Long distance dependencies in the Germanic sandwich

Jack Hoeksema\textsuperscript{1} and Ankelien Schippers\textsuperscript{2}

(1: Rijksuniversiteit Groningen, 2: Carl von Ossietzky Universität Oldenburg)
Outline

• Long-distance dependencies in English, German and Dutch
• A notorious sentence from the King’s Song
• Historical developments in German and Dutch
• *That*-trace effects
• Conclusions
Long-distance movement

• Movement over a finite clause boundary (typically signalled by the presence of *that* (*dat, dass*)
  1. \[ _{CP} \text{Who do you think} [ _{CP} \text{that Bill saw} \, t_{\text{who}} \, ?] \]

• Main features (cf. Chomsky, 1977):
  – Embedded finite clause contains a gap
  – Filler of this gap is in Spec-CP of a higher (matrix) clause
  – Movement is sensitive to so-called island constraints
  – Occurs in *wh*-questions, relatives, clefts, comparatives and topicalization construction
  – Occurs in quite restrictive contexts (cf. Ross, 1967)
The case of Dutch

- Line from the song composed for the inauguration of King Willem Alexander in 2013:

  “De dag die je wist dat zou komen is hier”
  *The day that you know that would come is here*
  *Den Tag der du wusste dass kommen wurde ist hier*

- Example of an LD-relative clause
- Created an enormous uproar in Dutch media, sentence considered to be “ungrammatical” and song “bad Dutch”
Some public opinions

• The Taalprof blog (by Peter Arno Coppen): the sentence is not wrong, similar constructions can be found in (older) literary works (see also Schippers, 2013)

• Reactions to the blog:

  * lul niet 'taalprof', tekst klopt van geen kanten, 'De dag die je wist dat zou komen', je kan t wel proberen te verdedigen met je goede gedrag, maar t slaat nergens op

  “don’t talk crap ‘language prof’, the text does not make any sense “the day that you knew that would come” – you can try and defend it with your good behavior, but it does not make any sense”

  * De tekst bevat veel te veel Anglicismen. Het lijkt of Ewbank eerst een Engelse tekst had en die door de vertaler van Google gehaald heeft

  “The text contains too many anglicisms. It seems as if Ewbank first had an English text and then ran it through Google Translate”
Interview with the culprits

Interviewer: *De dag die je wist dat zou komen*, that is not proper Dutch, is it?

John Ewbank: Well *de dag waarvan je wist dat die zou komen* didn’t work well (in this song)
Who is right?

• Everyone!
• LD-relatives are indeed highly unusual in present day Dutch (PDD), but are frequently attested in older stages of the language.
• The alternative construction with waarvan (called “resumptive prolepsis” is indeed much more common in PDD.
• LD-relatives indeed appear to be much more common in English (and thus may sound like an ‘anglicism’).
• Structurally speaking, there is nothing wrong with the sentence...
• ...but is does violate certain “soft” constraints on LD-movement.
Constraints on LD-movement

• Sensitive to the *that*-trace effect [* that t_{wh}*) (effectively banning LD-subject movement over a complementizer)
  – Leads to ungrammaticality in English, but also causes degraded acceptability in Dutch and German (cf. Maling & Zaenen, 1978; Bennis, 1980; Reuland, 1983; Featherston, 2005 and Den Dikken, 2007)

  \[ De \ dag \ die_i \ je \ wist \ dat \ t_i \ zou \ komen \]
  The day REL you know that would come

• Sensitive to so-called islands (not discussed here)
Corpus data on LD wh-movement

- Database of manually collected examples of LD movement in Dutch (ca. 2200, from 14\textsuperscript{th} century to now) and English (ca. 1100, from 18\textsuperscript{th} century to now) (cf. Hoeksema and Schippers, Schippers and Hoeksema and Schippers, 2012)
- Mostly from literary works
- Examples of all main types of LD-dependencies:
  - Wh-questions
  - Relatives (free and headed)
  - Topicalization constructions
  - Comparatives
## Dutch data

