DIALECTAL DITRANSITIVE PATTERNS IN BRITISH ENGLISH

LANGUAGE-INTERNAL VS. LANGUAGE-EXTERNAL CONSTRAINTS

Johanna Gerwin (Kiel) & Melanie Röthlisberger (Zürich)

(1) SETTING THE SCENE

DITRANSITIVES

- *She gave a book **to** him.* (PREP)
- *She gave him a book.* (DOC)

Dative alternation

- *She bought a book **for** ihm.* (PREP)
- *She bought him a book.* (DOC)

Benefactive alternation

DIALECTAL VARIANTS

- *She gave it him.* (altDOC)
- *She gave it to him.* (PREP)
- *She gave him it.* (DOC)

- *He had such a go at me for showing it my mum* (BNCreg)
- *If I find out I 'll give it you back.* (BNCreg)
- *I 'll fetch it you down anyway, next time* (BNCreg)

(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

- Ditransitive patterns as a Labovian sociolinguistic variable:
- How is pattern choice determined by language-external factors such as 'region/origin of the speaker', 'time', and 'spoken vs written mode'?

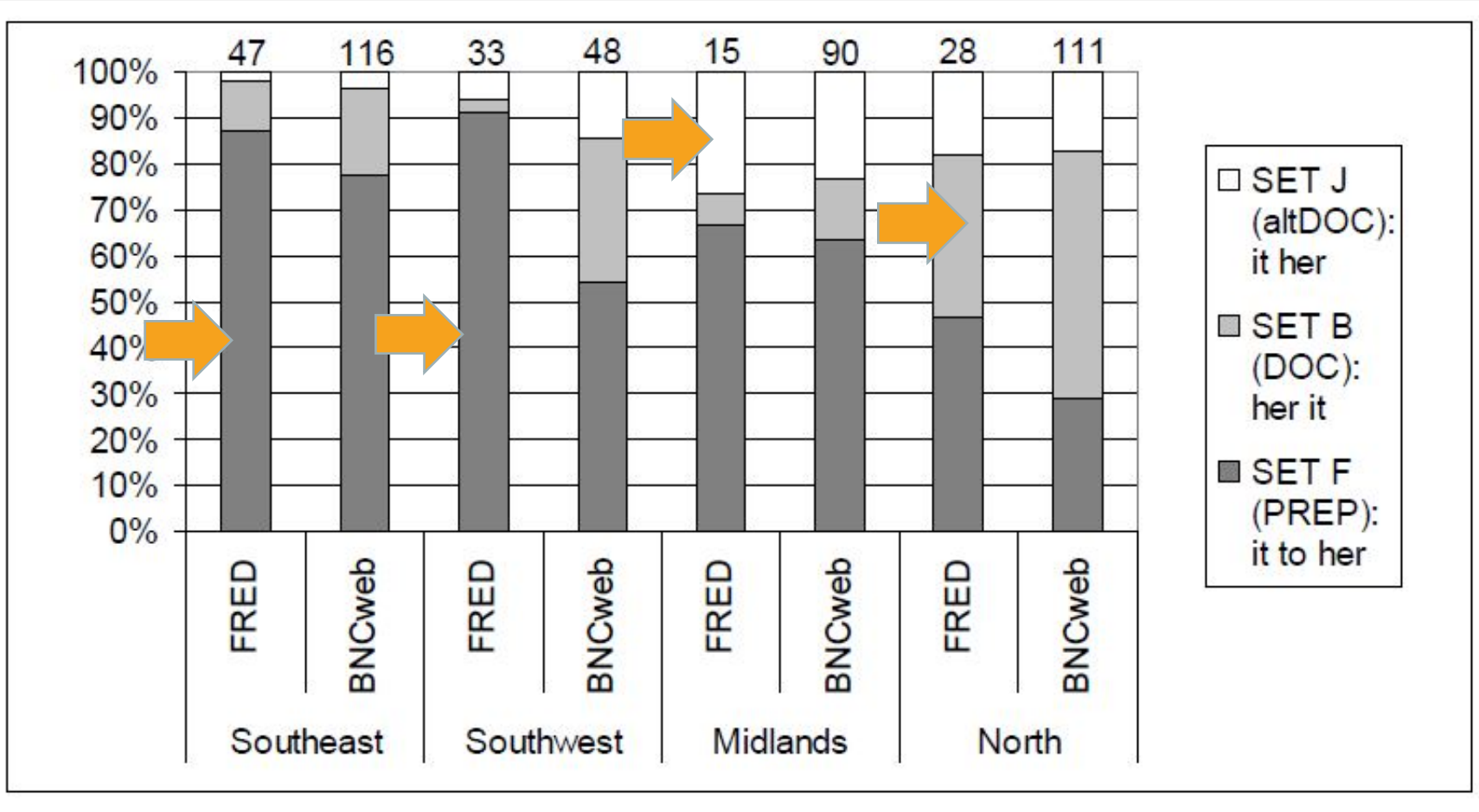
	1960/70s		1990s	
spoken data	2,962	FRED (2.5 mill.)	BNCweb (subcorpus BNCreg: 3.3 mill.)	5,147
written data	872	LOB (1 mill.)	F-LOB (1 mill.)	917

→ total:
9,899 tokens

(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

- regional preferences, especially with pronominal themes (*it*/*them*)

