

Lectal Contamination

How language-external variation becomes language-internal
through language contact

Dirk Pijpops & Freek Van de Velde

Research Foundation Flanders (FWO)
University of Leuven

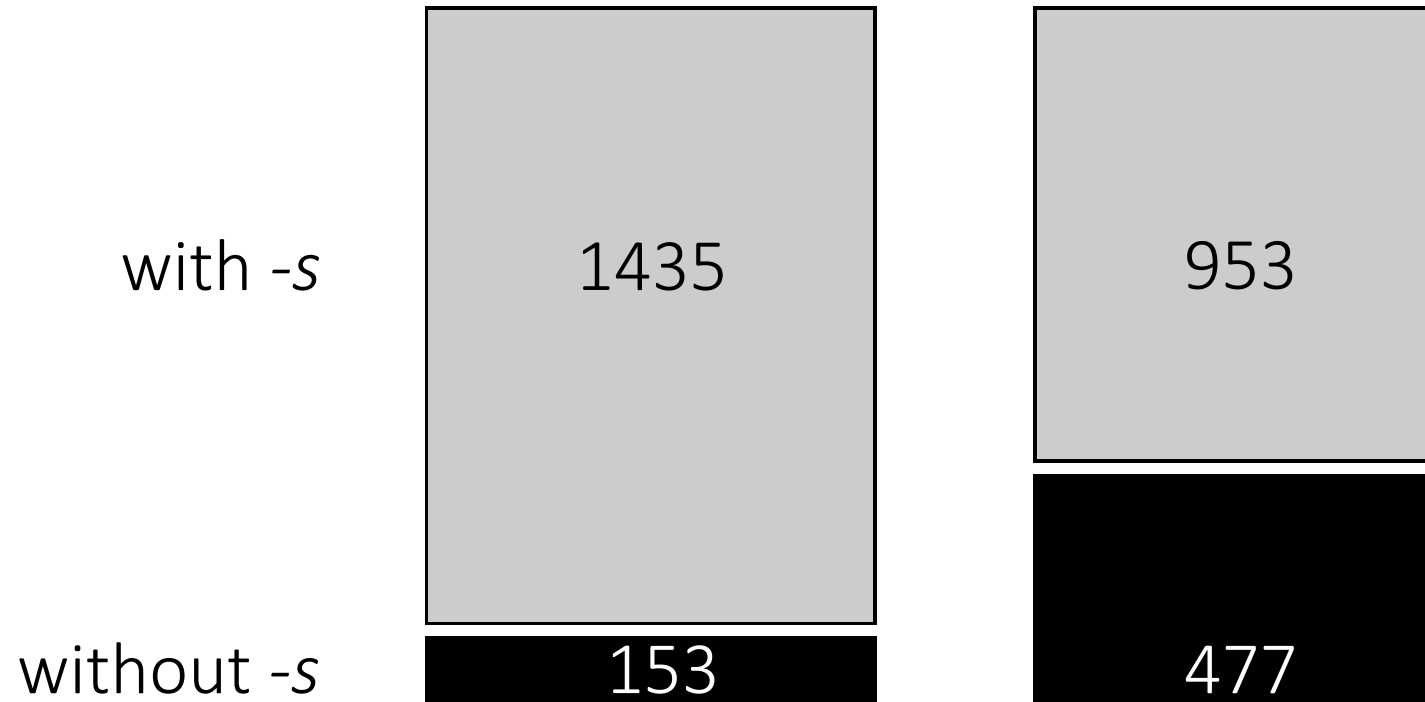
Biogenic Amine-Lectal Contamination

Dutch partitive genitive

- [Indefinite pronoun + Adjective]_{NP}
- *iets speciaal*
something special
- *wat bijzonder*
something peculiar
- *niets leuk*
nothing fun

Dutch partitive genitive

- [Indefinite pronoun + Adjective]_{NP}
- *iets speciaals*
something special
- *wat bijzonders*
something peculiar
- *niets leuks*
nothing fun

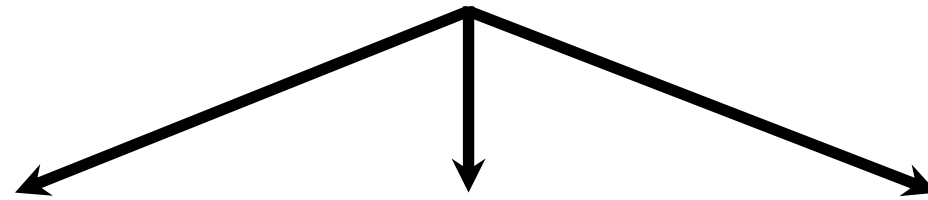


Other factors:

- Type of adjective
- Register
- Pronoun
- Frequency

$p < 0.0001$, Cramér's $V = 0.29$

143 partitive genitive phrases



Typically Netherlandic
phrases

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

...

Neutral phrases

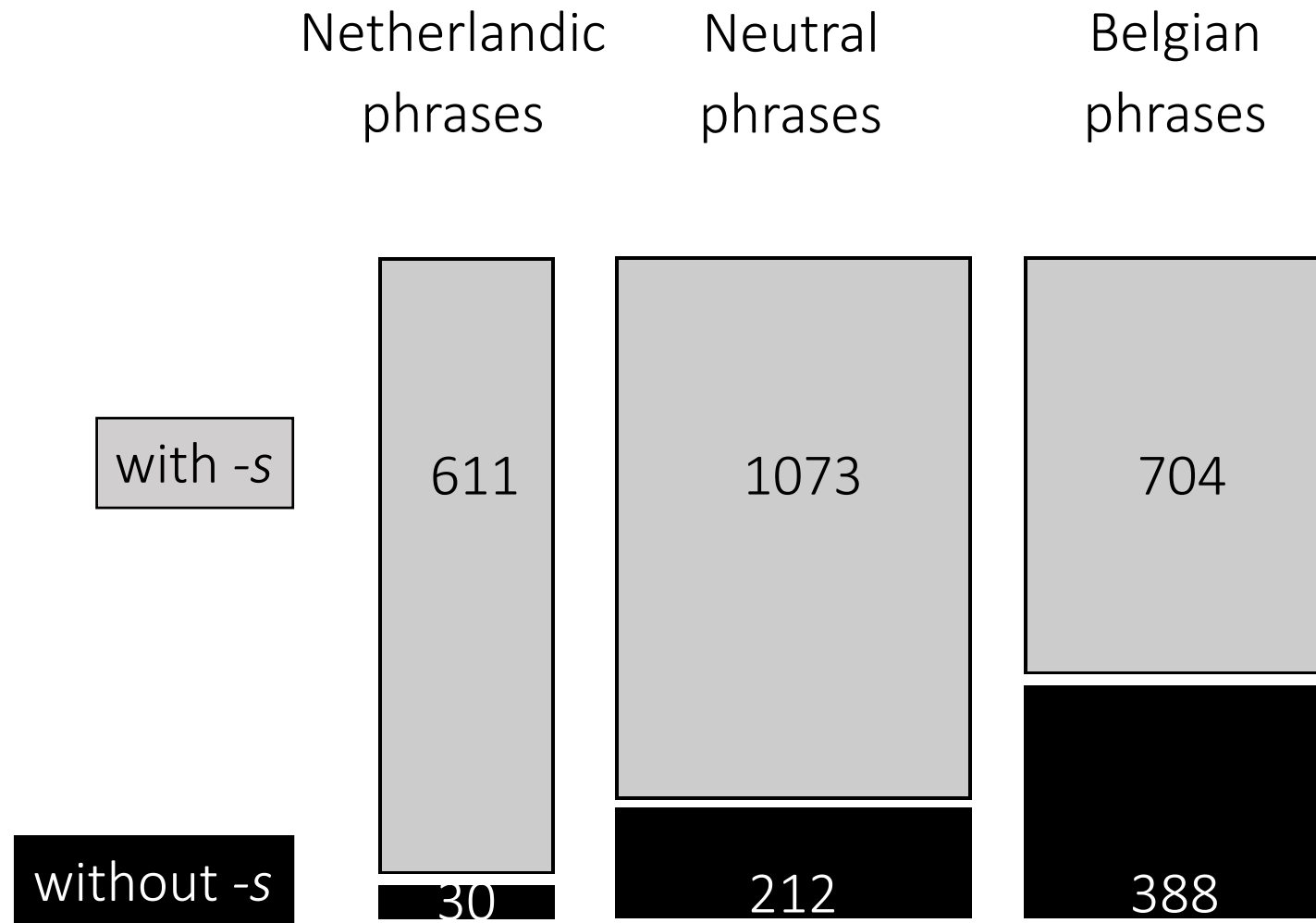
weinig concreet(s)
iets zinnig(s)
iets spannend(s)
niets erg(s)

