Constructional contamination in language

Freek Van de Velde & Dirk Pijpops
Göteborg, 9 May 2014

University of Leuven
RU Quantitative Lexicology and Variational Linguistics
Introduction

• Constructions are often defined as form-function pairings
• The underlying assumption is that the form of a construction is fixed, and uncontaminated. Otherwise, the indispensable link between form and function would be jeopardised.
• This is, however, not always the case
• Diachronically, a construction often derives from multiple lineages
• Synchronically, a construction often displays contamination effects at its fringes
Contamination of lineages

- Multiple Source Constructions (MSC)
- Van de Velde et al. (2013)
  
  "Innovations in language change may derive not just from one, but from different source constructions at once. That is, change often seems to involve some interaction between lineages or between different branches of a lineage"

- MSC in:
  - Phonology
  - lexical semantics
  - morphology
  - syntax

Contamination of lineages

- MSC in phonology: merger (IE ō and ā > Gmc. ō) plus subsequent changes:
  - PIE *pōd-s > PGmc. *fōt- > PdDu voet /u/
  - PIE *sueh₂d- > PGmc. swōti/u- > PdDu zoet /u/

\[ \begin{align*}
  &\text{o} \\
  \quad \quad \quad \quad \quad \rightarrow \quad \rightarrow \\
  &\text{ā} \\
\end{align*} \]
Contamination of lineages

- MSC in lexical semantics:
  - contaminations (*irregardless*)
  - folk etymology (*sparrow-grass*)
  - Diachronic polysemy through merger of distinct meanings (Geeraerts 1997)
  - ...

Contamination of lineages

- MSC in morphology:
  - suppletive verb forms: *(nous) allons* (< ambulare), *(je) vais* (< vadere), *(j') irai* (< ire*)
  - 'Constructional' suppletion (Van de Velde et al. 2013: 480):

![Image of verb classes and templates]

\[
\begin{align*}
\text{PIE} & : \text{present} & \text{o-grade} & \text{zero-grade} \\
\text{Cl.Gr.} & : \text{leíp-ô} & \text{Cl.Gr.} & : \text{lé-loip-a} & \text{Cl.Gr.} & : \text{é-lip-on} \\
\text{Germanic} & : \text{present} & \text{preterit-sg} & \text{preterite-pl} \\
\text{Go.} & : \text{bi-leib-an} & \text{Go.} & : \text{bi-laif} & \text{Old.D.} & : \text{be-lib-en}
\end{align*}
\]

Figure 3. Proto-Indo-European perfect/aorist > Germanic preterite*

*PIE: Proto-Indo-European; Cl.Gr.: Classical Greek; Go.: Gothic; Old.D.: Old Dutch

PIE perfect (o-grade) → Germanic preterite

PIE aorist (zero-grade) → Germanic preterite

Templates for verb classes, not (necessarily) concrete verbs
Contamination of lineages

• MSC in syntax:
  – *way-construction* (Traugott & Trousdale 2013):
    • and we were actually kicking our way through rubbish on the stairs (BNC)
    • a lady who giggled her way through Nightmare on Elm Street (BNC)

1. 'way' as the object of a transitive verb denoting creation or acquisition of a path
   a) *be next Marche folowand He suld take þat way* (1338, OED)
      'The following month of March he should take that way.'
   b) *The ship..may make her way 2. or 3. pointes from her caping* [i.e. 'course'].
      (1595, OED)

2. 'way' functioning as an adverbial with intransitive motion verbs:
   a) *Whoso myghte by þe grace of Godd go þis way he sulde noghte erre.* (1340, OED)
      'Whoever can by the grace of God go this way will not go astray.'
   b) *Sir Beawmaynes..sawe where the blak knyght rode his way wyth the dwarff, and so he rode oute of his syght.* (a1470, OED)
CASE STUDY: DUTCH PARTITIVE GENITIVE

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: (From: Weerman & de Wit 1999:1158)
- One remarkable resilient cx: Partitive genitive

Case study: Dutch partitive genitive

• Dutch partitive genitive

\[
\begin{align*}
\text{iets} & \quad \text{interesting-s} \\
\text{something} & \quad \text{interesting-GEN} \\
\text{‘something interesting’} & \\
\end{align*}
\]

\[
[N_{i} Q_{i} \text{ Adj-s }] \leftrightarrow [\text{modifier}_j \text{ head-quantity}_i]
\]

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

- Corpus: CONDIV (Grondelaers et al. 2000 for details)
- Lectally stratified (regional variety and register)
- 3018 partitive genitives after manual checking
- Binary response variable: [+s] / [-s]
- Effect of both structural and lectal variables
- Mixed models logistic regression (Baayen 2008, Gries 2013, Speelman, forthc.)
- Stepwise variable selection procedure

Explanatory variables

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

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

\[
\begin{align*}
\text{iets leuk(s)} & \quad \text{veel interessant(s)} \\
\text{‘something fun’} & \quad \text{‘a lot of interesting things’} \\
\text{[+s]} & \quad \text{[+s]} \\
\text{iets leuks} & \quad \text{veel interessants} \\
\text{[s]} & \quad \text{[-s]} \\
\text{iets leuk} & \quad \text{veel interessant} \\
\end{align*}
\]

\[
\begin{align*}
\text{iets erg gemakkelijk(s)} & \\
\text{‘something very easy’} & \\
\text{[+s]} & \\
\text{iets erg gemakkelijks} & \\
\text{[-s]} & \\
\text{iets erg gemakkelijk} & \\
\end{align*}
\]
### Model 2: with Phrase as a Random Factor

