ANNOTATION AND ANALYSIS OF OVERLAPPING SPEECH IN POLITICAL INTERVIEWS

Martine Adda-Decker, Claude Barras, Gilles Adda, Patrick Paroubek, Philippe Boula de Mareüil
LIMSI-CNRS Orsay FRANCE,
Benoît Habert
ICAR/ENS-LSH, Lyon FRANCE
TALK OUTLINE

• Framework
• Corpus
• Overlap segmentation
• Overlap tagset and annotation
• Analysis of overlapping speech
• Overlapping speech and disfluencies
• Conclusion and discussion
FRAMEWORK

- Overlapping speech in TV political interviews
- Questions addressed:
  - How to segment and annotate overlapping speech?
  - Typology of overlapping speech in relation with its intrusive nature
  - Speaker roles in different overlap types
  - Link with disfluencies
CORPUS

- 8 one-hour French TV shows
- 1 major figure (either politician or from civil society) interviewed by 3 journalists and 1 chairman
- Fine-grained transcriptions including discourse markers and disfluencies (filled pauses, repetitions, revisions)
- Overlaps transcribed using a customised version of Transcriber
  http://trans.sourceforge.net/
OVERLAP SEGMENTATION

Two overlapping cases:
case ①: no speaker change; case ②: speaker change.
OVERLAP SEGMENTATION

Overlap segmentation examples (cases ➊ and ➋) in the customised Transcriber annotation editor.
OVERLAP TAGSET AND ANNOTATION

Overlap tagset defined after an interactive process:

- good inter-annotator agreement
- mutually exclusive categories

4 overlap tags:

- **bck** *backchannel*: “hmm”, the listener follows the speaker;
- **cmp** *complementary*: the incoming speaker overlaps the main speaker but does not take the floor.
- **tst** *turn stealing*: the incoming speaker clearly interrupts the main speaker;
- **att** *anticipated turn taking*: the incoming speaker anticipates the end of the main speaker’s turn;
OVERLAP TAGSET AND ANNOTATION

Examples of the different overlap types, producing case ① (bck, cmp) and case ② (tst, att) overlaps.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bck</strong>: backchannel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A: it is simply /the fact/ /B: hmm/ that...</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>cmp</strong>: complementary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A: I have a last question /about/ /B: very short/ about your...</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>tst</strong>: turn stealing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A: and in /this case.../</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B: /I want to/ come back...</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>att</strong>: anticipated turn taking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A: and this leads to humanitarian /action?/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B: /well I/ think</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DISTRIBUTION OF SPEECH OVERLAPS PER TAG
PASSIVE/ACTIVE ROLES

3-4 overlaps per minute including less than 5% of the words of the corpus

Overlap frequency (# segments per minute), word rate and mean length for passive (P) and active (A) roles, for bck, cmp, tst and att.

<table>
<thead>
<tr>
<th>category</th>
<th>over. freq.</th>
<th>words %</th>
<th>mean length</th>
</tr>
</thead>
<tbody>
<tr>
<td>non intrusive overlaps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bck</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1.2</td>
<td>0.8</td>
<td>1.6</td>
</tr>
<tr>
<td>A</td>
<td>0.6</td>
<td>0.6</td>
<td>1.2</td>
</tr>
<tr>
<td>att</td>
<td>0.4</td>
<td>0.4</td>
<td>2.1</td>
</tr>
<tr>
<td>A</td>
<td>0.5</td>
<td>0.5</td>
<td>2.3</td>
</tr>
<tr>
<td>intrusive overlaps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cmp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>0.7</td>
<td>1.1</td>
<td>3.4</td>
</tr>
<tr>
<td>A</td>
<td>1.1</td>
<td>1.1</td>
<td>3.5</td>
</tr>
<tr>
<td>tst</td>
<td>1.1</td>
<td>1.7</td>
<td>3.3</td>
</tr>
<tr>
<td>A</td>
<td>1.9</td>
<td>1.9</td>
<td>3.8</td>
</tr>
</tbody>
</table>
ANALYSIS OF OVERLAPPING SPEECH
ATTACK/RESIST RATIO

Attack/resist ratio $R$

$$ R = \frac{A - P}{A + P} $$

Attack density $D$

$$ D = \left( \frac{100 \times A}{M + P + A} \right) $$

$A$ = active zone of the speaker; $P$ = passive; $M$ = mono-speaker (in words)

<table>
<thead>
<tr>
<th>set</th>
<th>$R$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>journalists</td>
<td>0.3</td>
<td>8.0</td>
</tr>
<tr>
<td>interviewees</td>
<td>-0.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Chairman</td>
<td>0.2</td>
<td>6.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>interviewees</th>
<th>$R$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IntPF0</td>
<td>-0.1</td>
<td>2.4</td>
</tr>
<tr>
<td>IntPF1</td>
<td>0.0</td>
<td>3.6</td>
</tr>
<tr>
<td>IntPF2</td>
<td>-0.2</td>
<td>3.4</td>
</tr>
<tr>
<td>IntPF3</td>
<td>-0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>IntCF1</td>
<td>-0.4</td>
<td>2.9</td>
</tr>
<tr>
<td>IntCF2</td>
<td>-0.7</td>
<td>1.2</td>
</tr>
<tr>
<td>IntPI1</td>
<td>-0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>IntPI2</td>
<td>-0.1</td>
<td>2.3</td>
</tr>
</tbody>
</table>
OVERLAPPING SPEECH AND DISFLUENCIES

Disfluency rates for the different segment types

**non-ov**: non overlapping speech

**over**: overlapping speech

**non-intr**: non-intrusive overlaps (**bck**, **att**)

**intr**: intrusive overlaps (**cmp**, **tst**)

![Box plot showing disfluency rates for non-overlapping and overlapping speech segments.](image-url)