...

Typically Belgian
phrases

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

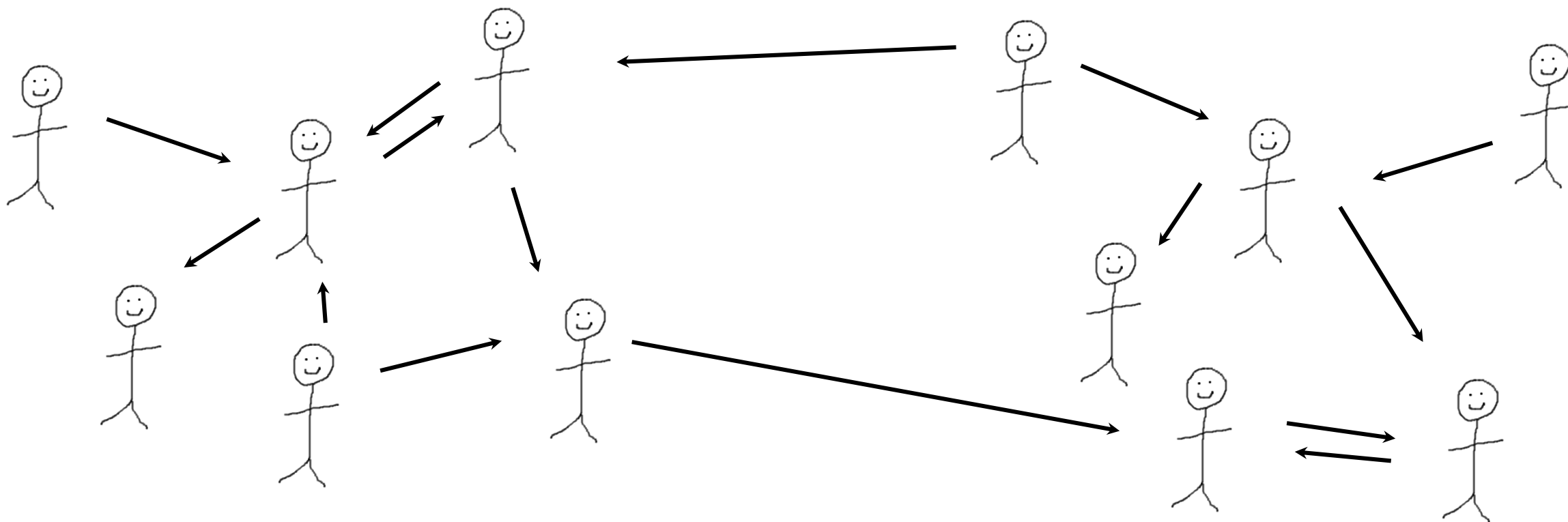
...



$p < 0.0001$, kendall's $\tau = 0.28$

Lectal contamination

1. Lectal difference causes lexically-specific effect
2. Lectal contact causes lexically-specific input



Morphological preference: variant *+s*
Lexical preference: *iets bijzonder(s)*

Morphological preference: variant *+∅*
Lexical preference: *iets speciaal(s)*



Hears:

iets speciaal

+∅

>

iets bijzonder

+∅

1. Lectal difference causes lexically-specific effect
2. Lectal contact causes lexically-specific input
3. Usage-based theory predicts that lexically-specific input causes lexically-specific output



Hears:

iets speciaal

>

iets bijzonder

+∅

+∅

Produces:

$P(+\emptyset \mid \textit{iets speciaal}(s))$

>

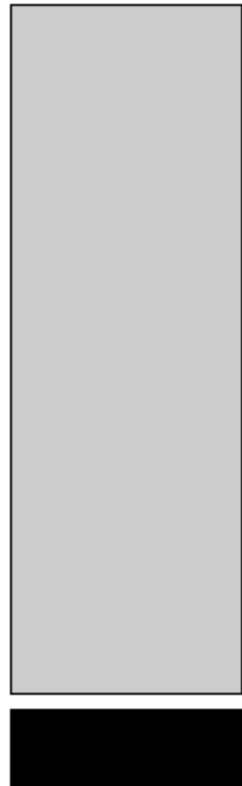
$P(+\emptyset \mid \textit{iets bijzonder}(s))$

1. Lectal difference causes lexically-specific effect
2. Lectal contact causes lexically-specific input
3. Usage-based theory predicts that lexically-specific input causes lexically-specific output
4. Even within a single lect, lexical preferences echo lectal differences