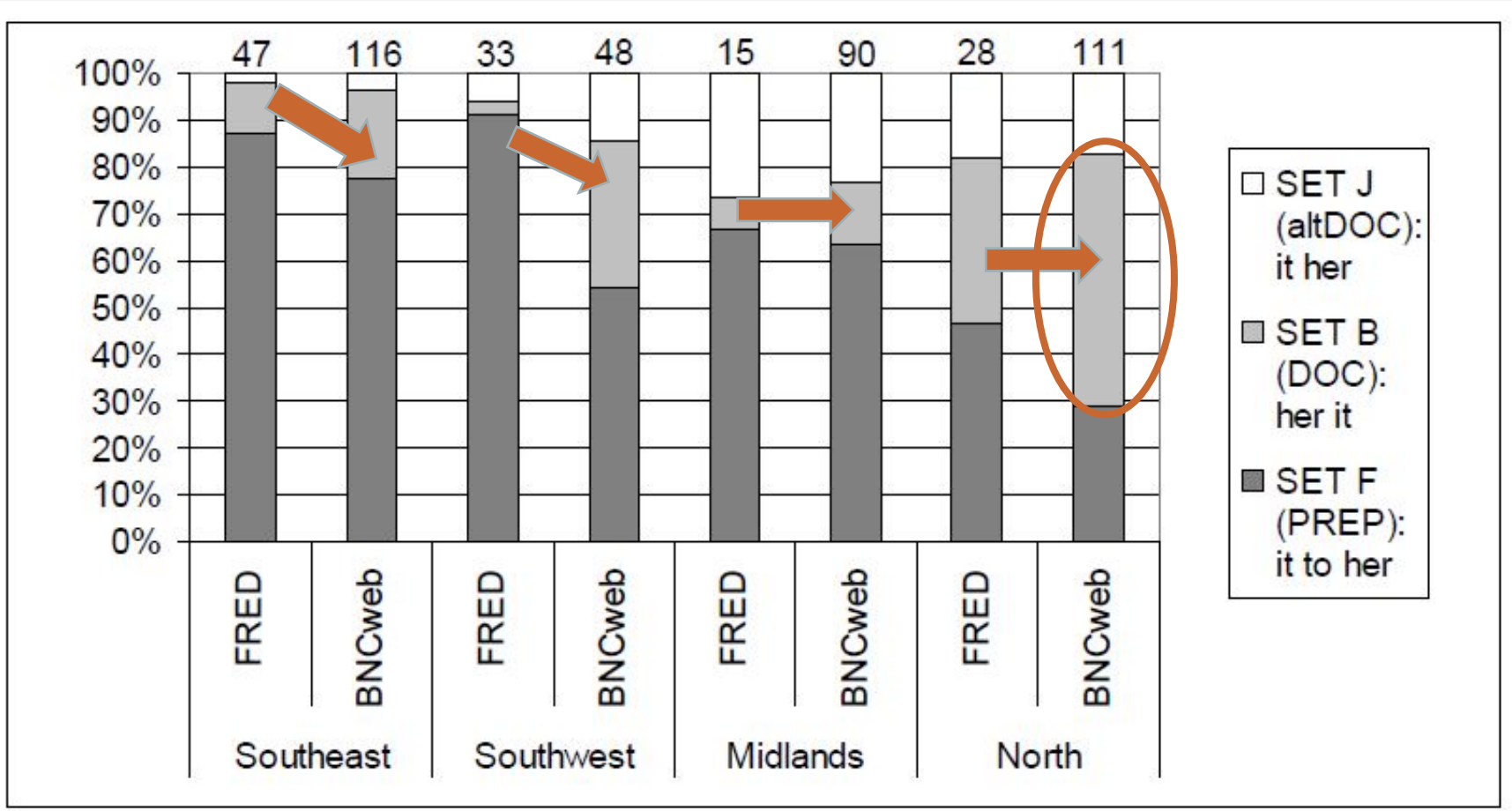


(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

- regional preferences, especially with pronominal themes (*it*/*them*)

- slight diachronic effect between FRED and BNCweb data



(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

- regional preferences, especially with pronominal themes (*it*/*them*)

(1) SETTING THE SCENE

RESEARCH QUESTIONS

(1) To which extent is pattern choice determined by the sociolinguistic factors 'region/origin of the speaker' and 'time' when well-established language-internal predictors are modelled simultaneously?

(2) Do language-internal predictors differ in their importance on dative choice across different British regions?

(2) DATA

FREIBURG ENGLISH DIALECT CORPUS (FRED)

- 2.5 million words
- 1970-1989
- 431 speakers (mainly NORMs born between 1890 & 1919)

BRITISH NATIONAL CORPUS ONLINE EDITION (BNCWEB)

- spoken-demographic part: ca. 10 million words
- regionally tagged data: 3.3. million words
- Early 1990s
- 124 speakers (from age 15)

Language-external predictor 'time'
→ real-time difference of about 20 years
→ apparent-time difference of about 2-3 generations of speakers

(2) DATA

BRITISH REGIONS

- Southeast (SE)
- Southwest (SW)
- Midlands (MID)
- North (N)
- Wales (WAL)
- Scotland (SCOT)

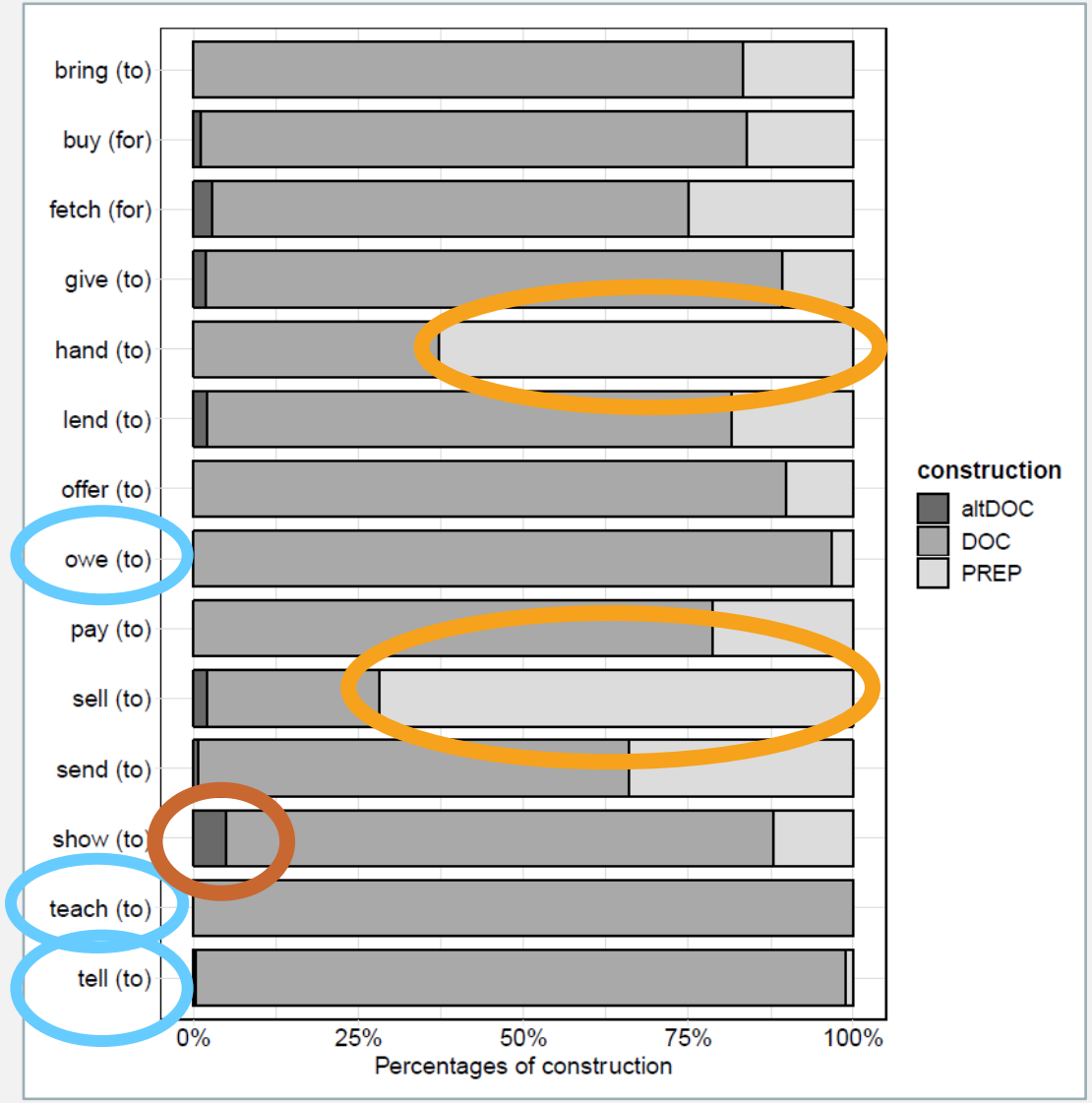
Language-external predictor 'region/origin of speaker'
→ 6 British regions

