Om-Omission and Filler-Gap Dependencies

Gosse Bouma

Centre for Language and Cognition
University of Groningen

Structure and Evidence in Linguistics, Stanford, April 2013
Optional Complementizer Om

What explains absence or presence of om?

De Indiërs aarzelen te investeren in Uganda
The Indians hesitate to invest in Uganda
*The Indians hesitate to invest in Uganda*

Moser had overwogen om zijn avontuur af te blazen
Moser had considered COMP his adventure PRT to cancel
*Moser had considered to cancel his adventure*
Optional Complementizer Om

What explains absence or presence of **om**?

De Indiërs aarzelen **om** te investeren in Uganda

The Indians hesitate **COMP** to invest in Uganda

*The Indians hesitate to invest in Uganda*

Moser had overwogen **zijn avontuur af te blazen**

Moser had considered **his adventure PRT to cancel**

*Moser had considered to cancel his adventure*
Optional Complementizer Om

What explains absence or presence of om?

De Indiërs aarzelen om te investeren in Uganda
The Indians hesitate COMP to invest in Uganda
The Indians hesitate to invest in Uganda

Moser had overwogen zijn avontuur af te blazen
Moser had considered his adventure PRT to cancel
Moser had considered to cancel his adventure

Filler-gap dependencies as predictor

Gap locations inside Om-te-infinitives generally considered ok
Optional Complementizer Om

What explains absence or presence of om?

De Indiërs aarzelen om te investeren in Uganda
The Indians hesitate COMP to invest in Uganda
*The Indians hesitate to invest in Uganda*

Moser had overwogen zijn avontuur af te blazen
Moser had considered his adventure PRT to cancel
*Moser had considered to cancel his adventure*

Filler-gap dependencies as predictor

Gap locations inside *Om-te-infinitives* generally considered ok
But hardly occurs in corpus data (this talk)
The Dutch complementizer Om

Om as optional complementizer in to-infinitive complements

De Kok *vraagt*\(_V\) (om) 1 procent van hun inkomen te geven aan het fonds
De Kok asks (COMP) 1 percent of their income to give to the fund
*De Kok asks to donate 1 percent of their income to the fund*

Ik ben niet *vrij*\(_A\) (om) daarover te spreken
I am not free (COMP) about-that to speak
*I am not free to speak about that*

Ik *hou er* niet *van*\(_P\) (om) Beverly Hills af te kammen
I like there not PRT (COMP) Beverly Hills PRT to disrespect
*I do not like to criticize Beverly Hills*

Huurders krijgen het *recht*\(_N\) (om) mee te praten
tenants obtain the right (COMP) with to talk
*Tenants obtain the right to have a say*
**Historical Development**

**IJbema 2002**

- *Om* originated as preposition
- Later used as complementizer in purpose modifier clauses
- Use as complementizer in *complement clauses* is recent development (rare before 1750)
Disapproval in Prescriptive Linguistics

Overview from Jansen 1987

- Brill (1852), no objections
- Woordenboek der Nederlandse Taal, 1869 (lemma *om*): *Om* behoort altijd een doel, eene bestemming, of eene strekking aan te wijzen” (*Om should always indicate a goal, purpose, or consequence*)
- WNT, 1934 (lemma *te*): no objections
- Van Es and Van Caspel (1971-75): *Om* is *superfluous*, typical of informal language
- ‘Nog in 1973 moet de redactie [van Onze Taal] inzenders die om als ‘slokdarmgeluid’ betitelen verdraagzaamheid voorhouden’ ('Even in 1973 the editors of Onze Taal had to plea for tolerance to members who described om as a guttural sound')
- Algemene Nederlandse Spraakkunst (1984): In spoken language there is a preference for *om*, leaving *om* out makes a formal impression
That-deletion in English complement clauses

The athlete realized that her goals would be difficult to achieve

Syntactic Complexity

Features that play a role in predicting presence of that:
- complexity of complement clause (cc), distance between governor and cc, frequency of governor, complexity of cc subject, subject starts with that, ...

Lexical bias (Roland et al 2006)

that-bias(governor) = ln(occurrences with that / occurrences without that)

Information Density (Jaeger 2010)

complement-bias(governor) = ln(occurrences with cc / 1 - occurrences with cc)
That-deletion in English complement clauses

The athlete realized that her goals would be difficult to achieve

Syntactic Complexity

Features that play a role in predicting presence of that:
- complexity of complement clause (CC), distance between governor and CC, frequency of governor, complexity of CC subject, subject starts with that, ...

