

Lectal contamination. How language-external variation becomes language-internal through language contact

Dirk Pijpops & Freek Van de Velde

QLVL, University of Leuven
Research Foundation Flanders (FWO)

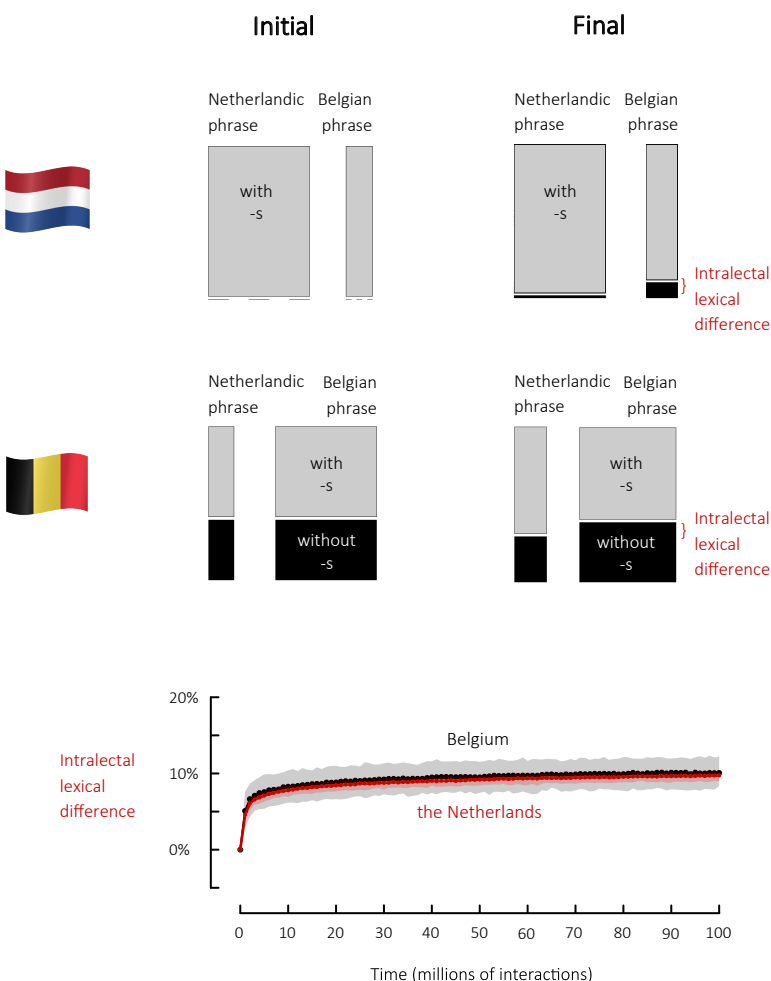
THEORY

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. **Intralectal lexical differences** echo lectal differences

CASE STUDY

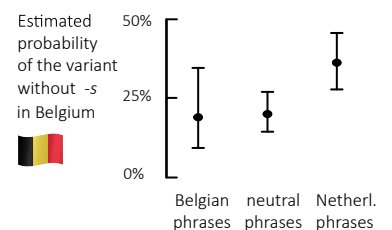
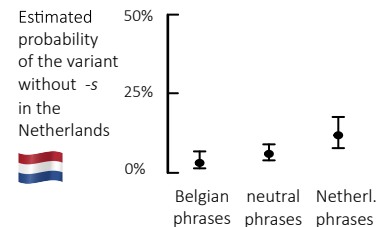
- Partitive genitive: [indefinite pronoun + adjective (-s)]_{NP}
- *iets bijzonder(s)* 'something peculiar', *niets speciaal(s)* 'nothing special', *wat leuk(s)* 'something fun'
- The Netherlands prefers with -s, Belgium without -s
- Other factors determining variation: type of adjective, register, choice of pronoun, frequency of the phrase

SIMULATION



REAL WORLD

Controlling for all other factors, and random lexical preference



IMPORTANCE

- Lexical diffusion of lectal difference
- Cannot study lects in isolation
- Need study of language contact



Dirk Pijpops holds Master's degrees in AI and Linguistics and is pursuing a PhD at the University of Leuven on argument realization, supervised by prof. Dirk Speelman. dirk.pijpops@kuleuven.be



Pijpops, Dirk and Freek Van de Velde. 2018. **Lectal contamination: How language-external variation becomes language-internal through language contact.** *Variationist Linguistics meets Contact Linguistics*. May 21, Ascona.