Netherlandic
phrases

Belgian
phrases



with -s

without -s

Netherlandic
phrases

Belgian
phrases

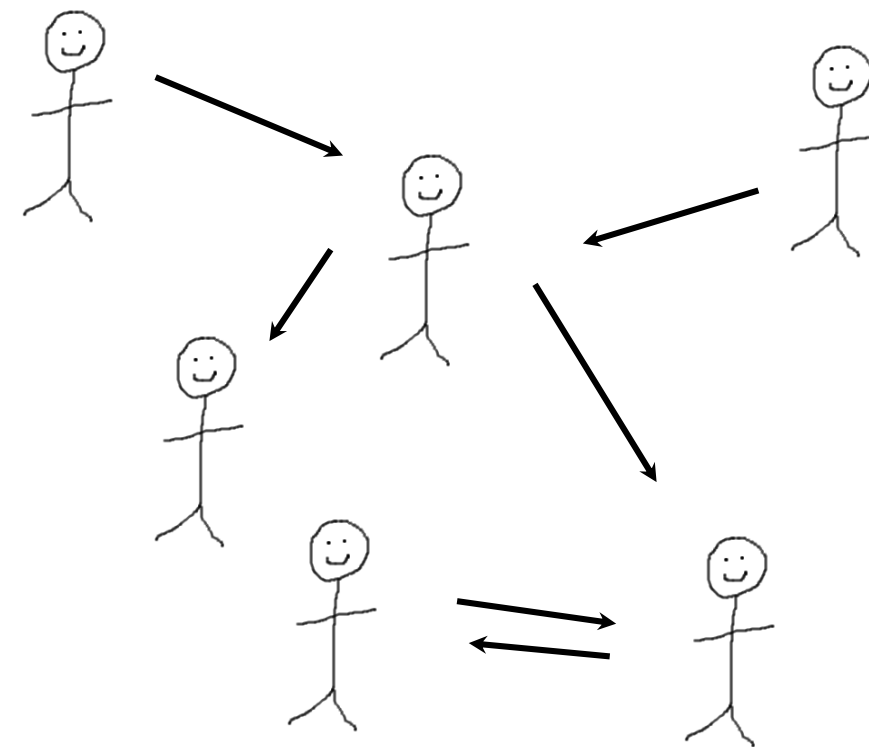
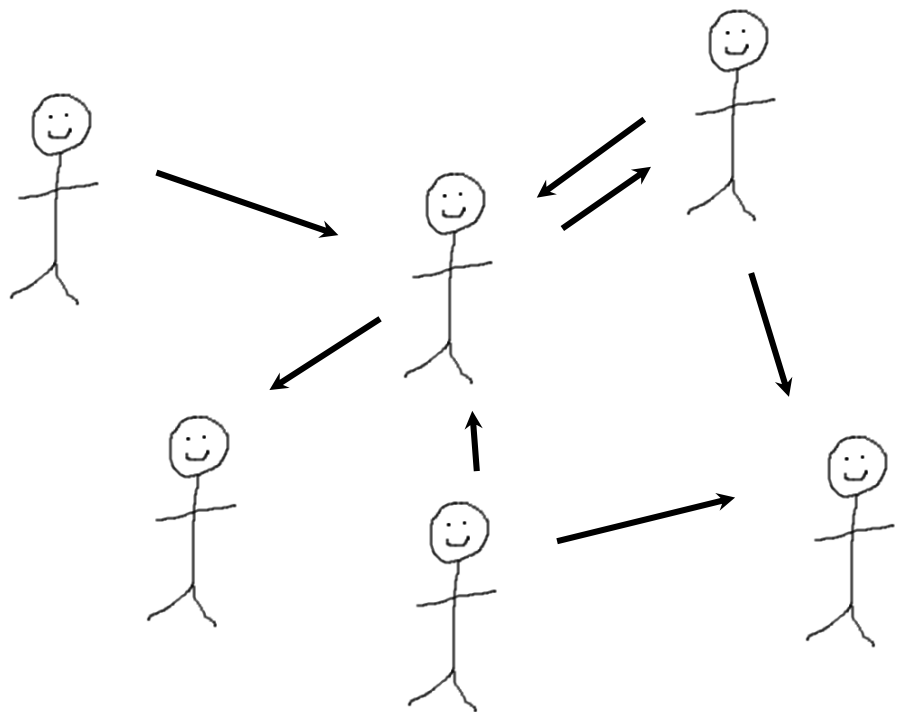


Lectal
Contami-
nation }

Lectal
Contami-
nation {

1. **Lectal difference** causes lexically-specific effect
2. **Lectal contact** causes lexically-specific input
3. **Usage-based theory** predicts that lexically-specific input causes lexically-specific output
4. **Lectal Contamination**: even within a single lect, lexical preferences echo lectal differences

Computer simulation



iets bijzonders: 694
iets bijzonder: 4
iets speciaals: 249
iets speciaal: 50

iets bijzonders



iets bijzonders: 694+1

iets bijzonder: 4

iets speciaals: 249

iets speciaal: 50

iets bijzonders



iets bijzonders: 695

iets bijzonder: 4

iets speciaals: 249

iets speciaal: 50

iets speciaals



iets bijzonders: 695

iets bijzonder: 4

iets speciaals: 249+1

iets speciaal: 50

iets speciaals



iets bijzonders: 695
iets bijzonder: 4
iets speciaals: 250
iets speciaal: 50

iets bijzonder

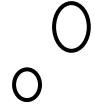


iets bijzonders: 695
iets bijzonder: 4+1
iets speciaals: 250
iets speciaal: 50

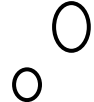
iets bijzonder



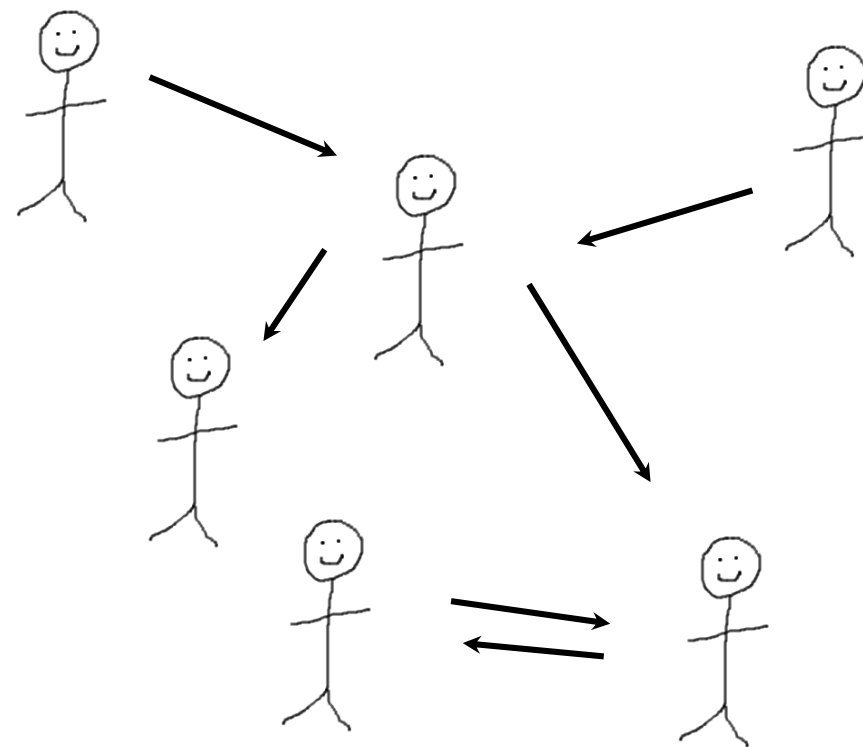
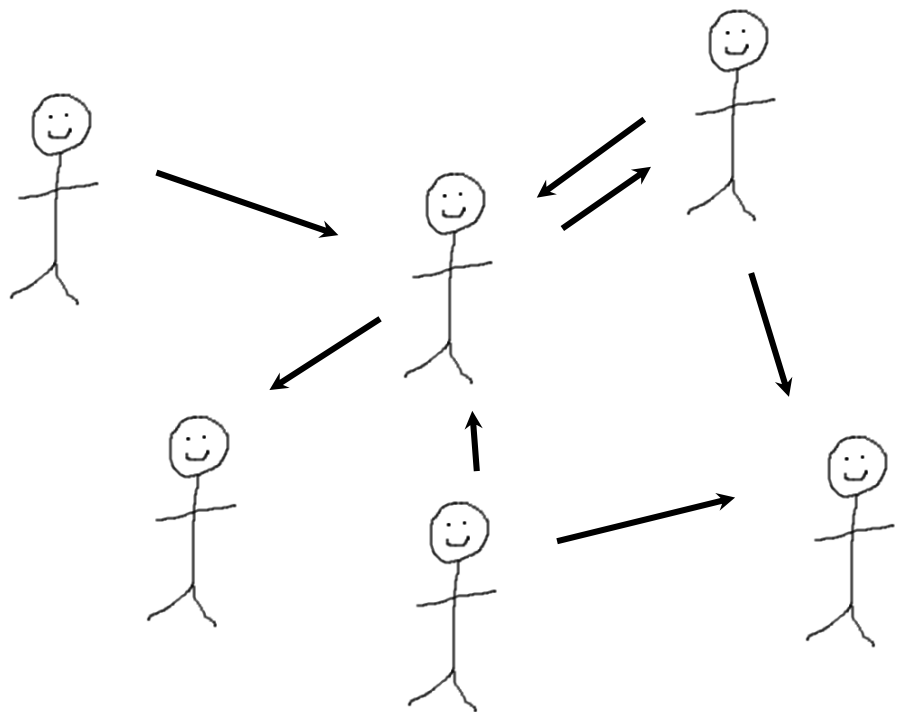
iets bijzonders: 695
iets bijzonder: 5
iets speciaals: 250
iets speciaal: 50



iets bijzonders: 695
iets bijzonder: 5
iets speciaals: 250
iets speciaal: 50



iets bijzonders: 69.5%
iets bijzonder: 0.5%
iets speciaals: 25%
iets speciaal: 5%



