

Characterizing discourse genres with prosodic features in a reference treebank of spoken French

Guillaume Desagulier¹ Anne Lacheret-Dujour²
Frédéric Isel² Seongmin Mun³

¹MoDyCo (UMR 7114) Paris 8, CNRS, Paris Nanterre & Institut Universitaire de France

²MoDyCo (UMR 7114) CNRS, Paris Nanterre

³MoDyCo (UMR 7114) CNRS, Paris Nanterre & Ajou University

AFLiCo JET 2018
Université Paris Nanterre
Friday, May 4, 2018



outline

- 1 Rhapsodie
- 2 goals
- 3 data and methods
- 4 results
- 5 discussion
- 6 conclusion

Rhapsodie : overview

- a treebank of spoken French
- 33000 words
- annotated for syntax and prosody
- 57 five-minute long samples (89 male and female speakers)

Rhapsodie : discourse profiles

The discourse profile of each sample is described by 6 variables :

- **event structure** (*dialogue vs. monologue*)
- **social context** (*public vs. private*)
- **subgenre** (*argumentation, description, narrative, oratory, and procedural*)
- **interactivity** (*interactive, non-interactive, and semi-interactive*)
- **channel** (*broadcasting and face-to-face*)
- **planning type** (*planned, semi-spontaneous, and spontaneous*)

Rhapsodie : prosodic profiles

The prosodic profile of each sample is described by 2 sets of 3 variables :

primary variables

- the mean number per second of pauses (*fPauses*)
- conversational overlaps (*fOverlap*)
- gap fillers (*fEuh*)

secondary variables

- mean numbers per second of prosodic prominences (*fProm*)
- intonational periods (*fIPE*)
- intonation packages (*fIPA*)

research questions

main question

Can discourse types in French be characterized and ultimately predicted by prosodic features ?

2 side questions

- does the fact that the corpus is relatively small, heterogeneous, and not necessarily balanced affect the representativeness of our results ?
- are the secondary prosodic features representative of discourse genres ?

data (excerpt)

- 57 corpus samples (rows)
- 12 variables
 - 6 primary variables (continuous)
 - 6 secondary variables (categorical)

sample ID	fPauses	fOverlap	fEuh	fProm	fPE	fPA	subgenre	interactivity	social context	event structure	channel	planning type
D0001	0.26	0.12	0.14	1.79	0.28	1.54	argumentation	interactive	private	dialogue	face-to-face	semi-spontaneous
D0002	0.42	0.11	0.10	1.80	0.33	1.75	argumentation	interactive	private	dialogue	face-to-face	semi-spontaneous
D0003	0.35	0.10	0.03	1.93	0.34	1.76	description	semi-interactive	private	dialogue	face-to-face	spontaneous
D2006	0.37	0.00	0.00	2.22	0.58	1.51	oratory	semi-interactive	public	dialogue	face-to-face	planned
D2007	0.38	0.08	0.05	2.08	0.59	1.87	argumentation	interactive	public	dialogue	broadcasting	spontaneous
M0001	0.40	0.00	0.25	1.92	0.34	1.52	procedural	non-interactive	private	monologue	face-to-face	spontaneous
M0002	0.40	0.00	0.21	2.17	0.23	1.75	description	semi-interactive	private	monologue	face-to-face	spontaneous
M0003	0.30	0.00	0.19	2.11	0.20	1.63	procedural	non-interactive	private	monologue	face-to-face	spontaneous
M0004	0.39	0.00	0.00	1.92	0.68	2.03	procedural	non-interactive	private	monologue	face-to-face	spontaneous
M0005	0.27	0.00	0.17	2.03	0.28	1.23	procedural	non-interactive	private	monologue	face-to-face	spontaneous

data visualization with RhapVis

<http://ressources.modyco.fr/sm/RhapVis/>
<http://ressources.modyco.fr/sm/VisualSystem/>

methods

2 steps :

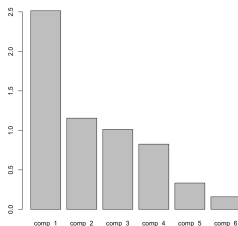
- ① summarizing the table with **principal component analysis**
- ② confirming the exploration with **ANOVAs**