- **AIC:** 2216
- **C-value:** 0.872
- **Number of phrases:** 140
- **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</th>
<th><strong>Confidence intervals</strong></th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td></td>
<td>0.07</td>
<td>-0.67 - 0.82</td>
<td>0.8482</td>
</tr>
<tr>
<td><strong>Type-Adjective</strong></td>
<td>other (Reference level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>deviant</td>
<td>1.96</td>
<td>1.45 - 2.46</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>colour</td>
<td>5.09</td>
<td>3.88 - 6.30</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td><strong>Variety</strong></td>
<td>Flanders (Reference level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>-1.69</td>
<td>-2.01 - -1.37</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td><strong>Register</strong></td>
<td>chat (Reference level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e-mail</td>
<td>-0.48</td>
<td>-0.77 - -0.19</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>mass-newspaper</td>
<td>-1.08</td>
<td>-1.42 - -0.74</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>quality-newspaper</td>
<td>-1.65</td>
<td>-2.22 - -1.08</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td><strong>Quantifier</strong></td>
<td>iets (Reference level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>niets</td>
<td>-0.05</td>
<td>-0.66 - 0.56</td>
<td>0.8809</td>
</tr>
<tr>
<td></td>
<td>veel</td>
<td>-1.14</td>
<td>-1.98 - -0.29</td>
<td>0.0083</td>
</tr>
<tr>
<td></td>
<td>wat</td>
<td>-2.00</td>
<td>-2.99 - -1.00</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>weinig</td>
<td>-2.50</td>
<td>-4.12 - -0.89</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>zoveel</td>
<td>-2.35</td>
<td>-4.37 - -0.34</td>
<td>0.0221</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td></td>
<td>-0.45</td>
<td>-0.79 - -0.10</td>
<td>0.0109</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>Flanders &amp; iets (Reference level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety - Quantifier</td>
<td>Netherlands - niets</td>
<td>-0.33</td>
<td>-1.03 - 0.38</td>
<td>0.3635</td>
</tr>
<tr>
<td></td>
<td>Netherlands - veel</td>
<td>0.98</td>
<td>0.02 - 1.94</td>
<td>0.0443</td>
</tr>
<tr>
<td></td>
<td>Netherlands - wat</td>
<td>1.22</td>
<td>0.19 - 2.25</td>
<td>0.0208</td>
</tr>
<tr>
<td></td>
<td>Netherlands - weinig</td>
<td>2.33</td>
<td>0.66 - 4.00</td>
<td>0.0062</td>
</tr>
<tr>
<td></td>
<td>Netherlands - zoveel</td>
<td>2.10</td>
<td>-0.94 - 5.13</td>
<td>0.1755</td>
</tr>
</tbody>
</table>
What is going on here?
Structural contamination effect: colour adjectives

veel geel
‘a lot of yellow (things)’

geel_{Adj} or geel_{Noun}

partitive genitive
\sim veel interessant
‘a lot of interesting things’

modifier – noun
\sim veel water
‘a lot of water’

[-s] or [+s]
always [-s]

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

\[ iets \text{ geel} \]
\[ \text{‘something yellow’} \]
\[ \text{geel}_{\text{Adj}} \text{ of geel}_{\text{Noun}} \]

partitive genitive
\[ \sim iets \text{ interessant} \]
\[ \text{‘something interesting’} \]

modifier \rightarrow \text{noun}
\[ \sim iets \text{ water} \]
\[ \text{‘something water’} \]

\([-s]\) or \([+s]\)

always \([-s]\)

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

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

other

with -s

without -s

colour
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,...’
  \[\text{something wrong} \text{ eat} \]
  [-s] or [+s]

- **adverbial construction**
  ‘If I have eaten something the wrong way,...’
  \[\text{something} \text{ wrong-eat} \]
  always [-s]

⇒ No bias towards [-s] preference?
Overall distribution of the two forms:

Mosaic plot: distribution of the variants over the verbs combined with the adjective *verkeerd* (‘wrong’)

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

Adverbial construction

Heb ik iets verkeerd verstaan?
‘Did I misunderstand something?’

Partitive genitive construction

als ik iets verkeerd gegeten heb,…
‘If I have eaten something wrong,…’

Always [-s]

preference for [-s]

Superficial resemblance
Contamination effect

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(s)*

- ‘something beautiful’
- ‘something interesting’

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

Indirect effect:

- *wat mooi(s)*
- *iets interessant(s)*

- 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)*, ...
- typically Flemish: *iets speciaal(s)*, *iets interessant(s)*, *niets concreet(s)*, *iets deftig(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

Conclusions (continued)

• 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)
Conclusions (continued)

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

• The audience
• My collaborators:
  – Hendrik De Smet
  – Lobke Ghesquière
  – Dirk Pijpops