---

**LREC 2008 Marrakech, 28th May 2008**
OVERLAPPING SPEECH AND DISFLUENCIES

Disfluency rates for the different segment types

- **non-ov**: non-overlapping speech
- **over**: overlapping speech
- **non-intr**: non-intrusive overlaps (bck, att)
- **intr**: intrusive overlaps (cmp, tst)

![Box plot showing disfluency rates for different segment types](image)

LREC 2008 Marrakech, 28th May 2008
OVERLAPPING SPEECH AND DISFLUENCIES

Disfluency rates for the different segment types

**non-ov**: non overlapping speech

**over**: overlapping speech

**non-intr**: non-intrusive overlaps (bck, att)

**intr**: intrusive overlaps (cmp, tst)
CONCLUSION AND DISCUSSION

• Annotation of overlapping speech
  – annotation process which preserve the interaction structure
  – reduced tagset to simplify the annotation
  – enables the study of overlap/disfluencies/speaker role

• Overlapping speech and disfluencies
  – twice more disfluencies on overlaps
  – very high %disfluencies for passive speakers in complementary overlaps
  – lower %disfluencies on backchannel than on non-overlapping speech

• Speaker role and disfluencies
  – high %disfluencies for journalists in passive/intrusive situation
OVERLAPPING SPEECH AND DISFLUENCIES

Disfluency rates for the different segment types

**bck**: backchannel; **cmp**: complementary; **tst**: turn stealing; **att**: anticipated turn taking

![Graph showing disfluency rates for different segment types](image-url)

LREC 2008 Marrakech, 28th May 2008
ANNOTATORS’ AGREEMENT

Overlap label distribution from 5 annotators (one show)

<table>
<thead>
<tr>
<th>label</th>
<th>count</th>
<th>bck</th>
<th>cmp</th>
<th>tst</th>
<th>att</th>
</tr>
</thead>
<tbody>
<tr>
<td>bck</td>
<td>63</td>
<td>91.1</td>
<td>8.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>cmp</td>
<td>50</td>
<td>9.2</td>
<td>75.8</td>
<td>15.0</td>
<td>0.0</td>
</tr>
<tr>
<td>tst</td>
<td>107</td>
<td>0.4</td>
<td>3.6</td>
<td>89.2</td>
<td>6.8</td>
</tr>
<tr>
<td>att</td>
<td>26</td>
<td>0.0</td>
<td>0.0</td>
<td>24.0</td>
<td>76.0</td>
</tr>
</tbody>
</table>

Kappa inter-annotation agreement 0.7 - 0.8
HOMOGENEOUS SPEECH REGIONS

H-Region:
maximum length segment keeping homogeneous speaker characteristics

<table>
<thead>
<tr>
<th>set</th>
<th>#H-regions</th>
<th>(%H-regions)</th>
<th>#words</th>
<th>(%words)</th>
<th>average length</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>4,000</td>
<td>(100)</td>
<td>83,000</td>
<td>(100)</td>
<td>20.7</td>
</tr>
<tr>
<td>mono-speaker</td>
<td>2,600</td>
<td>(65)</td>
<td>79,300</td>
<td>(95)</td>
<td>30.0</td>
</tr>
<tr>
<td>overlap</td>
<td>1,400</td>
<td>(35)</td>
<td>3,700</td>
<td>(5)</td>
<td>2.7</td>
</tr>
</tbody>
</table>
OVERLAPPING SPEECH
DISCOURSE MARKERS AND DISFLUENCIES

DM: discourse markers; FP: filled pauses; RV: revisions; RP: repetitions

<table>
<thead>
<tr>
<th>category</th>
<th>% DM</th>
<th>% FP</th>
<th>% RV</th>
<th>% RP</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>mono-speaker</td>
<td>2.4</td>
<td>2.0</td>
<td>2.5</td>
<td>2.5</td>
<td>6.9</td>
</tr>
<tr>
<td>overlaps</td>
<td>P</td>
<td>2.1</td>
<td>1.6</td>
<td>2.3</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>5.9</td>
<td>0.5</td>
<td>3.0</td>
<td>11.0</td>
</tr>
<tr>
<td>non-intrusive</td>
<td>P</td>
<td>2.4</td>
<td>1.6</td>
<td>2.0</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>7.2</td>
<td>0.6</td>
<td>0.9</td>
<td>5.2</td>
</tr>
<tr>
<td>intrusive</td>
<td>P</td>
<td>2.0</td>
<td>1.6</td>
<td>2.5</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>5.4</td>
<td>0.4</td>
<td>3.8</td>
<td>13.0</td>
</tr>
</tbody>
</table>
ANALYSIS OF OVERLAPPING SPEECH
ATTACK/RESIST RATIO

Attack/resist ratio $R$  

<table>
<thead>
<tr>
<th>set</th>
<th>$R$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>journalists</td>
<td>0.3</td>
<td>8.0</td>
</tr>
<tr>
<td>interviewees</td>
<td>-0.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Chairman</td>
<td>0.2</td>
<td>6.6</td>
</tr>
</tbody>
</table>

attack density $D$

<table>
<thead>
<tr>
<th>interviewees</th>
<th>$R$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinay</td>
<td>-0.1</td>
<td>2.4</td>
</tr>
<tr>
<td>Delors</td>
<td>0.0</td>
<td>3.6</td>
</tr>
<tr>
<td>Pasqua</td>
<td>-0.2</td>
<td>3.4</td>
</tr>
<tr>
<td>De Robien</td>
<td>-0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>Voynet</td>
<td>-0.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Brauman</td>
<td>-0.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Diouf</td>
<td>-0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>Brittan</td>
<td>-0.1</td>
<td>2.3</td>
</tr>
</tbody>
</table>