



# Constructional contamination effects

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

Dirk Pijpops<sup>1</sup> & Freek Van de Velde<sup>2</sup>

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



<sup>1</sup>University of Leuven

Master in AI: Speech & Language Technology



<sup>2</sup>University 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)

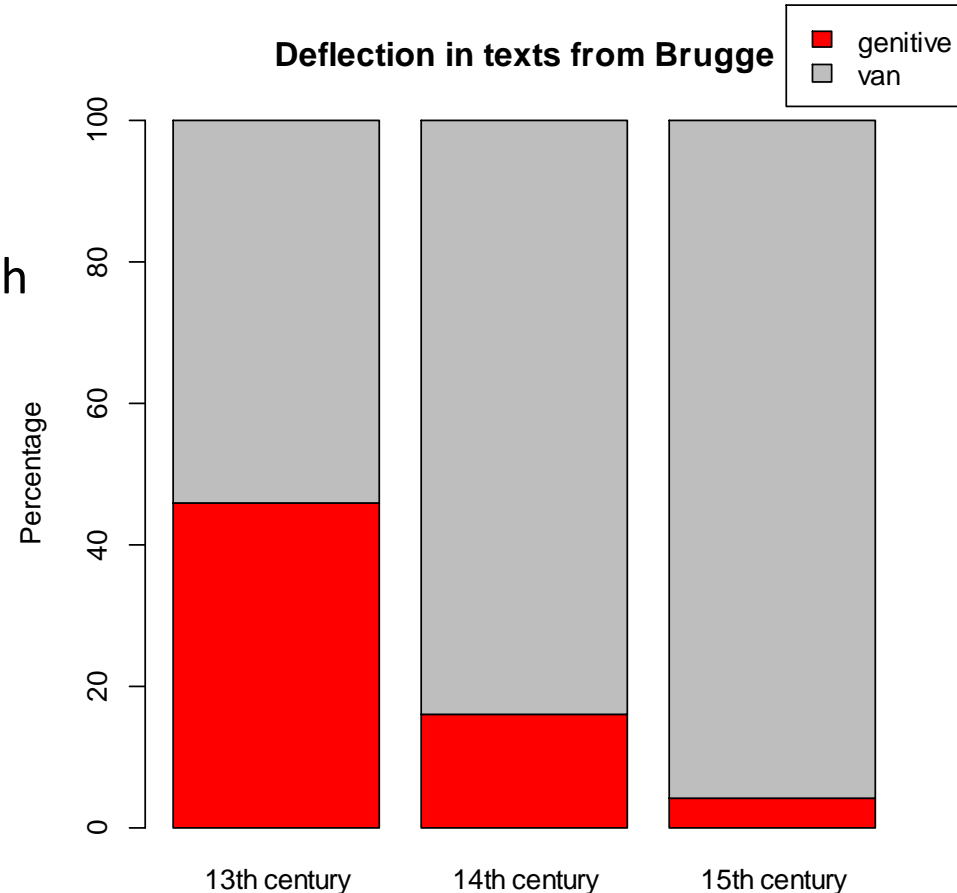
Pijpops, Dirk & Freek Van de Velde. 2014. 'A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion'. *Corpus Linguistics and Linguistic Theory* (DOI: 10.1515/cllt-2013-0027).

Van de Velde, Freek, Hendrik De Smet & Lobke Ghesquière. 2013. 'On multiple source constructions in language change'. *Studies in Language* 37(3): 473-489

# 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

Deflection in texts from Brugge



Harbert, W. 2007. *The Germanic languages*. Cambridge: Cambridge University Press.

Van der Horst, J.M. 2008. *Geschiedenis van de Nederlandse syntaxis*. Leuven: Leuven University Press.

Weerman, F. & P. de Wit. 1999. 'The decline of the genitive in Dutch'. *Linguistics* 37: 1155-1192.

# Case study: Dutch partitive genitive

- Dutch partitive genitive

*iets*                      *interessant-s*  
something                interesting-GEN  
'something interesting'

[<sub>NP</sub> Q<sub>i</sub> Adj<sub>j</sub>-s ]                ↔                [modifier<sub>j</sub> head-quantity<sub>i</sub>]

- Variation: The *s* can be expressed, or not: *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

Baayen, Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Gries, Stefan Th. 2013. *Statistics for linguistics with R. A practical introduction*. 2nd rev. edn. Berlin: de Gruyter.

Grondelaers, Stefan, Katrien Deygers, Hilde van Aken, Vicky Van Den Heede & Dirk Speelman. 2000. 'Het CONDIV-corpus geschreven Nederlands' [The Condiv corpus of spoken Dutch]. *Nederlandse Taalkunde* 5(4). 356-363.

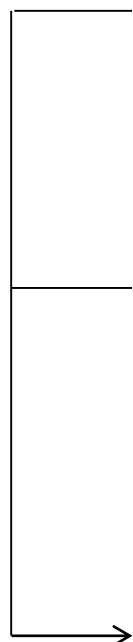
Speelman, Dirk. Forthcoming. 'Logistic regression in corpus linguistics'. In: Dylan Glynn & Justyna A. Robinson (eds.), *Polysemy and synonymy*. Amsterdam: John Benjamins.

