Constructional contamination effects
Evidence from mixed-effects logistic regression modeling of the Dutch partitive genitive

Dirk Pijpops¹ & Freek Van de Velde²

SLE 47 Workshop 'From methodology back to theory: how do current empirical methods feed back into linguistic theory?' Poznan, 12 September 2014

¹University of Leuven
Master in AI: Speech & Language Technology

²University of Leuven
RU Quantitative Lexicology and Variational Linguistics
Introduction

- Constructions are often defined as form-function pairings
- Under a naive view of how signs work, this pairing should be as fixed and predictable as possible, lest the semiotic link be jeopardised. If meaning $A$ corresponds to form $X$, $Y$ and $Z$, and form $X$ corresponds to meaning $A$, $B$ and $C$ (many-to-many mapping, instead of Humboldtian isomorphism), then language users are at a loss in communication
- entailing that constructions are uncontaminated by neighbouring constructions.
- This is, however, not always the case:
- Diachronically, a construction often derives from multiple lineages (Van de Velde, De Smet & Ghesquière 2013 on 'multiple source constructions')
- Synchronically, a construction often displays contamination effects at its fringes (Pijpops & Van de Velde 2014)

Case study: Dutch partitive genitive

- Like other West-Germanic languages, Dutch has undergone deflection (Van der Horst 2008:143)
- Especially in the nominal domain (Harbert 2007:90)
- Also targeting the genitive: see graph (From: Weerman & de Wit 1999:1158)
- One remarkable resilient cx:
  Partitive genitive

Case study: Dutch partitive genitive

• Dutch partitive genitive

\[ \text{iets \hspace{1em} interessant-s} \]
something \hspace{1em} interesting-GEN
‘something interesting’

\[
\left[ \text{NP Q}_i \text{ Adj-j-s} \right] \leftrightarrow \left[ \text{modifier}_j \text{ head-quantity}_i \right]
\]

• Variation: The \( s \) can be expressed, or not: \( \text{iets interessant}(s) \)
Alternation factors: Methodology

• Corpus: CONDIV (Grondelaers et al. 2000 for details)

• 3018 partitive genitives after manual checking

• Binary response variable: [+s] / [-s]

• Mixed models logistic regression (Baayen 2008, Gries 2013, Speelman, forthc.)

• Stepwise variable selection procedure

Explanatory variables

• Lectal variables
  – Variety: \textit{Netherlands, Flanders}
  – Register: \textit{chat, e-mail, mass-newspaper, quality-newspaper}

• Structural variables
  – Quantifier: \textit{iets (‘something’), niets (‘nothing’), veel (‘a lot’), wat (‘something’), weinig (‘little’), zoveel (‘so much’)}
  – Length-Adjective: number of syllables
  – Type-Adjective: \textit{other, deviant (verkeerd, goed, fout, beter), colour (blauw, rood, groen)}
  – Number-of-words-AP: \textit{iets erg leuk (‘something very fun’) vs. iets leuk (‘something fun’)}
  – Token frequency of different phrase types

• Random effect \textit{Phrase Type}: unique combination of quantifier and adjectival phrase
- AIC: 2216
- C-value: 0.872
- Total number of hits: 3018
- Hits with -s: 2388
- Hits without -s: 630

