

# Geographic variation as a window on probabilistic individual grammars

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

<sup>1</sup>QLVL, University of Leuven

<sup>2</sup>Research Foundation Flanders (FWO)



1 underlying mechanism

2 manifestations

Exemplar contamination



Constructional contamination

Lectal contamination

# Exemplar contamination

Original cause: strings appear (more often) in variant X



Entrenched in probabilistic grammar/exemplar memory



Strings appear (more often) in variant X, even where original cause is not causing it



Trace back original cause

# Lectal contamination

**lectal difference:** strings appear (more often) in variant X



Entrenched in probabilistic grammar/exemplar memory



Strings appear (more often) in variant X, even where **lectal difference** is not causing it



Trace back **lectal difference**

# Case study

- *iets speciaal*  
something special
- *wat bijzonder*  
something peculiar
- *niets leuk*  
nothing fun
- Dutch partitive genitive: [Quantifier + Adjective]<sub>NP</sub>

# Case study

- *iets speciaals*  
something special
- *wat bijzonders*  
something peculiar
- *niets leuks*  
nothing fun
- Dutch partitive genitive: [Quantifier + Adjective]<sub>NP</sub>

# What determines -s drop?

DE GRUYTER MOUTON

DOI 10.1515/cllt-2013-0027 — Corpus Linguistics and Ling. Theory 2014; aop

---

Dirk Pijpops and Freek Van de Velde\*

## **A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion**

DE GRUYTER MOUTON

Folia Linguistica 2016; 50(2): 543–581

---

Dirk Pijpops\* and Freek Van de Velde

## **Constructional contamination: How does it work and how do we measure it?**

The Netherlands

Belgium

with -s

1435

953

without -s

153

477

$p < 0.001$ , Cramér's  $V = 0.29$





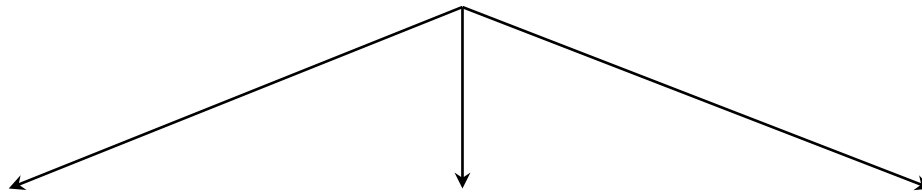
The Netherlands: variant with -s  
Belgium: variant without -s



In general:

Typically Netherlandic strings: variant with -s  
Typically Belgian strings: variant without -s

# 143 phrase types



Typically Netherlandic

*wat boeiend(s)*  
*iets bijzonder(s)*  
*wat leuk(s)*  
*iets leuk(s)*

...