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.



Simulation Design

The simulated agents are divided into two social groups. For each interaction, a speaker and hearer agent are selected, with there being a 99% chance that speaker and hearer stem from the same social group. During an interaction, the speaker may utter one of four forms to the hearer: *iets bijzonders*, *iets bijzonder*, *iets speciaals* and *iets speciaal*. Each agent maintains a memory of these forms and a counter of how often it has heard each form. The probability with which a speaker utters a form is directly proportional to these counts in its memory. The initial memories of the agents of each social group are chosen such that (i) one group starts with a lexical preference for the phrase *iets bijzonder(s)* and a relative morphological preference for the variant with *-s* ending, while the other group starts with a lexical preference for the phrase *iets speciaal(s)* and a relative morphological preference for the variant without *-s* ending; and (ii) these lexical and morphological preferences are fully independent of each other. The initial memory of all agents of one group is *iets bijzonders*: 80, *iets bijzonder*: 0, *iets speciaals*: 20, *iets speciaal*: 0. The initial memory of all agents of the other group is *iets bijzonders*: 12, *iets bijzonder*: 8, *iets speciaals*: 48, *iets speciaal*: 32. The graph shows the results of 10 runs of 100 million interactions of 100 agents. We find that intralectal lexical differences arise as a result of contact between the groups.

Real world data

From the ConDiv corpus (Grondelaers et al. 2000), all instances were extracted in which one of the pronouns *iets* 'something', *niets* 'nothing', *veel* 'a lot', *wat* 'something', *weinig* 'few' and *zoveel* 'so much' preceded one of 15 adjectives with or without *-s* ending. As for why these pronouns and adjectives were selected, see Pijpops & Van de Velde (2014: 9-10). The dataset was manually checked, which yielded 3018 genuine occurrences of partitive genitives, of which 2388 with *-s* ending, and 630 without *-s* ending. In a replication study, instances were extracted from the QLVL Twitter corpus, which contains Tweets gathered by Dr. Tom Ruetten between 2012 and 2013. The extraction and verification of the data followed the same procedure, yielding 1299 occurrences in total (1142 with *-s* ending 157 without *-s* ending). The results of the replication study confirmed those reported here.

Acknowledgments

We owe thanks to Tom Ruetten for providing the Twitter data. We also thank Hendrik De Smet, Lauren Fonteyn, and Hubert Cuyckens for insightful discussions and useful comments on earlier work on lectal and constructional contamination. In addition, we are grateful to Kris Heylen, Dirk Speelman and Eline Zenner for methodological assistance in the observational study, and Katrien Beuls for sharing important insights in agent-based simulation.

References

- Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2013. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-1.
- Bod, Rens. 2006. Exemplar-Based Syntax: How to Get Productivity from Examples. *Linguistic Inquiry* 37(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. *extralabel: Tools for using foies*.
- Dobrow, Ewa. 2012. Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguistic Approaches to Bilingualism* 2(1), 219-253.
- Dobrow, Ewa. 2014. Repeating utterances: A speaker's guide to sentence processing. *Cognition* 130(4), 617-653.
- Ferreira, Fernanda and Nikole Patson. 2007. The "good enough" approach to language comprehension. *Language and Linguistics Theory* 25, 171-183.
- Ferreira, Fernanda, Karl Bailey and Victoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11(1), 11-15.
- Fox, John, Sanford Weidong, Michael Friendly, Jingnan 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 Deygen, Hilde Van Allen, Vicky Van den Heede and Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpora of written Dutch]. *Nederlandse Taalkunde* 5(4), 356-363.
- Harrell, Frank. 2012. *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 regulatiedruk 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 Cleeremans. 2014. *Mind-Effects Regression Models in Linguistics*. Cham: Springer.
- van den Bosch, Arnt and Walter Dierken. 2011. *Impact: 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.