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Levels of categorical predictors</th>
<th>Estimates 2.5%</th>
<th>Estimates 97.5%</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>intercept</td>
<td>0.07</td>
<td>-0.67</td>
<td>0.82</td>
</tr>
<tr>
<td>Type-Adjective</td>
<td>other</td>
<td>Reference level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deviant</td>
<td>1.96</td>
<td>1.45</td>
<td>2.46</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>colour</td>
<td>5.09</td>
<td>3.88</td>
<td>6.30</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Variety</td>
<td>Flanders</td>
<td>Reference level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>-1.69</td>
<td>-2.01</td>
<td>-1.37</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Register</td>
<td>chat</td>
<td>Reference level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-mail</td>
<td>-0.48</td>
<td>-0.77</td>
<td>-0.19</td>
<td>0.0013</td>
</tr>
<tr>
<td>mass-newspaper</td>
<td>-1.08</td>
<td>-1.42</td>
<td>-0.74</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>quality-newspaper</td>
<td>-1.65</td>
<td>-2.22</td>
<td>-1.08</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Quantifier</td>
<td>iets</td>
<td>Reference level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>niets</td>
<td>-0.05</td>
<td>-0.66</td>
<td>0.56</td>
<td>0.8809</td>
</tr>
<tr>
<td>veel</td>
<td>-1.14</td>
<td>-1.98</td>
<td>-0.29</td>
<td>0.0083</td>
</tr>
<tr>
<td>wat</td>
<td>-2.00</td>
<td>-2.99</td>
<td>-1.00</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>weinig</td>
<td>-2.50</td>
<td>-4.12</td>
<td>-0.89</td>
<td>0.0023</td>
</tr>
<tr>
<td>zoveel</td>
<td>-2.35</td>
<td>-4.37</td>
<td>-0.34</td>
<td>0.0221</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction Variety -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td>Flanders &amp; iets</td>
<td>Reference level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands - niets</td>
<td>-0.33</td>
<td>-1.03</td>
<td>0.38</td>
<td>0.3635</td>
</tr>
<tr>
<td>Netherlands - veel</td>
<td>0.98</td>
<td>0.02</td>
<td>1.94</td>
<td>0.0443</td>
</tr>
<tr>
<td>Netherlands - wat</td>
<td>1.22</td>
<td>0.19</td>
<td>2.25</td>
<td>0.0208</td>
</tr>
<tr>
<td>Netherlands - weinig</td>
<td>2.33</td>
<td>0.66</td>
<td>4.00</td>
<td>0.0062</td>
</tr>
<tr>
<td>Netherlands - zoveel</td>
<td>2.10</td>
<td>-0.94</td>
<td>5.13</td>
<td>0.1755</td>
</tr>
</tbody>
</table>
What is going on here?
Structural contamination effect: colour adjectives

\[ veel \text{ geel} \]
‘a lot of yellow (things)’

\[ geel_{\text{Adj}} \text{ or } geel_{\text{Noun}} \]

partitive genitive
\simveal\text{ interessant}
‘a lot of interesting things’

[-s] or [+s]

modifier – noun
\simveal\text{ water}
‘a lot of water’

always [-s]

\Rightarrow \text{ Bias towards } [-s]
Structural contamination effect: colour adjectives

\( iets \ \text{geel} \)

‘something yellow’

\( \text{geel}_{\text{Adj}} \)

partitive genitive

\( \sim iets \ \text{interessant} \)

‘something interesting’

[-s] or [+s]

modifier – noun

\( \sim iets \ \text{water} \)

‘something water’

always [-s]

\[ \Rightarrow \text{still bias towards [-s] due to superficial resemblance to veel \ \text{geel}} \]
Colour adjectives: unambiguous cases (Q = 'iets')

iets + adj. (diff. in abs. numb. not visualized)

other

without -s

with -s

colour

p-value = 0.01122 (Fisher’s exact test)
Structural contamination effect: deviant adjectives

deviant adjectives:

- *verkeerd* ‘wrong’
- *goed* ‘good’
- *beter* ‘better’
- *fout* ‘incorrect’
Structural contamination effect: deviant adjectives

*Of heb ik hier *iets verkeerd* verstaan...*
or have I here *something wrong(ly)* understood

Partitive genitive
‘or did I understand something wrong?’

[iets verkeerd][verstaan]
[something wrong][understand]

[+s] or [-s]

Adverbial construction
‘or did I misunderstand something?’

[iets][verkeerd verstaan]
[something][wrongly understand]

always [-s]
Structural contamination effect: deviant adjectives

*Heb ik iets verkeerd gedaan?*

have I *something wrong(ly)* done

Partitive genitive

‘Did I do something wrong?’

[iets verkeerd][doen]

[something wrong][do]

[+s] or [-s]

Adverbial construction

‘Did I do something the wrong way?’

