Constructional contamination
An occasional rarity or a pervasive effect?

Dirk Pijpops, Isabeau De Smet & Freek Van de Velde

Research Foundation Flanders
QLVL, University of Leuven
What is constructional contamination?

Is it real?

If so, is it an occasional rarity or a pervasive effect?
Constructional contamination

• Mechanism based on shallow parsing & storage of ready-mades

• Lexical preferences resulting from that mechanism
TARGET CONSTRUCTION

- + ke
- + pa

CONTAMINATING CONSTRUCTION

- lolipapa
- "lollike" (99x)
- "lolipapa" (1x)
- tepoke (99x)
- "tepope" (1x)
- lazike (99x)
- "lazipa" (1x)
- ...
- ...
99x "lolike" 
99x "tepoke" 
99x "lazike" 
1x "lolip" 
1x "tepopa" 
1x "lazipa"

TARGET CONSTRUCTION
"lolike" > "lolip"
"tepoke" > "tepopa"
Is it real?

Case study 1: partitive genitive
TARGET: PARTITIVE GENITIVE

+ s  + ∅

something wrong

something fun

I had wrongly interpreted something

Ik had iets verkeerd geïnterpreteerd
Case study 1: partitive genitive

- Prediction: among the partitive genitives, the variant without -s will be much more dominant with adjectives that often appear as adverbs resembling partitive genitives without -s, viz. verkeerd 'wrong', goed 'good', beter 'better' and fout 'incorrect'
• Only look at strictly unambiguous partitive genitives
• Mixed-effects regression model
• Control for all factors known to influence alternation and random lexical preferences
Type of adjective

- Variety
- Quantifier
- Frequency
- Register

Random Factor: Phrase

Standard variable importance
(Strobl et al. 2008)
So is it an occasional rarity or a pervasive effect?

Case study 2: verbal clusters
Case study 2: verbal clusters

De deur moet door John gesloten zijn.
The door must by John closed be

... dat de deur door John gesloten is.
... that the door by John closed is.
• **PREDICTION 1**: The more often a participle is used as an adjective, the more often it will appear in the PARTICIPLE + AUXILIARY order in unambiguous verbal contexts.

• **PREDICTION 2**: This effect will be stronger among the auxiliaries that can be used as copula, viz. *zijn* 'be' and *worden* 'become', and weaker among other auxiliaries, such as *hebben* 'have'.
TARGET: PARTICIPLE + AUXILIARY

AUXILIARY + PARTICIPLE Order

PARTICIPLE + AUXILIARY Order

CONTAMINATING: ADJECTIVE + COPULA

1ST DEGREE CONTAMINATION: COMPLETE STRING OVERLAP

2ND DEGREE CONTAMINATION

... dat de deur door John is gesloten is closed

... dat de deur door John gesloten is closed is

... dat de deur al geruime tijd gesloten is closed is

... dat John de deur heeft gesloten has closed

... dat John de deur gesloten heeft closed has
Case study 2: verbal clusters

• **Prediction 1:** The more often a participle is used as an adjective, the more often it will appear in the PARTICIPLE + AUXILIARY order in unambiguous verbal contexts.

• **Prediction 2:** This effect will be stronger among the auxiliaries that can be used as copula, viz. *zijn* 'be' and *worden* 'become', and weaker among other auxiliaries, such as *hebben* 'have'.
Case study 2: verbal clusters

• Dataset from Gert De Sutter

• De Sutter distinguished between ambiguous & unambiguous verbal clusters

• Only looked at unambiguous verbal clusters

• Added variable *Adjectiveness* = \( arsin\left(\sqrt{\frac{\text{adjectival occurrences}}{\text{total occurrences}}}\right) \)
• Prediction 1: *Adjectiveness* will correlate positively with preference for the PARTICIPLE + AUXILIARY order

• Prediction 2: This effect will be stronger for auxiliaries *zijn* 'be' and *worden* 'become' than for *hebben* 'have'

![Graph showing the correlation between Adjectiveness and preference for participle + auxiliary order for *zijn* 'be' and *worden* 'become' compared to *hebben* 'have'.]
So is it an occasional rarity or a pervasive effect?

Case study 3: weak vs. strong preterites
Case study 3: weak vs. strong preterites

- Germanic languages: two morphological strategies to form preterite
  - strong inflection
    - vowel change (‘ablaut’)
    - *zwem-zwom* (‘swim’ – ‘swam’)
  - weak inflection
    - dental suffix
    - *speel-speelde* (‘play’ – ‘played’)

Case study 3: weak vs. strong preterites

- Contaminating construction: clitic realization of the 2nd person singular subject pronoun (cfr. Vosters 2012)

  Vandaag graaf-de een put. (Vosters 2012: 242)

  Today dig-2SG.PRS a hole

  ‘You will dig a hole today.’
Groef 'digged'

Graafde 'digged'

Vandaag graaf-de een put.
dig-2SG.PRS

TARGET: PRETERITE

CONTAMINATING: CLITIC 2ND SING
Case study 3: weak vs. strong preterites

• Two predictions:
  
  – (i) Weak preterites will be more prevalent in the regions known for their enclitic realization of the subject pronoun, compared to the other Dutch-speaking regions of the Low Countries.
  