Initial Netherlandic memory

iets bijzonders: 80

iets bijzonder: 0

iets speciaals: 20

iets speciaal: 0

with -s

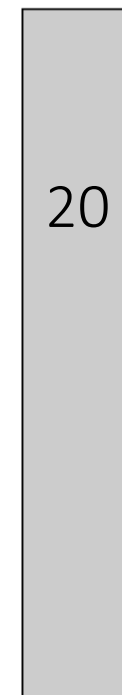
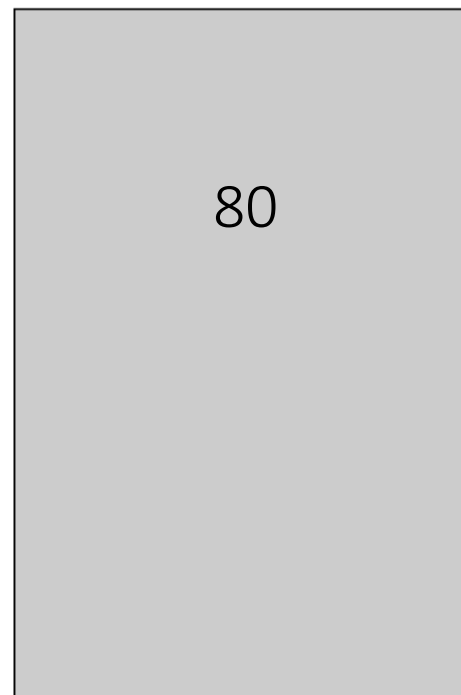
without -s

iets bijzonder(s)

iets speciaal(s)

80

20



Initial Belgian memory

iets bijzonders: 12

iets bijzonder: 8

iets speciaals: 48

iets speciaal: 32

with -s

without -s

iets bijzonder(s)

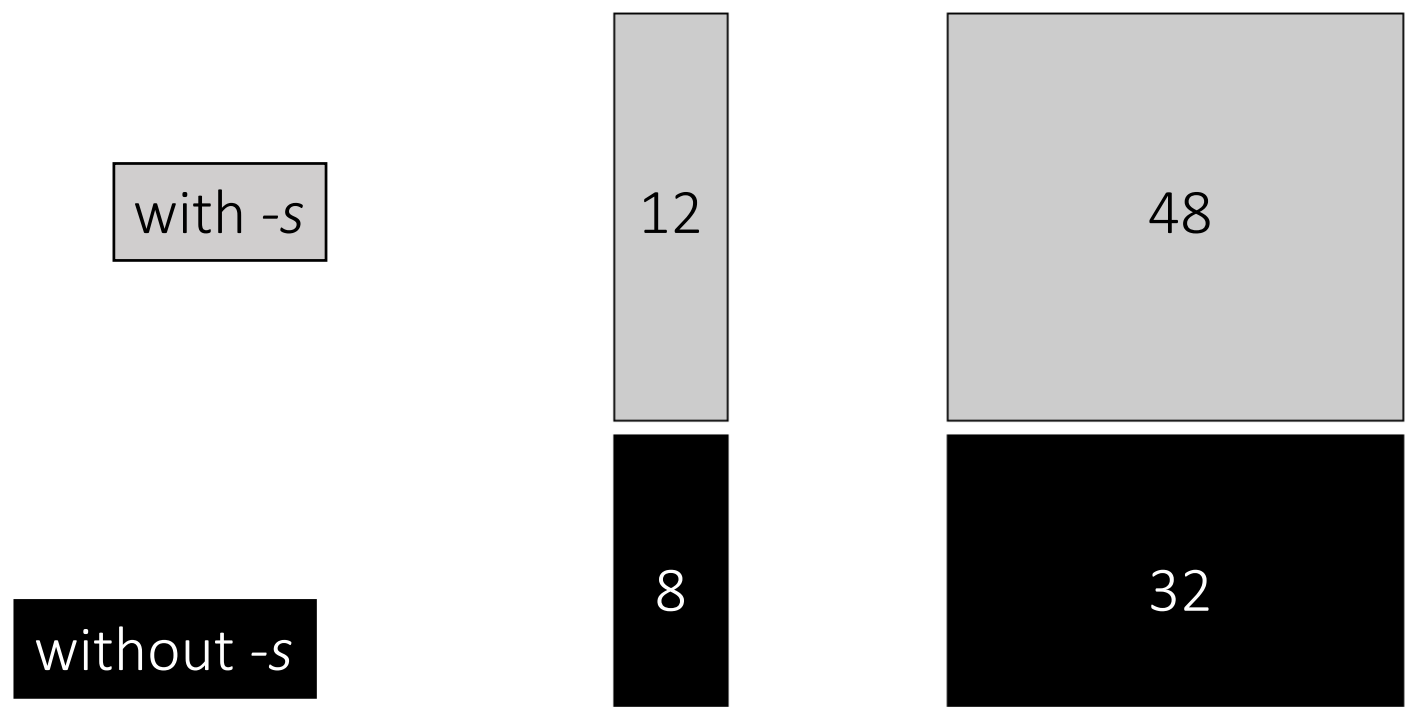
iets speciaal(s)

