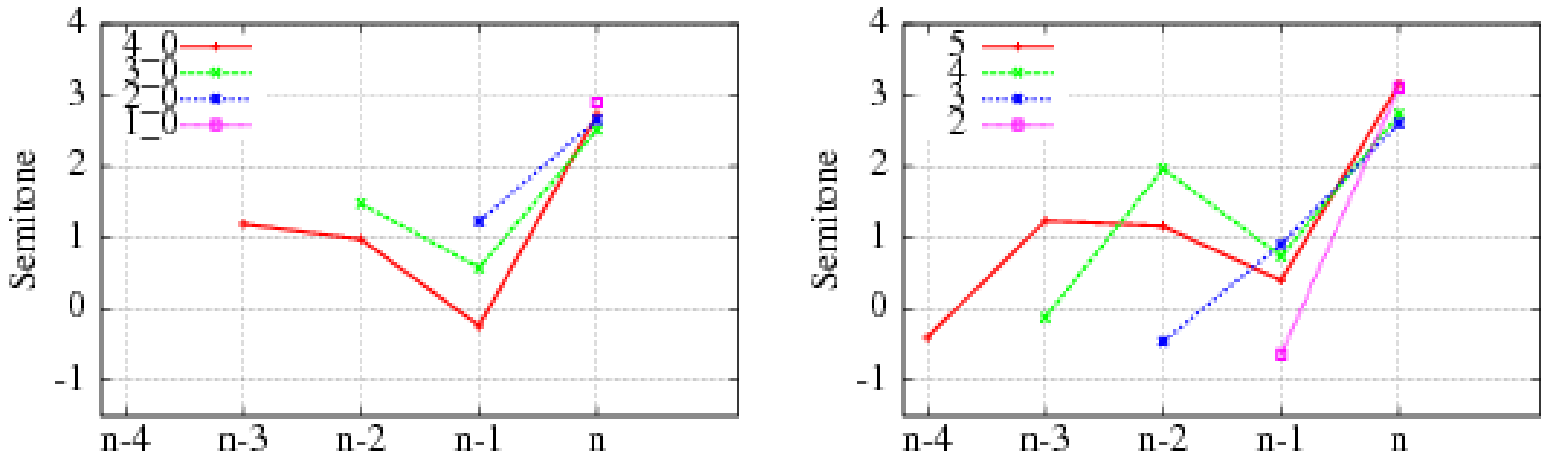
(iv) final schwa (n_1) profiles globally higher f_0 than n_0 profiles,

(v) delta between final syllable *n* and final schwa : 2-3 st.

(vi) weak initial accentuation

Mean f_0 profiles of *n*-syllabic noun phrases (no final schwa)

left: nouns (1-4 syll.) **right:** det + noun 13k occ. (2-5 syll.)



x-axis: vowel rank (w.r.t. final syllable vowel) - y-axis: f_0 (in semitones)

(i) noun phrase: f_0 minimal on 1st syllable

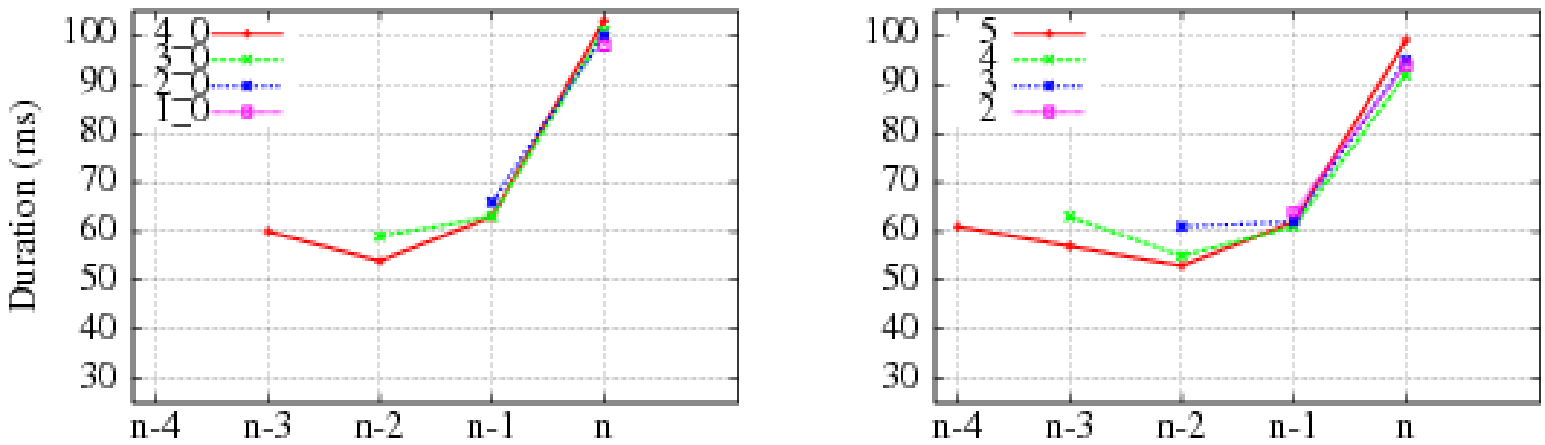
(ii) max. delta f_0 between 1st syllable (monosyllabic det.) & last syllable (

within a temporal window of some syllables, f_0 may provide cues for phrase boundaries, at least for the noun phrase case (`determiner noun`)