# Explanatory variables

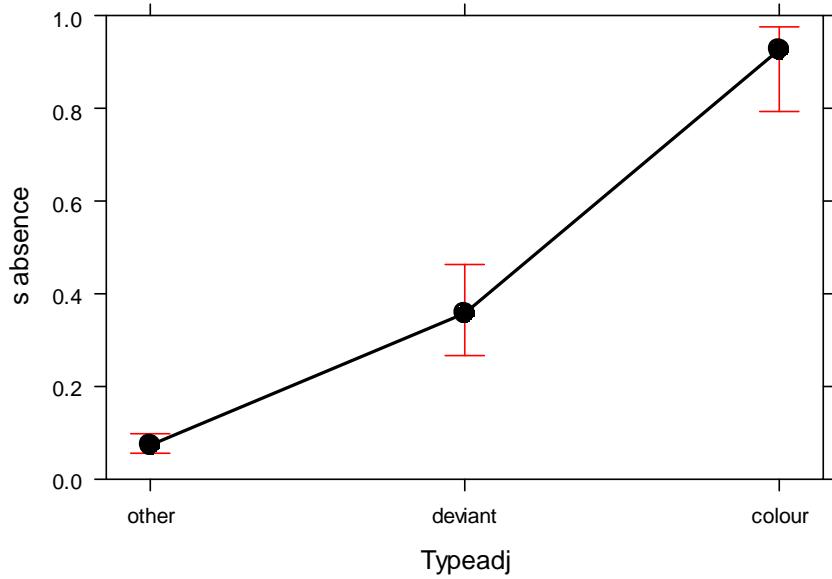
- Lectal variables
  - Variety: *Netherlands, Flanders*
  - Register: *chat, e-mail, mass-newspaper, quality-newspaper*
- Structural variables
  - Quantifier: *iets* ('something'), *niets* ('nothing'), *veel* ('a lot'), *wat* ('something'), *weinig* ('little'), *zoveel* ('so much')
  - Length-Adjective: 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*: unique combination of quantifier and adjectival phrase

- AIC:	2216	- Total number of hits:	3018
- C-value:	0.872	- Hits with -s:	2388
- Number of phrases:	140	- Hits without -s:	630

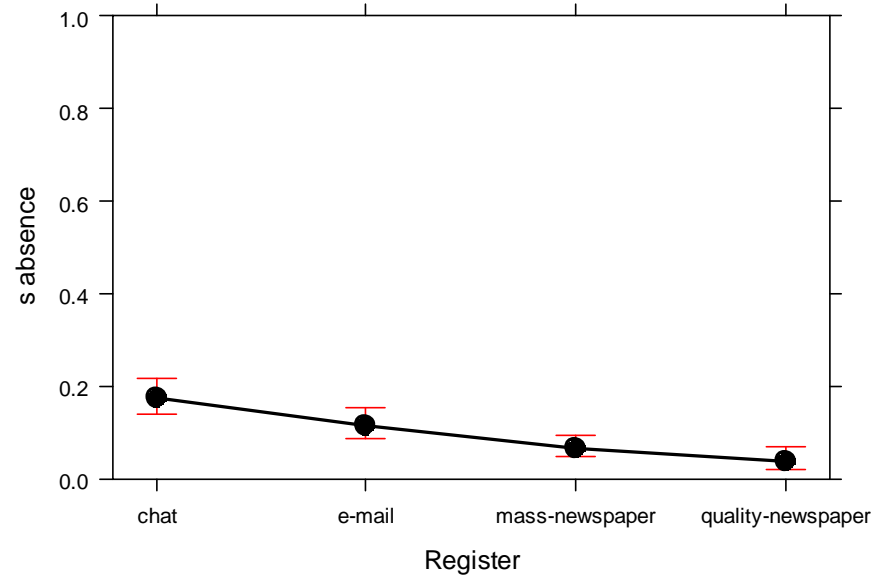
Predictors	Levels of categorical predictors	Estimates	Confidence intervals		P-values
			2,5%	97,5%	
	<b>intercept</b>	0.07	-0.67	0.82	0.8482
Type-Adjective	<i>other</i>	Reference level			
	<i>deviant</i>	1.96	1.45	2.46	< 0.0001
	<i>colour</i>	5.09	3.88	6.30	< 0.0001
Variety	<i>Flanders</i>	Reference level			
	<i>Netherlands</i>	-1.69	-2.01	-1.37	< 0.0001
Register	<i>chat</i>	Reference level			
	<i>e-mail</i>	-0.48	-0.77	-0.19	0.0013
	<i>mass-newspaper</i>	-1.08	-1.42	-0.74	< 0.0001
	<i>quality-newspaper</i>	-1.65	-2.22	-1.08	< 0.0001
Quantifier	<i>iets</i>	Reference level			
	<i>niets</i>	-0.05	-0.66	0.56	0.8809
	<i>veel</i>	-1.14	-1.98	-0.29	0.0083
	<i>wat</i>	-2.00	-2.99	-1.00	< 0.0001
	<i>weinig</i>	-2.50	-4.12	-0.89	0.0023
	<i>zoveel</i>	-2.35	-4.37	-0.34	0.0221
Frequency		-0.45	-0.79	-0.10	0.0109
Interaction Variety - Quantifier	<i>Flanders &amp; iets</i>	Reference level			
	<i>Netherlands - niets</i>	-0.33	-1.03	0.38	0.3635
	<i>Netherlands - veel</i>	0.98	0.02	1.94	0.0443
	<i>Netherlands - wat</i>	1.22	0.19	2.25	0.0208
	<i>Netherlands - weinig</i>	2.33	0.66	4.00	0.0062
	<i>Netherlands - zoveel</i>	2.10	-0.94	5.13	0.1755