12

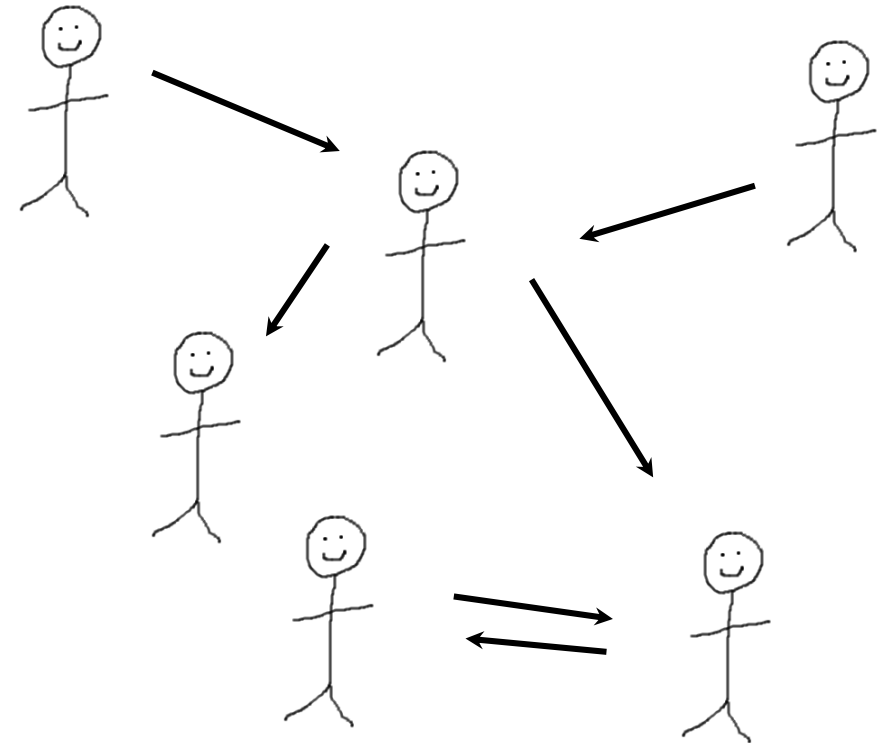
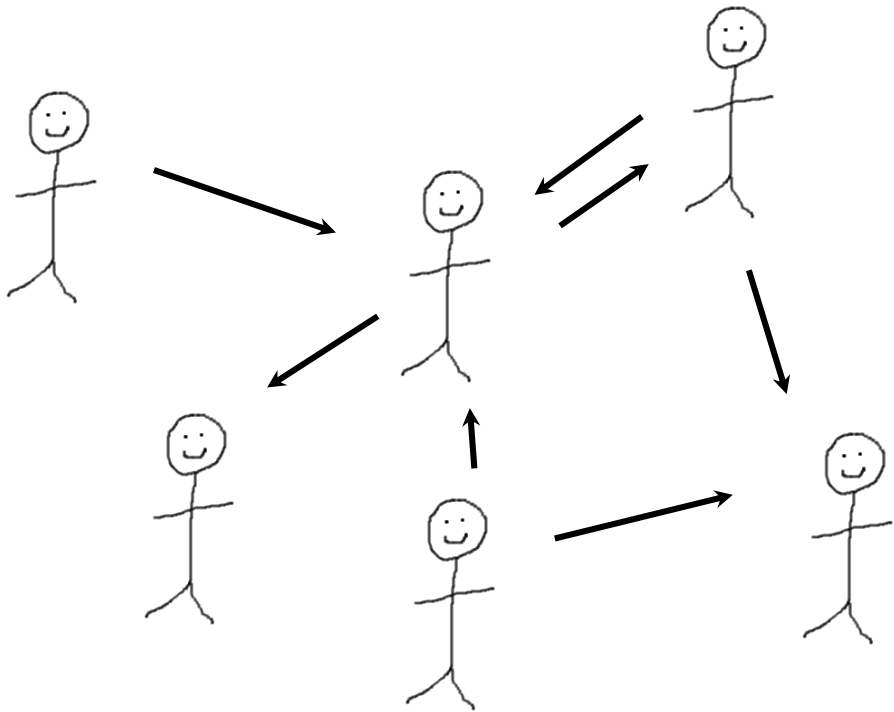
8

48

32



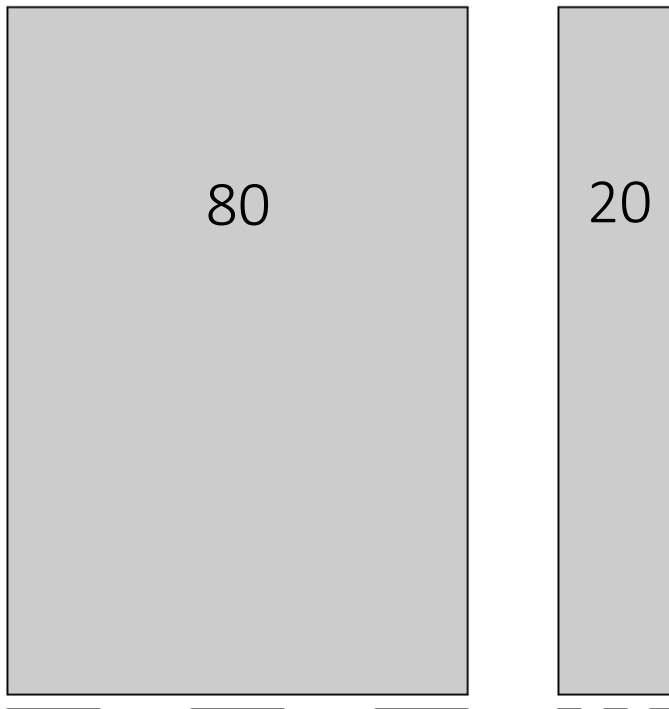
Without language contact





- 100 agents: 50 Netherlandics, 50 Belgians
- 10 series
- 100.000.000 interactions (1 million PIT)
- Record every 10.000 PIT
- 0 language contact

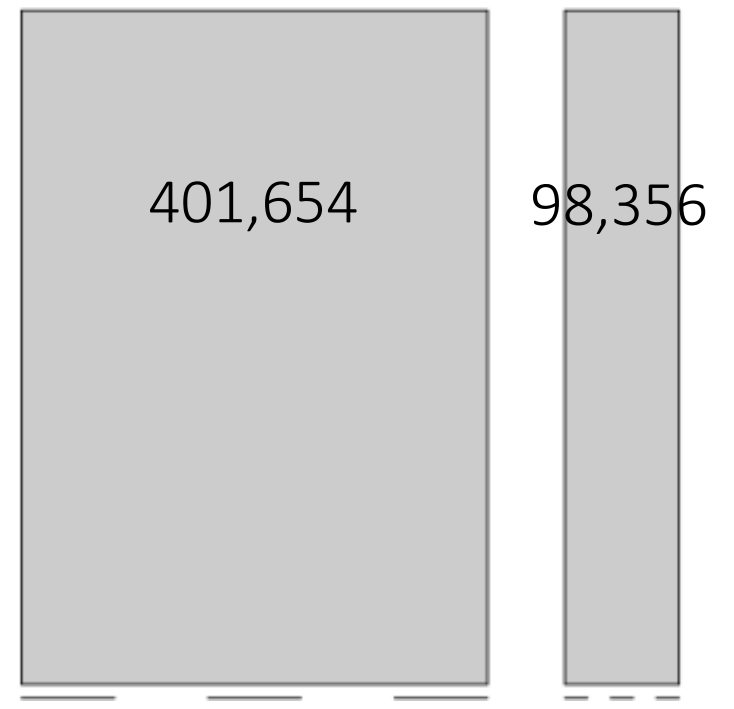
iets bijzonder(s) iets speciaal(s)



with -s

without -s

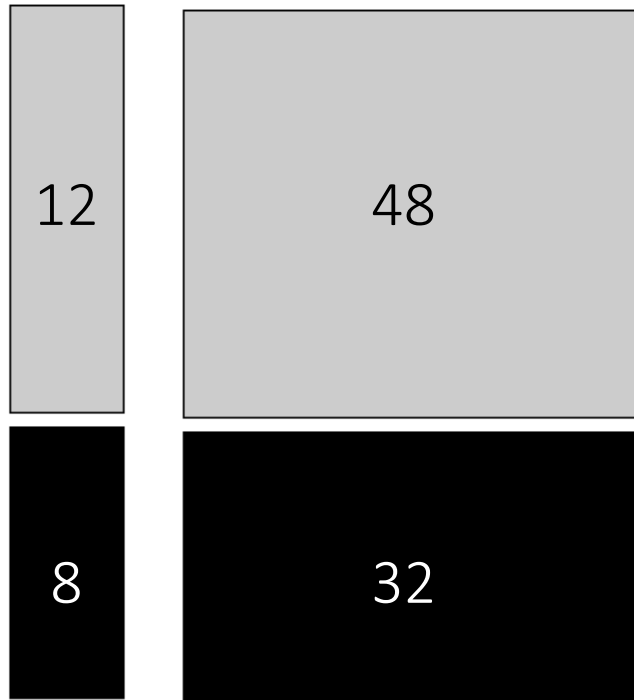
iets bijzonder(s) iets speciaal(s)



