The effect of sentence splitting on cohesion in German business translations

Mario Bisiada

4th Using Corpora in Contrastive and Translation Studies Conference

26 July 2014
Sentence-splitting in translation

Corpus details

Method

Findings
What is sentence splitting?

(1) Functional staffers often deferred to their higher-ups in corporate rather than their division vice president, since functional leaders were responsible for promotions. (HBR 6/08,60)

Das Funktionspersonal beugte sich oft den Fachvorgesetzten in der Konzernzentrale statt dem Leiter des eigenen Geschäftsbereichs. Denn die Funktionsleiter in der Zentrale waren für Beförderungen zuständig. (HBM 9/08,58)
Theoretical assumptions about sentence splitting

- ‘a translation strategy enforced by differences in structural conventions between the languages’\(^1\)

\(^1\)(Fabricius-Hansen 1996: 558)
\(^2\)(see Fabricius-Hansen 1996, 1999; Ramm 2008; Solfjeld 2008)
\(^3\)(Fabricius-Hansen 1999: 203f; Fabricius-Hansen and Ramm 2008)
Theoretical assumptions about sentence splitting

- ‘a translation strategy enforced by differences in structural conventions between the languages’\(^1\)
- translation from a high to a low informational density language (e.g. DE > EN) argued to trigger sentence splitting

\(^1\)(Fabricius-Hansen 1996: 558)
\(^2\)(see Fabricius-Hansen 1996, 1999; Ramm 2008; Solfjeld 2008)
\(^3\)(Fabricius-Hansen 1999: 203f; Fabricius-Hansen and Ramm 2008)
Theoretical assumptions about sentence splitting

- ‘a translation strategy enforced by differences in structural conventions between the languages’\(^1\)
- translation from a high to a low informational density language (e.g. DE > EN) argued to trigger sentence splitting
- research largely limited to German (hypotactic, hierarchical)–Norwegian/English (paratactic, incremental) direction of translation\(^2\)

---

\(^1\)(Fabricius-Hansen 1996: 558)
\(^2\)(see Fabricius-Hansen 1996, 1999; Ramm 2008; Solfjeld 2008)
\(^3\)(Fabricius-Hansen 1999: 203f; Fabricius-Hansen and Ramm 2008)
Theoretical assumptions about sentence splitting

- ‘a translation strategy enforced by differences in structural conventions between the languages’¹
- translation from a high to a low informational density language (e.g. DE > EN) argued to trigger sentence splitting
- research largely limited to German (hypotactic, hierarchical)–Norwegian/English (paratactic, incremental) direction of translation²
- ‘grammatical possibilities and peculiarities of German favour a more hierarchical sentence construction than is feasible in English’³.

¹(Fabricius-Hansen 1996: 558)
²(see Fabricius-Hansen 1996, 1999; Ramm 2008; Solfjeld 2008)
³(Fabricius-Hansen 1999: 203f; Fabricius-Hansen and Ramm 2008)
Theoretical assumptions about sentence splitting

- ‘a translation strategy enforced by differences in structural conventions between the languages’\(^1\)
- translation from a high to a low informational density language (e.g. DE > EN) argued to trigger sentence splitting
- research largely limited to German (hypotactic, hierarchical)–Norwegian/English (paratactic, incremental) direction of translation\(^2\)
- ‘grammatical possibilities and peculiarities of German favour a more hierarchical sentence construction than is feasible in English’\(^3\).

-translating into a high informational density language should require the opposite strategy to sentence splitting!

\(^1\)(Fabricius-Hansen 1996: 558)
\(^2\)(see Fabricius-Hansen 1996, 1999; Ramm 2008; Solfjeld 2008)
\(^3\)(Fabricius-Hansen 1999: 203f; Fabricius-Hansen and Ramm 2008)
Corpus architecture

Source: *Harvard Business Manager, Harvard Business Review*
Corpus architecture

**Source:** *Harvard Business Manager, Harvard Business Review*

### Size of the translation and comparable corpus

<table>
<thead>
<tr>
<th></th>
<th>English STs</th>
<th>TTs</th>
<th>CC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982–3</td>
<td>251,148</td>
<td>246,341</td>
<td>145,715</td>
<td>643,204</td>
</tr>
<tr>
<td>2008</td>
<td>258,589</td>
<td>260,261</td>
<td>88,312</td>
<td>607,162</td>
</tr>
</tbody>
</table>
## Corpus architecture