Neutral

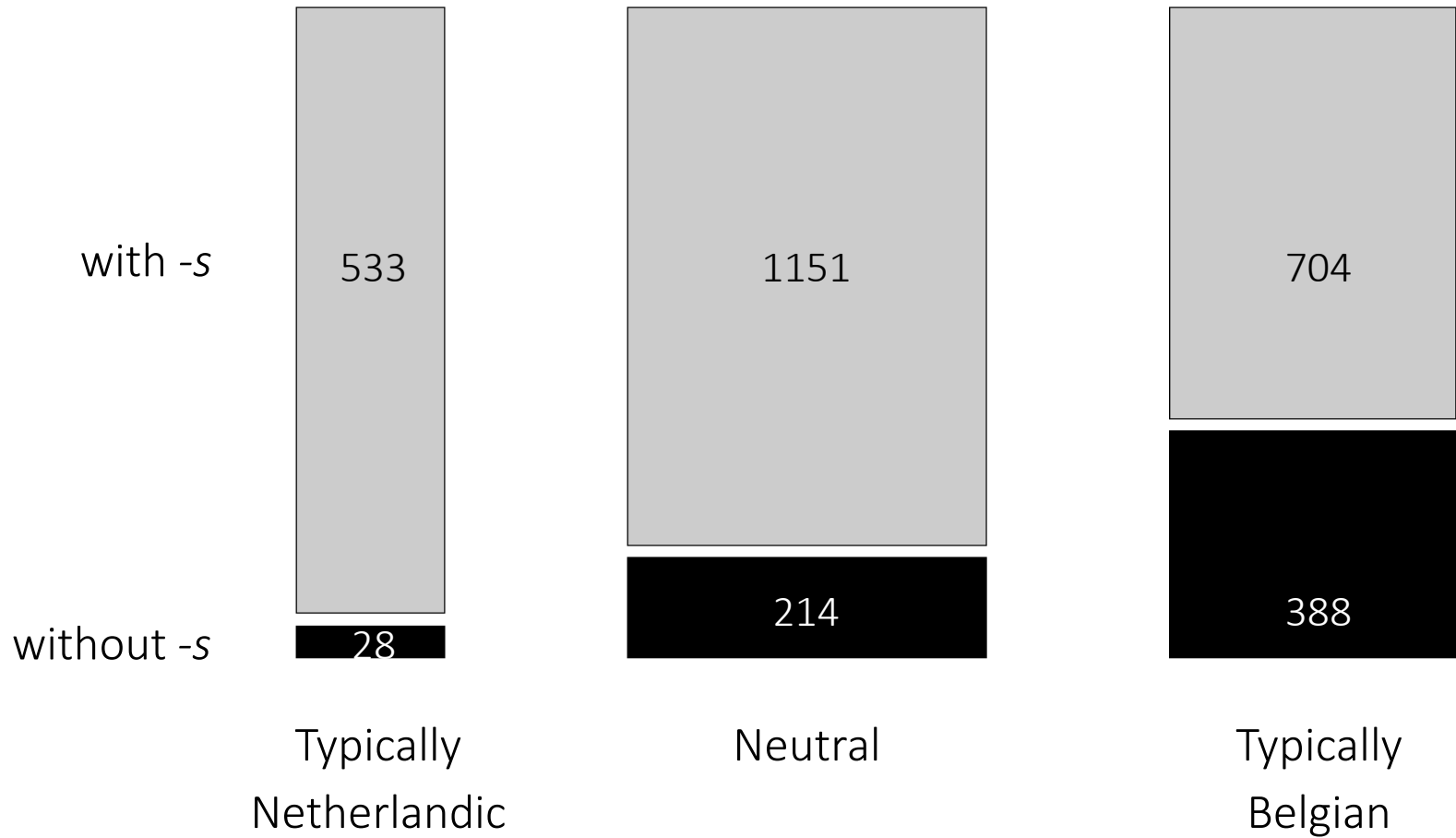
*weinig concreet(s)*  
*iets zinnig(s)*  
*iets spannend(s)*  
*niets erg(s)*

...

Typically Belgian

*iets interessant(s)*  
*niets speciaal(s)*  
*iets deftig(s)*  
*iets raar(s)*

...



$p < 0.001$ , kendall's  $\tau = 0.27$

The Netherlands: variant with -s  
Belgium: variant without -s



In general:

Typically Netherlandic strings: variant with -s  
Typically Belgian strings: variant without -s



Entrenched in probabilistic grammar



Even within a single lect:

Typically Netherlandic strings: variant with -s  
Typically Belgian strings: variant without -s