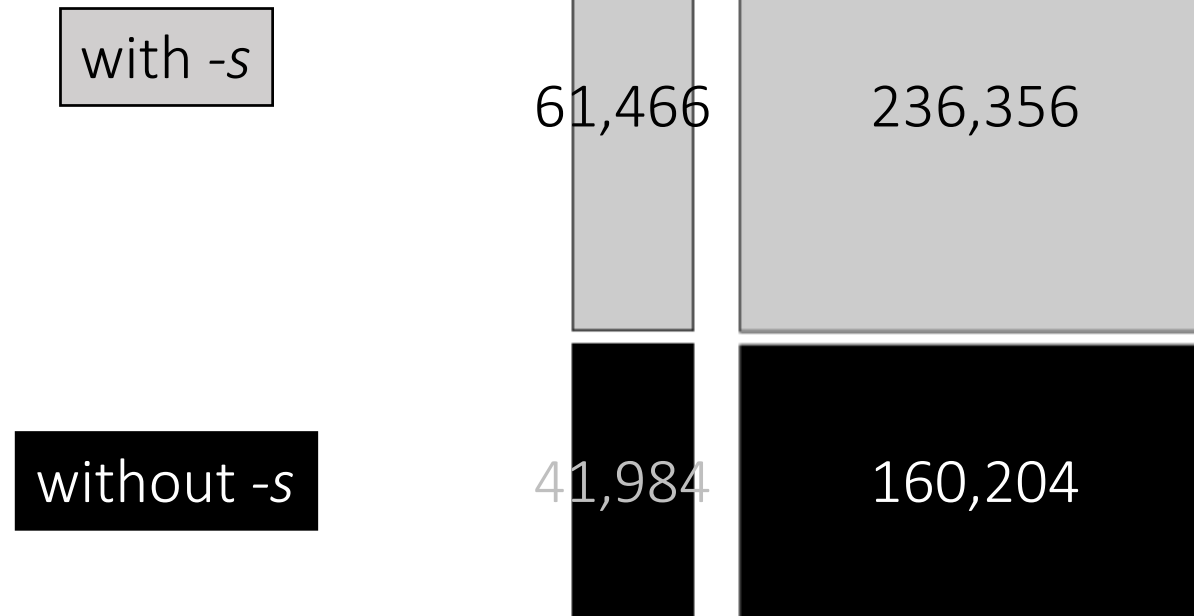


- 100 agents: 50 Netherlandics, 50 Belgians
- 10 series
- 100.000.000 interactions (1 million PIT)
- Record every 10.000 PIT
- 0 language contact

iets bijzonder(s) iets speciaal(s)



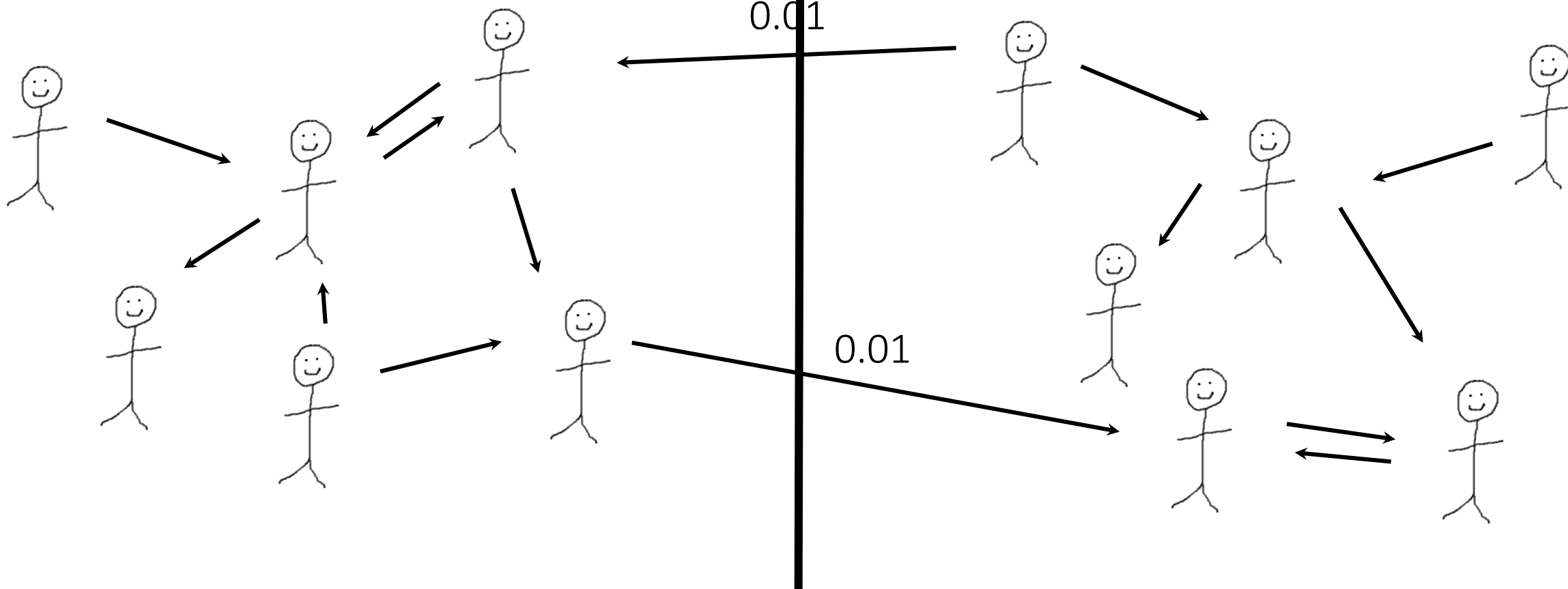
iets bijzonder(s) iets speciaal(s)



No language contact

=

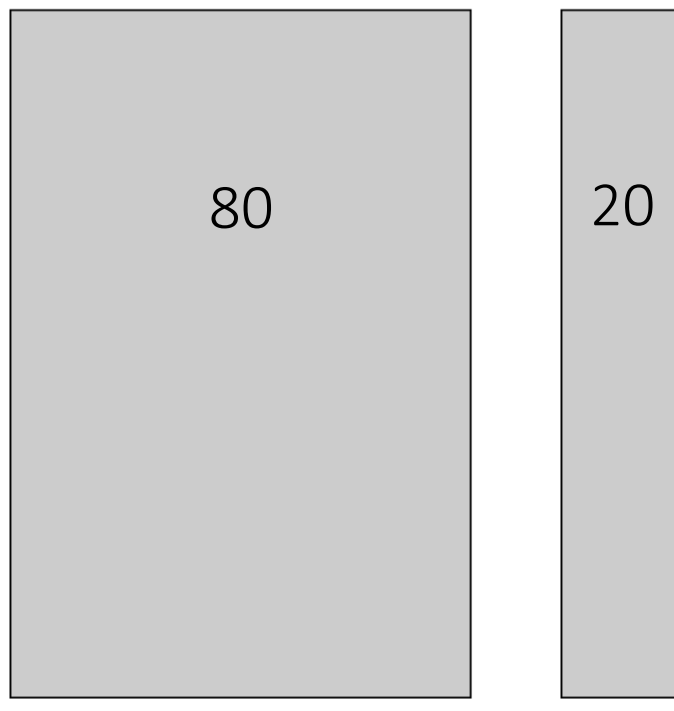
No lectal contamination





- 100 agents: 50 Netherlandics, 50 Belgians
- 10 series
- 100.000.000 interactions (1 million PIT)
- Record every 10.000 PIT
- 0.01 language contact

iets bijzonder(s) iets speciaal(s)



with -s

without -s

iets bijzonder(s) iets speciaal(s)



