Processing shapes grammar
But whose processing are we talking about?

Dirk Pijpops\textsuperscript{1,2} & Freek Van de Velde\textsuperscript{1}

\textsuperscript{1}QLVL Research Unit, University of Leuven
\textsuperscript{2}Research Foundation Flanders (FWO)

DGfS workshop 'Coding asymmetries' (Saarbrücken 8-10 March 2017)
Introduction

• Basic tenet of usage-based linguistics: processing -> usage -> grammar

**Performance-Grammar Correspondence Hypothesis** (Hawkins 2004: 3)

*Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.*
Introduction

• Building on Hawkins' theories, Rohdenburg proposes the **Cognitive Complexity Principle**:

  *In case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments.* (Rohdenburg 1996: 151)

  Construction + Ø  
  Construction + X<sub>morpheme/word/...</sub>  

  in less complex environments  
  in more complex environments
Introduction

• Why?
  • Processing-driven (cognitive complexity)
  • Whose processing?
  • Addressee's processing (Rohdenburg 1996:149)
    • The extra element aids the hearer
    • For the speaker, adding an extra element just adds to the cognitive burden

• Why the hearer? That is counter-intuitive, as:
  • Speaker's altruism is evolutionarily implausible (Kirby 1999)
  • Bottleneck in human communication is in encoding, not decoding (Levinson 2000: 28)
Ok. Now we have our straw men.
das Armdrücken
in
Saarbrücken
First a few words on clause structure in Dutch

- Works pretty much like German
- Topological approach with a bipolar structure (Klammerstruktur) (Zifonun 1997: 1498; Zwart 2011: 26)
- Ignoring the left-detached and the right-detached position, the schema for main clauses is:

<table>
<thead>
<tr>
<th>Prefield</th>
<th>1st pole</th>
<th>Midfield</th>
<th>2nd pole</th>
<th>Postfield (-&gt; extraposition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ik</td>
<td>zoek</td>
<td>(naar) een boek over taalkunde</td>
<td>gezocht</td>
<td></td>
</tr>
<tr>
<td>Ik</td>
<td>heb</td>
<td>(naar) een boek over taalkunde</td>
<td>gezocht</td>
<td>over taalkunde</td>
</tr>
<tr>
<td>Ik</td>
<td>heb</td>
<td>(naar) een boek</td>
<td>gezocht</td>
<td>een boek over taalkunde</td>
</tr>
<tr>
<td>Ik</td>
<td>heb</td>
<td>gezocht</td>
<td>naar een boek over taalkunde</td>
<td></td>
</tr>
<tr>
<td>XP</td>
<td>V-fin</td>
<td>XP</td>
<td>V-nonfin</td>
<td>any XP that starts with a relator (no bare NP)</td>
</tr>
</tbody>
</table>
The issue

• Verb *zoeken* occurs in two variants: with a DO and with a PO

  1. *We zoeken alternatieven.* (WR-P-P-G-0000254655.p.11.s.5)
     ‘We are looking for alternatives.’

  2. *Wij zoeken dan wel naar alternatieven.* (WR-P-P-G-0000488037.p.6.s.3)
     ‘We, then, look for alternatives.’

• The PO (in 2) is the 'bulkier' variant and may be expected to occur in cognitively more complex contexts, following Rohdenburg (1996)
1. Relevance for corpus linguists: Do psycholinguistic mechanisms of complexity affect language use itself? Do we find their influence in naturally occurring language use, outside of experimental settings? **Does processing shape (probabilistic) grammar?**

2. Relevance for psycholinguists: **Whose processing are we talking about?** The producer's or the addressee's?
Does processing shape (probabilistic) grammar?


- Why written language? To be **hyperconservative** (Ford & Bresnan 2013)

- Excluded tweets, text messages, chats, discussion lists: quality of syntactic parses deemed too low

- Extracted all instances of *zoeken 'to search'*', in which the object is overtly expressed: 61998 without *naar* vs. 17440 with *naar*
Does processing shape (probabilistic) grammar?

- 61998 without naar ↔ 17440 with naar

- As the object becomes more complex, the probability of naar increases (positive estimate for Object Length: 0.41)

- Highly significant: < 0.0001
Whose processing are we talking about?

- Producer-driven Hypothesis 1: *naar* allows the producer to extrapose long objects to the postfield
- Producer-driven Hypothesis 2: *naar* functions as a grammatical *uh*, buying time for the producer to formulate a complex object
- Addressee-driven Hypothesis: *naar* functions as a grammatical signpost for the addressee. It marks 'what follows now, is the object of the verb'

Remove the observations where the object is extraposed to the postfield

Remove: *Het stadsbestuur heeft daarom gezocht naar een efficiëntere en goedkopere oplossing*

Keep: *Nijmegen zoekt naar een oplossing*

**Prediction:** as the object becomes more complex, the probability of *naar* will **no longer** increase
Whose processing are we talking about?

