THE USE OF *ZO'N* VERSUS *ZULKE* 'SUCH' IN BELGIAN AND NETHERLANDIC DUTCH

Testing hypotheses relating to lexical biases, function, register and noun type

Dirk Pijpops

Lilith, University of Liège – QLVL, University of Leuven

dirk.pijpops@uliege.be

Zo'n vs. *zulke* 'such'

- Singular countable nouns, e.g. *zo'n toestand* 'such a situation', *zo'n man* 'such a man': *zo'n* utterly dominant

- Plural nouns & singular mass nouns, e.g. *zo'n/zulke toestanden* 'such situation', *zo'n/zulke diversiteit* 'such diversity': competition

  — Netherlandic Dutch: more *zulke*

  — Belgian Dutch: more *zo'n*

Literature: Ghesquière & Van de Velde (2011), Van Olmen & van der Auwera (2014), Van Olmen (2019)

THE USE OF *ZO'N* VERSUS *ZULKE* 'SUCH' IN BELGIAN AND NETHERLANDIC DUTCH

Testing hypotheses relating to lexical biases, function, register and noun type

Dirk Pijpops

Lilith, University of Liège – QLVL, University of Leuven

dirk.pijpops@uliege.be

THE USE OF *ZO'N* VERSUS *ZULKE* 'SUCH' IN BELGIAN AND NETHERLANDIC DUTCH

Testing hypotheses relating to **lexical biases**, function, register and noun type

Dirk Pijpops

Lilith, University of Liège – QLVL, University of Leuven

dirk.pijpops@uliege.be

Lexical biases

- Lexical biases = one word prefers one variant, another word prefers the other variant

- Lectal contamination = a mechanism that can create lexical biases in morphosyntactic variation

  Works under 4 conditions:

  1. Two morphosyntactic variants that have a different distribution in two lects

  2. These two lects prefer different words

  3. Language contact between the lects

  4. Language users store ready-mades

How does lectal contamination work?

1. Two morphosyntactic variants that have a different distribution in two lects

    e.g.   Netherlandic Dutch *iets boeiends/leuks/...* vs. Belgian Dutch *iets boeiend/leuk/...* 'something special/fun/...'

    Netherlandic Dutch *zulke aantallen/toestanden/...* vs. Belgian Dutch *zo'n aantallen/toestanden/...* 'such numbers/...'

2. These two lects prefer different words

    e.g.   Netherlandic Dutch *boeiend* 'interesting' vs. Belgian Dutch *speciaal* 'special'

    Netherlandic Dutch *aantallen* 'numbers' vs. Belgian Dutch *toestanden* 'situations'

⟹   When looking at material from both lects combined, words typical of lect A will appear more often in the variant typical of lect A

    e.g.   *iets boeiends > iets speciaals*

    zulke aantallen > zulke toestanden

How does lectal contamination work

⟹ When looking at material from both lects combined, words typical of lect A will appear more often in the variant typical of lect A

    e.g. *iets boeiend**s** > iets speciaal**s***

    ***zulke** aantallen > **zulke** toestanden*

3. Language contact: language users are exposed to language material from both lects

    e.g. Belgians hear proportionally more often *iets boeiend**s*** than *iets speciaal**s***

    Belgians hear proportionally more often ***zulke** aantallen* than ***zulke** toestanden*

4. Language users store ready-made forms: intralectal lexical biases emerge

    e.g. When Belgians use the word *boeiend* 'interesting', they will be more inclined to use the ***-s*** variant.

