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)

> "[I]nnovations 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
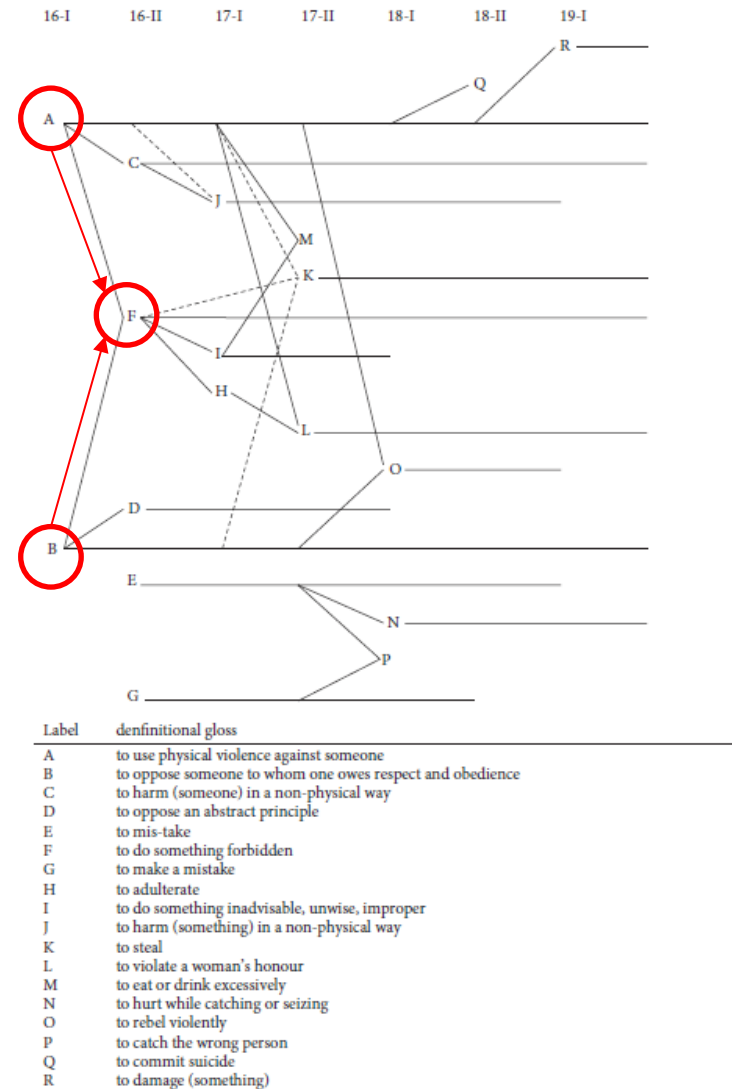 - lexical semantics
  - morphology
  - syntax

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

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 *$sueh_2d$-   >   PGmc. $swōti/u$-   >   PdDu $zoet$ /u/
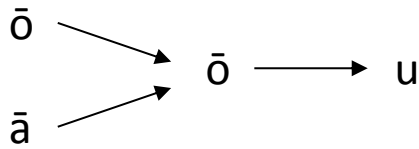
Contamination of lineages

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

Label	denfinitional gloss
A	to use physical violence against someone
B	to oppose someone to whom one owes respect and obedience
C	to harm (someone) in a non-physical way
D	to oppose an abstract principle
E	to mis-take
F	to do something forbidden
G	to make a mistake
H	to adulterate
I	to do something inadvisable, unwise, improper
J	to harm (something) in a non-physical way
K	to steal
L	to violate a woman's honour
M	to eat or drink excessively
N	to hurt while catching or seizing
O	to rebel violently
P	to catch the wrong person
Q	to commit suicide
R	to damage (something)

Figure 2. The diachronic structure of *vergrijpen* (adapted from Geeraerts 1997: 58–59)

Geeraerts, Dirk. 1997. *Diachronic prototype semantics: A contribution to historical lexicology.* Oxford: Clarendon Press.
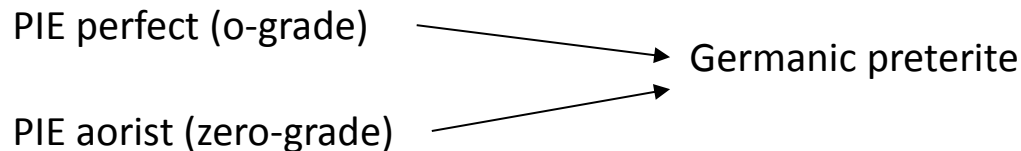
Contamination of lineages

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

	e-grade	o-grade	zero-grade
PIE	present Cl.Gr.*leíp-ō*	perfect Cl.Gr.*lé-loip-a*	aorist Cl.Gr.*é-lip-on*
Germanic	present Go.*bi-leib-an*	preterit-sg Go.*bi-laif*	preterite-pl OlD.*be-lib-en*

Figure 3. Proto-Indo-European perfect/aorist > Germanic preterite*
*PIE: Proto-Indo-European; Cl.Gr.: Classical Greek; Go.: Gothic; OlD: Old Dutch

