

UMR 6057 CNRS
 **Laboratoire Parole & Langage**
 Université de Provence Aix-en-Provence, France

EXAPP Conference
 University of Copenhagen
 March 20-22, 2013

**Phoneme deletion and fusion in
 conversational speech**

Christine Meunier
 Laboratoire Parole et Langage – Aix-Marseille Université

Context

- ▶ Growing interest for big corpora of natural speech
 - ▶ Buckeye corpus (Pitt et al., 2007): *American English*
 - ▶ Corpus of Interactional Data (Bertrand et al., 2008): *French*
 - ▶ Nijmegen corpora of casual speech (Ernestus): *French, Spanish, Czech, Dutch*
 - ▶ Etc.
- ▶ Consequences
 - ▶ Increase of non canonical forms (lexicon, syntax, phonetics...)
 - ▶ Interpretation of linguistic processes more complex
 - ▶ Phonetics
 - ▶ 'Massive reduction'(Johnson, 2004)
 - ▶ Phonetic underspecification (lack of phonetic cues)
 - ▶ How do we understand speech?
- ▶ Lead to a more holistic view of language mechanisms

General questions

- ▶ Why do we reduce? (or why don't we reduce in controlled speech?)
 - ▶ Flexibility: speaker's adaptation

- ▶ How do we reduce?
 - ▶ Reduction typology and characteristics

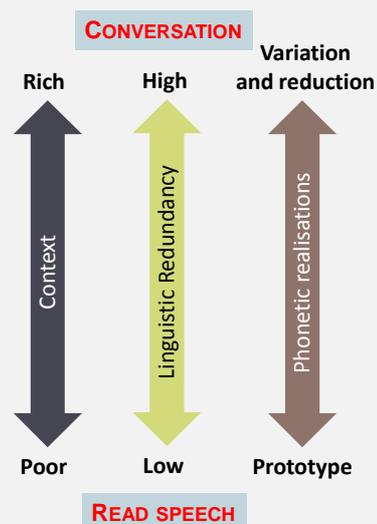
- ▶ Where (or when) do we reduce?
 - ▶ Interaction between phonetics and other linguistic domains



Speaker's flexibility and context (speech style)

Lindblom's H&H theory (1990):

- Speech production is adaptive
- Speakers can tune their performance according to communicative and situational demands



General questions

- ▶ Why do we reduce? (or why don't we reduce in controlled speech?)
 - ▶ Flexibility: speaker's adaptation

- ▶ How do we reduce?
 - ▶ Reduction typology and characteristics

- ▶ Where (or when) do we reduce?
 - ▶ Interaction between phonetics and other linguistic domains



Reduction typology (1)

- ▶ **Deletion**
 - ▶ French Schwa (Davidson 2006, Bürki et al., 2010)
 - ▶ French /e/ (Torreira & Ernestus, 2011)

- ▶ **Undershoot gestures:**
 - ▶ Unrealized closure for plosives (Duez, 1995)
 - ▶ Vowel centralisation (Lindblom, 1963; Gendrot & Adda-Decker 2005; Meunier & Espesser, 2011)

- ▶ **Assimilation**
 - ▶ Voicing (Duez, 1995; Ernestus 2000; Hallé & Adda-Decker, 2007)



Reduction typology (2)

- ▶ **Phonological and stereotyped reductions**
 - ▶ Phonological: French schwa
 - ▶ Stereotyped: **je** **sais** pas (*I don't know*) /ʃepa/, **c'****était** (*it was*) /ste/, **tu** **vois** (*you see*) /tywa/, **tu** **sais** (*you know*) /tse/, **plus** (*more*) /py/, ...
 - ▶ Frequent, affect specific words or sequences → predictable
 - ▶ Transcribers generally identify the reduction
 - ▶ Related to lexicon, phonological structure and indexical factors