    When Belgians use the word *aantallen* 'numbers', they will be more inclined to use the ***zulke*** variant
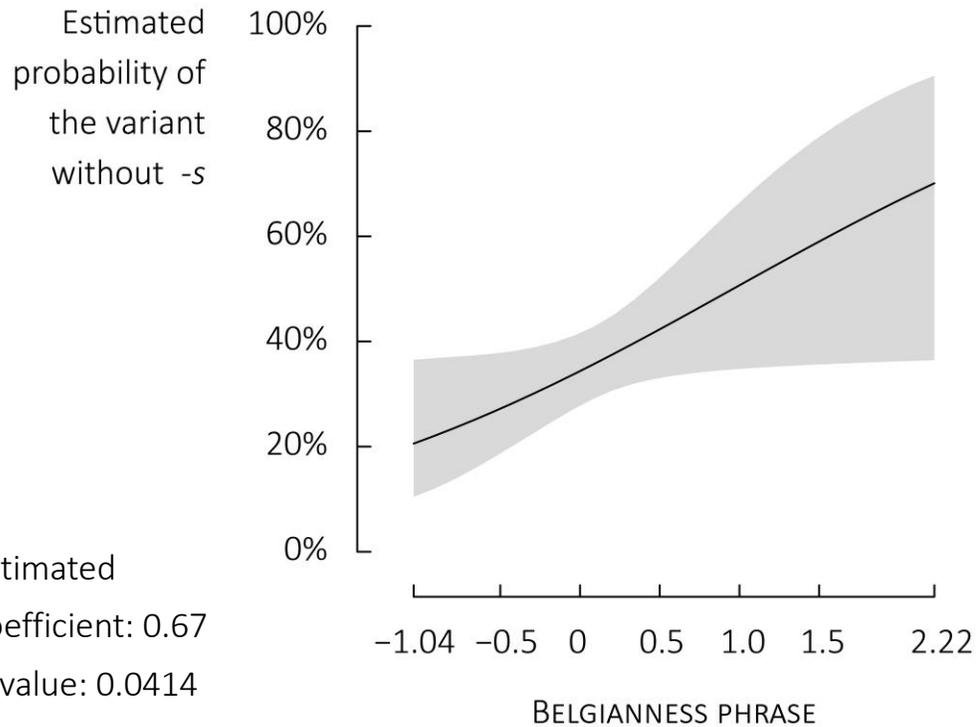
Why is lectal contamination exciting?

- Can explain how lexical biases in language variation emerge out of nowhere

- Simple mechanism

    — Not a special case of language variation

    — Not a special case of language contact

    — Strong evidence for the cognitive storage of ready-made language forms: Arnon & Snider (2010), Tremblay & Baayen (2010), Tremblay e.a. (2011), Dąbrowska (2014)
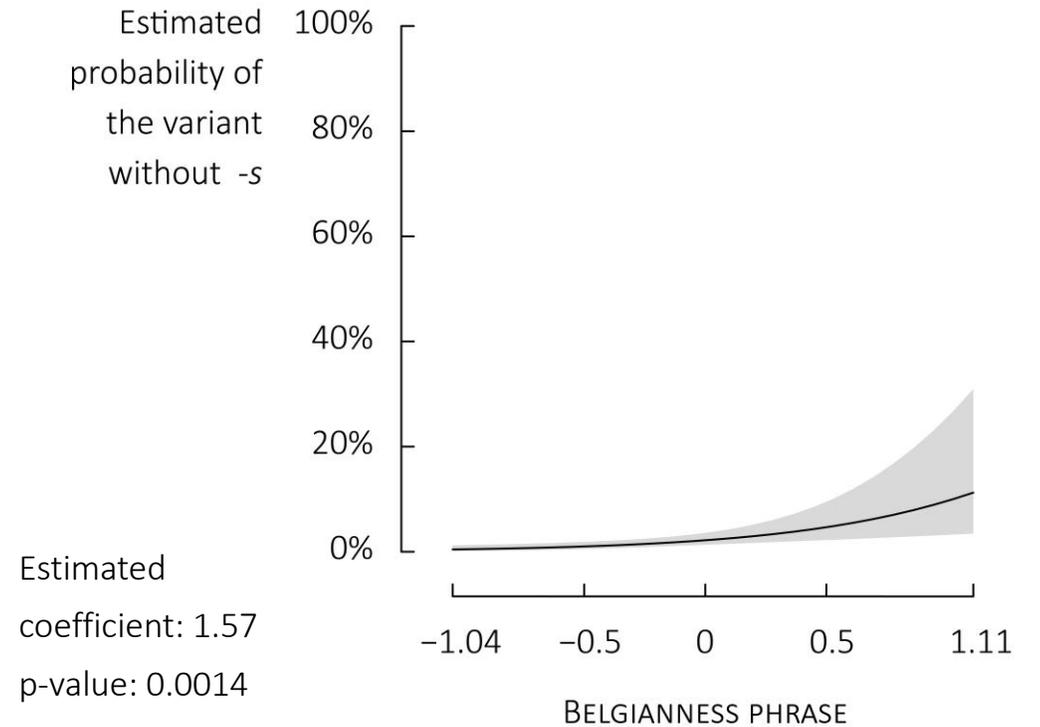
Evidence for lectal contamination