PIE perfect (o-grade) ⟶ Germanic preterite

PIE aorist (zero-grade) ⟶

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) *Þe 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)

Traugott, Elizabeth & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
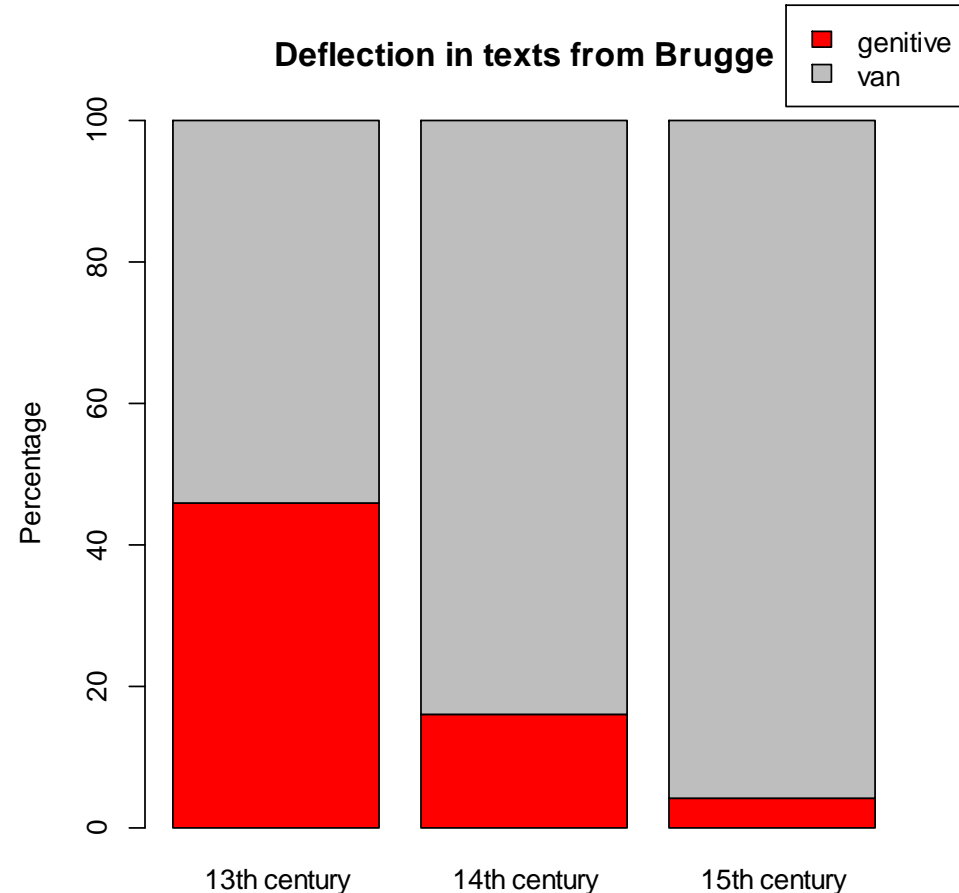
CASE STUDY: DUTCH PARTITIVE GENITIVE

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).

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

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.

Deflection in texts from Brugge

Legend: genitive (red), van (grey). Y-axis: Percentage (0–100). X-axis: 13th century, 14th century, 15th century.

Case study: Dutch partitive genitive

- Dutch partitive genitive

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

  [$_{NP}$ Q$_i$ Adj$_j$-s] $\leftrightarrow$ [modifier$_j$ head-quantity$_i$]

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

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

iets leuk(s)
'something fun'

⤲   ⤳

[+s]   [-s]
iets leuks   *iets leuk*

veel interessant(s)
'a lot of interesting things'

⤲   ⤳

[+s]   [-s]
veel interessants   *veel interessant*

iets erg gemakkelijk(s)
'something very easy'

⤲   ⤳

[+s]   [-s]
iets erg gemakkelijks *iets erg gemakkelijk*
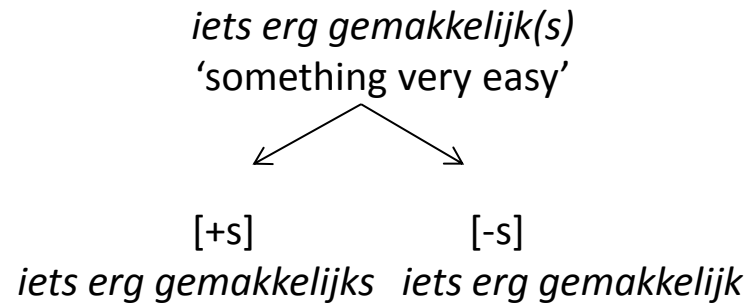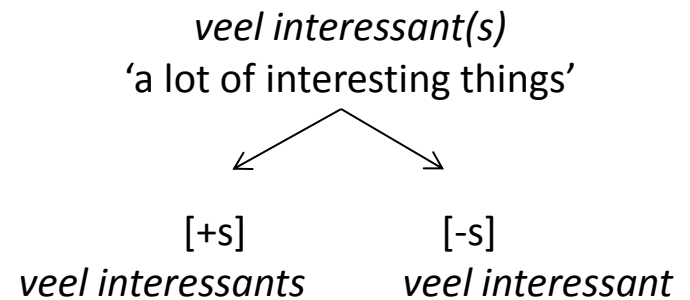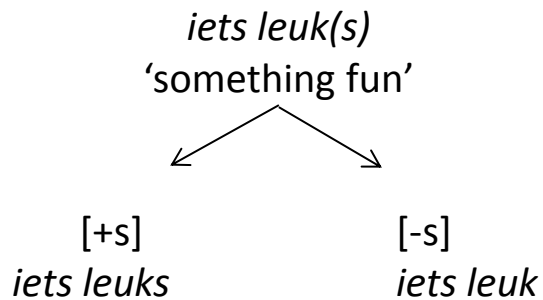
MODEL 2: WITH *PHRASE* AS A RANDOM FACTOR