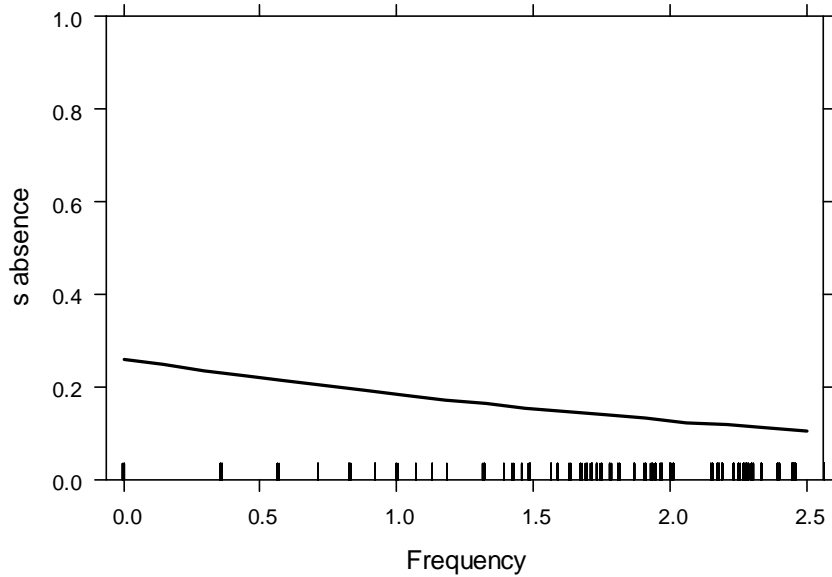
### Typeadj effect plot



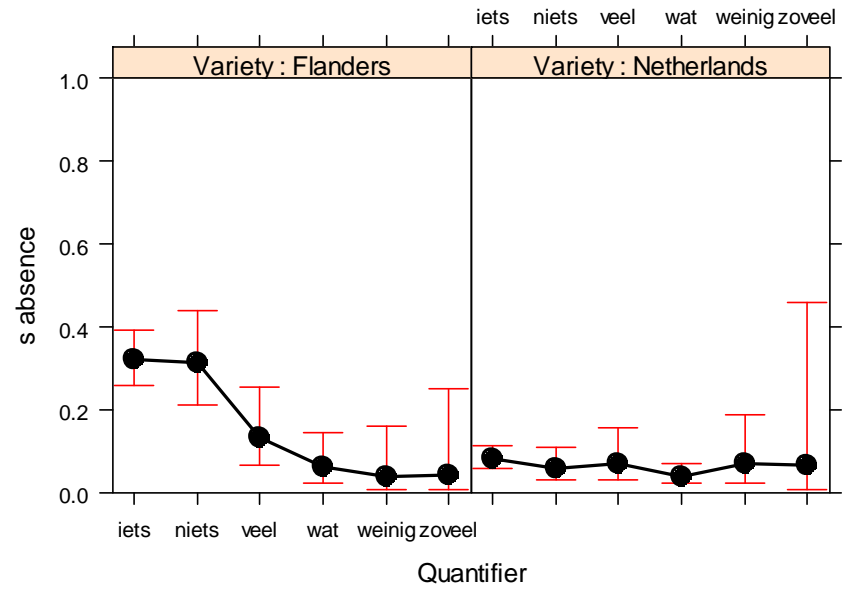
### Register effect plot



### Frequency effect plot

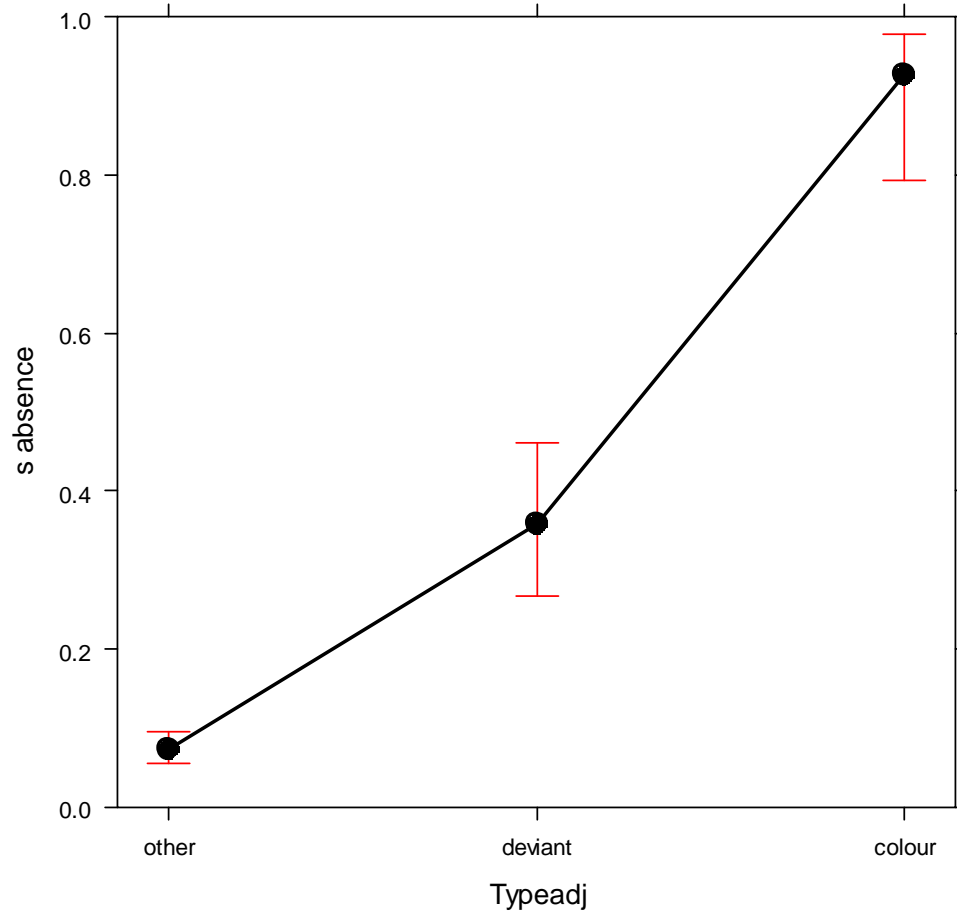


### Variety\*Quantifier effect plot





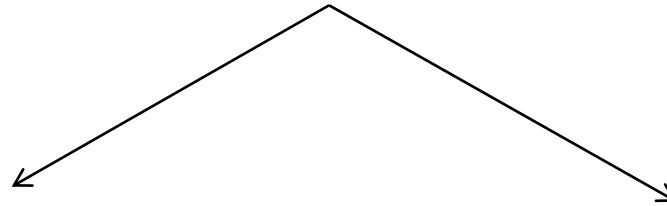
Typeadj effect plot