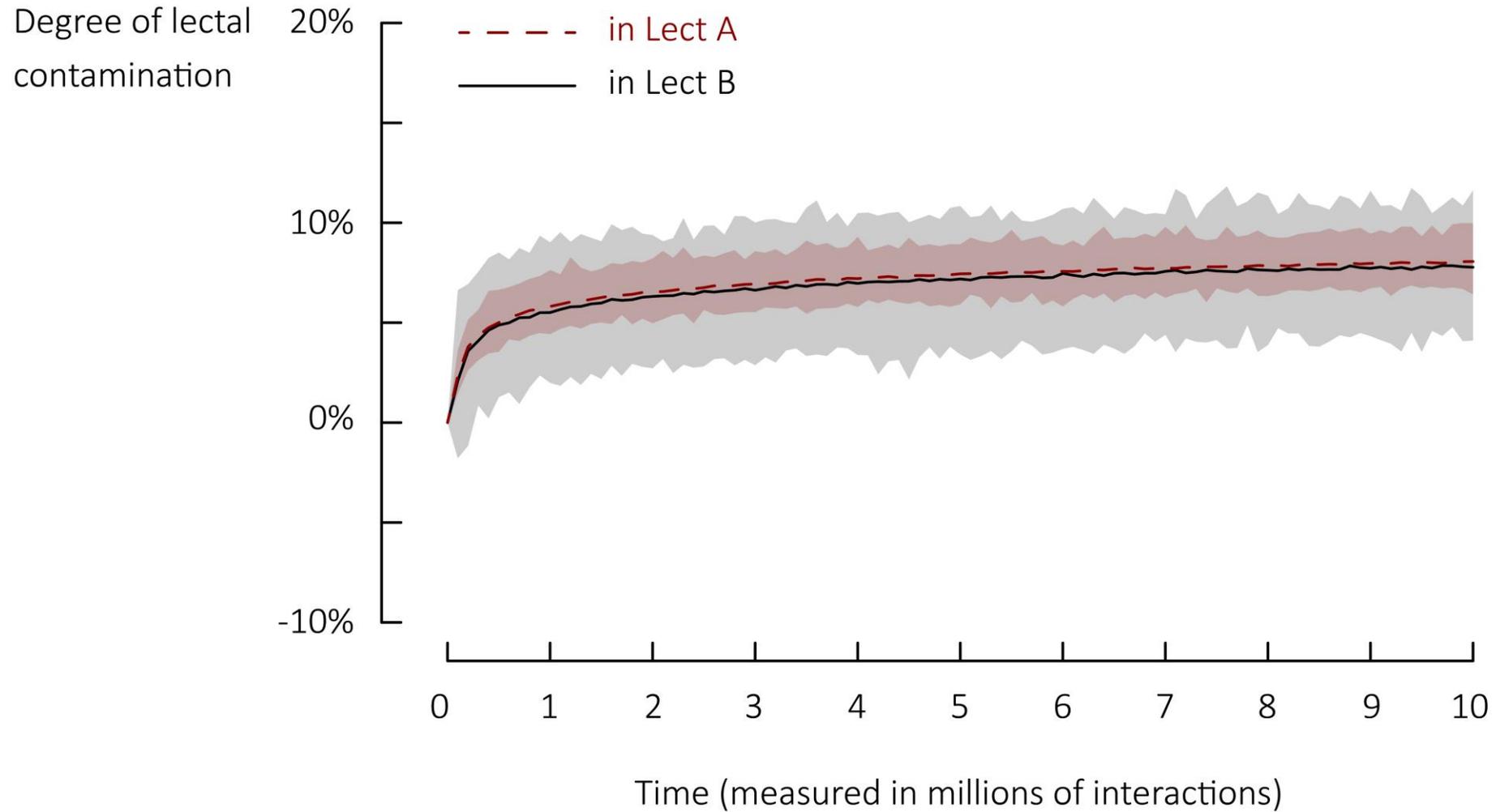- Observational evidence: partitive genitive, e.g. *iets leuk(s), iets bijzonder(s)*