- AIC: 2216
- C-value: 0.872
- Number of phrases: 140

- Total number of hits: 3018
- Hits with -s: 2388
- Hits without -s: 630

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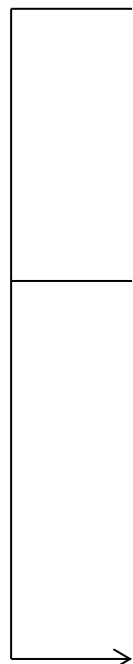
Typeadj effect plot
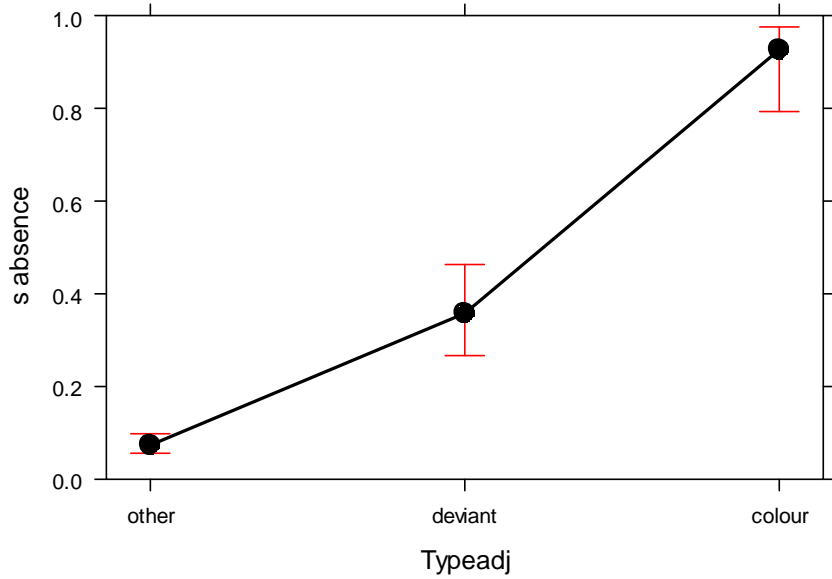
Register effect plot