PCA

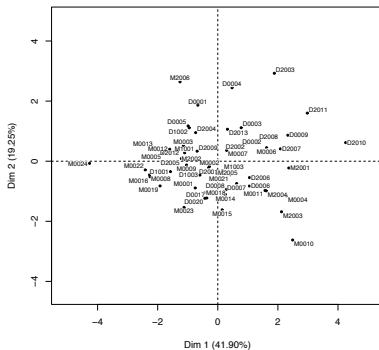
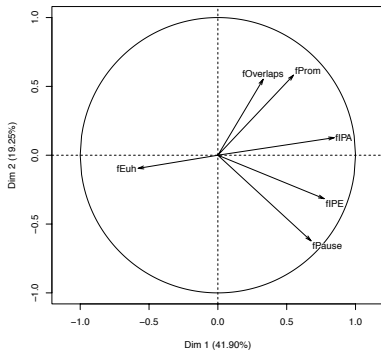
- the table was submitted to PCA
- 6 primary variables : active
- 6 secondary variables : illustrative

PCA : eigenvalue decomposition

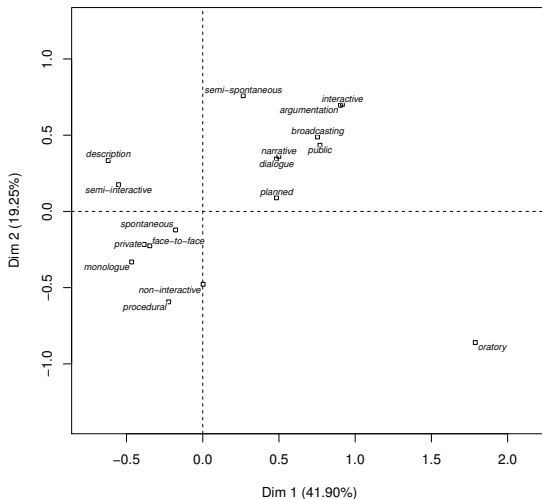
	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.51	41.90	41.90
comp 2	1.15	19.25	61.15
comp 3	1.01	16.88	78.03
comp 4	0.83	13.77	91.79
comp 5	0.33	5.56	97.35
comp 6	0.16	2.65	100



PCA graphs : variables & individuals



PCA graphs : variables & factors



PCA : interpretation

disconfirmed expectations

- ① distance between **semi-spontaneous** and **spontaneous** speech > distance between **planned** and **spontaneous** speech
- ② **planned speech** is characterized by a relatively high frequency of **overlaps**, but **spontaneous speech** is not

PCA : interpretation

disconfirmed expectation 1

- ① distance between **semi-spontaneous** and **spontaneous** speech > distance between **planned** and **spontaneous** speech

The semi-spontaneous samples are :

- **face-to-face**, **guided** interviews
- **well-planned** and highly **monitored**
- structured in a **question-and-answer** fashion
- the interviewer does not interrupt much but includes regular **back-channeling** devices such as *oui* ('yes') or *mh mh* ('hmm').

All the above have an impact on the production of **overlaps**.

PCA : interpretation

disconfirmed expectation 1

- ① distance between **semi-spontaneous** and **spontaneous** speech >
distance between **planned** and **spontaneous** speech

(1) L1: *le quartier vert \$- ce fameux -\$ quartier vert*
L2: *mh mh*

L1: 'the green district \$- this well-known -\$ green district'
L2: 'hmm'

(2) L1: *il y a \$- deux collèges là -\$ rue d'Alésia*
L2: *mh mh*

L1: 'there are \$- two junior high schools there -\$ rue d'Alésia'
L2: 'hmm'

(3) L1: *\$- donc -\$ euh*
L2: *oui*

L1: '\$-\$- so -\$ well'
L2: 'yes'

PCA : interpretation

disconfirmed expectation 2

- ② **planned speech** is characterized by a relatively high frequency of **overlaps**, but **spontaneous speech** is not

On the one hand...

- planned speech is prepared (wrt discourse topics and text structure)

On the other hand...

- interruptions and overlaps are common in planned speech when participants have **opposed viewpoints on controversial topics** (e.g. participants commonly contribute supplementary or contradictory arguments, initiate topic shifts or topic restarts)
- planned public dialogues tend to be strongly **interactive** and **argumentative**
- as evidenced by a high frequency of **prominences**

interim conclusion

- argumentative and narrative sequences are **prosodically marked**
- descriptive and procedural sequences are not
- a discourse genre is prosodically marked when it is :
 - characterized by a **high frequency of prosodic features** (overlaps + prominences + intonation packages)
 - **atypical** with respect to other speech genres
e.g. oratory speech is characterized by a high frequency of intonational periods and pauses and is consequently atypical