Belgians

Dutchmen

Estimated coefficient: 0.67
p-value: 0.0414

Estimated coefficient: 1.57
p-value: 0.0014

Evidence for lectal contamination

- Simulational evidence: agent-based simulation implements 4 conditions

  1. Two morphosyntactic variants that have a different distribution in two lects

  2. These two lects prefer different words

  3. Language contact between the lects

  4. Language users store ready-mades

  → Lectal contamination emerges consistently

Evidence for lectal contamination

Degree of lectal contamination

20%

in Lect A

in Lect B

10%

0%

-10%

0   1   2   3   4   5   6   7   8   9   10

Time (measured in millions of interactions)

Does lectal contamination affect *zo'n* vs. *zulke*?

- It should affect *zo'n* vs. *zulke* if all 4 conditions hold for *zo'n* vs. *zulke*:

1. Two morphosyntactic variants that have a different distribution in two lects: Van Olmen & Van der Auwera (2014)

2. These two lects prefer different words: Speelman et al. (2008), Ruette (2012), Daems et al. (2015)

3. Language contact between the lects: Centraal Bureau voor de Statistiek (2020), van Agtmaal-Wobma et al. (2007).

4. Language users store ready-mades

→ They store e.g. *zulke hoge cijfers, zo'n personen, zo'n diversiteit,...* as a ready-made in memory
(Arnon & Snider 2010, Tremblay & Baayen 2010, Tremblay e.a. 2011, Dąbrowska 2014)

Does lectal contamination affect *zo'n* vs. *zulke*?

- It should affect *zo'n* vs. *zulke* if all 4 conditions hold for *zo'n* vs. *zulke*:

    1. Two morphosyntactic variants that have a different distribution in two lects: Van Olmen & Van der Auwera (2014)

    2. These two lects prefer different words: Speelman et al. (2008), Ruette (2012), Daems et al. (2015)

    3. Language contact between the lects: Centraal Bureau voor de Statistiek (2020), van Agtmaal-Wobma et al. (2007).

    4. Language users store ready-mades

        → Hypothesis: The more 'Belgian' a phrase is, the more often it will appear in the 'Belgian' variant *zo'n* in Belgian Dutch and Netherlandic Dutch

ANALYSIS

Data

- Corpora

    — Corpus of Spoken Dutch: all speaker info known

    — Sonar Corpus of Written Dutch: only material for which writer info is known

- List of potential mass nouns from *Referentiebestand Nederlands* (RBN 2014)

- Extracted all instances of *zulke/zo'n* followed by a plural noun or a potential mass noun within the next 5 words

- Manually checked, excluded:

    — Instances with singular countable nouns

    — Instances of the recognitional and approximating function

    — Instances where both variants were used

    — Clear instances of copying (Sonar)

    — Metalinguistic comments (CGN)

    — *Zo een open mogelijk beleid* 'a policy that is as open as possible'

    — Head noun is over 5 words away

Netherlandic regression model

- 4415 *zulke* vs. 1115 *zo'n*

- $\text{BELGIANNESS PHRASE} = \log(\frac{(count\ of\ phrase\ in\ Belg.WRPPG + 1)/(size\ of\ the\ Belgian\ WRPPG - (count\ of\ phrase\ in\ Belg.WRPPG + 1))}{(count\ of\ phrase\ in\ Neth.WRPPG + 1)/(size\ of\ the\ Neth.WRPPG - (count\ of\ phrase\ in\ Neth.WRPPG + 1))})$