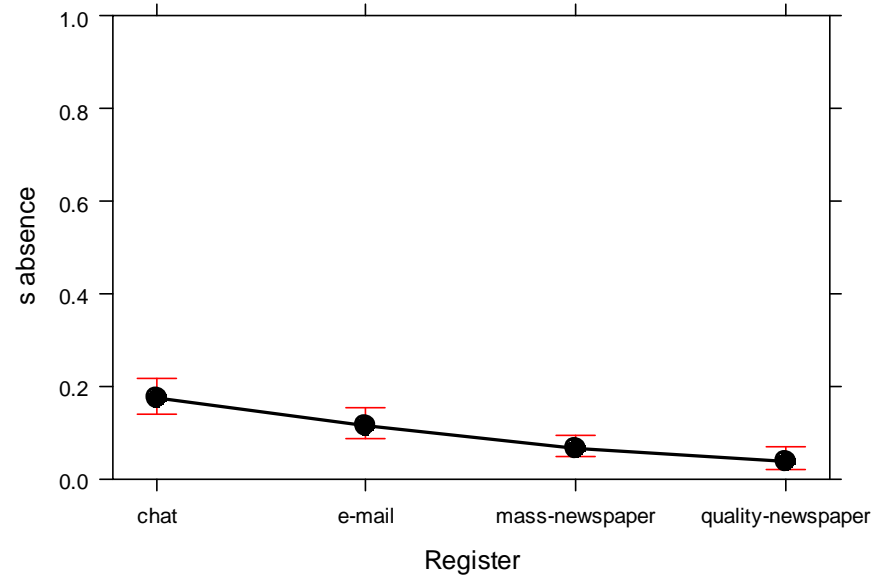
Frequency effect plot
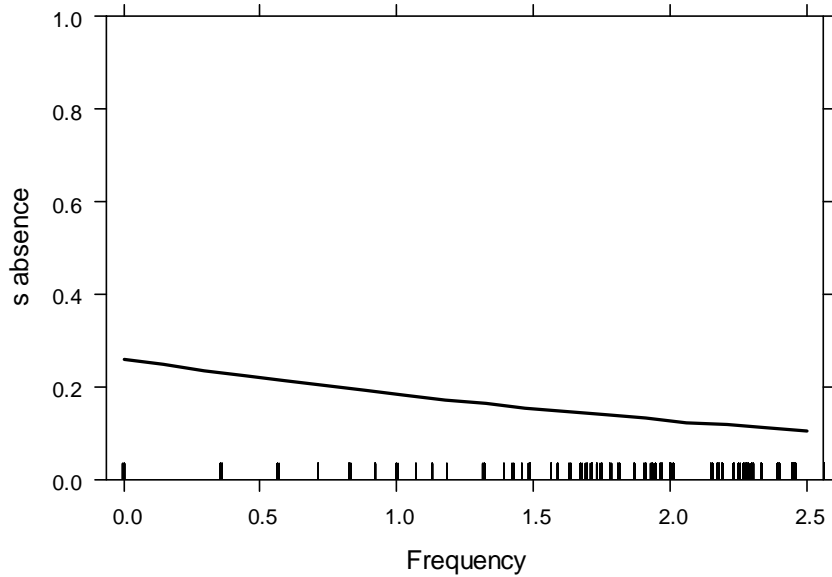
Variety*Quantifier effect plot
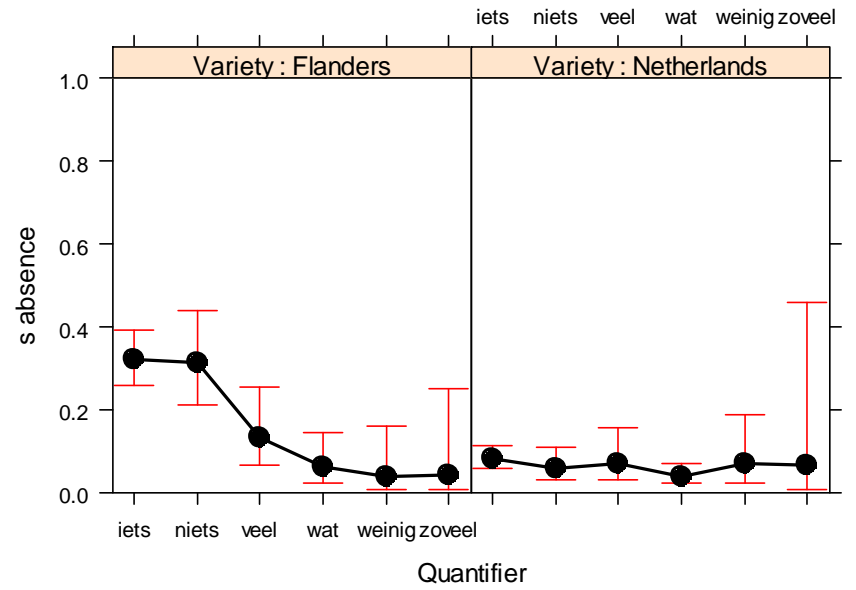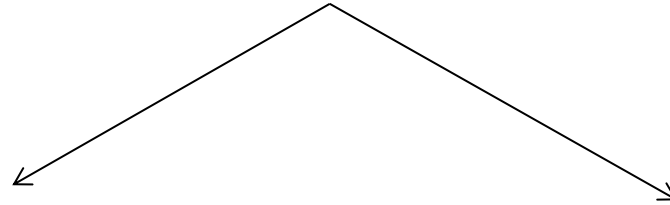
Typeadj effect plot
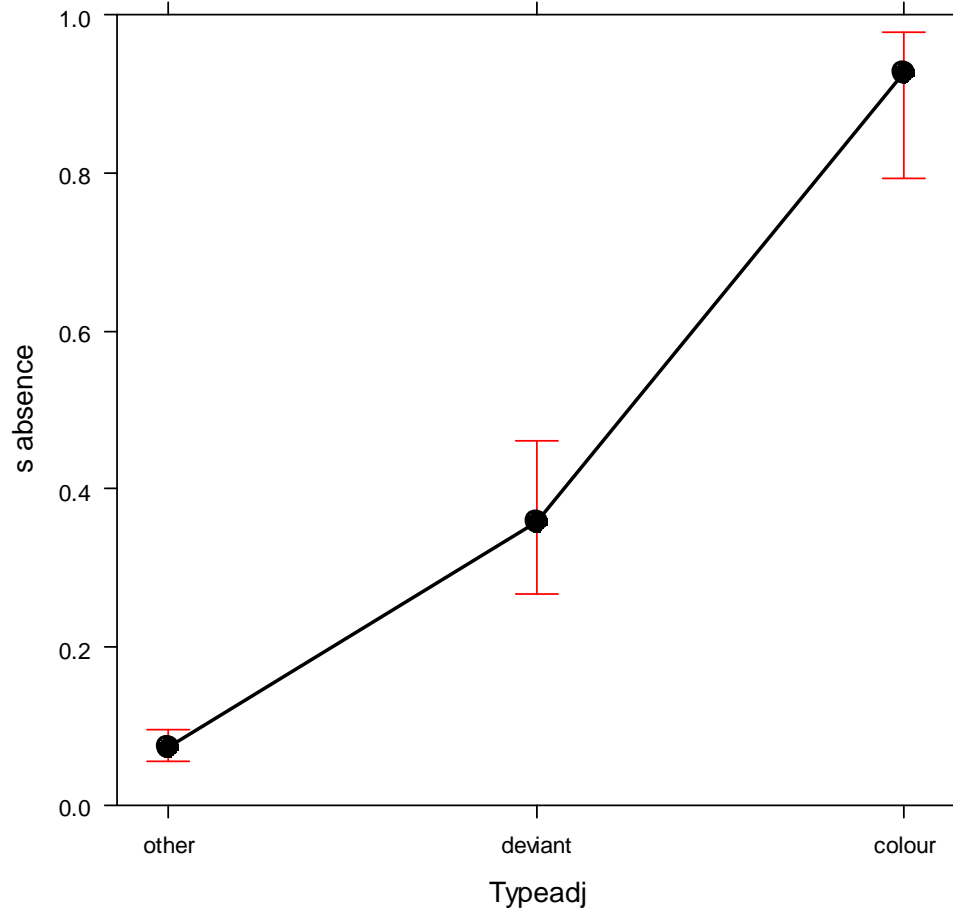
What is going on here?

