INFORMATIONAL LOAD AS A TRIGGER FOR DISFLUENCIES IN INTERPRETING

Koen Plevoets, Bart Defrancq
Overview
Intro
Research questions
Data
Method
Analysis
Conclusion
Intro

**Interpreting**: The rendition of utterances in another language

- Consecutive
- Simultaneous
- ...

`Real-time` translation
Intro

‘Interpreting is a cognitively demanding activity’

• Multitasking: Division of attention to different concurring tasks
• ‘Tightrope hypothesis’: Interpreters work at the limits of their processing capacities
  (Gile 1999)
Intro

Effort Model
(Gile 1985; 1997)

Interpreting = L + P + M + C

- L: Listening effort
- P: Production effort
- M: Memory effort
- C: Coordination effort
Intro

Cognitive Load Model  
(Seeber 2011; 2013)

Interpreting: “Real-time combination of language comprehension and language production task”

Both tasks have “demand vectors”:
  • Auditory verbal
  • Cognitive-verbal
  • Verbal-response (only for production)
Figure 1: A model of the simultaneous interpretation process.
Moser (1978)
Setton (1999)
Intro

Information overload unsettles interpreting

Errors and omissions

• Delivery rate (Gerver 1969; Pio 2003)
• Propositional density (Dillinger 1994; Tommola & Helevä 1998)

BUT: Vague definition of ‘interpreting error’
(Barik 1975; Gerver 1976)
Intro

**Information overload** unsettles interpreting

**Disfluencies:** Pauses, *uh*(m),…

(Goldman-Eisler 1967; Mead 2000; Tissi 2000; Cecot 2001)

**Vast psycholinguistic literature**

(Levett 1983; Arnold et al. 2000; 2003; Bortfeld et al. 2001; Clark & Fox Tree 2002; Watanabe et al. 2008,...)
### Intro

**Setton (1999: 247)**

<table>
<thead>
<tr>
<th></th>
<th>Attention to input</th>
<th>Attention to formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long silent pause</td>
<td>High</td>
<td>-</td>
</tr>
<tr>
<td>Short pausing</td>
<td>Normal listening</td>
<td>Routine planning</td>
</tr>
<tr>
<td>Filled pause</td>
<td>Normal listening</td>
<td>Routine planning</td>
</tr>
<tr>
<td>Mixed: Short &amp; filled pauses &amp; voice effects</td>
<td>Normal listening</td>
<td>Routine planning</td>
</tr>
<tr>
<td>Long filled pause</td>
<td>Relaxed or off</td>
<td>Planning/Searching</td>
</tr>
<tr>
<td>Fluent unmodulated string</td>
<td>Relaxed or off</td>
<td>Off</td>
</tr>
</tbody>
</table>
Intro

Naturalistic data: **Corpus-based**

(Gile 1998)

Next to classical study of source influence on target, also ‘Bakerian’ comparison of interpreting with non-interpreting

(Baker 1993)
Research questions

1. How does informational load manifest itself in interpreting vs. non-interpreting?
2. How does the input load differ from the output load in interpreting?
Data

2 corpora:

• **European Parliament Interpreting Corpus** – Ghent
• **Spoken Dutch Corpus** – component g
Data

**European Parliament Interpreting Corpus – Ghent**

Plenary sessions of the European Parliament 2006-2008

French, Spanish, Dutch, and English

190 000 tokens… and rising
Data

**European Parliament Interpreting Corpus – Ghent**

Transcribed according to VALIBEL-corpus (Bachy et al. 2007)
POS-tagged and chunked by means of LeTs (Van de Kauter et al. 2013)
Sentence-aligned with WinAlign
Data

Spoken Dutch Corpus – component g
(Oostdijk 2000)
Parliamentary debates

1998-2003
POS-tagged
Data

**Spoken Dutch Corpus – component g**
(Oostdijk 2000)

360 000 tokens

- Flanders: 140 000
- The Netherlands: 220 000
## Data

<table>
<thead>
<tr>
<th>Language</th>
<th>Type</th>
<th>Nr. of files</th>
<th>Nr. of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPICg</td>
<td>FRA (source)</td>
<td>108</td>
<td>1458</td>
</tr>
<tr>
<td></td>
<td>DUT (target)</td>
<td>108</td>
<td>1437</td>
</tr>
<tr>
<td>SPCg</td>
<td>240</td>
<td>19046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(FI)</td>
<td>155</td>
<td>8293</td>
</tr>
<tr>
<td></td>
<td>(NL)</td>
<td>85</td>
<td>10753</td>
</tr>
</tbody>
</table>
Method

Predict the nr. of \textit{uh(m)’s} in each sentence on the basis of:

- Lexical density
- Proportion of numbers (Gile 2009)
- Delivery rate
Method

**Lexical density**: Nr. of content words / (nr. of content words + nr. of function words)

**Proportion of numbers**: Nr. of numerals / total nr. of words

**Delivery rate**: Total nr. of words / total nr. of minutes
Method

**Multilevel rate model**
(Faraway 2006: 61-63, 221-230)

Predict nr. of *uh(m)*’s per sentence, conditioned on the sentence’s total nr. of words (‘offset’)

Random factor: files (108+240)
Method

First-level predictors:
• Lexical density
• Proportion of numbers

Second-level predictor:
• Delivery rate
Analysis

1. ‘Bakerian’ comparison of the output in the 3 ‘languages’ (\textit{NED\_or}, \textit{NED\_in} & \textit{FRA})
2. Comparison of the input and output in the interpretations
Analysis

1. ‘Bakerian’ comparison of the output in the 3 ‘languages’ (NED_or, NED_in & FRA)
2. Comparison of the input and output in the interpretations
Cf. Goldman-Eisler (1967); Gerver (1975)
Conclusion

Interpreters produce more $uh(m)$’s than non-interpreters

Rate of $uh(m)$ in interpreting is enhanced by lexical density and numbers in the source

In non-interpreting, $uh(m)$ is more related to the delivery rate
Conclusion

Negative effect of numbers in the target may point to omission

Prospect of applying Semantic Vector Spaces to interpreting

Underlines advantages of naturalistic corpus data
Thank you!

koen.plevoets@ugent.be
bart.defrancq@ugent.be