- NOUN TYPE: plural nouns, singular mass nouns

- FUNCTION: anaphoric, kataphoric, exophoric, intensifying

- Interaction between NOUN TYPE and FUNCTION

- COMPLEXITY PHRASE: natural logarithm of the number of words of the phrase

- REGISTER: formal, informal

- CORPUS: sonar, cgn

- Random intercepts: PHRASE

Netherlandic regression model

- Model fitting:

  — Phrase: all phrases that occur only once, were pooled. (Wolk et al. 2013: 399)

  — VIFs < 5 (Levshina 2015: 160)

 — C-index: 0.974 (Hosmer & Lemeshow 2000: 162)

Belgian model

- 88 *zulke* vs. 209 *zo'n:* maximum 4 parameters (Speelman 2014: 530)

- $\text{BELGIANNESS PHRASE} = \log\left(\dfrac{(count\ of\ phrase\ in\ Belg.\ WRPPG + 1)/(size\ of\ the\ Belgian\ WRPPG - (count\ of\ phrase\ in\ Belg.\ WRPPG + 1))}{(count\ of\ phrase\ in\ Neth.\ WRPPG + 1)/(size\ of\ the\ Neth.\ WRPPG - (count\ of\ phrase\ in\ Neth.\ WRPPG + 1))}\right)$

- NOUN TYPE: plural nouns, singular mass nouns

- REGISTER: formal, informal

- Interaction between NOUN TYPE and REGISTER

- Model fitting

    — VIFs < 5 (Levshina 2015: 160)

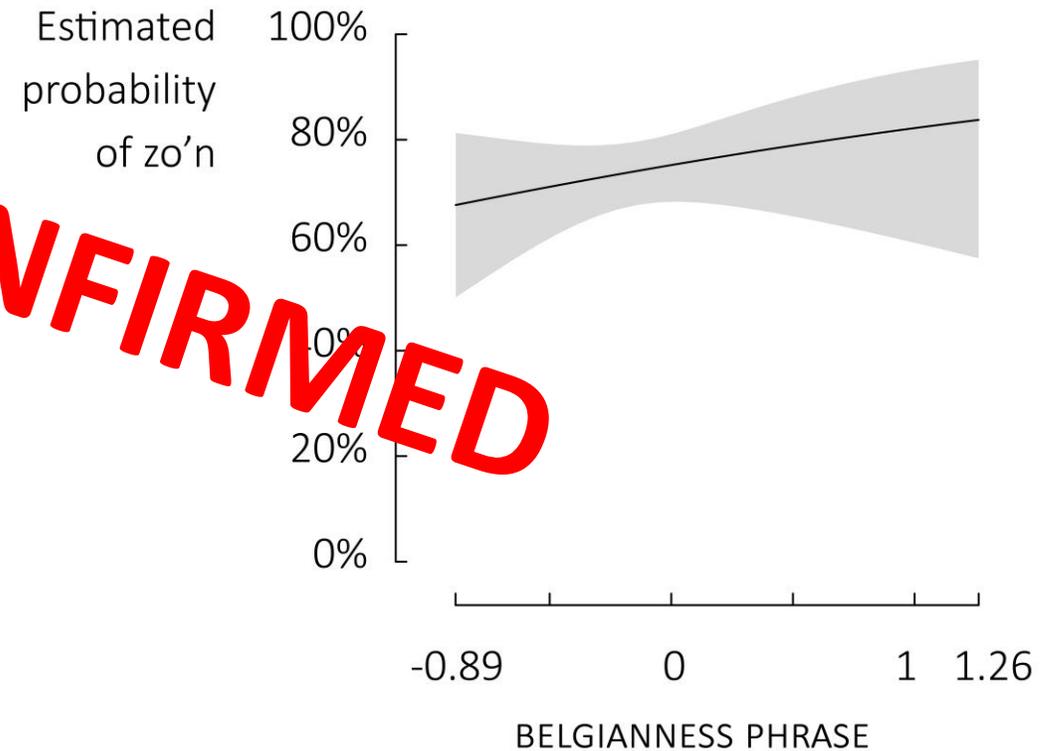   — C-index: 0.763 (Hosmer & Lemeshow 2000: 162).