- Producer-driven Hypothesis 1: *naar* allows the producer to extrapose long objects to the postfield
- Producer-driven Hypothesis 2: *naar* functions as a grammatical *uh*, buying time for the producer to formulate a complex object
- Addressee-driven Hypothesis: *naar* functions as a grammatical signpost for the addressee. It marks 'what follows now, is the object of the verb'

Remove the observations where the object is extraposed to the postfield

Remove: *Het stadsbestuur* heeft daarom *gezocht naar een efficiëntere en goedkopere oplossing*

Keep: *Nijmegen zoekt naar een oplossing*

**Prediction:** as the object becomes more complex, the probability of *naar* will **still** increase
Whose processing are we talking about?

- 61998 without $naar \leftrightarrow 10949$ with $naar$

- As the object becomes more complex, the probability of $naar$ still increases (be it less so, positive estimate for Object Length: 0.25)

- Highly significant: < 0.0001
Whose processing are we talking about?

- **Producer-driven Hypothesis 1**: *naar* allows the producer to extrapose long objects to the postfield.

- **Producer-driven Hypothesis 2**: *naar* functions as a grammatical *uh*, buying time for the producer to formulate a complex object.

- **Addressee-driven Hypothesis**: *naar* functions as a grammatical signpost for the addressee. It marks 'what follows now, is the object of the verb'.

**Prediction confirmed**: as the object becomes more complex, the probability of *naar* will still increase.
Whose processing are we talking about?

• Producer-driven Hypothesis 2.1: *naar* buys time to formulate a complex object. However, if it limits the producer’s future choice of verb, he/she’d rather not express it.

• Producer-driven Hypothesis 2.2: *naar* buys time to formulate a complex object. Even if it limits the producer’s future choice of verb, that’s a price he/she is willing to pay.

• Addressee-driven Hypothesis: *naar* functions as a grammatical signpost for the addressee. It marks ‘what follows now, is the object of the verb’.

Remove the observations where the verb proceeds the object

Remove:  *Nijmegen zoekt naar een oplossing*

Keep:  *Naar politiek als roeping, of zelfs maar als ethos, zoekt de lezer tevergeefs*

Prediction: as the object becomes more complex, the probability of *naar* will no longer increase, or even decrease
Whose processing are we talking about?

• Producer-driven Hypothesis 2.1: *naar* buys time to formulate a complex object. However, if it limits the producer’s future choice of verb, he/she’d rather not express it.

• Producer-driven Hypothesis 2.2: *naar* buys time to formulate a complex object. Even if it limits the producer’s future choice of verb, that’s a price he/she is willing to pay.

• Addressee-driven Hypothesis: *naar* functions as a grammatical signpost for the addressee. It marks ‘what follows now, is the object of the verb’.

Remove the observations where the verb precedes the object

Remove: *Nijmegen zoekt naar een oplossing*

Keep: *Naar politiek als roeping, of zelfs maar als ethos, zoekt de lezer tevergeefs*

**Prediction:** as the object becomes more complex, the probability of *naar* will still increase
Whose processing are we talking about?

- Producer-driven Hypothesis 2.1: *naar* buys time to formulate a complex object. However, if it limits the producer’s future choice of verb, he/she’d rather not express it.

- Producer-driven Hypothesis 2.2: *naar* buys time to formulate a complex object. Even if it limits the producer’s future choice of verb, that’s a price he/she is willing to pay.

- Addressee-driven Hypothesis: *naar* functions as a grammatical signpost for the addressee. It marks ‘what follows now, is the object of the verb’.

Remove the observations where the verb precedes the object

Remove:  *Nijmegen zoekt naar een oplossing*

Keep:  *Naar politiek als roeping, of zelfs maar als ethos, zoekt de lezer tevergeefs*

**Prediction:** as the object becomes more complex, the probability of *naar* will **still** increase, perhaps even more so
Whose processing are we talking about?

- 35089 without *naar* ↔ 4288 with *naar*

- As the object becomes more complex, the probability of *naar* decreases
  (negative estimate for Object Length: -0.13)

- Highly significant: < 0.0001

Prediction confirmed: as the object becomes more complex, the probability of *naar* decreases
Processing shapes grammar

But whose processing are we talking about?

The producer's

This dovetails with findings in psycholinguistic experiments, e.g. Ferreira & Dell’s *that*-omission study (2000), and references cited therein.
Thanks!

Dirk Pijpops & Freek Van de Velde
References