Lexical bias (Roland et al 2006)

that-bias(governor) = ln(ccs with that) / ccs without that

Information Density (Jaeger 2010)

complement-bias(governor) = ln(occurrences with CC) / 1 – occurrences with CC
**That-deletion in English complement clauses**

The athlete realized that her goals would be difficult to achieve

### Syntactic Complexity

Features that play a role in predicting presence of *that*:

- complexity of complement clause (CC), distance between governor and CC, frequency of governor, complexity of CC subject, subject starts with *that*, ...

### Lexical bias (Roland et al 2006)

\[
\text{that-bias(governor)} = \ln \frac{\text{CCs with } that}{\text{CCs without } that}
\]

### Information Density (Jaeger 2010)

\[
\text{complement-bias(governor)} = \ln \frac{\text{occurrences with CC}}{1 - \text{occurrences with CC}}
\]
That-deletion in English complement clauses

The athlete realized that her goals would be difficult to achieve.

**Syntactic Complexity**

Features that play a role in predicting presence of *that*:
- complexity of complement clause (CC), distance between governor and CC, frequency of governor, complexity of CC subject, subject starts with *that*, ...

**Lexical bias (Roland et al 2006)**

\[
\text{that-bias(governor)} = \ln \frac{\text{CCs with } that}{\text{CCs without } that}
\]

**Information Density (Jaeger 2010)**

\[
\text{complement-bias(governor)} = \ln \frac{\text{occurrences with CC}}{1 - \text{occurrences with CC}}
\]
Corpus Study

Data
- Twente Newspaper Corpus (approx 400M words)
- Corpus of Spoken Dutch (10M words)

Annotation
- Automatically parsed with the HPSG inspired Alpino parser for Dutch (van Noord 2006)
- Output is dependency analysis (with phrasal nodes)
Ik spreek af vandaag thuis te blijven
I arrange to stay at home today
In OESO-verband is afgesproken om die subsidies vanaf 1 januari te schrappen

*In OECD context it was agreed to stop those subsidies as of January, 1st*
Spoken vs. Written

Percentage Om in spoken and written material for selected verbs

<table>
<thead>
<tr>
<th>Verb</th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td>vergeet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weiger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>besluit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overweeg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dwing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>raad_aan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beslis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spreek_achteruit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neem_voor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vraag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>maak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>verplicht</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nodig_uiterlijk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vind</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Om-bias does not correlate with Complement-bias
Hard to go beyond lexical bias

- Approx 25% of the to-infinitival complements are introduced by *om*
- *om*-bias of the governing *verb* is by far strongest predictor
- *Syntactic complexity* plays a small but significant role
- *Semantic association* plays a small but significant role
- Results (far) less clear than for English *that*-deletion
Filler-gap dependencies inside Om-te Infinitives?

Island Constraints vs. Processing Difficulties

Our goal is to assess certain arguments that have been made to the effect that grammatical constraints must be involved in island phenomena ... it is our contention that independently motivated processing factors can successfully explain a substantial amount of the judgment variation that has been used to motivate island constraints within grammar.


Role of Corpus Research

- Can automatically annotated treebanks provide any insights in this discussion?
- Non-local filler-gap dependencies are not very frequent in text
- Parse results (automatic annotation) are not very accurate for such cases
Filler-gap dependencies inside Om-te Infinitives?

Island Constraints vs. Processing Difficulties

Our goal is to assess certain arguments that have been made to the effect that grammatical constraints must be involved in island phenomena ... it is our contention that independently motivated processing factors can successfully explain a substantial amount of the judgment variation that has been used to motivate island constraints within grammar.


Role of Corpus Research

- Can automatically annotated treebanks provide any insights in this discussion?
- Non-local filler-gap dependencies are not very frequent in text
- Parse results (automatic annotation) are not very accurate for such cases
Filler-gap dependencies inside Om-te Infinitives?

Extraction from OTI

- Linguistic literature suggests that such extractions are ok
- Hans Bennis (2000), *Adjectives and Argument Structure*
  - Waar<i> is Jan bang *(om)* over *t*<i> te praten
  - Wat<i> heeft Jan geprobeerd *(om)* *t*<i> te lezen
  - “it must be mentioned that the complementizer is preferably dropped”

Challenges for Corpus Study

- Wh-questions hardly occur in newspaper corpus
- Relative clauses do occur frequently
- But most filler-gap dependencies are ’local’...
Filler-gap dependencies inside Om-te Infinitives?