(2) DATA

DATA SELECTION

- only two post-verbal objects
- prototypical declarative sentences
- 14 alternating verbs

→7070 ditransitives
→2661 in FRED
→4449 in BNC



(2) DATA

Language-internal predictor 'verb semantics'
→ 3 different senses

- TRANSFER:
  All the time they was bringin' them in their living (FRED: KEN 011)

- COMMUNICATION:
  It tells you all about it in the book and also here (BNCreg: KDL 207)

- ABSTRACT/IDIOMATIC:
  And you put your foot behind the stone, and give it a good (e ‚cracking sound') crack (FRED: PEE 001)

(2) DATA

Language-internal predictor 'recipient animacy'
→ 2-way distinction: animate – inanimate

Language-internal predictor 'syntactic weight'

→ number of typed characters
→ including spaces in multi-word objects
→ excluding prepositions *to* and *for*

Language-internal predictors 'recipient pronominality' & 'theme pronominality'
→ pronoun v. full noun phrase

(2) DATA

Language-internal predictor 'particle'
→ intervening, following, none at all

EXAMPLES:

- Particle intervening between two objects

  *All the time they was bringing 'em **in** their living* (FRED: KEN 011)
  *Now look you've got to give that ticket **back** to Mr […]* (BNCreg: KCT 14186)

- Particle following two objects

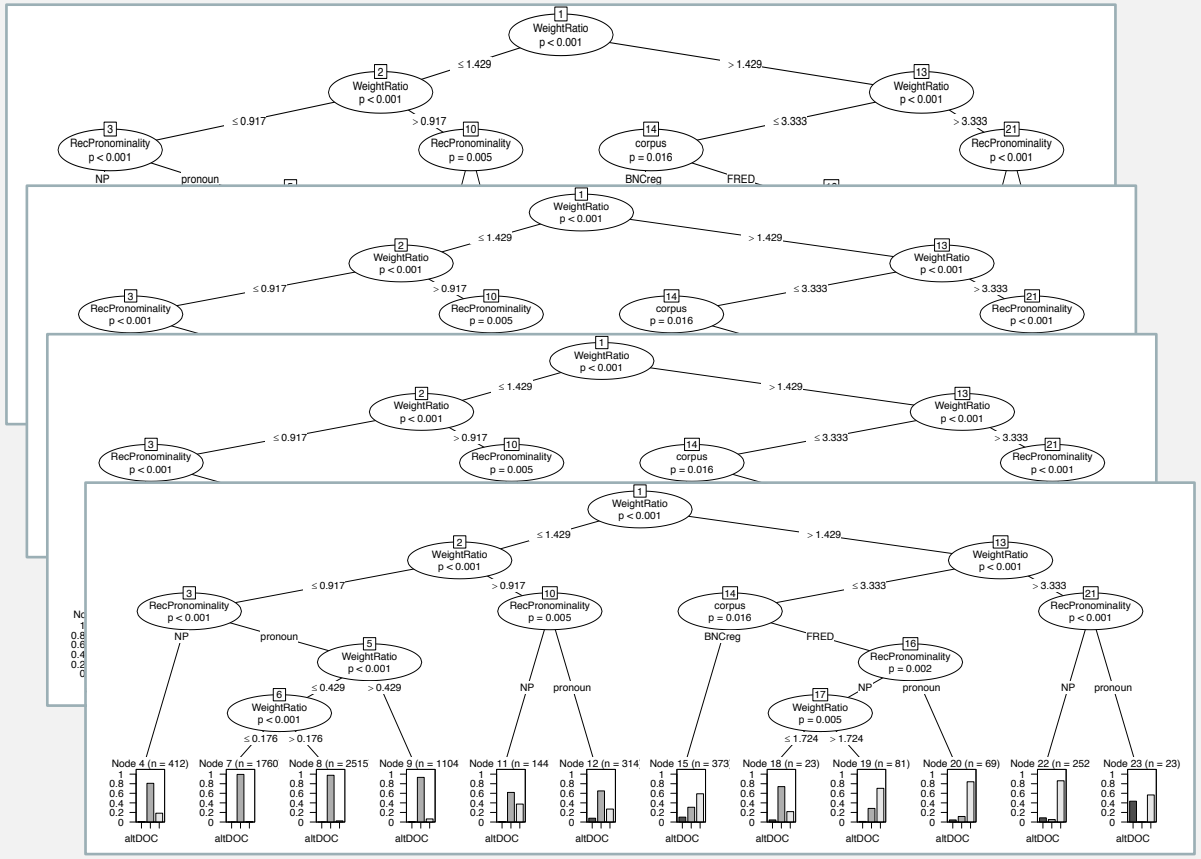
  *My mum used to give me a penny **back**.* (FRED: SFK 032)

(3) METHODOLOGY

CONDITIONAL RANDOM FORESTS (CRFS)

- Party package
- trial-and-error
- *ntree* ctrees on random subsample of the data
- bagging over testset to predict out-of-bag-sample
- permutation scheme

- FOR THIS STUDY:
  - 1 random forest on full dataset
  - 6 separate conditional random forests for each region



(3) METHODOLOGY

COMPARING THE RANKING OF 6 CRFS

1. fit one forest per region
2. define the ranking of predictors (e.g. 1 = most important, 2 = second important)
3. calculate spearman‘s rank correlation coefficient as measure of similarity
4. transform similarity to distance: 1 – similarity coefficient
5. use Multidimensional scaling analysis to visualize distance between regions (smacof package for non-metric MDS solution)