What is going on here?

# Structural contamination effect: colour adjectives

*veel geel*  
'a lot of yellow (things)'  
*geel*<sub>Adj</sub> or *geel*<sub>Noun</sub>



partitive genitive  
~ *veel interessant*  
'a lot of interesting things'

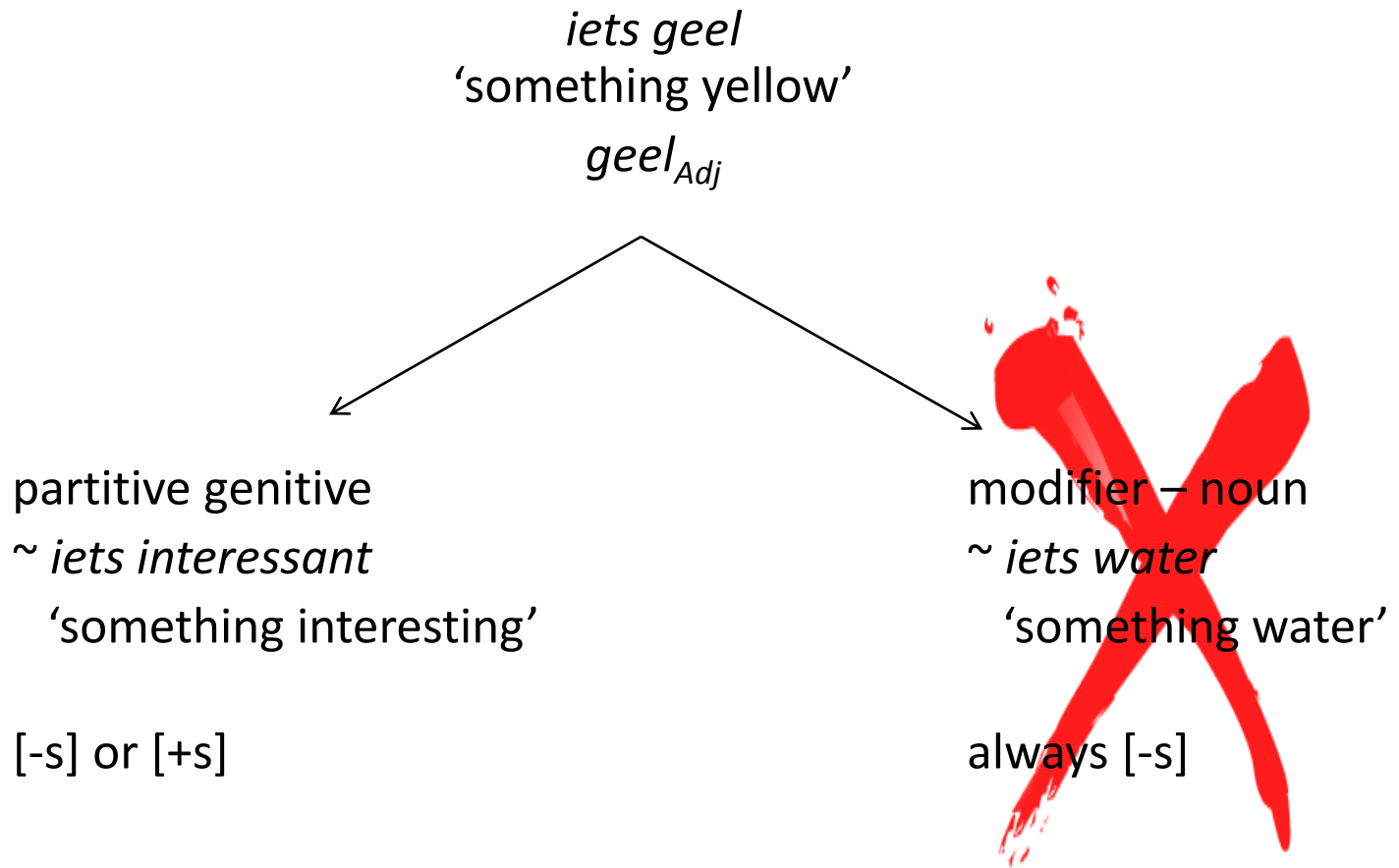
[-s] or [+s]