<table>
<thead>
<tr>
<th>period</th>
<th>WH</th>
<th>HR</th>
<th>FR</th>
<th>TOP</th>
<th>COMP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1620</td>
<td>4</td>
<td>45</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>62</td>
</tr>
<tr>
<td>1620 – 1669</td>
<td>9</td>
<td>41</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>1670 – 1719</td>
<td>6</td>
<td>138</td>
<td>5</td>
<td>46</td>
<td>4</td>
<td>199</td>
</tr>
<tr>
<td>1720 – 1769</td>
<td>11</td>
<td>157</td>
<td>2</td>
<td>41</td>
<td>15</td>
<td>226</td>
</tr>
<tr>
<td>1770 – 1819</td>
<td>19</td>
<td>172</td>
<td>1</td>
<td>43</td>
<td>7</td>
<td>242</td>
</tr>
<tr>
<td>1820 – 1869</td>
<td>24</td>
<td>130</td>
<td>13</td>
<td>39</td>
<td>11</td>
<td>217</td>
</tr>
<tr>
<td>1870 - 1919</td>
<td>59</td>
<td>94</td>
<td>26</td>
<td>20</td>
<td>15</td>
<td>214</td>
</tr>
<tr>
<td>1920 - 1969</td>
<td>135</td>
<td>48</td>
<td>31</td>
<td>10</td>
<td>25</td>
<td>249</td>
</tr>
<tr>
<td>1970 - 2015</td>
<td>478</td>
<td>55</td>
<td>91</td>
<td>18</td>
<td>71</td>
<td>713</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>744</td>
<td>881</td>
<td>172</td>
<td>242</td>
<td>152</td>
<td>2191</td>
</tr>
</tbody>
</table>
Diachronic developments in long-distance extraction

Graph 1: relative frequencies Dutch LD-constructions 1620-2015

- Two most frequent constructions (wh-questions and relatives) show an opposite development: wh-questions increase compared to relatives.
- Major change around second half of the 19th century.
- Topicalization constructions patterns with headed relatives.
The demise of LD relatives and topicalization

• Hoeksema & Schippers (2012), Schippers & Hoeksema (2010) and Schippers (2012): LD relatives and topicalization constructions disappear at the expense of resumptive prolepsis (here illustrated for LD relatives)

De dag [CP waarvan je wist dat die zou komen is hier]
the day where of you knew that [RP] would come is here
(RP = resumptive pronoun)

• Construction shows none of the hallmarks of LD movement (no gap, no that-trace, no sensitivity to islands) but is functionally similar to it (cf. Salzmann, 2006).

• Only possible when the proleptic object (i.e. de dag) is individual denoting and d-linked

• This explains why it is excluded for comparatives and free relatives (where the proleptic object would be null)
Percentage of resumptive prolepsis in Dutch headed relatives with long-distance dependencies
Wh-questions

- Resumptive prolepsis is also possible for wh-questions, so why do these not decline?
- Schippers (2012): corpus data suggests that wh-questions are not d-linked in most cases
- Furthermore, wh-questions increasingly show less lexical variation and are not very productive in that respect (cf. Schippers, 2012; Verhagen, 2006 and Dąbrowska, 2008)
- In ca. 50-90% of the cases (depending on the corpus one looks at) have the form of:
  WH denk je [dat … ]
  WH think you that
German

• No corpus data available, but descriptive grammars and other sources suggest LD constructions have also by an large disappeared (see Schippers 2012, section 4.4 for a literature overview)

• Timing of the decrease appears to coincide with Dutch change: middle of the 19th century.

• Main difference between Dutch and German: in German, LD wh-questions have also disappeared (especially in (northern) Germany).

• German, unlike Dutch, has alternatives for LD wh-questions: so-called “partial” wh-movement and parentheticals (extraction from embedded V2 clauses)
German alternatives to LD wh-questions

- Long-distance movement
  \[\%Wen \text{ meinst du dass Gunther } t_{\text{wen}} \text{ angerufen hat?}\]
- Partial wh-movement
  \[\text{Was meinst du wen Gunther } t_{\text{wen}} \text{ angerufen hat}\]
- (integrated) parenthetical ("extraction from embedded V2", cf. Reis, 1995 & 1996)
  \[\text{Wen (meinst du), hat Gunther } t_{\text{wen}} \text{ angerufen}\]
- Whereas partial wh-movement and parenthetical questions are fully productive in German, they are much less common in Dutch and generally do not function as functional alternatives to LD wh-movement.
What about English?