RESULTS

Hypothesis: the more 'Belgian' a phrase is, the more often it will appear in the 'Belgian' variant *zo'n* in Netherlandic Dutch and Belgian Dutch

Netherlandic model
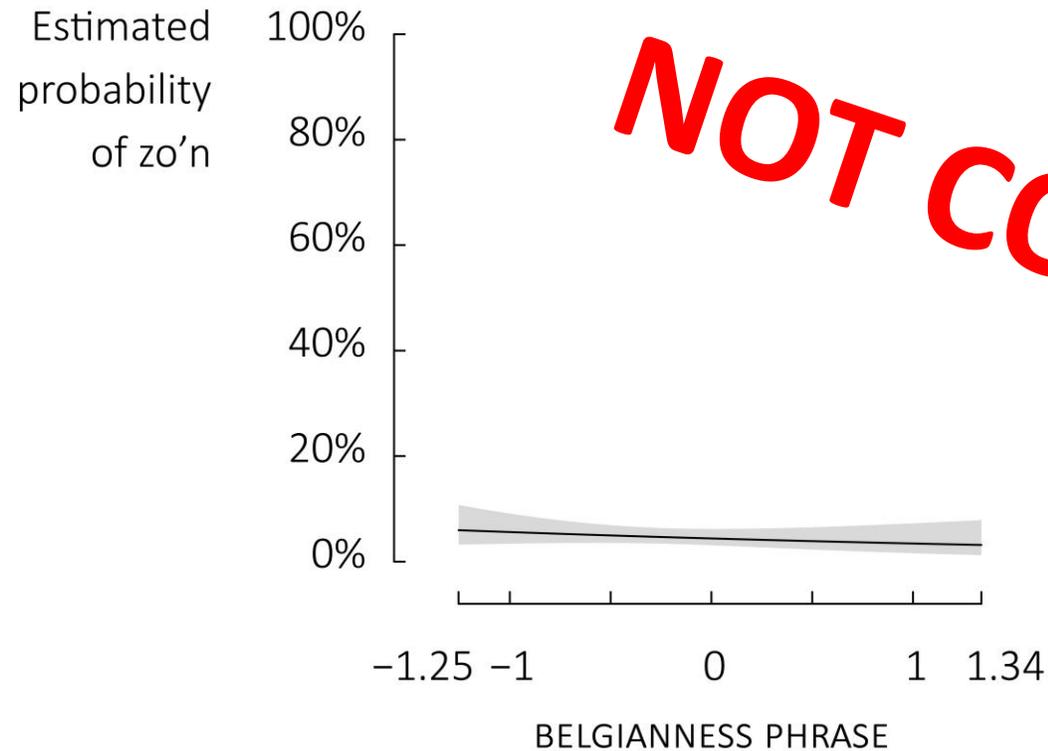
Belgian model

Estimate: -0.28, p = 0.3295

Estimate: 0.42, p = 0.3663

CONCLUSIONS

Lectal contamination: not confirmed

- Option 1: Lectal contamination does not exist

    → Explain away the other evidence:

    — Partitive genitive

    — Agent-based simulation

- Option 2: Lectal contamination exists, but is not at work for *zo'n* vs. *zulke*

    → At least one of the 4 conditions does not hold for *zo'n* vs. *zulke*:

    1. Two morphosyntactic variants that have a different distribution in two lects

    2. These two lects prefer different words

    3. Language contact between the lects

    4. **Language users store ready-mades**

Lectal contamination

- Option 3: Lectal contamination is at work for *zo'n* vs. *zulke,* but we failed to observe it

  — Operationalization of BELGIANNESS PHRASE not ideal

  — Belgian model based on very few data

  — Failed to control for important variables (Belgian model)

  — …

Want to know more?

- Papers under review: dirk.pijpops@uliege.be

- Related mechanism: **constructional contamination**

  — 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 and Freek Van de Velde. 2018. Constructional contamination in morphology and syntax: Four case studies. *Constructions and Frames* 10(2). 269–305.

  — Hilpert, Martin and Susanne Flach. A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive. In Marcin Grygiel (ed.), *Contrast and analogy in language: Perspectives from Cognitive Linguistics*. Amsterdam: John Benjamins.