modifier – noun  
~ *veel water*  
'a lot of water'

always [-s]

⇒ Bias towards [-s]

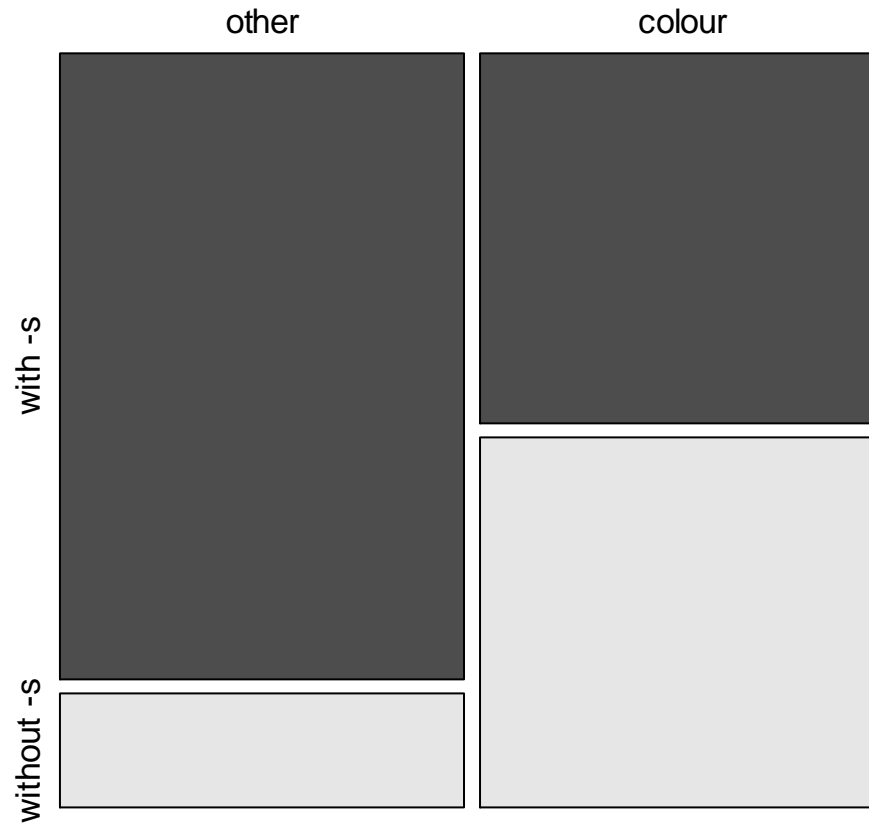
# Structural contamination effect: colour adjectives



⇒ still bias towards [-s] due to superficial resemblance to *veel geel*

# Colour adjectives: unambiguous cases (Q = 'iets')

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



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?'

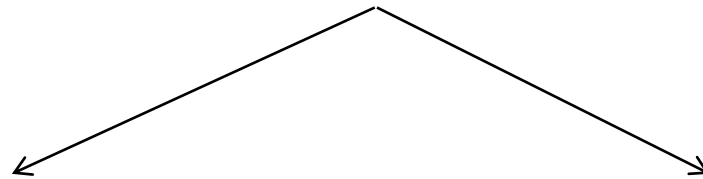
[iets][verkeerd verstaan]

[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]

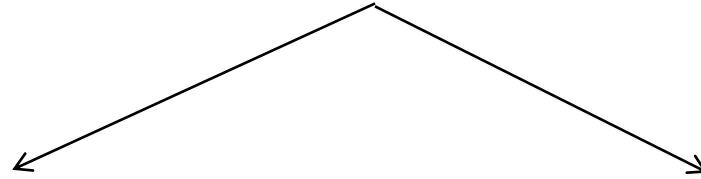
always [-s]

⇒ 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]