Structural contamination effect: colour adjectives

veel geel
'a lot of yellow (things)'
$geel_{Adj}$ or $geel_{Noun}$

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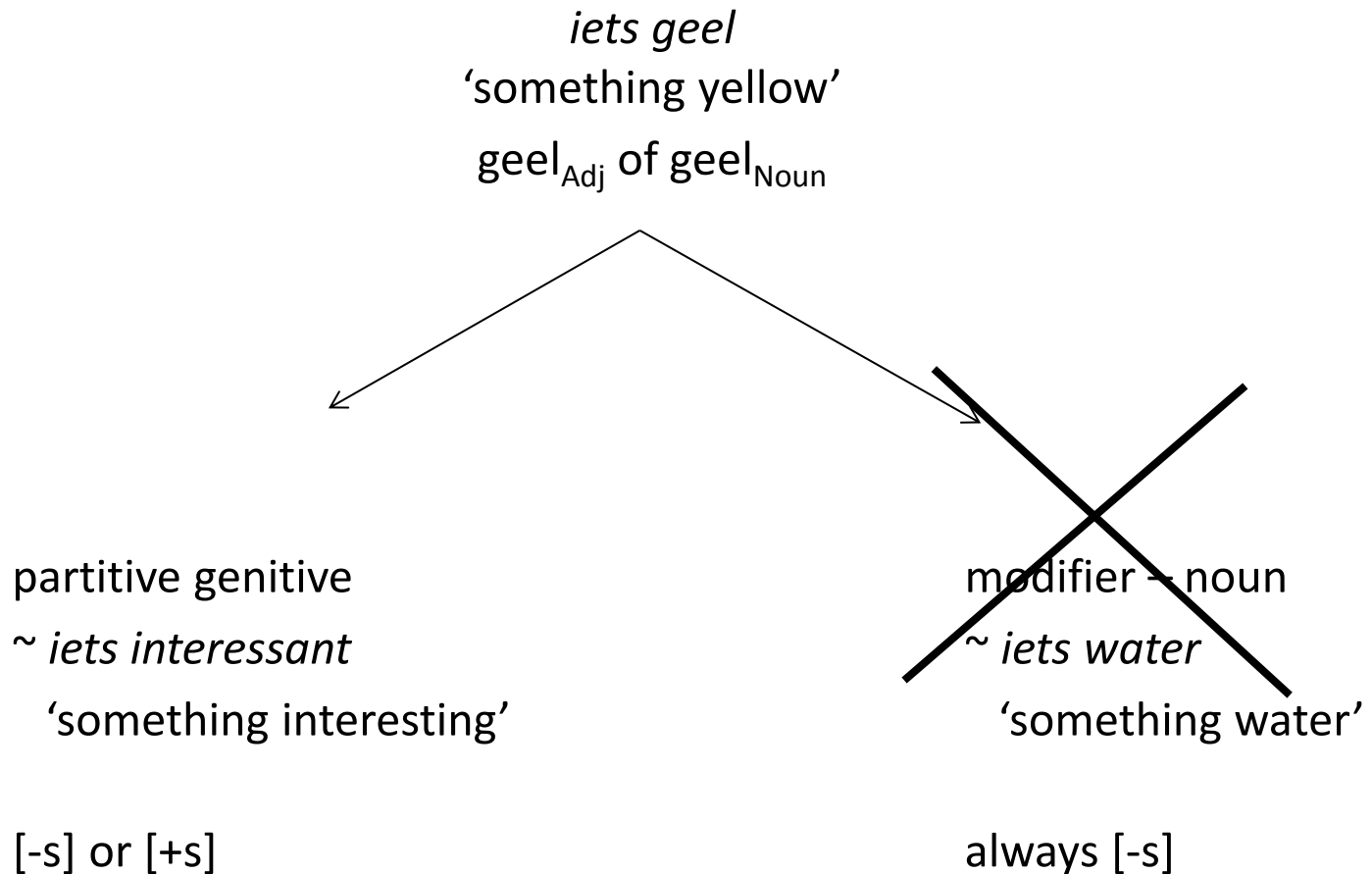
~ *veel interessant*
  'a lot of interesting things'