Source: *Harvard Business Manager, Harvard Business Review*

### Size of the translation and comparable corpus

<table>
<thead>
<tr>
<th></th>
<th>English STs</th>
<th>TTs</th>
<th>CC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1982–3</strong></td>
<td>251,148</td>
<td>246,341</td>
<td>145,715</td>
<td>643,204</td>
</tr>
<tr>
<td><strong>2008</strong></td>
<td>258,589</td>
<td>260,261</td>
<td>88,312</td>
<td>607,162</td>
</tr>
</tbody>
</table>

### Size of pre-edited corpus

<table>
<thead>
<tr>
<th></th>
<th><strong>2006–11</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>English Source texts</td>
<td>104,678</td>
</tr>
<tr>
<td>Pre-edited German translations</td>
<td>106,829</td>
</tr>
<tr>
<td>Published German translations</td>
<td>104,448</td>
</tr>
<tr>
<td><strong>Total size</strong></td>
<td><strong>315,955</strong></td>
</tr>
</tbody>
</table>

‘Please translate as true to the original as possible in clear, understandable, and lively German. Do not leave out sentences or parts of sentences that seem hard to understand to you. Avoid nominalizations, subject-specific terminology, the passive, and impersonal constructions using man. Where possible, dissolve nested sentences, especially those containing dass.’ (my translation)
Method of analysis

- Sentence-alignment (using the text alignment tool in +Tools) → One-to-two sentence correspondence maintained
Method of analysis

- Sentence-alignment (using the text alignment tool in +Tools) → One-to-two sentence correspondence maintained
- File converted to *tab-delimited* TXT format → each line contains the corresponding ST and TT items

Sentence-splitting and cohesion in German translations
Method of analysis

- Sentence-alignment (using the text alignment tool in +Tools) → One-to-two sentence correspondence maintained
- File converted to *tab-delimited* TXT format → each line contains the corresponding ST and TT items
- Search for ‘.’ (full stop followed by space) yields only split/joined TT sentences
Method of analysis

- Sentence-alignment (using the text alignment tool in +Tools) → One-to-two sentence correspondence maintained
- File converted to \textit{tab-delimited} TXT format → each line contains the corresponding ST and TT items
- Search for ‘.’ (full stop followed by space) yields only split/joined TT sentences

Other full stops occur either at the end of a line or followed by a tab and so were not returned by the search.
Sentence splitting and cohesion in German translations

Method of analysis

- Sentence-alignment (using the text alignment tool in +Tools) → One-to-two sentence correspondence maintained
- File converted to tab-delimited TXT format → each line contains the corresponding ST and TT items
- Search for ‘.’ (full stop followed by space) yields only split/joined TT sentences
  ↪ Other full stops occur either at the end of a line or followed by a tab and so were not returned by the search.
  ↪ Full stops in abbreviations (‘U.S.’) or enumerations (‘1., 2.’) were manually removed in advance.
Two-step diachronic corpus analysis

Step 1—Translation corpus

- Differences in frequency of sentence-splitting between 1982–3 and 2008.
- Repeat analysis for the pre-edited corpus — What is the influence of editors on sentence splitting?
Method of analysis

Two-step diachronic corpus analysis

Step 1—Translation corpus

- Differences in frequency of sentence-splitting between 1982–3 and 2008.
- Repeat analysis for the pre-edited corpus — What is the influence of editors on sentence splitting?

Step 2—Comparable corpus

Are the observed phenomena characteristic of translated text or can they also be found in non-translated texts?
Analysis: Translation and pre-edited corpora

<table>
<thead>
<tr>
<th>Year</th>
<th>Joined</th>
<th>Split</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982–3</td>
<td>8</td>
<td>85</td>
</tr>
<tr>
<td>2008</td>
<td>12</td>
<td>292</td>
</tr>
</tbody>
</table>
Analysis: Translation and pre-edited corpora

Sentence splitting and cohesion in German translations

<table>
<thead>
<tr>
<th>Translator’s action</th>
<th>Editor’s action</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence split</td>
<td>—</td>
<td>542</td>
</tr>
<tr>
<td>Sentence split</td>
<td>split reverted</td>
<td>26</td>
</tr>
<tr>
<td>—</td>
<td>sentence split</td>
<td>296</td>
</tr>
<tr>
<td>2 sentences joined</td>
<td>—</td>
<td>20</td>
</tr>
<tr>
<td>2 sentences joined</td>
<td>joining reverted</td>
<td>0</td>
</tr>
</tbody>
</table>

Graph showing occurrences of sentence splitting and joining in 1982-3 and 2008:
- 1982-3: 8 sentence splits, 85 occurrences
- 2008: 12 sentence splits, 292 occurrences
(2)  a. Many younger employees find they can complete tasks faster than older workers, perhaps partly because of technological proficiency.

b. Viele dieser jungen Angestellten stellen fest, dass sie Aufgaben schneller erledigen als ihre älteren Kollegen. Das mag zum einen an ihrem technischen Können liegen.