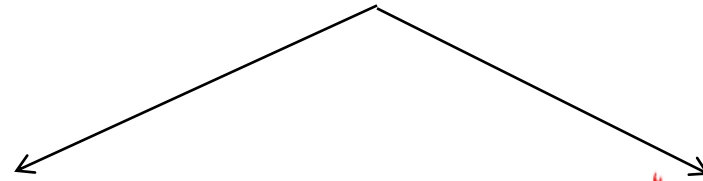




# Structural contamination effect: deviant adjectives

*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

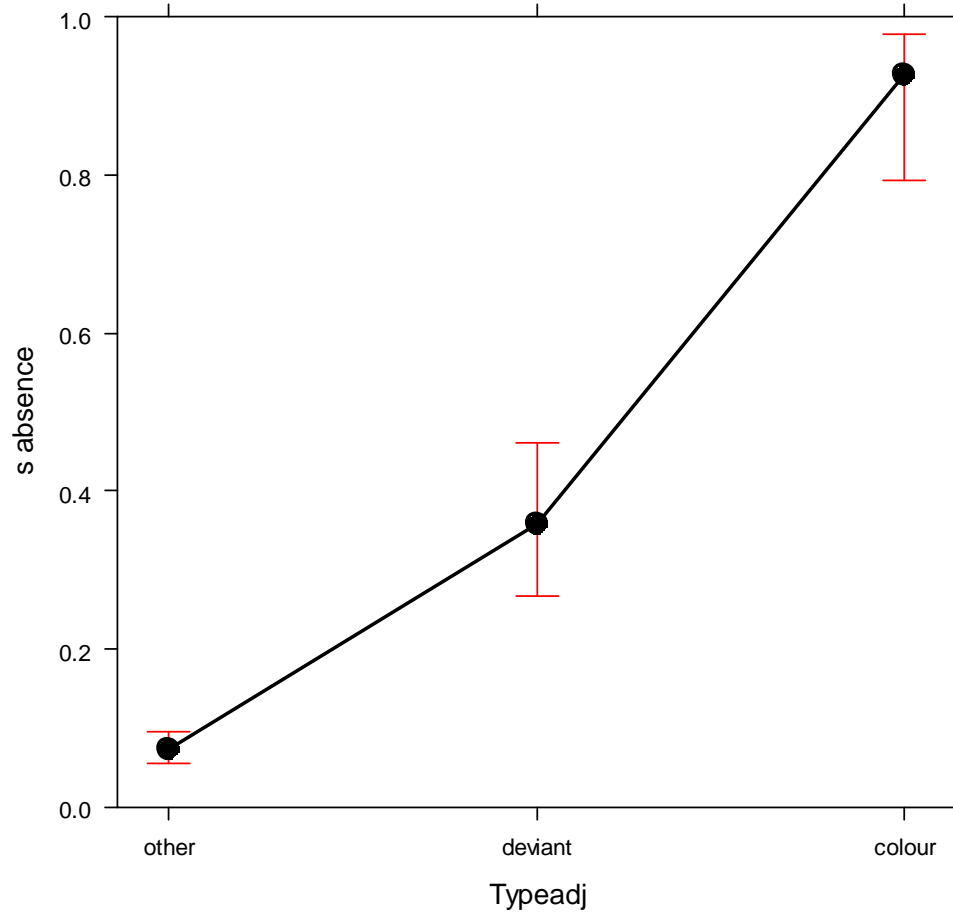
'If I have eaten something the wrong way,...'

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

always [-s]

⇒ No bias towards [-s] preference?

Typeadj effect plot



What is going on here?

Overall distribution of the two forms:



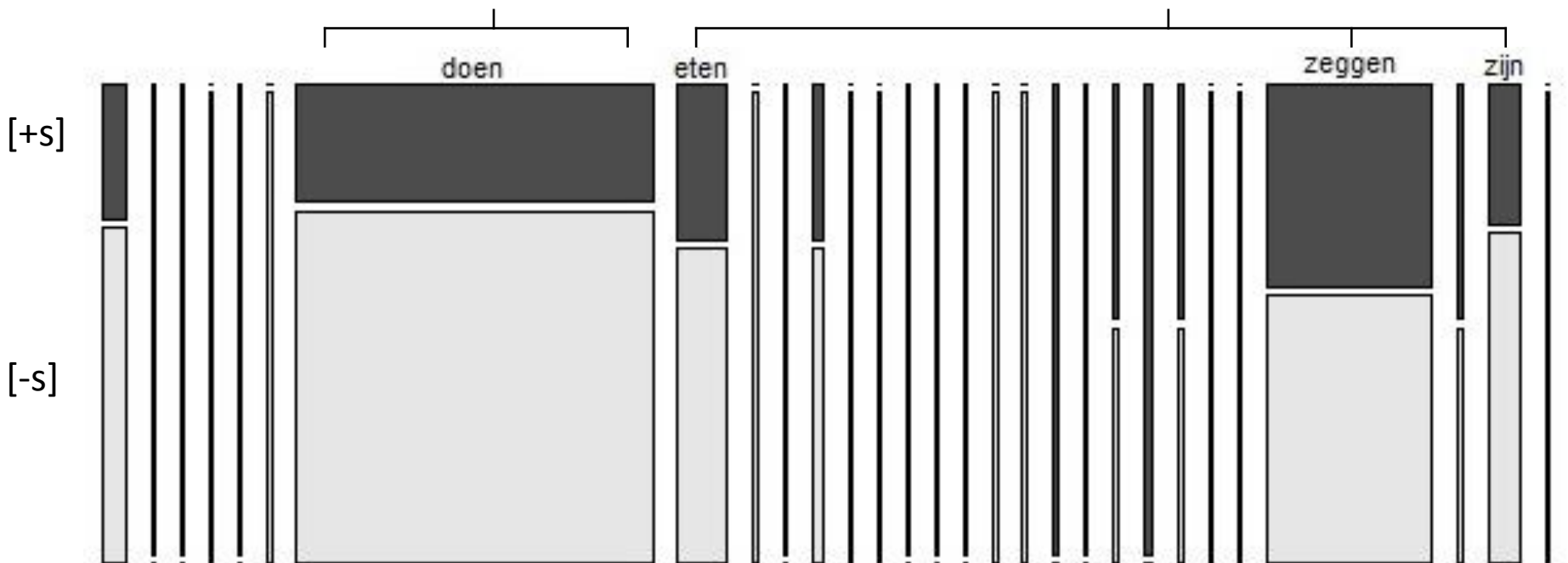
with -s



without -s

possible syntactic ambiguity

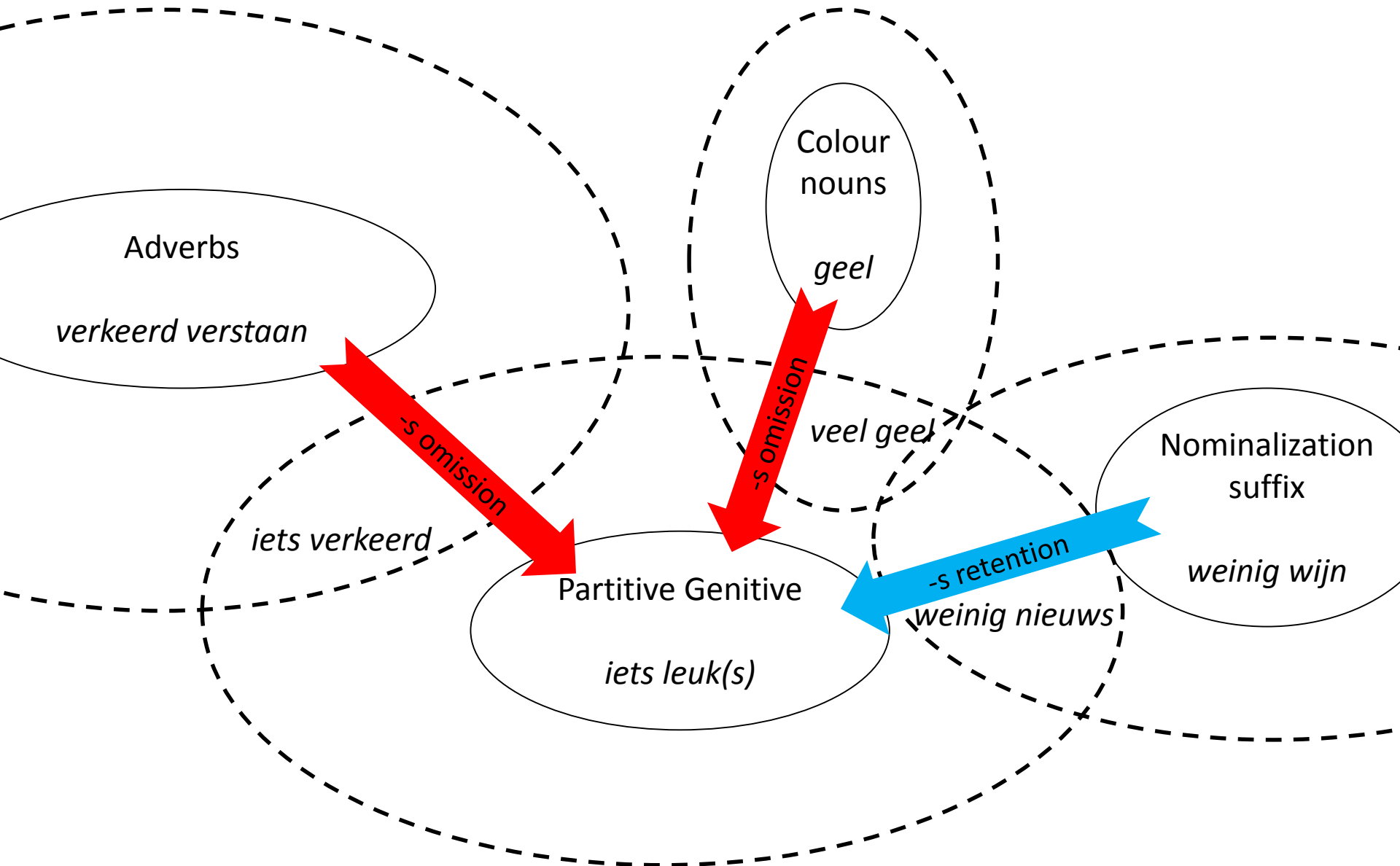
no syntactic ambiguity



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



# Contamination effects

Direct cause:

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



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*

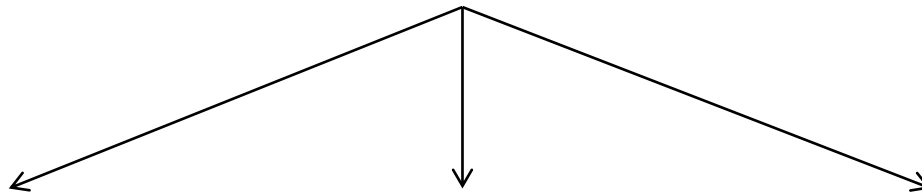
*iets interessant*

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)*

...