  — Hilpert, Martin. 2019. *Construction Grammar and its Application to English*. 2nd edn. Edinburgh: Edinburgh University Press.

    → Chapter 3

References

Algeo, J. (2006). *British or American English?: a handbook of word and grammar patterns*. Cambridge: Cambridge University Press.

Arnon, I., & Snider, N. (2010). More Than Words: Frequency Effects for Multi-Word Phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.4.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49–69). Oxford: Oxford University Press.

Centraal Bureau voor de Statistiek. (n.d.). Retrieved March 13, 2020, from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83926NED/barv?dl=17256&ts=1584114358740

Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, *25*(4), 617–653.

Daems, J., Heylen, K., & Geeraerts, D. (2015). Wat dragen we vandaag: een hemd met blazer of een shirt met jasje? *67*(2), 307–342.

Diessel, H. (2015). Usage-based construction grammar. In E. Dąbrowska & D. Divjak (Eds.), *Handboek of Cognitive Linguistics* (pp. 296–322). Berlin: Mouton De Gruyter.

Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D., & Taylor, S. (2016). *Effect Displays for Linear, Generalized Linear, and Other Models.* R package version 3.2.

Geeraerts, D., Grondelaers, S., & Speelman, D. (1999). *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P. J. Meertens-Instituut.

Ghesquière, Lobke and Freek Van de Velde. 2011. A corpus-based account of the development of English such and Dutch zulk: Identification, intensification and (inter)subjectification. *Cognitive Linguistics* 22(4). 765–797.

Harrell, F. J. (2017). *Hmisc: Harrell Miscellaneous.* R package version 4.0-3.

Hilpert, M., & Flach, S. (n.d.). A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive. In M. Grygiel (Ed.), *Contrast and analogy in language: Perspectives from Cognitive Linguistics*. Amsterdam: John Benjamins.

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd edn.). New York: Wiley.

Meyer, D., Zeileis, A., & Hornik, K. (2020). *vcd: Visualizing Categorical Data.* R package version 1.4-6.

Oostdijk, N., Goedertier, W., Van Eynde, F., Boves, L., Martens, J.-P., Moortgat, M., & Baayen, H. (2002). Experiences from the Spoken Dutch corpus project. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, 340–347.

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013b). The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In P. Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing* (pp. 219–247). Heidelberg: Springer.

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 and Freek Van de Velde. 2018. Constructional contamination in morphology and syntax: Four case studies. *Constructions and Frames* 10(2). 269–305.

Ruette, T. (2012). *Aggregating Lexical Variation: towards large-scale lexical lectometry* (Unpublished doctoral dissertation). University of Leuven.

Speelman, D., Heylen, K., & Geeraerts, D. (2018). *Mixed-Effects Regression Models in Linguistics*. Heidelberg: Springer.

Tremblay, A., & Baayen, R. H. (2010). Holistic Processing of Regular Four-word Sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 151–173). London : Continuum.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence from Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, *61*(2), 569.

van Agtmaal-Wobma, E., Harmsen, C., Dal, L., & Poulain, M. (2007). *Belgen in Nederland en Nederlanders in België*. Report Centraal Bureau voor de Statistiek.

Van de Velde, Freek and Dirk Pijpops. 2018. Grensoverschrijdend syntactisch gedrag. In Timothy Colleman, Johan De Caluwe, Veronique De Tier, Anne-Sophie Ghyselen, Liesbet Triest, Roxane Vandenberghe & Ulrike Vogl (eds.), *Woorden om te bewaren. Huldeboek voor Jacques Van Keymeulen*, 433–449. UGent, Vakgroep Taalkunde, Afdeling Nederlands; Gent.

van Eerten, L. (2007). Over het Corpus Gesproken Nederlands. *Nederlandse Taalkunde*, *12*(3), 194–215.

Van Olmen, Daniël. 2019. A diachronic corpus study of prenominal zo'n "so a" in Dutch: Pathways and (inter)subjectification. *Functions of Language* 26(2). 216–247.

Van Olmen, Daniël and Johan van der Auwera. 2014. Over zo'n en zo meer. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 215–228. Leuven: Leuven University Press.