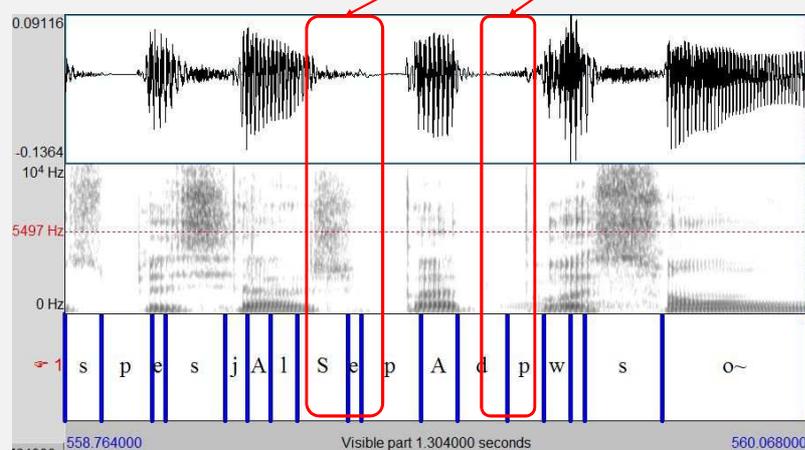
- ▶ **Opagues reductions**
 - ▶ Frequent but not specific to certain words or sequences → not predictable
 - ▶ Transcribers rarely identify them
 - ▶ May be related to lexicon and indexical factors but also to larger domains (prosody, discourse, pragmatics, etc.)



Phonological or stereotyped reductions

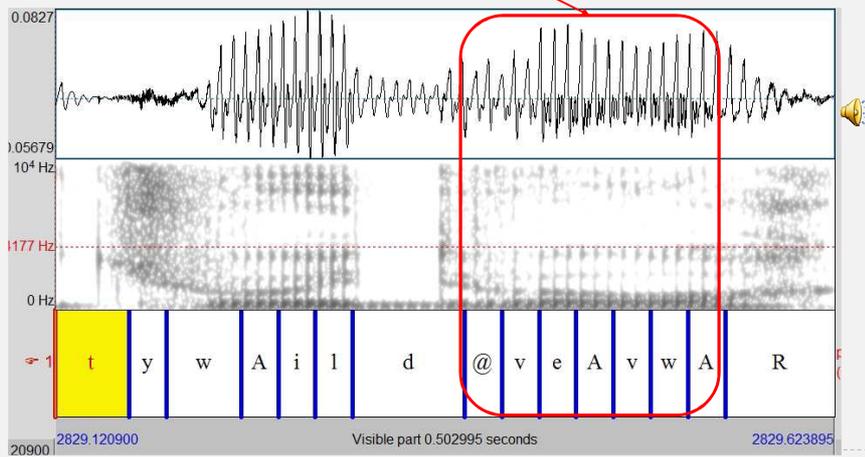
- ▶ « un resto spécial [je sais, **ché**] pas d(e) poisson » 

(a special restaurant, I don't know, a fish restaurant...)



Opaques reductions(1)

- ▶ « assez âgé tu (v)ois il **devait avoir** la quarantaine » 🗨️
(quite old, you know, he should be forty...)



General questions

- ▶ Why do we reduce? (or why don't we reduce in controlled speech?)
 - ▶ Flexibility: speaker's adaptation
- ▶ How do we reduce?
 - ▶ Reduction typology and characteristics
- ▶ Where (or when) do we reduce?
 - ▶ Interaction between phonetics and other linguistic domains (lexicon, prosody, discourse, pragmatics, etc)



Dependency

- ▶ **Prosody**
 - ▶ Vowel centralisation for unaccented vowels (Lindblom, 1963)
 - ▶ Vowel centralisation correlated with speech rate (Gendrot & Adda-Decker, 2005, 2007; Meunier & Espesser, 2011)
- ▶ **Word category**
 - ▶ Vowel centralisation for function words (Bergem, 1993; Meunier & Espesser, 2011)
 - ▶ Neighborhood density (Wright, 1997)
- ▶ **Word frequency** (Bybee, 2002; Johnson 2004; Pluymakers et al., 2005)
- ▶ **Position within words**
 - ▶ Centralisation for vowels in final word position (Meunier & Espesser, 2011)