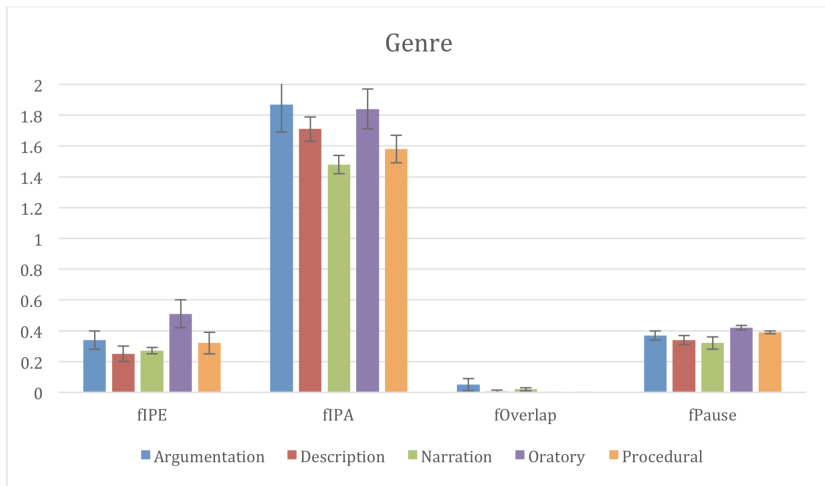
refined hypothesis

the internal structure of intonation packages is revealing of the nature of discourse genres

ANOVA – method

- we ran a non-parametric one-way ANOVA using the Kruskal-Wallis H test
- the ANOVA was conducted on our six between-item factors (situational variables) : Genres, i.e. Discourse sequences, Type of speech, Event structure, Channel, Planning and Interactivity.

ANOVA – Genre



ANOVA – focus on *fPauses* and *fOverlap*

fPauses

- a significant main effect of *Genre* ($p < 0.05$)
→ *fPauses* varies according to *Genre*
- the lowest *fPause* score was found in *Narration* ($M = 0.32$; $SD = 0.04$)
- the highest *fPause* score was observed in *Oratory* ($M = 0.42$; $SD = 0.01$)

fOverlap

- a significant main effect of *Genre* ($p < 0.001$)
→ *fOverlap* also varies according to *Genre*
- the highest *fOverlap* scores were found in both *Argumentation* ($M = 0.05$; $SD = 0.04$) and *Narration* ($M = 0.02$; $SD = 0.01$)
- no overlap was found in both *Oratory* and *Procedural* samples

ANOVA – focus on *fIPE* and *fIPA*

fIPE

- a significant main effect of *Genre* ($p < 0.001$)
→ *fIPE* varies according to *Genre*
- the lowest *fIPE* score was found in *Description* ($M = 0.25$; $SD = 0.05$)
- the highest *fIPE* score was observed in *Oratory* ($M = 0.51$; $SD = 0.09$)

fIPA

- a significant main effect of *Genre* ($p < 0.01$)
→ *fIPA* also varies according to *Genre*
- the lowest *fIPA* score was found in *Narration* ($M = 1.48$, $SD = 0.06$)
- the highest *fIPA* scores were found in both *Argumentation* ($M = 1.87$; $SD = 0.13$) and *Oratory* ($M = 1.84$; $SD = 0.13$)

ANOVA – summary

- *fPauses* and *fOverlap* are **objective** prosodic features
- *fIPE* and *fIPA* are **constructed** prosodic features
- all vary significantly across the **discourse types**
- these **prosodic preferences** of discourse type emerged despite the **limited size** and **heterogeneity** of the Rhapsodie corpus

back to our goals

Our aim was to :

- ① identify global trends regarding the interface between **prosodic constructions** and **situational variables**
- ② pinpoint **prosodic constructions** associated with certain **discourse types**
- ③ show that it is also possible to build a **predictive model** based on a small corpus

what we found

data

- **intuitively** : discourse sequences where the speaker is **strongly involved** are marked by **prominences** (e.g. *Argumentation*)
- **counterintuitively** : interactive sequences and dialogues have distinct prosodic profiles (as shown by statistically distinct FIPE frequencies in these two modalities)

methods

- the results obtained with PCA and ANOVA are broadly consistent,
- yet, there is a difference concerning the narrative modality for FIPA.
- can be due to limited data, or...
- just because a feature is frequent in a modality (PCA) does not mean that it is the best representation of this modality (ANOVA)

what we found

- primary and secondary prosodic data are **complementary** to describe the prosodic phenomena exemplified in the Rhapsodie corpus
- the accuracy of a feature for the prosodic description of situational variables varies according to
 - variables
 - statistical tools
- **each tool brings its own insight**