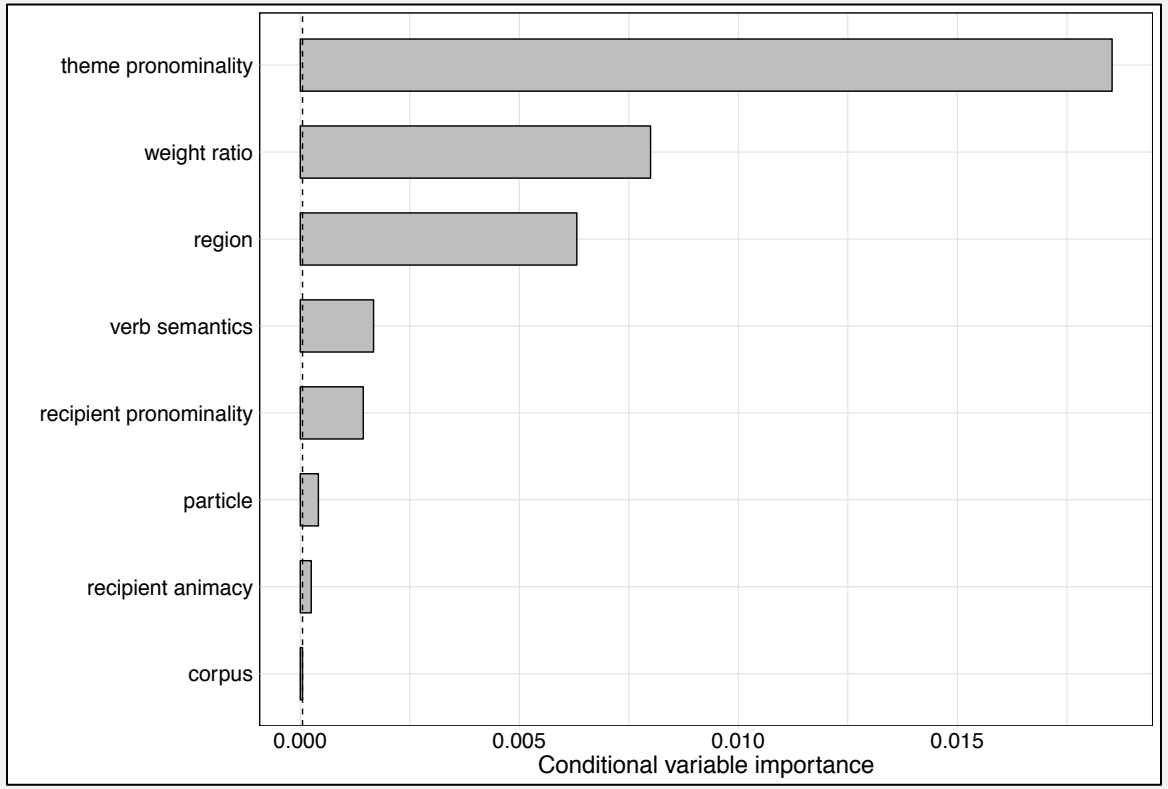
(4) RESULTS

CRF ON FULL DATASET

RESEARCH QUESTION (1):

To which extent is pattern choice determined by the sociolinguistic factors 'region/origin of the speaker' and 'time' when well-established language-internal predictors are modelled simultaneously?

- ntree = 2000
- mtry = 3
- 93.2 % accuracy



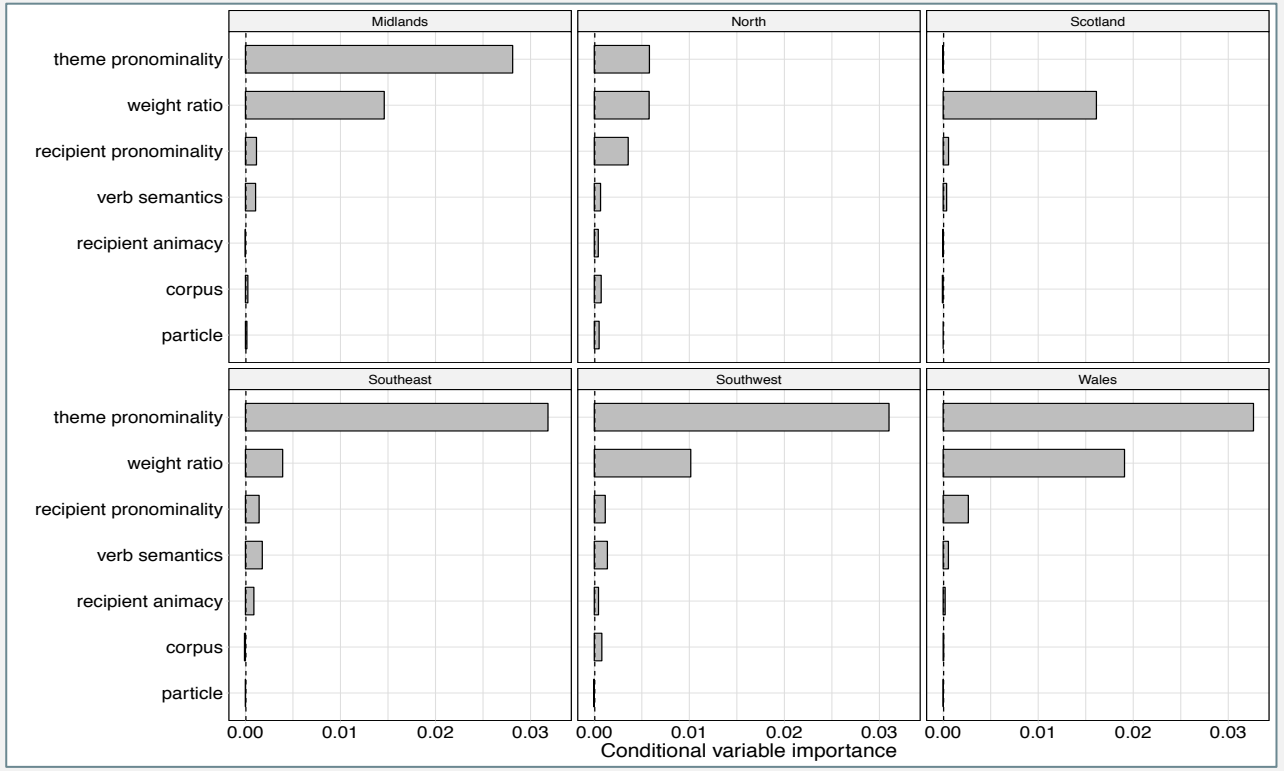
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region



