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

Guillaume Desagulier\textsuperscript{1}  
Frédéric Isel \textsuperscript{2} 
Anne Lacheret-Dujour \textsuperscript{2}  
Seongmin Mun \textsuperscript{3}

\textsuperscript{1}MoDyCo (UMR 7114) Paris 8, CNRS, Paris Nanterre & Institut Universitaire de France  
\textsuperscript{2}MoDyCo (UMR 7114) CNRS, Paris Nanterre  
\textsuperscript{3}MoDyCo (UMR 7114) CNRS, Paris Nanterre & Ajou University

AFLiCo JET 2018  
Université Paris Nanterre  
Friday, May 4, 2018
Characterizing discourse genres with prosodic features

Desagulier et al.
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 ($f_{Pauses}$)
- conversational overlaps ($f_{Overlap}$)
- gap fillers ($f_{Euh}$)

**secondary variables**
- mean numbers per second of prosodic prominences ($f_{Prom}$)
- intonational periods ($f_{IPE}$)
- intonation packages ($f_{IPA}$)
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?
• 57 corpus samples (rows)
• 12 variables
  • 6 primary variables (continuous)
  • 6 secondary variables (categorical)

<table>
<thead>
<tr>
<th>sample ID</th>
<th>fPauses</th>
<th>fOverlap</th>
<th>fEuh</th>
<th>fProm</th>
<th>fIPE</th>
<th>fIPA</th>
<th>subgenre</th>
<th>interactivity</th>
<th>social_context</th>
<th>event_structure</th>
<th>channel</th>
<th>planning_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>D0001</td>
<td>0.26</td>
<td>0.12</td>
<td>0.14</td>
<td>1.79</td>
<td>0.28</td>
<td>1.54</td>
<td>argumentation</td>
<td>interactive</td>
<td>private</td>
<td>dialogue</td>
<td>face-to-face</td>
<td>semi-spontaneous</td>
</tr>
<tr>
<td>D0002</td>
<td>0.42</td>
<td>0.11</td>
<td>0.10</td>
<td>1.80</td>
<td>0.33</td>
<td>1.75</td>
<td>argumentation</td>
<td>interactive</td>
<td>private</td>
<td>dialogue</td>
<td>face-to-face</td>
<td>semi-spontaneous</td>
</tr>
<tr>
<td>D0003</td>
<td>0.35</td>
<td>0.10</td>
<td>0.03</td>
<td>1.93</td>
<td>0.34</td>
<td>1.76</td>
<td>description</td>
<td>semi-interactive</td>
<td>private</td>
<td>dialogue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
<tr>
<td>D2006</td>
<td>0.37</td>
<td>0.00</td>
<td>0.00</td>
<td>2.22</td>
<td>0.58</td>
<td>1.51</td>
<td>oratory</td>
<td>semi-interactive</td>
<td>public</td>
<td>dialogue</td>
<td>face-to-face</td>
<td>planned</td>
</tr>
<tr>
<td>D2007</td>
<td>0.38</td>
<td>0.08</td>
<td>0.05</td>
<td>2.08</td>
<td>0.59</td>
<td>1.87</td>
<td>argumentation</td>
<td>interactive</td>
<td>public</td>
<td>dialogue</td>
<td>broadcasting</td>
<td>spontaneous</td>
</tr>
<tr>
<td>M0001</td>
<td>0.40</td>
<td>0.00</td>
<td>0.25</td>
<td>1.92</td>
<td>0.34</td>
<td>1.52</td>
<td>procedural</td>
<td>non-interactive</td>
<td>private</td>
<td>monologue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
<tr>
<td>M0002</td>
<td>0.40</td>
<td>0.00</td>
<td>0.21</td>
<td>2.17</td>
<td>0.23</td>
<td>1.75</td>
<td>description</td>
<td>semi-interactive</td>
<td>private</td>
<td>monologue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
<tr>
<td>M0003</td>
<td>0.30</td>
<td>0.00</td>
<td>0.19</td>
<td>2.11</td>
<td>0.20</td>
<td>1.63</td>
<td>procedural</td>
<td>non-interactive</td>
<td>private</td>
<td>monologue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
<tr>
<td>M0004</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
<td>1.92</td>
<td>0.68</td>
<td>2.03</td>
<td>procedural</td>
<td>non-interactive</td>
<td>private</td>
<td>monologue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
<tr>
<td>M0005</td>
<td>0.27</td>
<td>0.00</td>
<td>0.17</td>
<td>2.03</td>
<td>0.28</td>
<td>1.23</td>
<td>procedural</td>
<td>non-interactive</td>
<td>private</td>
<td>monologue</td>
<td>face-to-face</td>
<td>spontaneous</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>eigenvalue</th>
<th>percentage of variance</th>
<th>cumulative percentage of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp 1</td>
<td>2.51</td>
<td>41.90</td>
<td>41.90</td>
</tr>
<tr>
<td>comp 2</td>
<td>1.15</td>
<td>19.25</td>
<td>61.15</td>
</tr>
<tr>
<td>comp 3</td>
<td>1.01</td>
<td>16.88</td>
<td>78.03</td>
</tr>
<tr>
<td>comp 4</td>
<td>0.83</td>
<td>13.77</td>
<td>91.79</td>
</tr>
<tr>
<td>comp 5</td>
<td>0.33</td>
<td>5.56</td>
<td>97.35</td>
</tr>
<tr>
<td>comp 6</td>
<td>0.16</td>
<td>2.65</td>
<td>100</td>
</tr>
</tbody>
</table>
PCA graphs: variables & individuals
PCA graphs: variables & factors

Dim 1 (41.90%)
Dim 2 (19.25%)

- argumentation
- description
- narrative
- oratory
- procedural
- interactive
- semi-spontaneous
- broadcasting
- public
- dialogue
- monologue
- planned
- face-to-face
- planned
- semi-spontaneous
- spontaneous
- private
- non-interactive
- procedural
- oratory

Characterizing discourse genres with prosodic features
Desagulier et al.
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* $-$ *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

*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

Characterizing discourse genres with prosodic features
Desagulier et al.
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 ($p < 0.001$)
  $\rightarrow$ $fIPE$ 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 ($p < 0.01$)
  $\rightarrow$ $fIPA$ 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

- $f_{\text{Pauses}}$ and $f_{\text{Overlap}}$ are objective prosodic features
- $f_{\text{IPE}}$ and $f_{\text{IPA}}$ 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