Lexical duration profiles: based on vocalic durations

mean vocalic segment duration for each vowel rank $k = 1...n$

left: nouns (no final schwa) **right:** noun phrase (no final schwa)



x-axis: vowel rank (w.r.t. final vowel) - y-axis: vocalic segment duration (ms)

(i) final vowel duration $\sim$ 100 ms on average

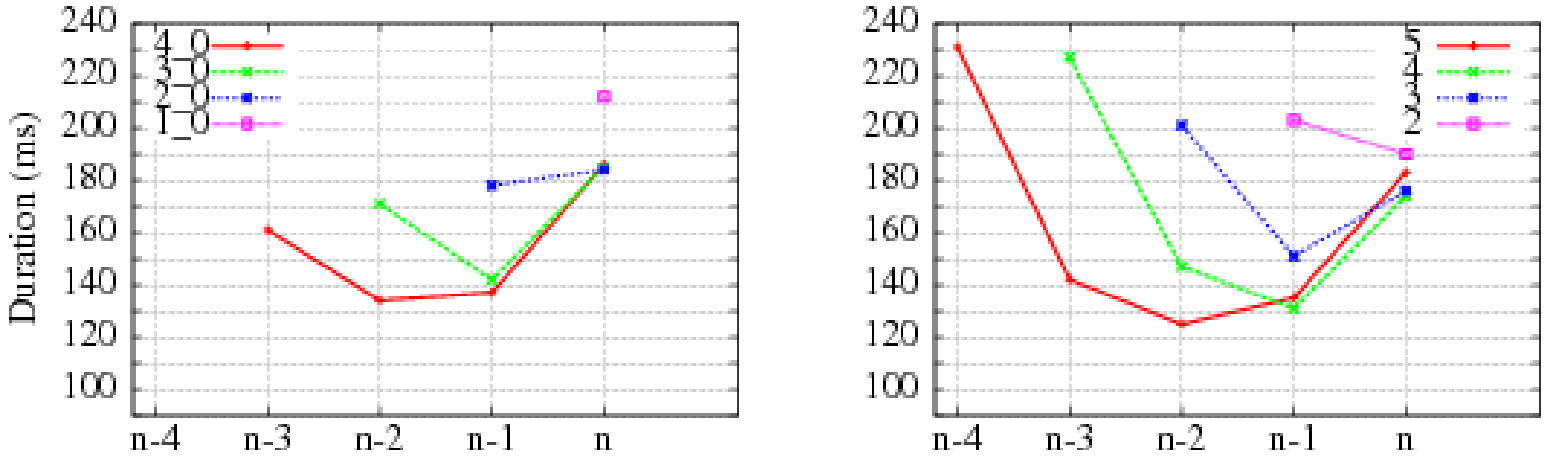
(ii) all other vowels $\sim$ 60 ms on average

high segment duration: cue for word ending (noun)

Lexical inter-vocalic duration (IVD) profiles

mean IVD for each vowel rank $k = 1...n$ (between preceding & present vowels)

left: nouns (no final schwa) **right:** noun phrase (no final schwa)



x-axis: vowel rank (w.r.t. final vowel) - y-axis: IVD duration (ms)

(i) high inter-vocalic duration $\sim$ 180 ms on final vowels

(ii) very high IVD $\sim$ 220 ms on phrase-initial vowels

high IVD: cue for prosodic word boundaries (in particular noun phrase start)

Conclusions

Are there acoustic cues signaling word boundaries in French?

- **Hypotheses** concerning influential factors:
  syllabic word length, presence/absence of word-final schwa, syntax
- 13 hours of broadcast news speech - 165k words - male speakers
- Automatic tools for annotation: f_0, duration, vowels, syllabic rank, POS
- Original methodology to study prosodic regularities of French words via average lexical profiles

Word boundary information evidenced via average f_0, VD, IVD profiles:

word final syllable f_0 rises
long word final syllable lengths
long IVD on phrase boundaries

Conclusions & perspectives

Measurable cues contributing to word boundary location can be found!

Future studies:

other POS sequences, more prosodic words, more detailed f_0 patterns

other speaking styles (especially spontaneous speech), other languages

Findings for ASR:

acoustic modelling

post-processing step for error recovery (improved boundary location)

Thank you for your attention