Extraction from OTI

- Linguistic literature suggests that such extractions are ok
- Hans Bennis (2000), *Adjectives and Argument Structure*
  - Waar is Jan bang (om) over ti te praten
  - Wat heeft Jan geprobeerd om ti te lezen
  - “it must be mentioned that the complementizer is preferably dropped”

Challenges for Corpus Study

- Wh-questions hardly occur in newspaper corpus
- Relative clauses do occur frequently
- But most filler-gap dependencies are ’local’...
Gap Location in Relative Clauses

15.361 Relative Clauses in 100,000 sentences (Wikipedia sample)
Distribution of Gaps in Relatives dominating an (O)TI

Relatives with Object gap in (O)TI complement

een tekstbericht dat Tankink besloot voorlopig te bewaren

*a text-message that Tankink decided to store for the moment*

het werk wat veel jongeren vertikken om te doen

*the work that most youth refuse to do*

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences containing a verb with an (O)TI complement</td>
<td>285,000</td>
<td>100</td>
</tr>
<tr>
<td>Relatives containing a verb with an (O)TI compl.</td>
<td>22,059</td>
<td>7.74</td>
</tr>
<tr>
<td>Relatives with ’local’ filler-gap dependency</td>
<td>17,449</td>
<td>6.12</td>
</tr>
<tr>
<td>Relatives with non-local filler-gap dependency</td>
<td>4,610</td>
<td>1.62</td>
</tr>
</tbody>
</table>
**Distribution of Gaps in Relatives dominating an (O)TI**

### Relatives with Object gap in (O)TI complement

- Een tekstbericht *dat* Tankink besloot *voorlopig te bewaren*  
  *a text-message that Tankink decided to store for the moment*

- Het werk *wat* veel jongeren vertikken *om te doen*  
  *the work that most youth refuse to do*

<table>
<thead>
<tr>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>285,000</td>
<td>100</td>
</tr>
<tr>
<td>22,059</td>
<td>7.74</td>
</tr>
<tr>
<td>17,449</td>
<td>6.12</td>
</tr>
<tr>
<td>4,610</td>
<td>1.62</td>
</tr>
</tbody>
</table>
Distribution of Gaps in Relatives dominating an (O)TI

Relatives with Object gap in (O)TI complement

een tekstbericht dat Tankink besloot voorlopig te bewaren
a text-message that Tankink decided to store for the moment

het werk wat veel jongeren vertikken om te doen
the work that most youth refuse to do

<table>
<thead>
<tr>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences containing a verb with an (O)TI complement</td>
<td>285,000</td>
</tr>
<tr>
<td>Relatives containing a verb with an (O)TI compl.</td>
<td>22,059</td>
</tr>
<tr>
<td>Relatives with 'local' filler-gap dependency</td>
<td>17,449</td>
</tr>
<tr>
<td>Relatives with non-local filler-gap dependency</td>
<td>4,610</td>
</tr>
</tbody>
</table>
Distribution of Gaps in Relatives dominating an (O)TI

Relatives with Object gap in (O)TI complement

een tekstbericht dat Tankink besloot voorlopig te bewaren
a text-message that Tankink decided to store for the moment

het werk wat veel jongeren vertikken om te doen
the work that most youth refuse to do

<table>
<thead>
<tr>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences containing a verb with an (O)TI complement</td>
<td>285,000</td>
</tr>
<tr>
<td>Relatives containing a verb with an (O)TI compl.</td>
<td>22,059</td>
</tr>
<tr>
<td>Relatives with 'local' filler-gap dependency</td>
<td>17,449</td>
</tr>
<tr>
<td>Relatives with non-local filler-gap dependency</td>
<td>4,610</td>
</tr>
</tbody>
</table>
Distribution of Gaps in Relatives dominating an (O)TI

Relatives with Object gap in (O)TI complement

een tekstbericht dat Tankink besloot voorlopig te bewaren
*a text-message that Tankink decided to store for the moment*

het werk wat veel jongeren vertikken om te doen
*the work that most youth refuse to do*

<table>
<thead>
<tr>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences containing a verb with an (O)TI complement</td>
<td>285,000</td>
</tr>
<tr>
<td>Relatives containing a verb with an (O)TI compl.</td>
<td>22,059</td>
</tr>
<tr>
<td>Relatives with 'local' filler-gap dependency</td>
<td>17,449</td>
</tr>
<tr>
<td>Relatives with non-local filler-gap dependency</td>
<td>4,610</td>
</tr>
</tbody>
</table>
Relatives with Object gap in (O)TI