3,563

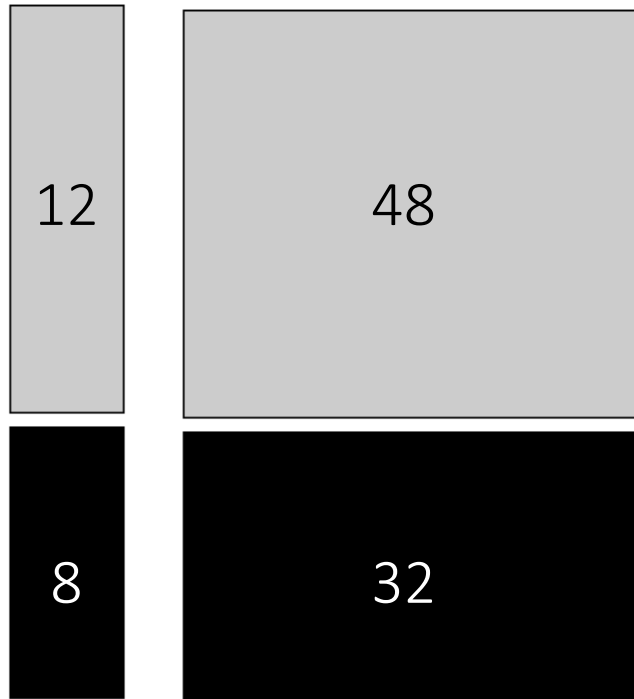
12,565

Lectal
Contami-
nation

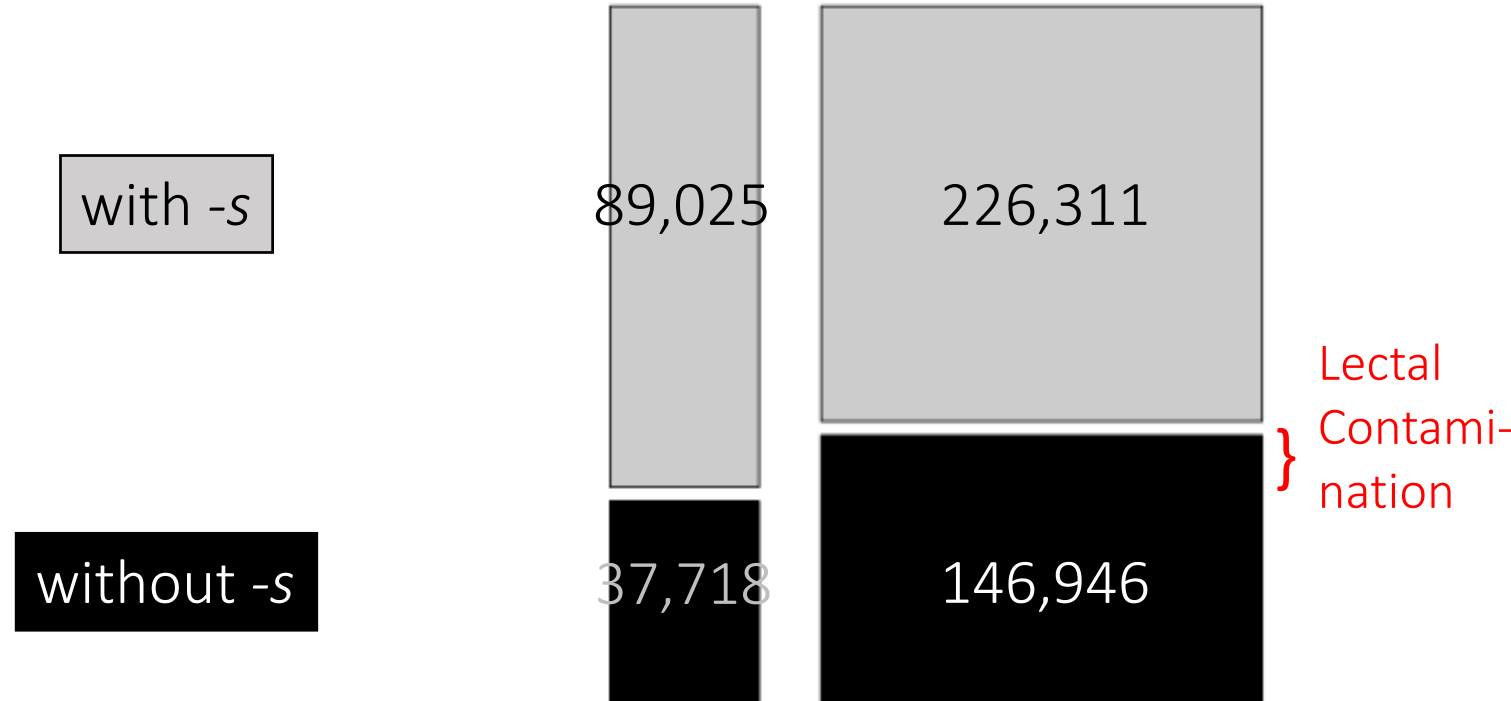


- 100 agents: 50 Netherlandics, 50 Belgians
- 10 series
- 100.000.000 interactions (1 million PIT)
- Record every 10.000 PIT
- 0.01 language contact

iets bijzonder(s) iets speciaal(s)



iets bijzonder(s) iets speciaal(s)



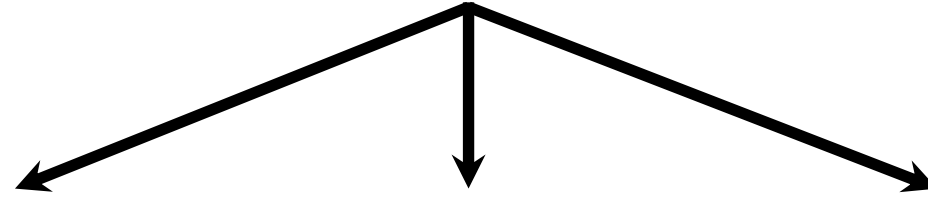
- **100 agents: 50 Netherlandics, 50 Belgians**
- 10 series
- 100.000.000 interactions (1 million PIT)
- Record every 10.000 PIT
- 0.01 language contact

1. Lectal difference
2. Lectal contact
3. Usage-based theory
4. Lectal Contamination

Observational data

- ConDiv corpus
- Indefinite pronouns: *iets* 'something', *niets* 'nothing', *wat* 'something', *veel* 'a lot', *weinig* 'few', *zoveel* 'so much'
- Adjectives: *aardig* 'nice', *apart* 'apart', *belangrijk* 'important', *beter* 'better', *bijzonder* 'particular', *blauw* 'blue', *concreet* 'concrete', *deftig* 'decent', *dergelijk* 'similar', *erg* 'awful', *geel* 'yellow', *gek* 'crazy', *goed* 'good', *groen* 'green', *interessant* 'interesting', *klein* 'small', *lekker* 'tasty', *leuk* 'fun', *mooi* 'beautiful', *nieuw* 'new', *nuttig* 'useful', *oranje* 'orange', *positief* 'positive', *purper* 'purple', *raar* 'weird', *rood* 'red', *spannend* 'exciting', *speciaal* 'special', *verkeerd* 'wrong', *verschrikkelijk* 'horrible', *vreemd* 'weird', *warm* 'warm', *wit* 'white', *zinnig* 'sensible', *zwart* 'black'
- 143 unique phrases
- Manual checking
- 3018 instances, 2388 with -s, 630 without -s

143 phrases



Typically Netherlandic
phrases

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

...