The Netherlands: variant with -s  
Belgium: variant without -s

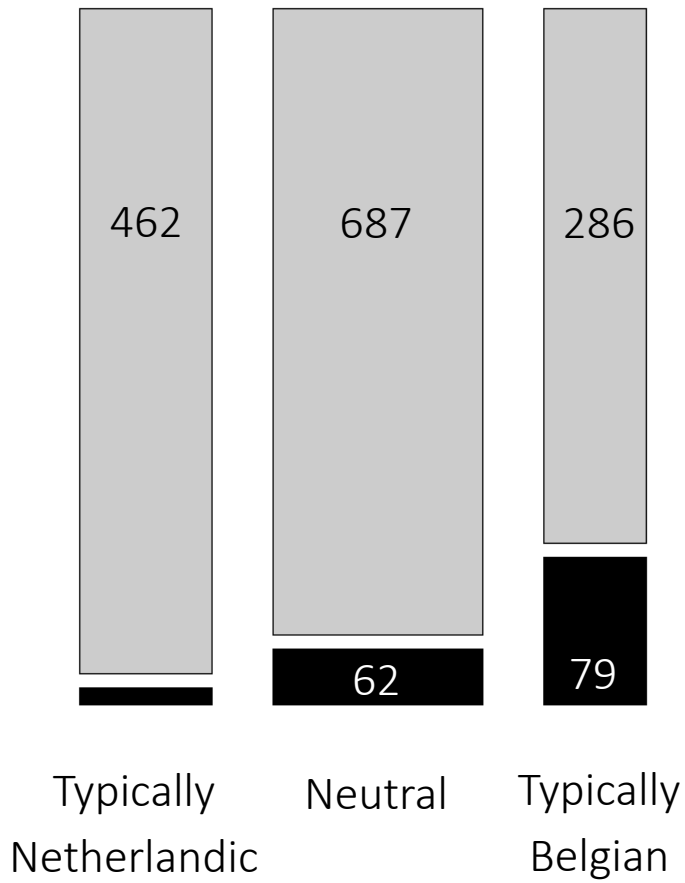


*wat bijzonders*      *wat bijzonder*



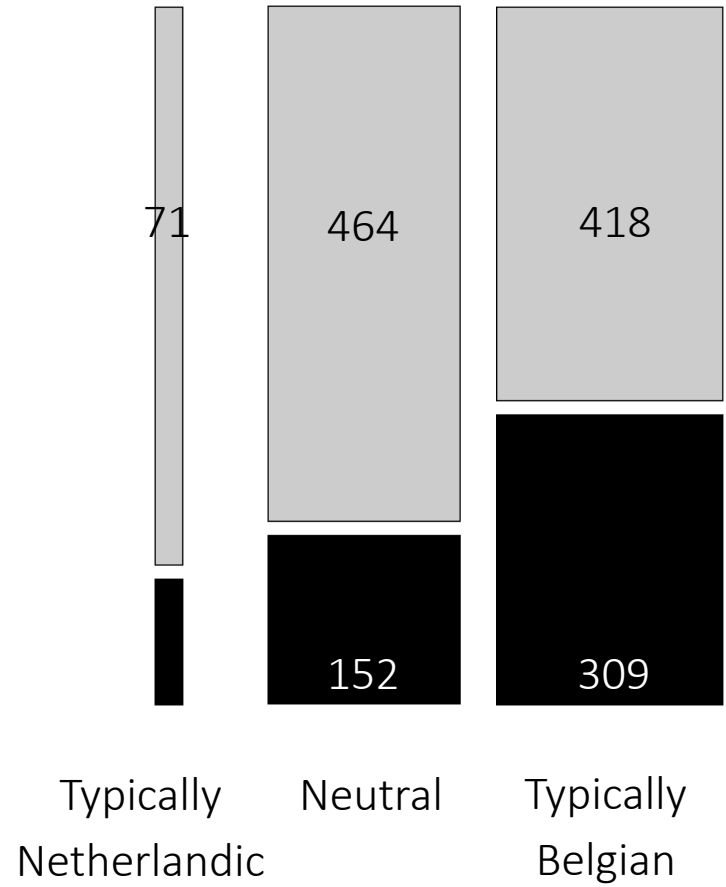
*wat bijzonders*      *wat bijzonder*

## The Netherlands



$p < 0.001$  , kendall's  $\tau = 0.21$

## Belgium

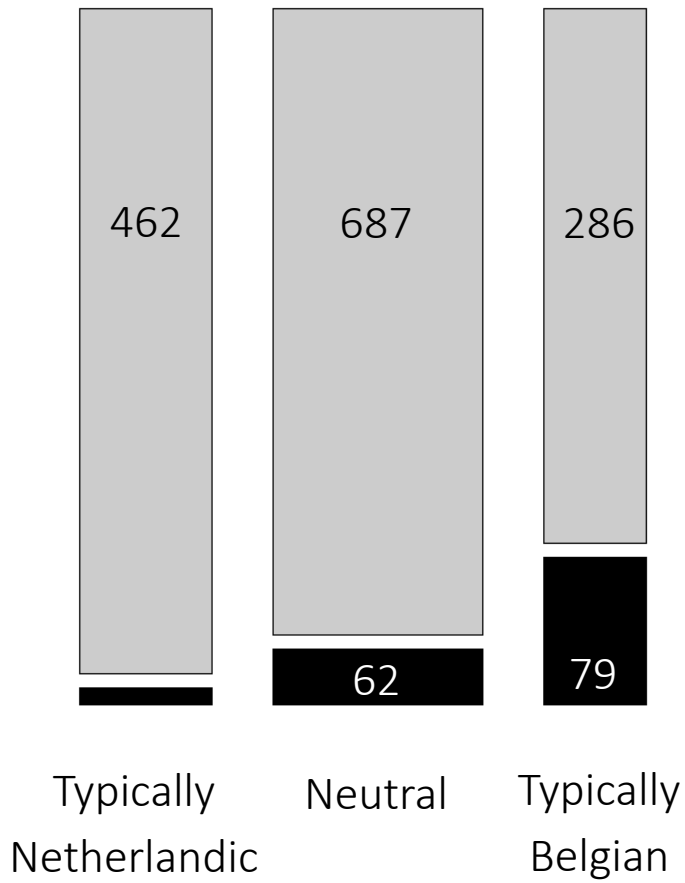


$p < 0.001$ , kendall's  $\tau = 0.19$

# Methodological Relevance

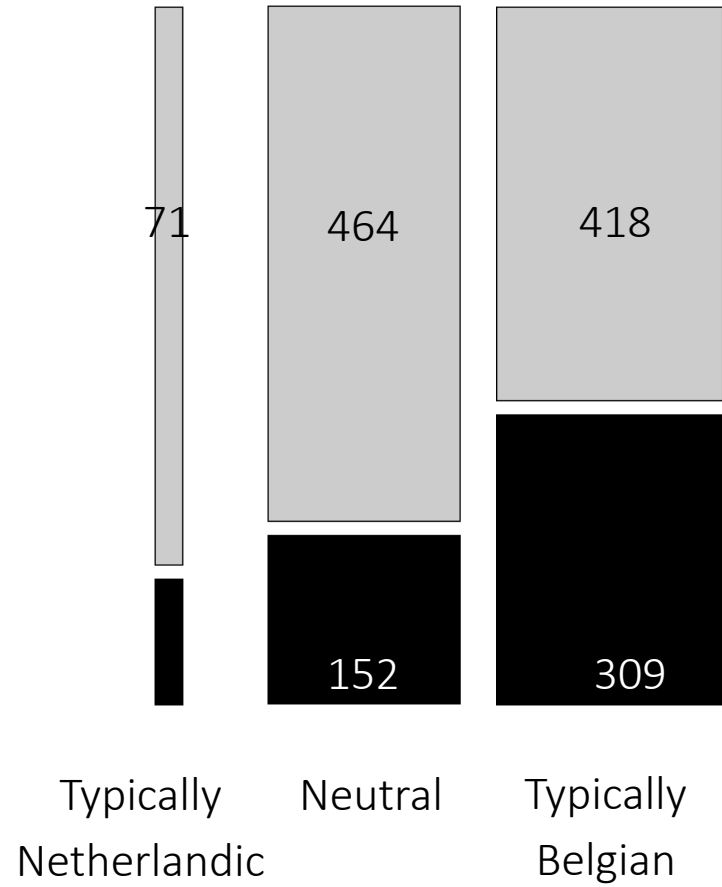
- Lectal variation has language-internal effects
- Way to find lectal effects, even if you do not have data on a particular lect