- Examples automatically extracted from corpus
- 849 occurrences (manually checked)

```xml
//node[@rel="rhd"] and
  @index = ..//node[@cat="ti" and @rel="vc"]
    //node[@rel="obj1"]/@index
  and not(@index = //node[@rel="su"]/@index)
```
Optional Om Corpus Study  Filler Gap  Om and Gaps  Collocations

All (10) relatives with Object gap in OTI (1)

het werk wat veel jongeren vertikken om te doen
the work which many youth refuse to do

de sociale problemen die de kunst geacht wordt om op te lossen
the social problems that art is supposed to solve

een nieuwszender die de kabelmaatschappijen niet verplicht zullen zijn om in het pakket op de kabel door te geven
a news channel that cable companies will not be obliged to include in their offers

de plichten die ze het beneden hun waardigheid achten om thuis te vervullen
the duties that they consider to be below their dignity at home

wat ik vergeten ben om te vertellen
what I forgot to tell

De maatregelen welke men van plan is om door te voeren
Measurements which one is planning to implement
All (10) relatives with Object gap in OTI (2)

(all from same day/article:)

Bijna heb ik gedaan wat ik me heb voorgenomen om te doen
Almost, I did what I had planned to do

mensen die doen wat ze zich hebben voorgenomen om te doen
doen wat je je hebt voorgenomen om te doen
doen wat je je voorneemt om te doen
Relatives with Object gap in TI (839 cases)

Het is dit kantoortje dat Eva Joly besluit voortaan zelf schoon te houden
It is this office that Eva Joly decides to clean herself from now on

dingen die de overheid besloot niet te doen
things that the government decided not to do

de Golf GTI, die VW-importeur Pon ( nog ) niet van plan is naar Nederland te halen
the Golf GTI, that VW-dealer Pon (not) plans to bring to the Netherlands

brandbommen, die men van plan was op het terrein van het azc te gooien
fire bombs that one was planning to throw on the premises of the refugee center

het presidentiële vliegtuig dat hij had beloofd direct van de hand te zullen doen
the presidential plane that he had promised to do away immediately

de problemen die Schröder belooft met harde hand te zullen bestrijden
the problems which Schröder promises to attack fiercely
Filler-gap dependencies in (O)TI

<table>
<thead>
<tr>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>839</td>
</tr>
<tr>
<td>OTI</td>
<td>10</td>
</tr>
</tbody>
</table>

- Non-local filler-gap dependency accurately predicts absence of *om*
- But non-local filler-gap dependencies are very scarce in general
  - 0.3% of all relevant data
- So hardly useful for statistical model predicting *om*
Filler-gap dependencies in (O)TI

**Statistics**

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI</td>
<td>839</td>
<td>98.8</td>
</tr>
<tr>
<td>OTI</td>
<td>10</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**Filler-gap dependency predicts absence of Om**

- non-local filler-gap dependency accurately predicts absence of *om*
- But non-local filler-gap dependencies are very scarce in general
  - 0.3% of all relevant data
- So hardly useful for statistical model predicting *om*
Collocational effects

Relative frequency of verbs governing TI in filler-gap dependencies and in general TI corpus

- Filler-gap
- General
Achten (consider, be supposed to)

De eerbied [die] we *geacht* worden *jegens het koningshuis te koesteren*  
*The respect that we are supposed to cherish concerning the royal family*

de vrouw *die* hij *geacht* wordt *te schaduwen*  
*the woman that he is supposed to shadow*

antwoorden *die* ook minister Zalm wordt *geacht* *met overtuiging uit te spreken*  
*answers which minister Zalm also is supposed to utter with conviction*

degeneren *die* hij *geacht* wordt *in de gaten te houden*  
*those which he is supposed to keep an eye on*

concurrentiestrijd *die* de uitvoerders worden *geacht* *aan te gaan*  
*competition which the responsibles are supposed to engage in*

Edelgassen doen toch [wat] ze *geacht* worden *niet te doen*  
*Noble gasses after all do what they are supposed not to do*

werk *dat* in Nederland de leraar *geacht* wordt *er ook nog eens bij te doen*  
*work which, in the Netherlands, the teacher is supposed to do along the side*
Conclusions

Syntax and Lexicon

- Syntactic variation clearly governed by **lexical items** (*that*-deletion, *om*-omission)
- Distribution of filler-gap dependencies shows **collocational effects**

Large (Automatically) Annotated Treebanks

- Provide detailed syntactic information
- Allow study of frequency effects in **syntactic** constructions
- Allow search for **rare syntactic phenomena**
  - Relative clauses with non-local filler gap dependencies
  - ...containing a specific clausal complement
  - ...and a specific verbal head
Greetings from Groningen!