[iets][verkeerd doen]

[something][wrong-do]

always [-s]

⇒ Bias towards [-s]
Structural contamination effect: deviant adjectives

Als ik *iets verkeerd* gegeten heb, heb ik buikpijn.
If I *something wrong* eaten have, have I stomach-ache

Partitive genitive
‘If I have eaten something wrong,…’

[iets verkeerd][eten]
[something wrong][eat]
[-s] or [+s]

Adverbial construction
‘If I have eaten something the wrong way,…’

[iets][verkeerd eten]
[something][wrong-eat]
always [-s]
Structural contamination effect: deviant adjectives

Als ik *iets verkeerd* gegeten heb, heb ik buikpijn.
If I *something wrong* eaten have, have I stomach-ache

Partitive genitive
‘If I have eaten something wrong,…’

[iets verkeerd][eten]
[something wrong][eat]

[-s] or [+s]

Adverbial construction
‘If I have eaten something the wrong way,…’

[iets][verkeerd eten]
[something][wrong-eat]

always [-s]

⇒ No bias towards [-s] preference?
What is going on here?
Mosaic plot: distribution of the variants over the verbs combined with the adjective *verkeerd* ('wrong')

=> Data still show preference for [-s], even where there’s no syntactic ambiguity!
Contamination effects

Direct cause:

*iets verkeerd (verstaan)* often appears without –s

\[ \downarrow \]

Indirect effect on superficially similar or identical occurrences:

*iets verkeerd (eten)*

Preference for [-s]
Lectal contamination

Direct cause: Variety

typically Netherlandic         typically Flemish

*wat mooi-s*                        *iets interessant*

‘something beautiful’            ‘something interesting’

more often appear [+s]            more often appear [-s]

Indirect effect:

*wat mooi-s*                        *iets interessant*

preference for [+s]               preference for [-s]
Operationalisation

140 phrase types

typically Netherlandic

- iets bijzonder(s)
- wat zinnig(s)
- wat mooi(s)
- iets leuk(s)

neutral

- weinig concreet(s)
- iets zinnig(s)
- iets spannend(s)
- niets erg(s)

... neutral ...

typically Flemish

- iets speciaal(s)
- iets interessant(s)
- niets concreet(s)
- iets deftig(s)

... typically Flemish ...
Lectal contamination

Direct cause: Variety

typically Netherlandic  typically Flemish

*wat mooi-s*  *iets interessant*

‘something beautiful’  ‘something interesting’

more often appear [+s]  more often appear [-s]

\[+s\]

\[-s\]

Indirect effect:

*wat mooi-s*  *iets interessant*

preference for [+s]  preference for [-s]
Lectal contamination

The Netherlands

Mosaic plot: distribution of the variants over the typically Netherlandic, neutral and typically Flemish phrases in only the Netherlandic material
(Kendall’s $\tau = -0.2146$, p-value < 0.0001)

Flanders

Mosaic plot: distribution of the variants over the typically Netherlandic, neutral and typically Flemish phrases in only the Flemish material
(Kendall’s $\tau = -0.1943$, p-value < 0.0001)
Conclusions

• Constructions are not discretely stored, but entertain links to each other
• These links come in various sorts:
  1. Vertical links between related constructions: inheritance hierarchies, where more abstract, higher-order constructions 'sanction' or 'license' lower-order constructions
  2. Horizontal links between related constructions: related constructions in a functional domain are mutually defined by differential values they take on a set of grammatical parameters (see Van de Velde 2014)
  3. Relations between unrelated constructions: superficial similarities between constructions yield contamination effects.
• This supports an 'exemplar-based' view on language (Bybee 2010): Prior use of constructions leaves a (context-rich) trail in the mind of the language users

• We need a usage-based perspective (Kemmer & Barlow 2000; Bybee 2006, 2010; Bybee & Beckner 2010; Von Mengden & Coussé 2014), recognising:
  – ‘Emergent’ nature of grammar (Hopper 1987, 1998)
  – Importance of variation, including variation along sociolinguistic axes (Geeraerts & Kristiansen, forthc.)
  – The importance of frequency in routinisation or ‘entrenchment’ of linguistic patterns
  – Emphasis on empirical data, e.g. from corpus inquiry (Tummers et al. 2005; Geeraerts 2006; Gries & Stefanowitsch 2006)

Bybee, Joan. 2006. 'From usage to grammar: the mind’s response to repetition'. Language 82(4): 711-733.
Thanks!

• Dirk Pijpops: pijnpsd@outlook.com
• Freek Van de Velde: freek.vandevelde@arts.kuleuven.be