- English does not show the same asymmetry as Dutch
- LD movement still appears to be productive across the board (judging from the literature and native speakers' judgements)

<table>
<thead>
<tr>
<th>Construction</th>
<th>Frequency</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-questions</td>
<td>459</td>
<td>40.2</td>
</tr>
<tr>
<td>Headed relatives</td>
<td>412</td>
<td>36</td>
</tr>
<tr>
<td>Free relatives</td>
<td>176</td>
<td>15.4</td>
</tr>
<tr>
<td>Topicalization</td>
<td>21</td>
<td>1.8</td>
</tr>
<tr>
<td>Comparatives</td>
<td>79</td>
<td>7</td>
</tr>
<tr>
<td>It-cleft</td>
<td>4</td>
<td>0.3</td>
</tr>
<tr>
<td>Though-movement</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>1143</td>
<td>100</td>
</tr>
</tbody>
</table>
English compared to Dutch (20/21st century)

**Graph 2: Relative Frequencies per Construction 1900-2015: English vs. Dutch**

<table>
<thead>
<tr>
<th>Construction Type</th>
<th>% within language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-question</td>
<td>60</td>
</tr>
<tr>
<td>Headed relative</td>
<td>40</td>
</tr>
<tr>
<td>Free relative</td>
<td>10</td>
</tr>
<tr>
<td>Topicalization</td>
<td>0</td>
</tr>
<tr>
<td>Comparative</td>
<td>0</td>
</tr>
</tbody>
</table>

- English
- Dutch
Is LD-movement really productive in English?

- LD-movement appears more productive in English... However:
- Corpus data reveals that LD-extraction almost invariably involves *that*-less clauses (95% of the cases)!
- Do these examples really involve LD-movement proper?
- Generalized *that*-trace effect?
That-trace in English

Schippers (2012): acceptability ratings on subject and object extraction with and without *that*:

Subject - Ø
Who did Robert think will blame Sandra?
Subject - that
Who did Sandra think that will blame Robert?
Object - Ø
Who did Robert think Sandra will blame?
Object – that
Who did Sandra think that Robert will blame?

• Allows us to investigate the effect of the presence vs. absence of the complementizer.
• Comparison with filler items without extraction (declaratives) shows whether extraction and presence of complementizer interact:

Declarative – Ø
Roland believes Deanna may plant a tree
Declarative – that
Hector hopes that Maxine will sell her boat
Results *that*-trace

**EXTRACTION AND THAT IN ENGLISH**

- Extraction causes a significant drop in acceptability (difference between declaratives and extraction constructions always $p \leq 0.000$).
- Under extraction - presence of *that* causes a significant drop in acceptability (already for object extraction!)
- Under subject movement, presence of *that* creates the strongest effect
corpus data

- 1143 occurrences from English
- subject extractions: 341 without complementizer, 0 with complementizer
- all other extractions: 751 without complementizer, 51 with.

Comparison with COCA and BNC:
- **COCA**: *think that he*: 2059, *think he*: 25232: 92% *that*-drop
- **BNC**: *think that he*: 166, *think he*: 3067: 95% *that* drop
- *think + that* in LD data: 2, without *that* 428: 99.5% *that*-drop
- Difference COCA and LD data: chi square 29.8, p < 0.0001
- Difference BNC and LD data: Fisher’s exact test, p < 0.0001
- For other verbs we checked, *say* and *hope*, also significant differences between LD cases and general usage data
Conclusions

• Dutch in between German (almost no long-distance movement) and English. English represents the most archaic system, German is the most innovative, and Dutch is in between.
• German has alternatives for WH-movement that can be found in Dutch, but are not yet common
• German and Dutch have replaced long-distance relativization by resumptive prolepsis
• Both Dutch and English have a weak island effect for extraction from headed finite CPs, which is strongest for subjects (“that-trace effect”) but can also be noted for extraction of nonsubjects
• English has the option of dropping the complementizer, Dutch replaces extraction from dat-clauses whenever possible by alternatives
References


Schippers, A (2013). De dag die ik niet wist dat zou komen. MS. Url: sites.google.com/site/ankelieneschippers/home/koningslied.