## The Netherlands



$p < 0.001$  , kendall's  $\tau = 0.21$

## Belgium



$p < 0.001$  , kendall's  $\tau = 0.19$



# Methodological Relevance

- Lectal variation has language-internal effects
- Way to find lectal differences, even if you do not have data on a particular lect

⇒ Replicate

# Replication

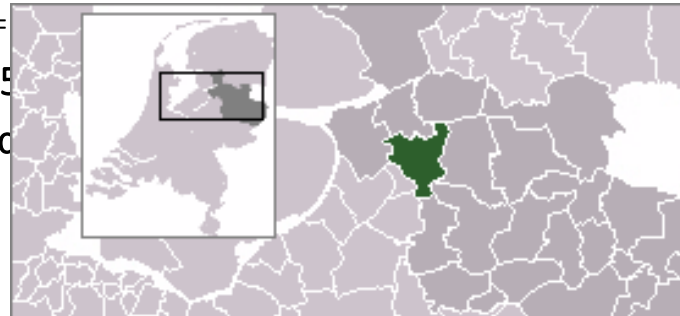
- Can the lectal contamination effects (lectal persistence) be replicated in another lectally stratified corpus?
- QLVL Twitter Corpus (Thanks Tom Ruetten, for providing the data)
- Metadata on geographic location:

<tweet user="jaapotten" norm\_loc="Zwolle, Netherlands, Zwolle, Netherlands"

rep\_loc="Zwolle" date=

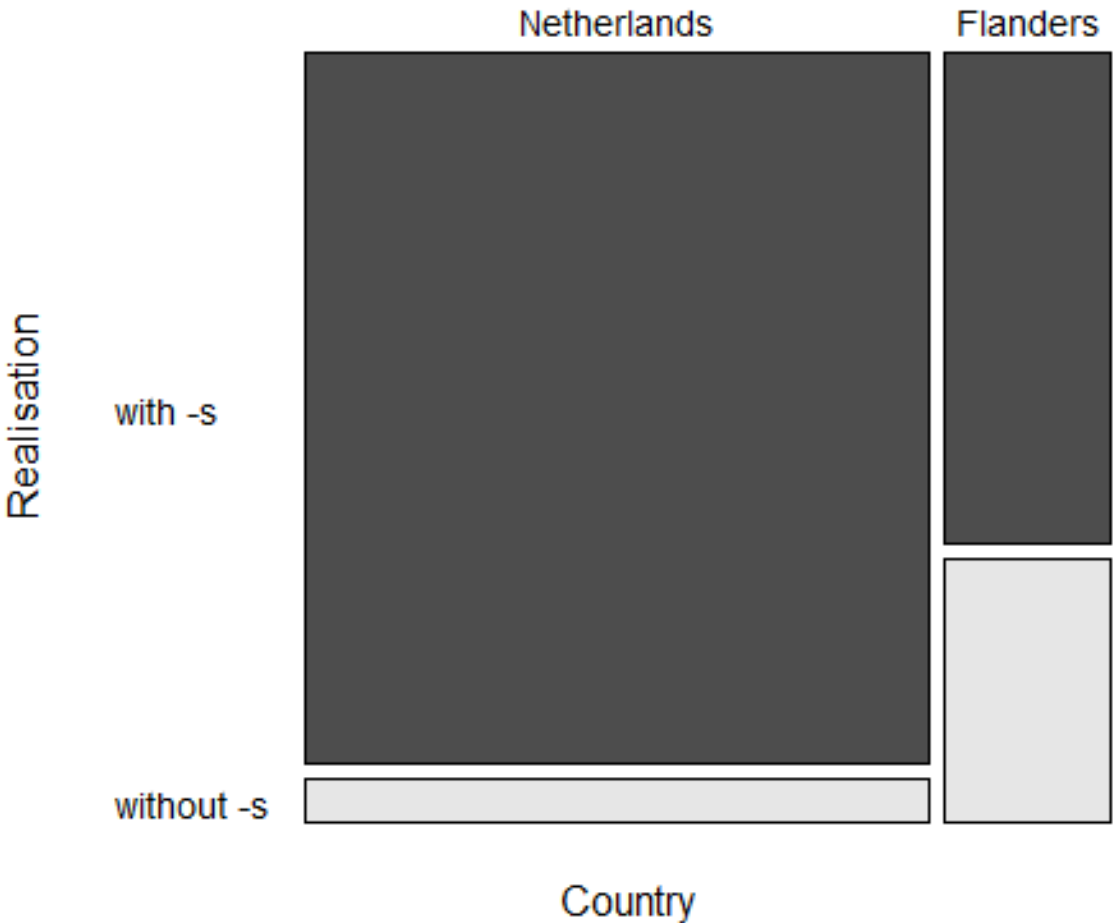
id="3126057372705095

sneeuw blijft liggen op c



ie vanavond toch afgelast,

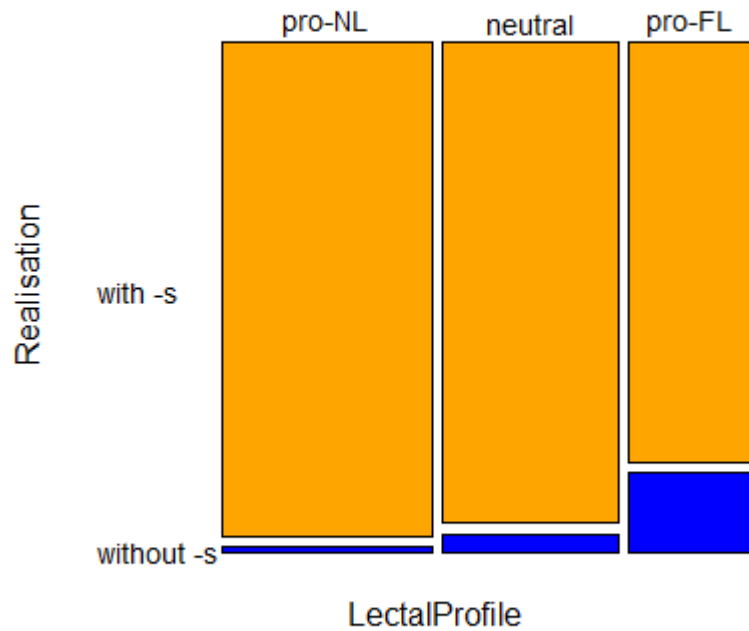
# Twitter data (n = 1299)



$\chi^2$  p < 0.001, Cramér's V 0.36

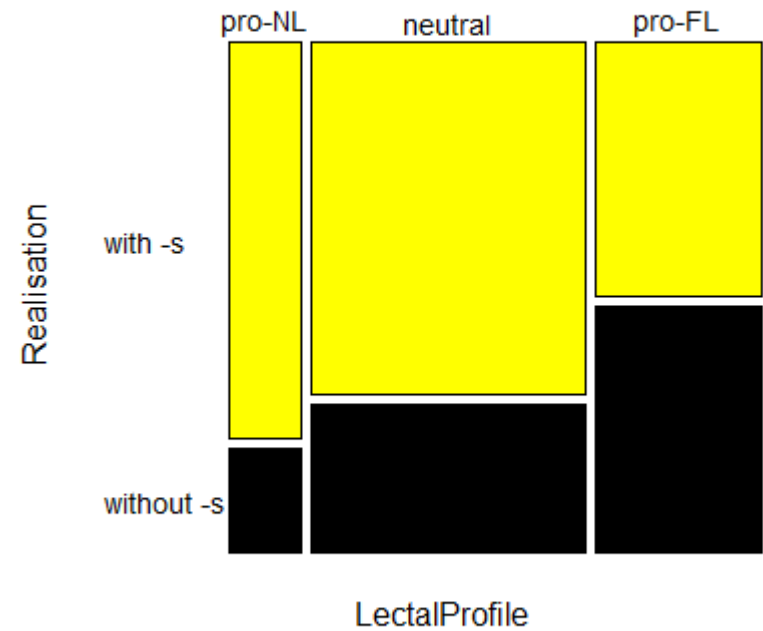
# Lexical diffusion: lectal persistence

**Netherlandic data Twitter (n = 1024)**



kendall's tau = 0.21 ( $p < 0.001$ )

**Flemish data Twitter (n = 275)**



kendall's tau = 0.21 ( $p < 0.001$ )

## Border cities (black):

- Bergen-Op-Zoom
- Bilzen
- Brasschaat
- Moelingen
- Eeklo
- Eindhoven
- Etten-Leur
- Geleen
- Heel
- Heerlen
- Knokke(-Heist)
- Lommel
- Maastricht
- Oisterwijk
- Rosendaal
- Sittard
- Tilburg
- Turnhout
- Valkenswaard
- Weert