(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region

Factor	Midlands	North	Scotland	Southeast	Southwest	Wales
Theme pronominality	1	1	6	1	1	1
Weight ratio	2	2	1	2	2	2
Recipient pronominality	3	3	2	4	4	3
Verb semantics	4	5	3	3	3	4
Recipient animacy	7	7	5	5	6	5
Corpus	5	4	7	7	5	6
Particle	6	6	4	6	7	7

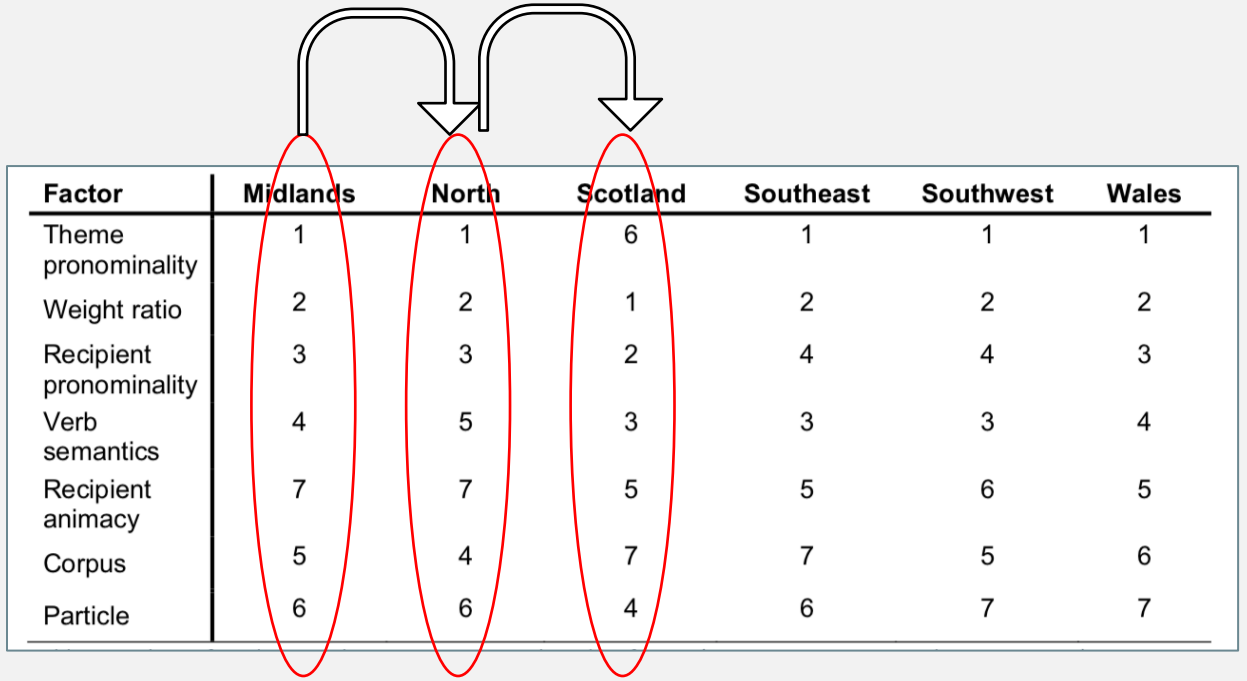
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman's rank correlation coefficient between all regions (15 comparisons)



Factor	Midlands	North	Scotland	Southeast	Southwest	Wales
Theme pronominality	1	1	6	1	1	1
Weight ratio	2	2	1	2	2	2
Recipient pronominality	3	3	2	4	4	3
Verb semantics	4	5	3	3	3	4
Recipient animacy	7	7	5	5	6	5
Corpus	5	4	7	7	5	6
Particle	6	6	4	6	7	7

(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman's rank correlation coefficient between all regions (15 comparisons)
4. Take inverse for distance (1 – coefficient)

	Midlands	North	Scotland	Southeast	Southwest
North	0.036				
Scotland	0.714	0.857			
Southeast	0.179	0.321	0.607		
Southwest	0.071	0.143	0.786	0.107	
Wales	0.107	0.179	0.679	0.071	0.071

(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

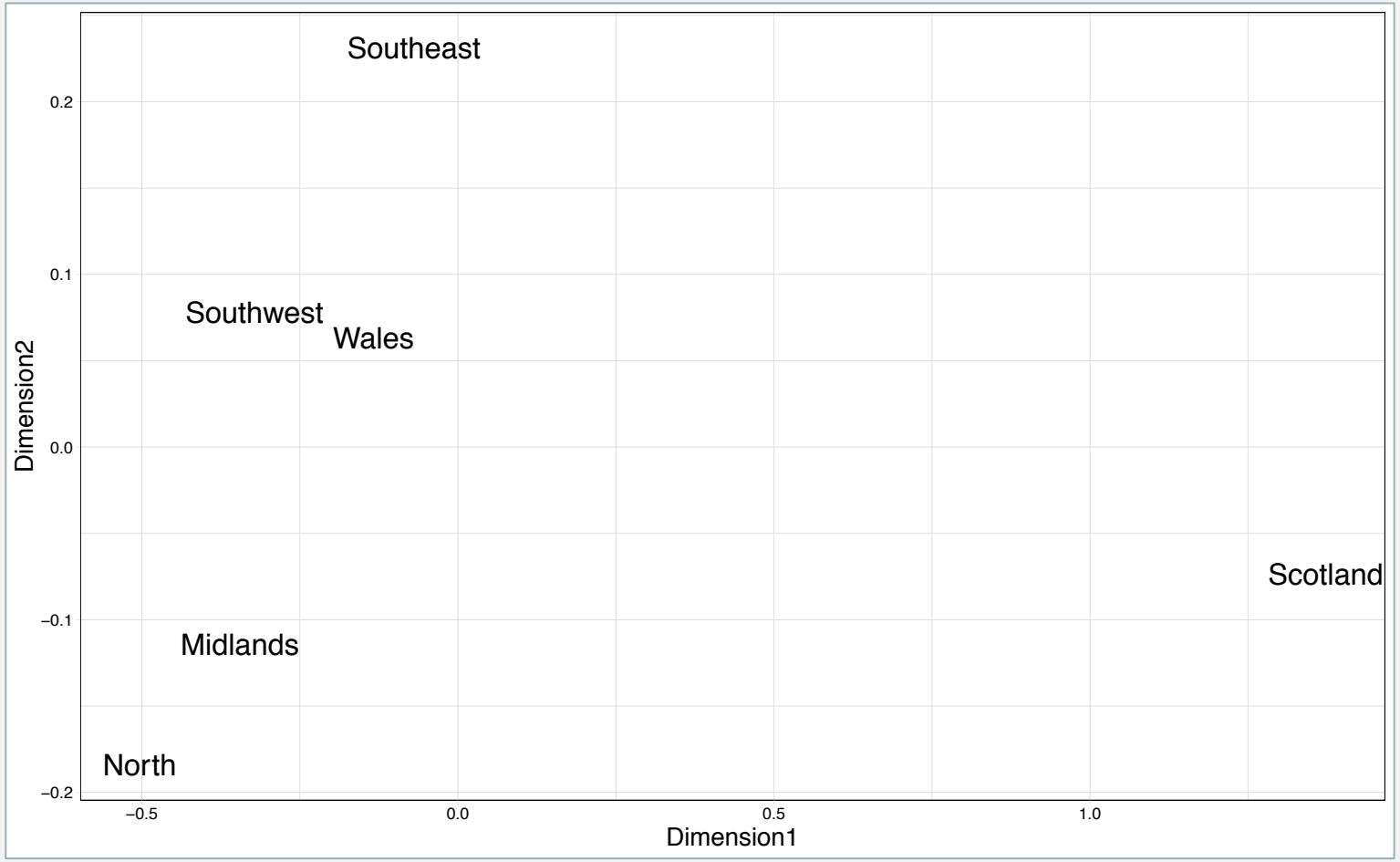
Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman's rank correlation coefficient between all regions (15 comparisons)
4. Take inverse for distance (1 – coefficient)
5. Fit MDS: Reduce number of comparisons to 2