General questions

- ▶ *Why* do we reduce? (or why don't we reduce in controlled speech?)
 - ▶ Flexibility: speaker's adaptation
- ▶ *How* do we reduce?
 - ▶ Reduction typology and characteristics
- ▶ *Where* (or when) do we reduce?
 - ▶ Interaction between phonetics and other linguistic domains (prosody, discourse, pragmatics, etc)
- ▶ **In this study:**
 - ▶ Explanatory work describing two types of reduction: deletion and fusion (*'How'*)
 - ▶ Hypothesis: may be related to different linguistic processes (*'Where'*)



Method: deletion and fusion distinction

▶ Deletion

- ▶ One or more phonemes are not realized
- ▶ Categorical
- ▶ The absence of the segment is clearly identified (perceptually and on the speech signal)

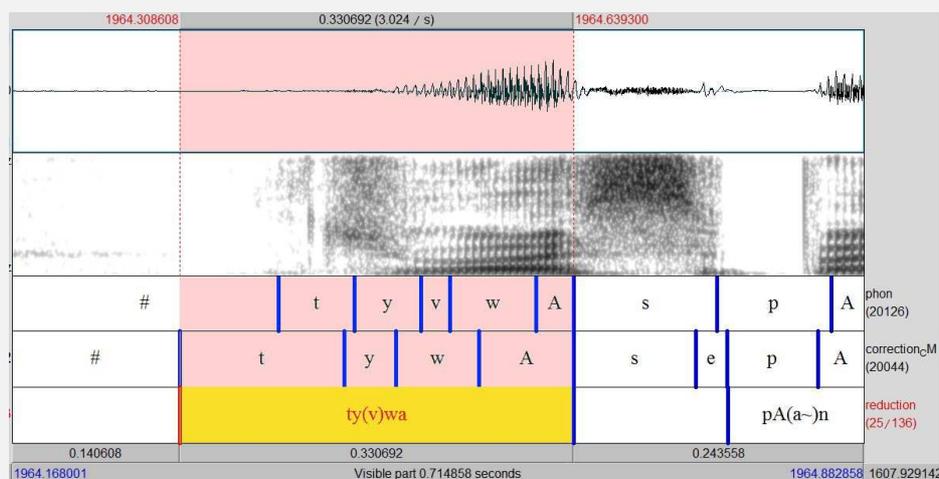
▶ Fusion

- ▶ Two or more phonemes are merged
- ▶ Gradual
- ▶ Several phonemes (at least two) cannot be distinguished individually and it is not possible to determine which is realized and which is absent (even if they are perceived with a larger context)



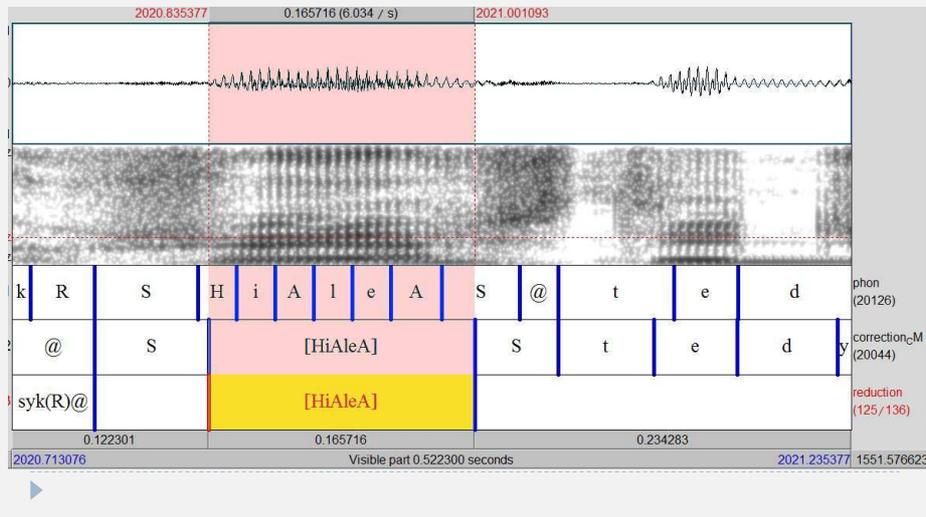
Deletion selection

🔊 « Tu vois » (*you see*) /tyvwa/ → [tywa]



Fusion selection

🔊 « Je suis allé acheter » (*I went to buy*) [ʒʁialeaʃte]



Method: speech material

CID: French **C**orpus of **I**nteractional **D**ata (Bertrand & al., 2008).