‘That may be due to...’
(3) a. The gift of time—in the form of hours spent on coaching and building networks—is seen as crucial to the collaborative culture at Nokia.


‘This gift is seen as...’
(4)  

a. Johnson & Johnson offered Guidant shareholders $68 a share in late 2004, **which** wasn’t much of a premium over the stock’s trading price.


‘(Therein) / In that / That contained...’
## Translation & comparable corpus

### Development of referential items in the TC and CC

<table>
<thead>
<tr>
<th></th>
<th>1982–3</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n i/htw</td>
<td>n i/htw</td>
</tr>
<tr>
<td>TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Das</td>
<td>18 3.6</td>
<td>94 18.1</td>
</tr>
<tr>
<td>Dies/e/es/er</td>
<td>25 5.0</td>
<td>109 21.0</td>
</tr>
<tr>
<td>Pronominal adverb*</td>
<td>0 0.0</td>
<td>28 5.4</td>
</tr>
</tbody>
</table>

*darin, daraus, darum, darüber hinaus, darauf(hin), darunter, daran
## Development of referential items in the TC and CC

<table>
<thead>
<tr>
<th></th>
<th>1982–3</th>
<th></th>
<th>2008</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>i/htw</td>
<td>n</td>
<td>i/htw</td>
</tr>
<tr>
<td><strong>TC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Das</em></td>
<td>18</td>
<td>3.6</td>
<td>94</td>
<td>18.1</td>
</tr>
<tr>
<td><em>Dies/e/es/er</em></td>
<td>25</td>
<td>5.0</td>
<td>109</td>
<td>21.0</td>
</tr>
<tr>
<td><strong>Pronominal adverb</strong></td>
<td>0</td>
<td>0.0</td>
<td>28</td>
<td>5.4</td>
</tr>
<tr>
<td><strong>CC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Das</em></td>
<td>64</td>
<td>43.9</td>
<td>86</td>
<td>97.4</td>
</tr>
<tr>
<td><em>Dies/e/es/er</em></td>
<td>429</td>
<td>294.4</td>
<td>158</td>
<td>178.9</td>
</tr>
<tr>
<td><strong>Pronominal adverb</strong></td>
<td>25</td>
<td>17.2</td>
<td>13</td>
<td>14.7</td>
</tr>
</tbody>
</table>

*darin, daraus, darum, darüber hinaus, darauf(hin), darunter, daran*
Relative and modal clause constructions (*which, by*+part., gerund) are commonly split and expressed by sentence-initial anaphoric reference. → Change in the marking of cohesion?
Tentative conclusions

Relative and modal clause constructions (*which, by*+part., gerund) are commonly split and expressed by sentence-initial anaphoric reference. → Change in the marking of cohesion?

Expression of logical interrelationships in translation

- less often on the clause level by hypotactic or paratactic conjunctions or relative pronouns
- increasingly on the sentence-level by sentence-initial anaphoric reference
- apparently limited to translations
Relative and modal clause constructions (*which*, *by+part.*, gerund) are commonly split and expressed by sentence-initial anaphoric reference. → Change in the marking of cohesion?

Expression of logical interrelationships in translation

- less often on the clause level by hypotactic or paratactic conjunctions or relative pronouns
- increasingly on the sentence-level by sentence-initial anaphoric reference
- apparently limited to translations

→ Sentence splitting seems to be a phenomenon of translation that occurs irrespective of the informational density of the target language.
Thank you for your attention!

Read more:
Bisiada, Mario. 2014. “‘Lösen Sie Schachtelsätze möglichst auf’: The impact of editorial guidelines on sentence splitting in German business article translations’. *Journal of Applied Linguistics*. Advance online access.
→ doi:10.1093/applin/amu035
References


Solfjeld, Kåre. 2008. ‘Sentence splitting—and strategies to preserve discourse structure in German–Norwegian translations’ in Fabricius-Hansen, Cathrine and Wiebke Ramm (eds.).