	Midlands	North	Scotland	Southeast	Southwest
North	0.036				
Scotland	0.714	0.857			
Southeast	0.179	0.321	0.607		
Southwest	0.071	0.143	0.786	0.107	
Wales	0.107	0.179	0.679	0.071	0.071

(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS



(5) SUMMARY

RESEARCH QUESTION (1)

To which extent is pattern choice determined by the sociolinguistic factors 'region/origin of the speaker' and 'time' when well-established language-internal predictors are modelled simultaneously?

- Region = third most important factor
- Time ('corpus') = least important factor

RESEARCH QUESTION (2)

Do language-internal predictors differ in their importance on dative choice across different British regions?

- Yes.
- Differences between regions in the way language-internal predictors impact dative choice
- Close correlation with geographic distance between regions

(5) SUMMARY

- **Importance of theme pronominality**
  due to higher frequency of altDOC datives?

- **Scotland behaves differently from the other regions**
  influence of Danelaw, language contact?

- **Fundamental Dialectology Principle (Nerbonne & Kleiweg 2007: 154)**
  importance of geographic proximity to predict linguistic similarity

- **On the methodological level**
  Important to include other possible (dialectal) variants in the analysis to contribute to an integrated approach to syntactic variability in non-standard varieties

(6) BIBLIOGRAPHY

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.
Arnold, Jennifer, Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language* 76(1). 28–55.
Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.
Bernaisch, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* 35(1). 7–31. doi:10.1075/eww.35.1.02ber.
Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.
Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245–259.
Bresnan, Joan & Tatiana Nikitina. 2003. On the Gradience of the Dative Alternation. 2003.
Burnard, Lou. 2012. Reference Guide for the British National Corpus (XML Edition).
Collins, Peter. 1995. The indirect object construction in English: An informational approach. *Linguistics* 33(1). 35–49.
Garretson, Gregory. 2003. Optimal Typology of Determiner Phrases Coding Manual Excerpt Section 5: "Applying Tags to the Examples." (29 November, 2018).
Gast, Volker. 2007. *I gave it him* - on the motivation of the "alternative double object construction" in varieties of British English. In Anna Siewierska & Willem Hollmann (eds.), *Functions of Language - Special issue: Ditransitivity*, 31–56.
Gerwin, Johanna. 2014. *Ditransitives in British English dialects*. Berlin/New York: Mouton de Gruyter.
Givón, Talmy. 1984. Direct object and dative shifting: Semantic and pragmatic case. In Frans Plank (ed.), *Objects: Towards a theory of grammatical relations*, 151–182. New York: Academic Press.
Goldberg, Adele. 1992. The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics* 3(1). 37–74.
Grafmiller, Jason & Benedikt Szmrecsanyi. to appear. Mapping out particle placement in varieties of English: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3).
Grafmiller, Jason, Benedikt Szmrecsanyi, Melanie Röthlisberger & Benedikt Heller. 2018. General introduction: A comparative perspective on probabilistic variation in grammar. *Glossa: a journal of general linguistics* 3(1). 94. 1–10.
Green, Georgia. 1974. *Semantics and syntactic regularity*. Bloomington: Indiana University Press.
Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg & Ronald Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65(2). 203–257.
Heller, Benedikt. 2018. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. Leuven: KU Leuven PhD dissertation.
Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45(1). 3–27.
Hernández, Nuria. 2006. User's Guide to FRED (Freiburg English Dialect Corpus). MS. http://www.freidok.uni-freiburg.de/volltexte/2489/pdf/Userguide_neu.pdf (22 September, 2016).
Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning : A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674.
Jeszenszky, Péter, Philipp Stoeckle, Elvira Glaser & Robert Weibel. 2017. Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography* 5(2). 86–108. doi:10.1017/jlg.2017.5.
Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355–384.
Kruskal, Joseph & Myron Wish. 1978. *Multidimensional scaling*. Newbury Park/London/New Delhi: SAGE Publications.
Leeuw, Jan de & Patrick Mair. 2009. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software* 31(3). 1–30.
Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
Levin, Beth & Malka Rappaport Hovav. 2005. *Argument Realisation*. Cambridge: Cambridge University Press.
Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.
Montemagni, Simonetta. 2008. The space of Tuscan dialectal variation: A correlation study. *International Journal of Humanities and Arts Computing* 2. 135–152.
Mukherjee, Joybrato & Sebastian Hoffmann. 2006. Describing verb-complementation profiles for New Englishes: A pilot-study of Indian English. *English World-Wide* 27(2). 147–173.
Nerbonne, John & Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2). 148–166.
R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Röthlisberger, Melanie. 2018. Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation. Leuven: KU Leuven PhD dissertation.
Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). 673–710.
Scherrer, Yves & Philipp Stoeckle. 2016. A quantitative approach to Swiss German - Dialectometric analyses and comparison of linguistic levels. *Dialectologia et Geolinguistica* 24. 92–125.
Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences* 14(9). 411–417.
Siewierska, Anna & Willem Hollmann. 2007. Ditransitive clauses in English with special reference to Lancashire dialect. In Mike Hannay & Gerard J. Steen (eds.), *Structural-Functional Studies in English Grammar*, 81–102. Amsterdam/Philadelphia: John Benjamins.
Spruit, Marco René, Wilbert Heeringa & John Nerbonne. 2009. Associations among linguistic levels. *Lingua* 119(11). 1624–1642.
Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(307). http://www.biomedcentral.com/1471-2105/9/307.
Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8(25). http://www.biomedcentral.com/1471-2105/8/25.
Szmrecsanyi, Benedikt. 2012a. Typological profile: L1 varieties. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*, 826–843. Berlin/Boston: Mouton de Gruyter.
Szmrecsanyi, Benedikt. 2012b. Geography is overrated. In Sandra Hansen, Christian Schwarz, Philipp Stoeckle & Tobias Streck (eds.), *Dialectological and folk dialectological concepts of space: Current methods and perspectives in sociolinguistic research on dialect change*, 215–231. Berlin / Boston: De Gruyter.
Szmrecsanyi, Benedikt, Douglas Biber, Jesse Egbert & Karlien Franco. 2016. Towards more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change* 28(1). 1–29.
Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137.
Szmrecsanyi, Benedikt, Jason Grafmiller & Laura Rosseel. submitted. Variation-Based Distance and Similarity Modeling: a case study in World Englishes.
Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *Was*/*were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.
Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(1). 81–105.
Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419. doi:10.1075/dia.30.3.04wol.