- ▶ Audio and video recordings of French speakers.
- ▶ 8 dialogues of 1 hour each (6 men, 10 women)
- ▶ Familiar conversation
- ▶ Annotations
 - ▶ Orthographic (enriched)
 - ▶ Phonetic: phonetisation and automatic alignment
 - ▶ For this study, alignment has been corrected and reductions have been annotated manually

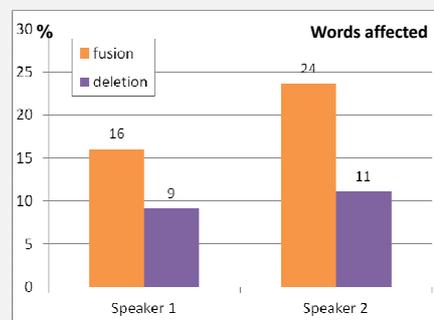
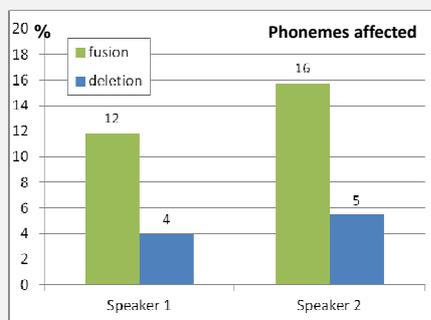
Method: this study

- ▶ Selected speakers and sequences:
 - ▶ 2 speakers
 - ▶ Recording duration : 204 seconds
 - ▶ Mean phoneme durations: 72ms
 - ▶ Number of phonemes: 1322
 - ▶ Number of tokens: 535
 - ▶ Number of deletions: 52
 - ▶ Number of fusions: 69



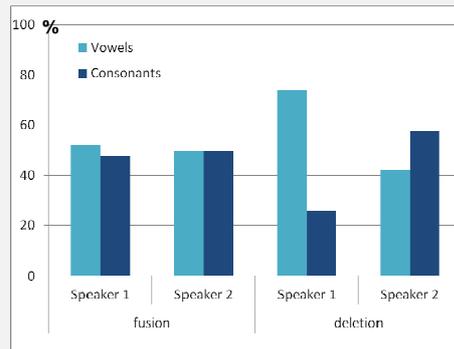
Results: reduction frequency

- ▶ Proportion of phonemes affected by fusion + deletion: 18.5%
- ▶ Obviously fusion affects more phonemes than deletion
- ▶ More reduction for speaker 2
- ▶ Proportion of words affected by fusion + deletion: 30%
- ▶ fusion affects more words than deletion (less obvious)

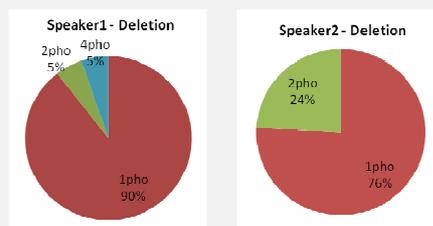


Results: phonetic context (phoneme category)

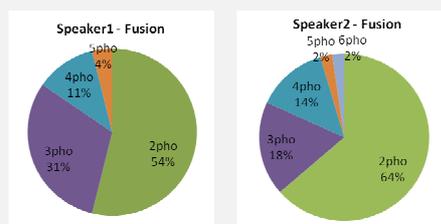
- ▶ Fusion
 - ▶ For both speakers, vowels and consonants are affected equally
 - ▶ /a/, /i/, /l/ and /e/ are the most affected phonemes for both speakers
- ▶ Deletion
 - ▶ Speakers 1 and 2 differ according to the type of phoneme affected
 - ▶ More data needed to interpret