  – (ii) Verbs that are more often realized with an enclitic subject tend to weaken more than verbs that are less often realized with an enclitic subject.
Prediction I: more weak forms in Antwerp, Flemish-Brabant and East-Flanders compared to the other Dutch speaking regions
Prediction I: more weak forms in Antwerp, Flemish-Brabant and East-Flanders compared to the other Dutch speaking regions (p=0.031)
Prediction II: more weak forms for verbs that are more likely to appear with clitic

\[ \text{graaf-de} \]
\[ \text{dig-2SG.PRS} \]
\[ 'Do you dig?' \]

\[ \text{?slinkt-te} \]
\[ \text{lessen-2SG.PRS} \]
\[ 'Do you lessen?' \]
Prediction II: more weak forms for verbs that are more likely to appear with enclitic (p>0.05)

graaf-de
dig-2SG.PRS
‘Do you dig?’

?slinkt-te
lessen-2SG.PRS
‘Do you lessen?’
So is it an occasional rarity or a pervasive effect?

Case study 4: long vs. bare infinitives
Case study 4: long vs. bare infinitives

- Auxiliaries can be classified according to the type of complement they take:
  - participle
  - infinitival complement
    - bare infinitive: *Dat moet Ø/*te werken. (‘That must Ø work.’)
    - long infinitive (or: to-infinitive): *Dat lijkt *Ø/te werken. (‘That seems to work.’)
Case study 4: long vs. bare infinitives

• Posture verbs (zitten ‘sit’, staan ‘stand’, liggen ‘lie’)
  – finite auxiliary takes long infinitive: *Hij zit te/*Ø slapen. (‘He is sleeping’.)
  – infinite auxiliary
    ▪ Infinitivus Pro Participo (IPP or ‘Ersatzinfinitiv’)
    ▪ when used in the perfect, auxiliaries may occur in the infinitive instead of the past participle
      ▪ *Hij heeft de hele les zitten Ø slapen. (‘He has been sleeping throughout the entire class.’)
Case study 4: long vs. bare infinitives

- Posture verbs (zitten ‘sit’, staan ‘stand’, liggen ‘lie’)
  - finite auxiliary takes long infinitive: Hij zit te/*Ø slapen. (‘He is sleeping’.)
    - Exception: if the auxiliary is present simple plural in a subordinate clause, bare infinitive is possible too (Haeseryn et al. 1997: 970; Klooster 2001: 61)
    - Als die jongens de hele les zitten Ø slapen, zullen ze niet veel opsteken. (‘If those boys are sleeping through the entire class, then they won’t learn much’) (Haeseryn et al. 1997: 970)
  - infinite auxiliary
    - Infinitivus Pro Participio (IPP or ‘Ersatzinfinitiv’)
    - when used in the perfect, auxiliaries may occur in the infinitive instead of the past participle
    - Hij heeft de hele les zitten Ø slapen. (‘He has been sleeping throughout the entire class.’)
Als die jongens de hele les...
...zitten te slapen...

Hij heeft de hele les zitten slapen.

1ST DEGREE CONTAMINATION

2ND DEGREE CONTAMINATION
...zaten te slapen...
...zaten slapen...
Prediction: Group I is strongly affected by constructional contamination, group II less so and group III even less so, or not at all.

Group (i): superficial formal identity (1st degree contamination)
e.g. *Als die jongens de hele les zitten Ø slapen, zullen ze niet veel opsteken.*
(‘If those boys are sleeping throughout the entire class, then they won’t learn much’)

Group (ii): superficial formal resemblance (2nd degree contamination)
e.g. *Als die jongens de hele les zaten Ø slapen, hebben ze niet veel opgestoken.*
(‘If those boys were sleeping throughout the entire class, they haven’t learned much.’)

Group (iii): no resemblance
e.g. *De jongen zit al heel de les (te) slapen.*
(‘The boy has been sleeping the entire class’)

*Note: Zittie and zaten are incorrect words in Dutch.*
Prediction: Group I is strongly affected by constructional contamination, group II less so and group III even less so, or not at all.

Out of 2766 bare infinitives...

Group (i): superficial formal identity (1st degree contamination)
7 instances (↔ 2622 long infinitives)

Group (ii): superficial formal resemblance (2nd degree contamination)
3 instances (↔ 11978 long infinitives)

Group (iii): no resemblance
1 instance (↔ 13576 long infinitives)
Conclusions

• Constructional contamination is a pervasive effect

• It follows naturally from a usage-based view on language processing, in particular shallow parsing and ready-mades

• If we can so easily find four case studies in a single language, you should be able to find many more in other languages
Special thanks to

- **Gert De Sutter**, for generously sharing dataset of verbal clusters
- **Tom Ruette**, for giving us access to his Twitter-corpus
References


Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth & Steve Taylor. 2016. Effect Displays for Linear, Generalized Linear, and Other Models. _R_ package version 3.2.