THANK YOU!

JOHANNA GERWIN

gerwin@anglistik.uni-kiel.de

MELANIE RÖTHLISBERGER

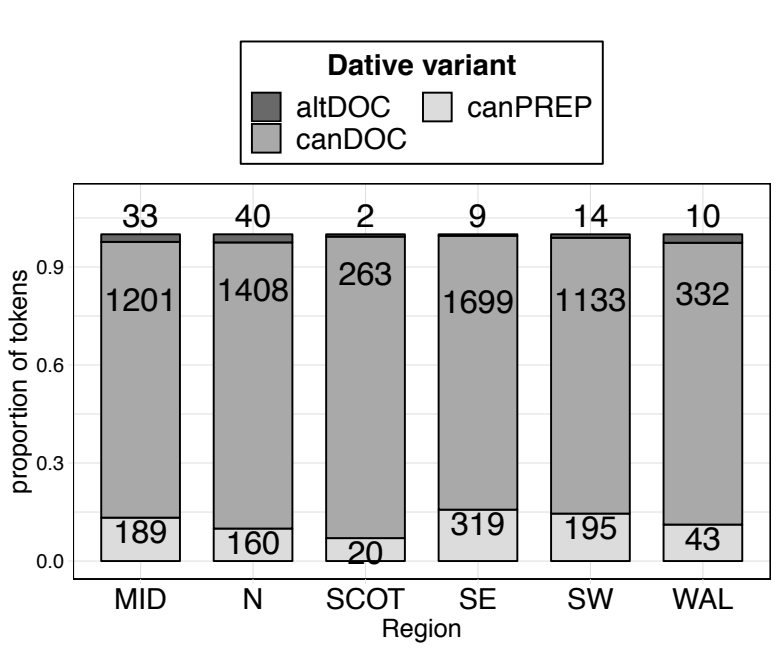
melanie.roethlisberger@es.uzh.ch

@m_rothlisberger

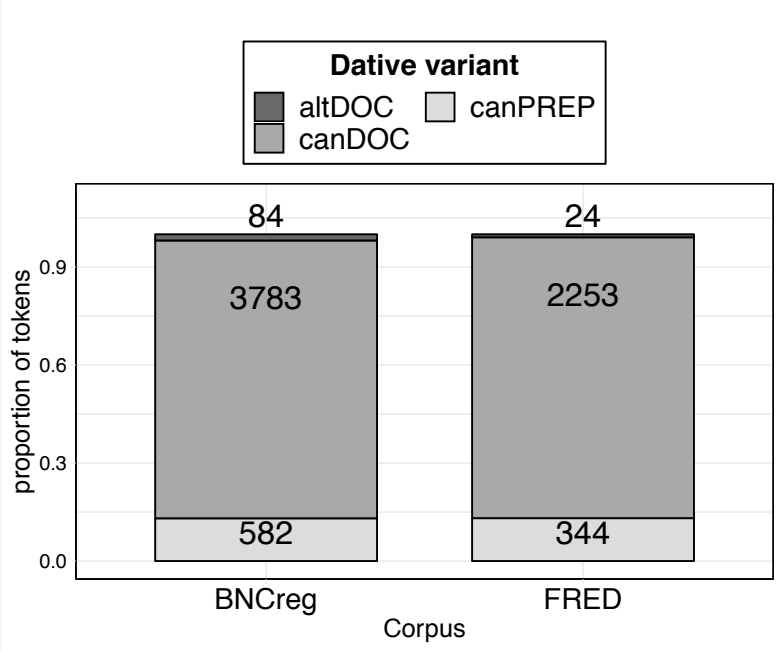
More information in our upcoming paper!