[-s] or [+s]

modifier – noun

~ *veel water*
  'a lot of water'

always [-s]

$\Rightarrow$ Bias towards [-s]

Structural contamination effect: colour adjectives

iets geel
'something yellow'
geel$_{Adj}$ of geel$_{Noun}$

partitive genitive

~ *iets interessant*

'something interesting'

[-s] or [+s]

modifier + noun

~ *iets water*

'something water'

always [-s]

$\Rightarrow$ still bias towards [-s] due to superficial resemblence to *veel geel*

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

iets + adj. (diff. in abs. numb. not visualized)
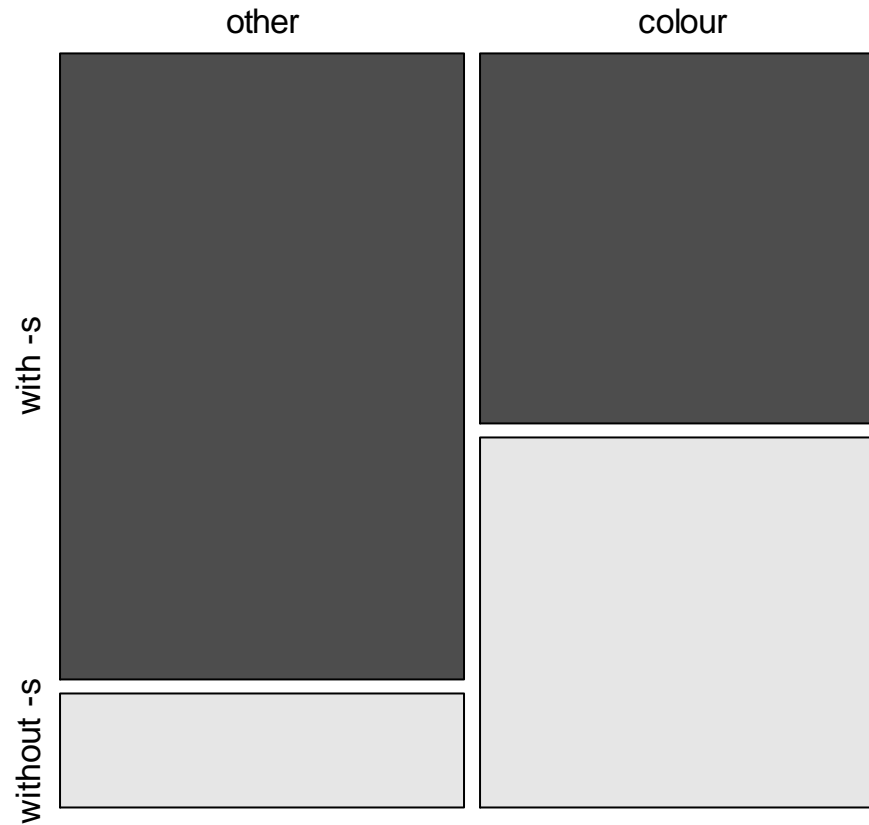
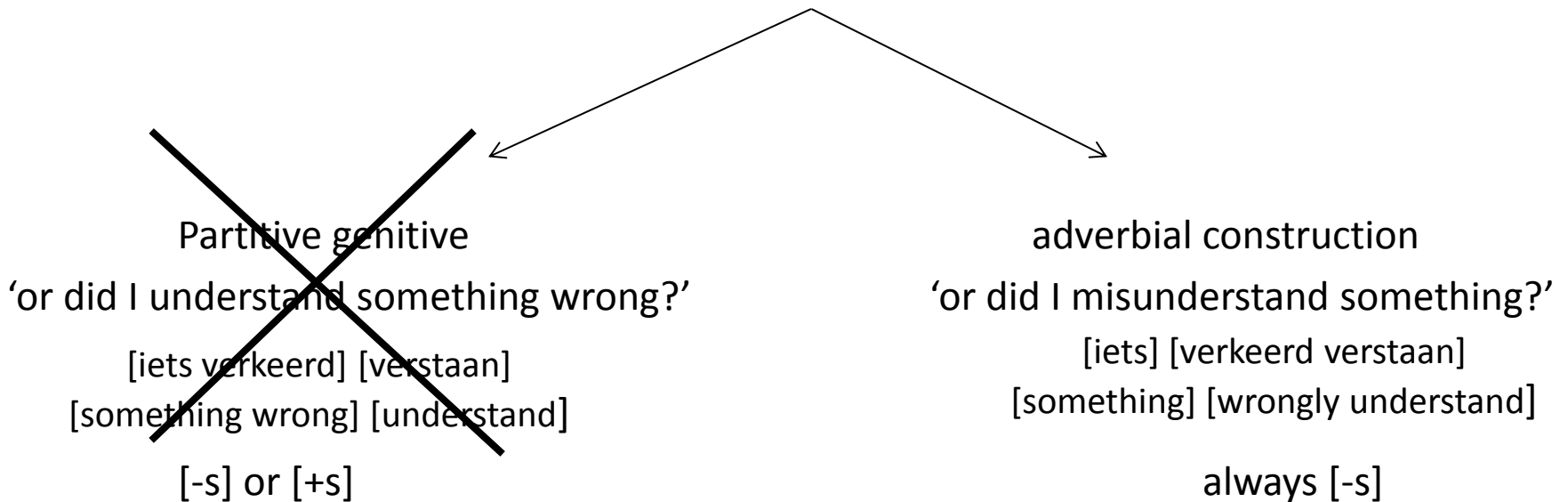# Structural contamination effect: deviant adjectives

deviant adjectives:

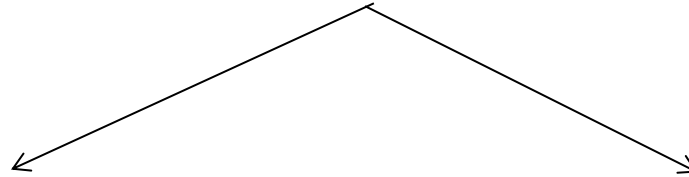|	
verkeerd	'wrong'
goed	'good'
beter	'better'
fout	'incorrect'

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	adverbial construction
'Did I do something wrong?'	'Did I do something the wrong way?'
[iets verkeerd] [doen]	[iets] [verkeerd doen]
[something wrong] [do]	[something] [wrong-do]
[-s] or [+s]	always [-s]

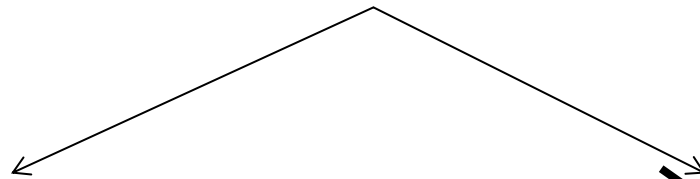$\Rightarrow$ Bias towards [-s]

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,...'

[something wrong] [eat]

[-s] or [+s]

~~adverbial construction~~

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

~~[something] [wrong-eat]~~

~~always [-s]~~
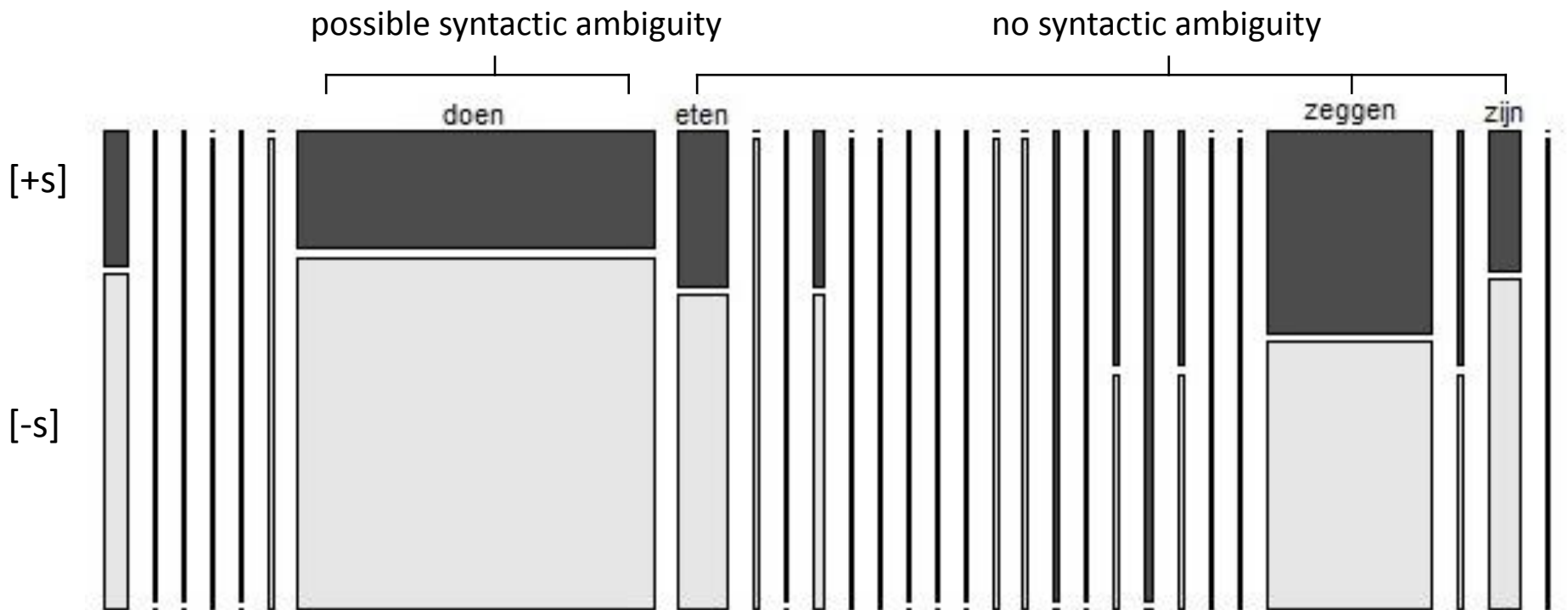
⇒ No bias towards [-s] preference?

