

Using chunks for disfluency detection and studying variation

K. Peshkov, L. Prévot, S. Rauzy, B. Pallaud

*Aix Marseille Université & CNRS
Laboratoire Parole et Langage*

Variamu presentation, 10.03.2014, Hong-Kong



The presence of **disfluencies** is characteristic for the spoken speech.

The term regroups diverse phenomena which perturb normal speech flow, producing utterances where “the syntagmatic continuity is broken” [Blanche-Benveniste et al., 1990].

- Repetitions
- Truncated words
- Filled pauses
(ex. *euh*, *mh* in French)

- Systematic account of the phenomenon [Shriberg, 1994]
- Type of system
 - Statistical [Charniak and Johnson, 2001]
 - Hybrid: statistical + rule-based [Germesin et al., 2008, Liu et al., 2003]
- Knowledge sources
 - Transcripts (or ASR outputs) only [Heeman and Allen, 1999, Snover et al., 2004]
 - Prosody only [Shriberg et al., 1997, Liu et al., 2003]
 - Multiple knowledge sources [Liu et al., 2005, Germesin et al., 2008]
 - Text-based system outperforms prosody-only system, the combination of prosody and text performs even better [Liu et al., 2003]
- Disfluencies in French
 - Descriptive studies [Henry and Pallaud, 2003, Mareüil et al., 2005]
 - [Bouraoui, 2008]

- Transcription only
 - independent characterization of text- and prosody-based cues
 - high-quality morpho-syntactic tagging is available
- **Building chunks and then characterizing each as fluent or disfluent based on its morpho-syntactic pattern**

- **Chunking** consists in including in one unit all the constituents situated to the left of each syntactic head [Abney, 1991]
- Limitation: ability to capture only local syntactic irregularities
- Examples of target disfluencies
 - (1) *les les les gens*
the the the people
 - (2) *ouais ben ouais je crois*
yeah well yeah I think

- 1 Chunking
- 2 Pattern categorization
 - "Spokenhood" ratio
 - Pattern simplification
 - Syntactic well-formedness annotation
- 3 Evaluation
- 4 Application: pilot study of variation

Corpus of Interactional Data

- Corpus of Interactional Data (CID) [Bertrand et al., 2008]
- 8 dialogues of one hour
- Subjects are native speakers of French, sharing a conversational history



[Rauzy and Blache, 2009]

- Tagset includes 50 labels
- F-measure reached by the tagger = 0.948
- Manually corrected

- 26 rules
- two types
 - 1 rules that specify tags which are always added in the same chunk as the following token
 - determiner (*D*) + anything
 - preposition (*S*) + anything
 - conjunction (*C*) + anything
 - personal pronoun (*Pp*) + anything
 - auxiliary verb (*Va*) + anything
 - 2 rules which represent ordered pairs of POS tags which must be in the same chunk
 - adjective (*A*) + noun (*N*)
 - demonstrative pronoun (*Pd*) + verb (*V*)
 - adverb (*R*) + adjective (*A*)
 - proper noun (*Np*) + proper noun (*Np*)
 - verb (*V*) + verb (*V*)

First characterization: spokenhood

$$\rho_i = \frac{\text{Written Frequency}(i)}{\text{Spoken Frequency}(i)}$$

- Appear in spoken data only, many disfluency related patterns should be there (Disfluent)
- Majority in spoken data (Spoken)
- Majority in written data (Canonical)

Category	Chunk	ρ	Example
"Disfluent"	I I I Pp V	∞	ouais ben ouais j' imagine yeah well yeah I imagine
Spoken	I D N	109,75	bon ce truc well that thing
Canonical	D A N	0,49	un vrai conflit a real conflict

Tag sequence simplification

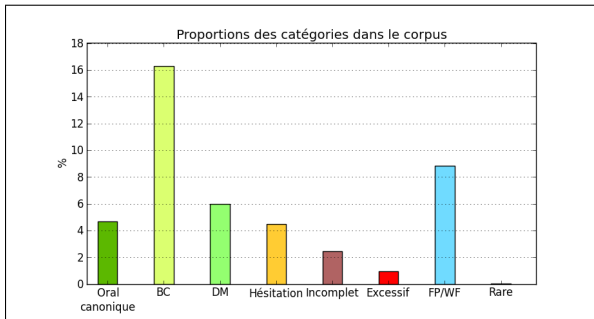
- $C \rightsquigarrow I$: assimilate conjunctions to interjection for the sake of simplicity since in this context they play a very similar role;
- $\{V Vi\}, \{Va V\} \rightsquigarrow V$: simplification of verb structures, ($V =$ Verb ; $Vi =$ Infinitive verb ; $Va =$ auxiliary verb);
- $\{Pp Px\}, \{Pp Po\} \rightsquigarrow P$: simplification of pronoun structures, ($Pp =$ Personal Subject Pronoun ; $Px =$ Reflexive pronoun ; $Po =$ Personal Object Pronoun);
- $\{XX\}, \{XXX\}, \dots \rightsquigarrow X^+$: simple regular expression patterning.