Results: phonetic context (phonetic size)



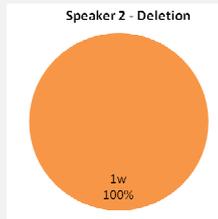
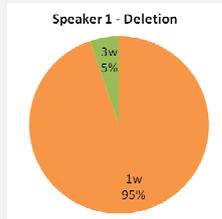
- ▶ Deletion affects more often a single phoneme



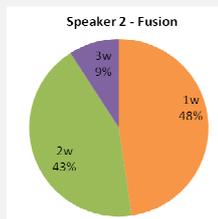
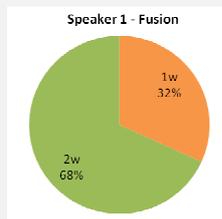
- ▶ Fusion affects at least 2 phonemes and often more



Results: lexical context (number of words affected for each type)

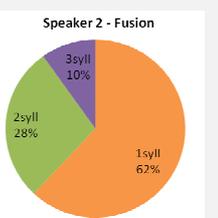
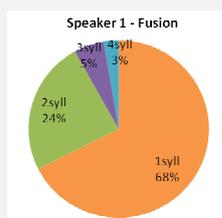
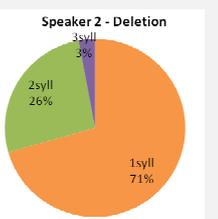
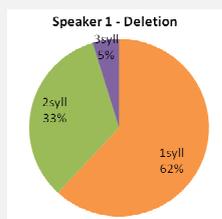


- ▶ Deletion affects a single word
- ▶ 64% of deletion concern **stereotyped** reduction (68% for S1 and 61% for S2)



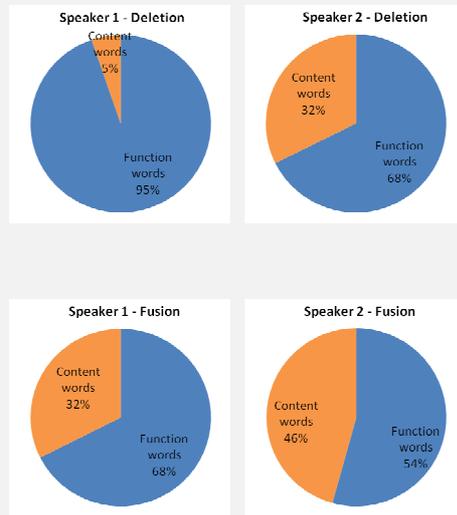
- ▶ Fusion often affects more than a word: 'lexical overlapping'

Results: lexical context (word size)



- ▶ For deletion or fusion, words affected are most often monosyllabic or bisyllabic.
- ▶ Representative of what is produced in spontaneous speech

Results: lexical context (word category)



- ▶ Function words are more often affected by deletion than by fusion (stereotyped reduction)
- ▶ Reductions of Speaker 1 more often affect function words

Summary

- ▶ Deletion and fusion are quite frequent (fusion>deletion)
- ▶ Reduction frequency is speaker-dependant (S2>S1)
- ▶ Deletion:
 - ▶ A single phoneme within a single word
 - ▶ Often a stereotyped reduction
- ▶ Fusion:
 - ▶ Several phonemes are merged and overlapp several words
 - ▶ Often an opaque reduction
- ▶ Phoneme and words affected: no particularity → representative of what is produced in spontaneous speech

Conclusions and hypotheses

▶ Deletion

- ▶ Seems to be related to lexicon or lexicalized forms
- ▶ Phonological encoding of words, exemplars

▶ Fusion

- ▶ Seems to be related to larger domain (prosody or discourse)
- ▶ Gesture undershoot and overlapping



- *What is the phonetic nature of fusion?*
- *Do we simply restore lacking phonetic information with context (top down process)?*
- *Or is there sufficient articulatory/acoustic information in fusion to drive perception?*



Reduction

- ▶ « faire venir une **assist**ante sociale » 
- (to bring a social worker...)