Neutral phrases

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

...

Typically Belgian
phrases

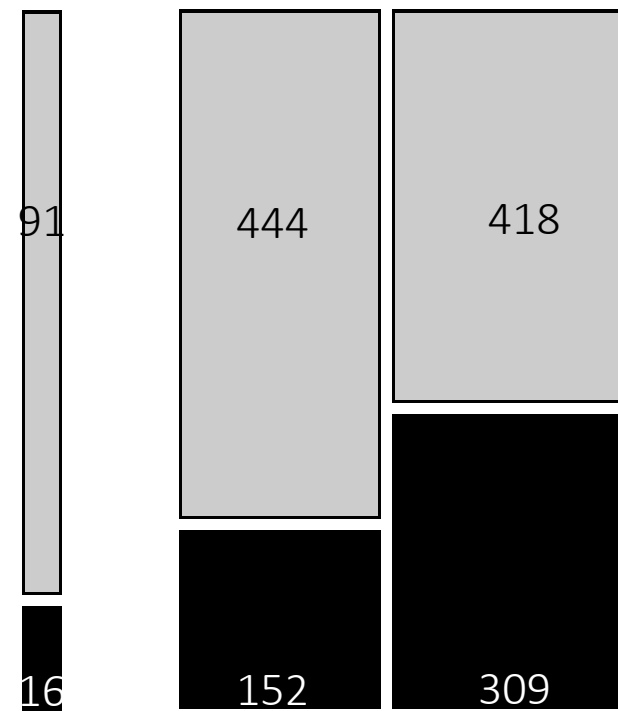
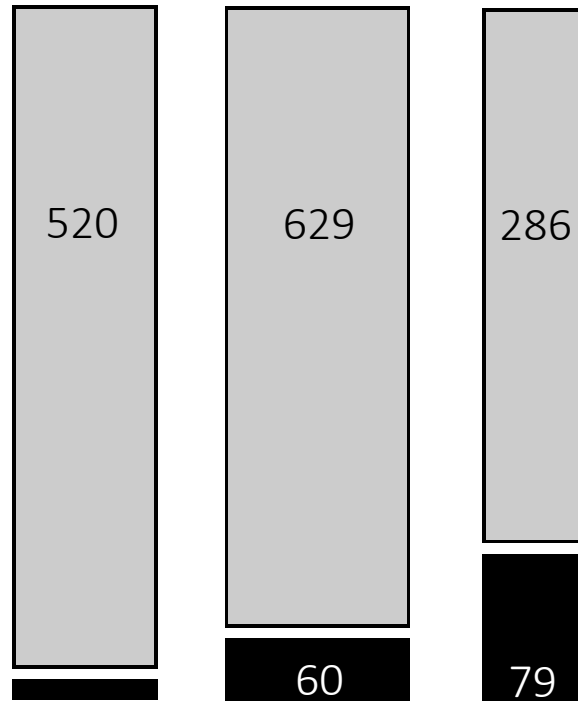
iets interessant(s)
niets speciaal(s)
iets deftig(s)
iets raar(s)

...



Netherlandic phrases Neutral phrases Belgian phrases

Netherlandic phrases Neutral phrases Belgian phrases



$p < 0.0001$, kendall's $\tau = 0.22$

$p < 0.0001$, kendall's $\tau = 0.20$

Mixed effects logistic regression

*-s absence ~ AdjectiveType + Country + Register + Pronoun + Frequency
+ 1|Phrase*

+ LectalProfile + Country:LectalProfile

-



Netherlandic
phrases neutral
phrases Belgian
phrases



Netherlandic
phrases neutral
phrases Belgian
phrases

Replication

- Twitter corpus (compiled by Tom Ruetten)
- Same pronouns & adjectives
- 1299 instances, 1142 with -s, 157 without -s

Why is lectal contamination important?

- Simple idea, and therefore dangerous
 - Not a special case of language variation
 - Not a special case of language contact
 - Base assumptions of usage-based theory
- Deeply language-internal effect of a mere difference between lects
 - Cannot study lects in isolation
 - Language of the Netherlands is influenced by Belgian language use
 - Need the study of language contact

How to deny lectal contamination?

- This specific case study
 - Belgians have no prestige
 - Too little language contact
 - More Dutch-speaking Netherlandics than Dutch-speaking Belgians

How to deny lectal contamination?

- In general
 1. Lectal difference causes lexically-specific effect
 2. Lectal contact causes lexically-specific input
 3. Usage-based theory predicts that lexically-specific input causes lexically-specific output: Dabrowska (2012,2014); Ferreira et al. (2002); Bod (2006); Ferreira & Patson (2007); van den Bosch & Daelemans (2013)
 4. Even within a single lect, lexical preferences echo lectal differences

Want to hear more?

- Come have a chat with me
- Handout
- Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2). 543–581.
- Pijpops, Dirk, Isabeau De Smet & Freek Van de Velde. Constructional contamination in morphology and syntax. Four case studies. Forthcoming in *Constructions and Frames*.

Thanks!

References

- Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2013. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.4.
- Bod, Rens. 2006. Exemplar-Based Syntax: How to Get Productivity from Examples. *Linguistic Review* 23(3). 291–320.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Broekhuis, Hans. 2013. *Syntax of Dutch. Adjectives and Adjective Phrases*. Amsterdam: Amsterdam University Press.
- Chang, Winston. 2014. *extrafont: Tools for using fonts*.
- Dąbrowska, Ewa. 2012. Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253.
- Dąbrowska, Ewa. 2014. Recycling utterances: A speaker’s guide to sentence processing. *Cognitive Linguistics* 25(4). 617–653.
- Ferreira, Fernanda and Nikole Patson. 2007. The “good enough” approach to language comprehension. *Language and Linguistics Compass* 1. 71–83.
- Ferreira, Fernanda, Karl Bailey and Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11(1). 11–15.
- Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth and Steve Taylor. 2016. *Effect Displays for Linear, Generalized Linear, and Other Models*. R package version 3.2.
- Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede and Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]. *Nederlandse Taalkunde* 5(4). 356–363.
- Harrell, Frank. 2013. *rms: Regression Modeling Strategies*. R package version 4.0-0.
- Pijpops, Dirk and 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*. Published online, ahead of print.
- Pijpops, Dirk and Freek Van de Velde. 2015. Ethnolect speakers and Dutch partitive adjectival inflection. A corpus analysis. *Taal en Tongval* 67(2). 343–371.
- Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2). 543–581.
- Pijpops, Dirk and Katrien Beuls. 2015. Agent-gebaseerde modellering in de historische taalkunde. Een model van regularisatiedruk op de Nederlandse werkwoorden. *Handelingen der Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 69. 5–23.
- Pijpops, Dirk, Katrien Beuls and Freek Van de Velde. 2015. The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational linguistics in the Netherlands Journal* 5. 81–102.
- Speelman, Dirk, Kris Heylen and Dirk Geeraerts. 2018. *Mixed-Effects Regression Models in Linguistics*. Cham: Springer.
- van den Bosch, Antal and Walter Daelemans. 2013. Implicit Schemata and Categories in Memory-based Language Processing. *Language and Speech* 56(3). SAGE Publications Ltd. 309–328.
- van der Horst, Joop. 2008. *Geschiedenis van de Nederlandse syntaxis [History of Dutch syntax]*. Leuven: Universitaire Pers Leuven.

Acknowledgments

We owe thanks to Tom Ruetten for providing us with his Twitter corpus