Second characterization

manual tagging of the 50 most frequent patterns of the Spoken and Disfluent categories

- 1 **Spoken Canonical:** Spoken form, fluent (4.67% of the total number of chunks)
- 2 **Hesitation:** I+ Canonical, that is a sequence of discourse markers or conjunctions followed by a Canonical chunk (4.48%)
- 3 **BC:** Sequence of discourse markers, surrounded by silent pauses longer than 200 milliseconds (16.26%)
- 4 **DM:** Discourse markers and conjunctions: I+ (6%)
- 5 **Incomplete:** Chunks without a head (2.44%)
- 6 **Excessive:** Chunks with abnormally long patterns (0.93%)
- 7 **Rare:** Other chunks, considered to be disfluent (0.03%)
- 8 **FP/WF:** All chunks containing a filled pause or a word fragment (8.85%)

Second characterization



Version	Precision	Recall	F-score
Baseline (FP and WF)	72.5 %	47.4 %	57.3 %
Morpho-syntax only	69.7 %	56.2 %	62.2 %
Combined	70.2 %	57.7 %	63.3 %

Table : Evaluation against manually annotated data

Variation (1): Speaker style

- Intuitive characterization by an expert prosodist
- Expert's evaluation is in line with results

Speaker	A	B	C	D	Mean
Expert's opinion	fluent	fluent	disfluent	disfluent	
Canonical	60.7 %	63.2 %	49.8 %	51.0 %	55.9 %
Disfluency	9.6 %	6.8 %	14.4 %	11.4 %	12.3 %

Variation (2): Discourse activities

- Narratives vs. conversational zones
- Hypothesis: narratives are less disfluent
- No clearly emerging tendencies, disfluencies are distributed rather uniformly between narrative and conversational zones.

Conclusion

- method of detecting disfluencies using syntactic patterns
- F-score = 63.3
- improves the coverage compared to filled pauses and truncated words only
- may be useful for speaker style characterization

Thank you!



Abney, S. (1991).

Parsing by chunks.

Principle-based parsing, 44:257–278.



Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. (2008).

Le CID - corpus of interactional data - annotation et exploitation multimodale de parole conversationnelle.

Traitement Automatique des Langues, 49(3):1–30.



Blanche-Benveniste, C., Bilger, M., Rouget, C., Van Den Eynde, K., Mertens, P., and Willems, D. (1990).

Le français parlé (études grammaticales).

Sciences du langage.



Bouraoui, J.-L. M. (2008).

Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint: le cas du Contrôle Aérien.

PhD thesis, Université Paul Sabatier-Toulouse III.



Brill, E. (1995).

Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging.

Computational Linguistics, 21(4):543–565.



Charniak, E. and Johnson, M. (2001).

Edit detection and parsing for transcribed speech.

In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–9. Association for Computational Linguistics.



Germesin, S., Becker, T., and Poller, P. (2008).

Hybrid multi-step disfluency detection.

Machine Learning for Multimodal Interaction, pages 185–195.



Heeman, P. A. and Allen, J. F. (1999).

Speech repairs, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue.

Computational Linguistics, 25(4):527–571.



Henry, S. and Pallaud, B. (2003).

Word fragments and repeats in spontaneous spoken french.

In ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech.



Liu, Y., Shriberg, E., and Stolcke, A. (2003).

Automatic disfluency identification in conversational speech using multiple knowledge sources.

In Proceedings Eurospeech, volume 1, pages 957–960. Geneva, Switzerland.



Liu, Y., Shriberg, E., Stolcke, A., and Harper, M. (2005).

Comparing hmm, maximum entropy, and conditional random fields for disfluency detection.

In Proceedings of the INTERSPEECH'05. Citeseer.



Mareüil, P. B. d., Habert, B., Bénard, F., Adda-Decker, M., Barras, C., Adda, G., and Paroubek, P. (2005).

A quantitative study of disfluencies in french broadcast interviews.
In Disfluency in Spontaneous Speech.



Rauzy, S. and Blache, P. (2009).

Un point sur les outils du lpl pour l'analyse syntaxique du français.
In Workshop ATALA: Quels analyseurs syntaxiques pour le français ?,
Paris, France.



Shriberg, E., Bates, R., and Stolcke, A. (1997).

A prosody-only decision-tree model for disfluency detection.
In Proc. Eurospeech, volume 5, page 2383–2386.



Shriberg, E. E. (1994).

Preliminaries to a theory of speech disfluencies.
PhD thesis, University of California.



Snover, M., Dorr, B., and Schwartz, R. (2004).

A lexically-driven algorithm for disfluency detection.

In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 157–160.

Association for Computational Linguistics.