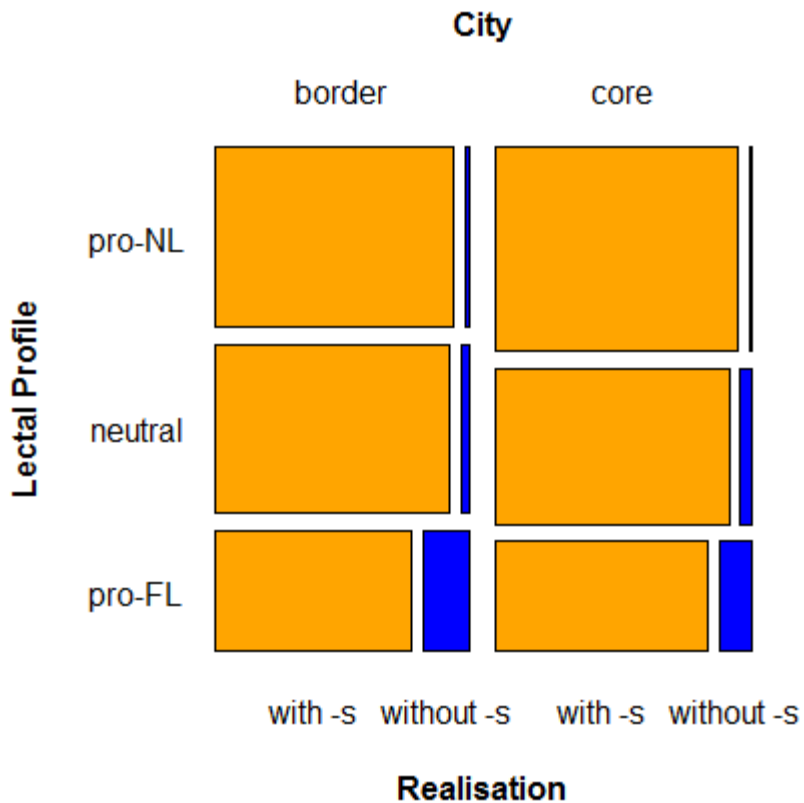
## non-border cities (grey)



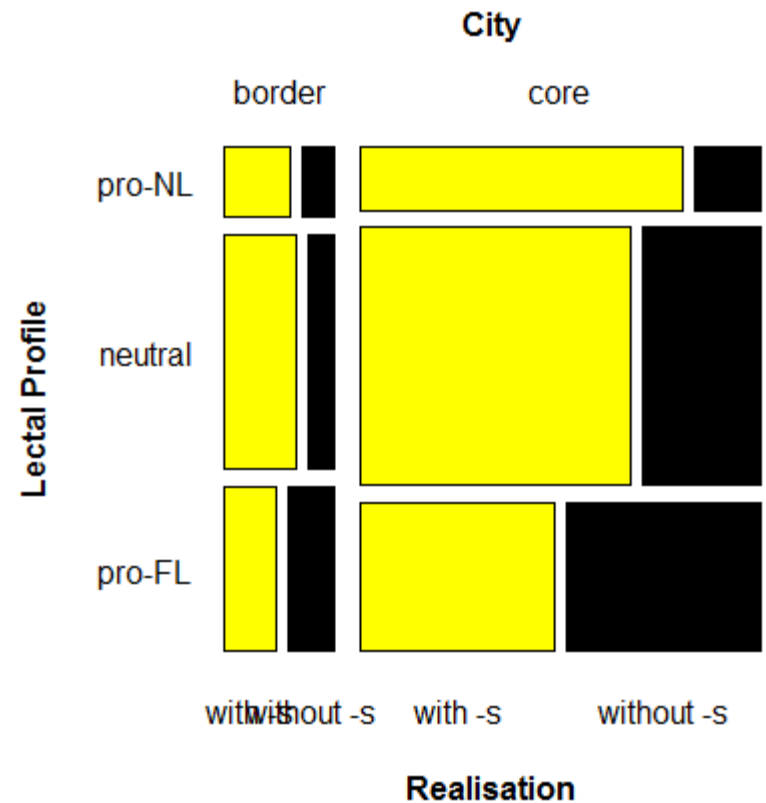
# Data Twitter (n = 1299)



## Netherlandic data Twitter (n = 1024)



## Flemish data Twitter (n = 275)



- (Flemish) border cities show less effect of lectal profile than core cities

# Conclusion

- In line with exemplar theories of language: prior use of constructions leaves a (context-rich) trail in the mind of the language users
- Similar effects reported in Delvaux & Soquet (2007), see also MacFarlane & Hay (2015:260)
- New instances of constructions show allegiance to their perceived parentage: “blood is thicker than water”: ‘Echoing’ effects of lectal provenance