Overall distribution of the two forms:

with -s

without -s

possible syntactic ambiguity

no syntactic ambiguity

doen eten zeggen zijn

[+s]

[-s]

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?'

Always [-s]

Partitive genitive construction

*als ik **iets verkeerd** gegeten heb,…*

'If I have eaten something wrong,…'

preference for [-s]

Superficial resemblence

Contamination effect

Direct cause:

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

↓

↓

Indirect effect on superficially similar or identical occurences:

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]
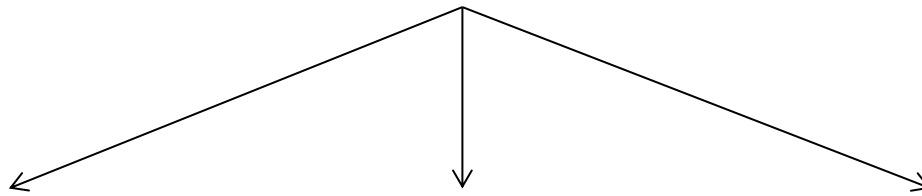
↓

↓

Indirect effect:

wat mooi(s)   *iets interessant(s)*
preference for [+s]   preference for [-s]

Operationalisation

140 phrase types

typically Netherlandic	neutral	typically Flemish
iets bijzonder(s)	*weinig concreet(s)*	*iets speciaal(s)*
wat zinnig(s)	*iets zinnig(s)*	*iets interessant(s)*
wat mooi(s)	*iets spannend(s)*	*niets concreet(s)*
iets leuk(s)	*niets erg(s)*	*iets deftig(s)*
…	…	…

Lectal contamination

The Netherlands
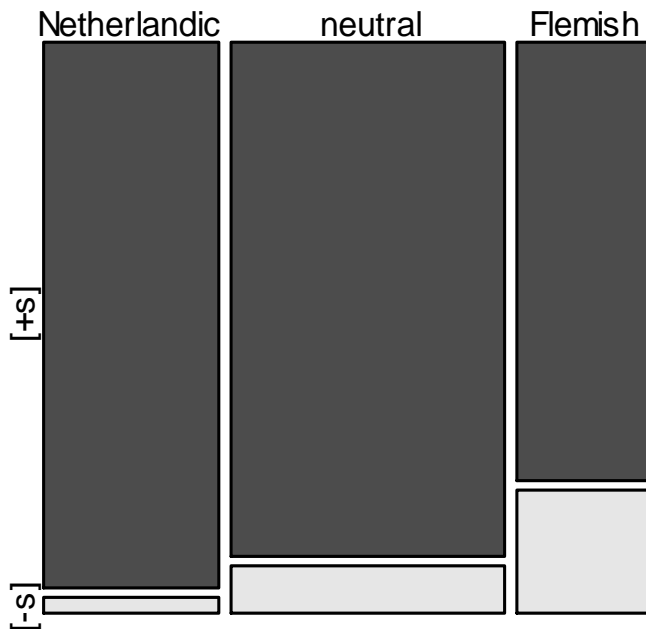
Netherlandic neutral Flemish

[+s]

[-s]

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

Flanders
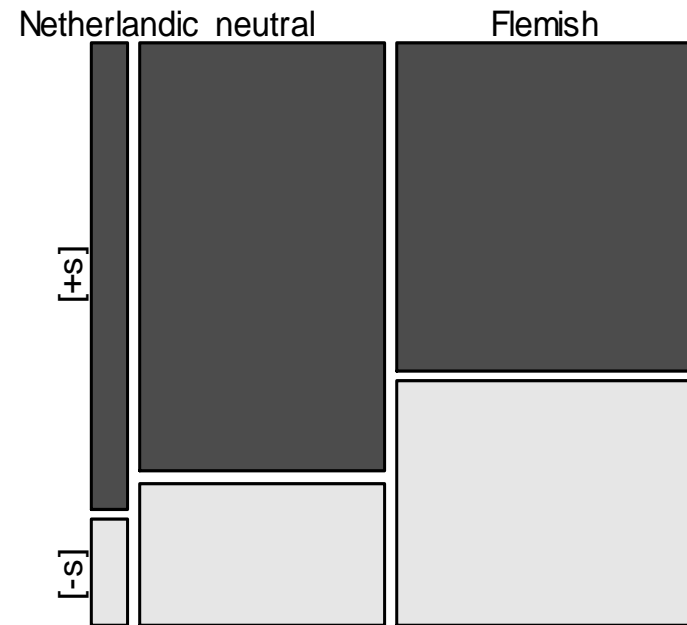
Netherlandic neutral Flemish

[+s]

[-s]

Mosaic plot: distribution of the variants over the typically
Netherlandic, neutral and typically Flemish phrases in only the
Flemish material
(Kendall's τ = - 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. Forthcoming 2014. 'Degeneracy: the maintenance of constructional networks'. In: Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Constructions all the way everywhere: the extending scope of construction grammar*. Berlin: Mouton de Gruyter.

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.

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*. Mahwau: 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 to:

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