typically Flemish

*iets speciaal(s)*

*iets interessant(s)*

*niets concreet(s)*

*iets deftig(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*

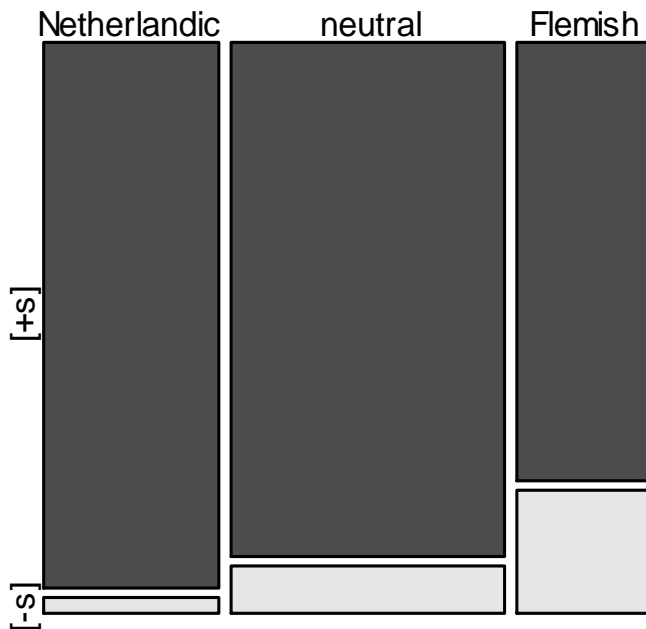
*iets interessant*

preference for [+s]

preference for [-s]

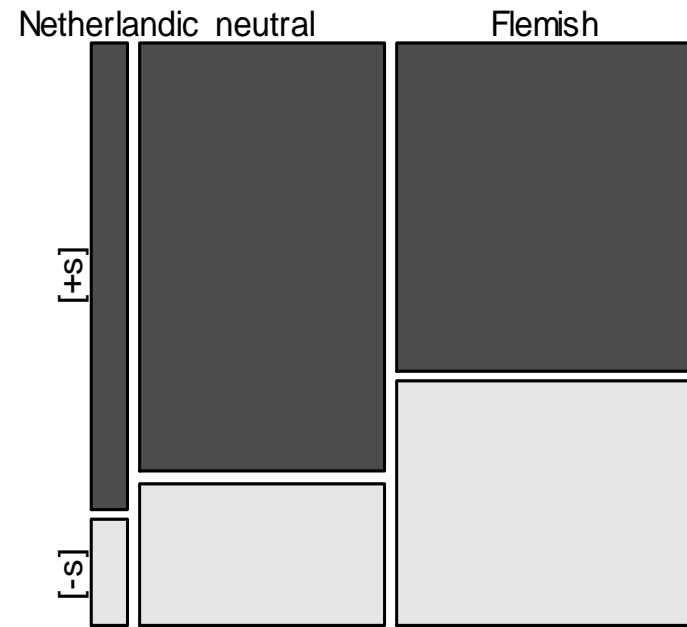
# Lectal contamination

## 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

Bybee, Joan. 2010. *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Van de Velde, Freek. 2014. 'Degeneracy: the maintenance of constructional networks'. In: Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *The extending scope of construction grammar*. Berlin: Mouton de Gruyter. 141-180.

- 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.

Bybee, Joan. 2010. *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Clay Beckner. 2010. Usage-based theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press. 827-855.

Geeraerts, Dirk. 2006. 'Methodology in cognitive linguistics'. In: Gitte Kristiansen, Michel Achard, René Dirve & Francisco Ruiz de Mendoza Ibañez (eds.), *Cognitive linguistics: current applications and future perspectives*. Berlin: Mouton de Gruyter. 21-49.

Geeraerts, Dirk & Gitte Kristiansen. Forthcoming. 'Variationist linguistics'. In: Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*. Berlin: Mouton de Gruyter.

Gries, Stefan Th. & Anatol Stefanowitsch (eds.). 2006. *Corpora in cognitive linguistics. Corpusbased approaches to syntax and lexis*. Berlin: Mouton de Gruyter.

Hopper, Paul J. 1987. 'Emergent Grammar'. *Berkeley Linguistic Society* 13: 139-157.

Hopper, Paul J. 1998. 'Emergent Grammar'. In: Michael Tomasello (ed.), *The new psychology of language. Cognitive and functional approaches to language structure*. Mahwah: Lawrence Erlbaum. 155-175.

Kemmer, Suzanne & Michael Barlow. 2000. Introduction: A usage-based conception of language. In: Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, vii-xxviii. Stanford: CSLI.

Tummers, José, Kris Heylen & Dirk Geeraerts. 2005. 'Usage-based approaches in cognitive linguistics: a technical state of the art'. *Corpus Linguistics and Linguistic Theory* 1(2): 225-261.

Von Mengden, Ferdinand & Evie Coussé. 2014. 'Introduction. The role of change in usage-based conceptions of language'. In: Evie Coussé & Ferdinand Von Mengden (eds.), *Usage-Based approaches to language change*. Amsterdam: John Benjamins. 1-20.

# Thanks!

- Dirk Pijpops: pijpopsd@outlook.com
- Freek Van de Velde: freek.vandevelde@arts.kuleuven.be  
<http://wwwling.arts.kuleuven.be/qlvl/freek.htm>