APPENDIX

Proportional distributions by region

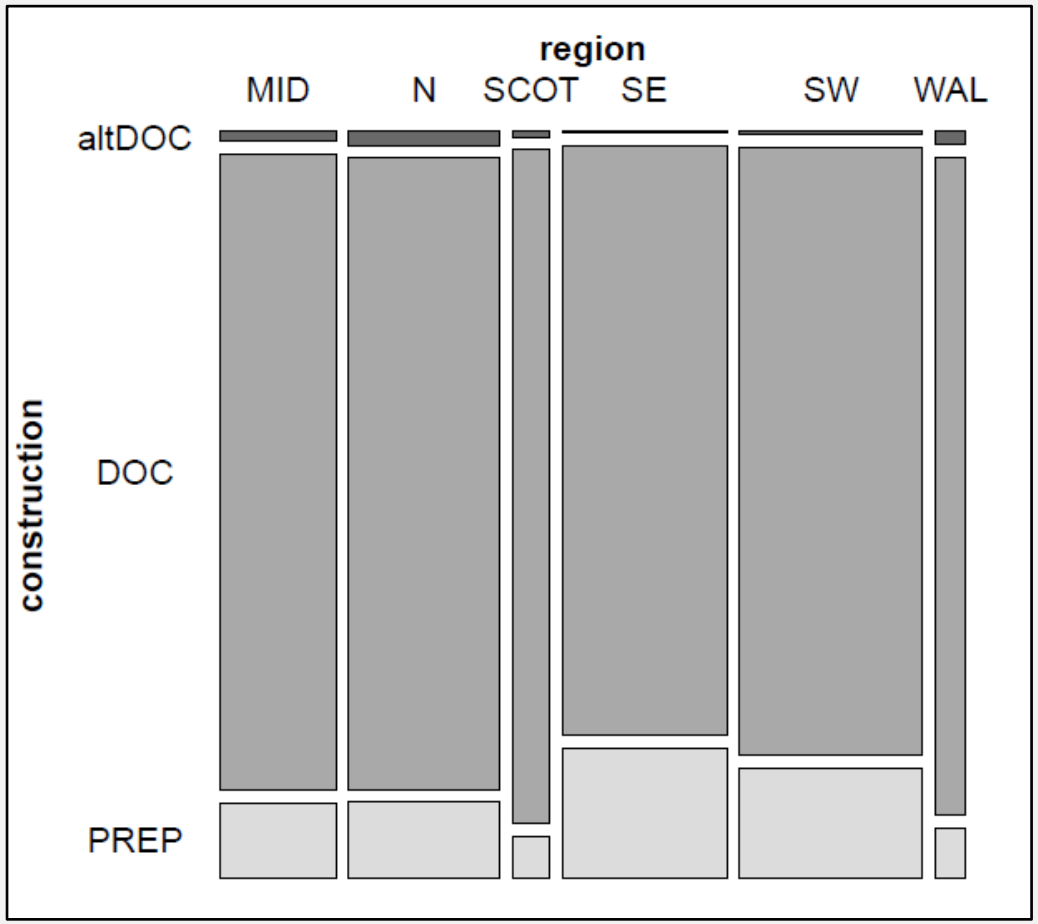


Proportional distributions by corpus

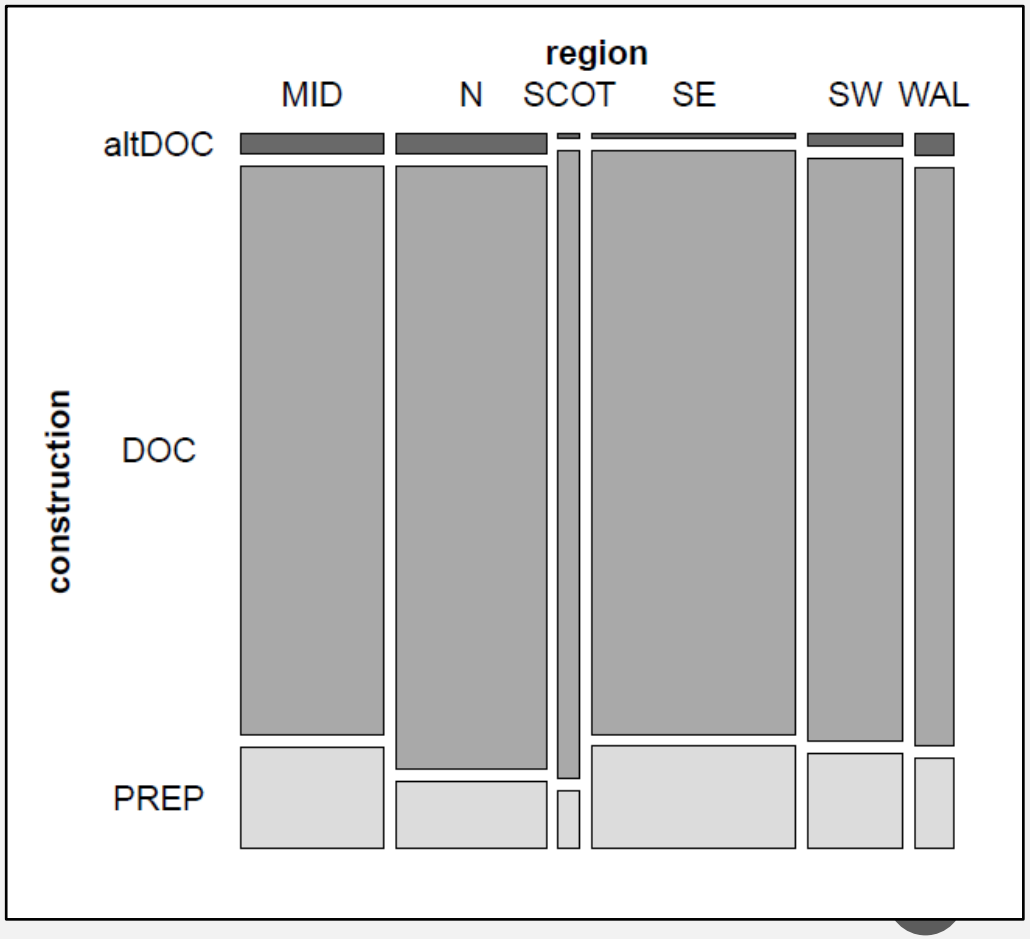


APPENDIX

REGIONAL DISTRIBUTION OF DITRANS. PATTERNS IN FRED

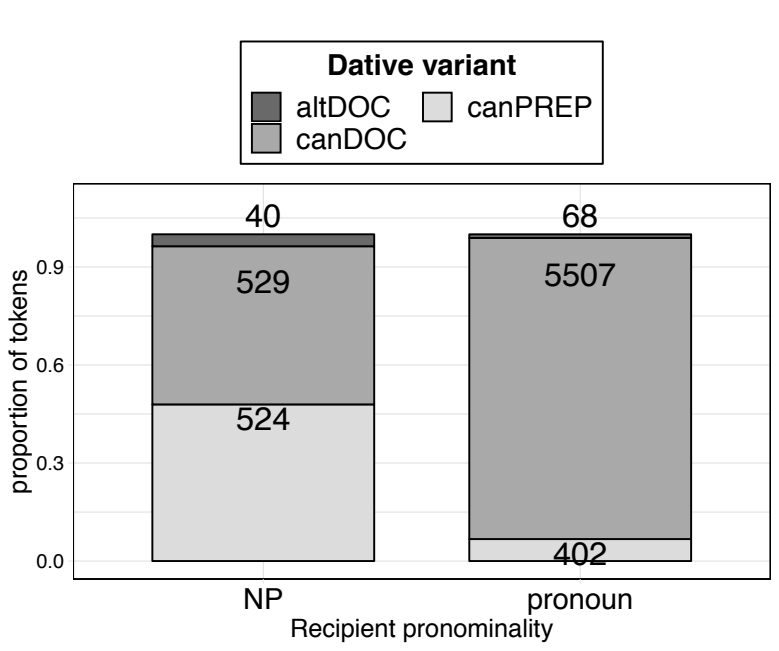


… AND IN THE BNC$_{REG}$

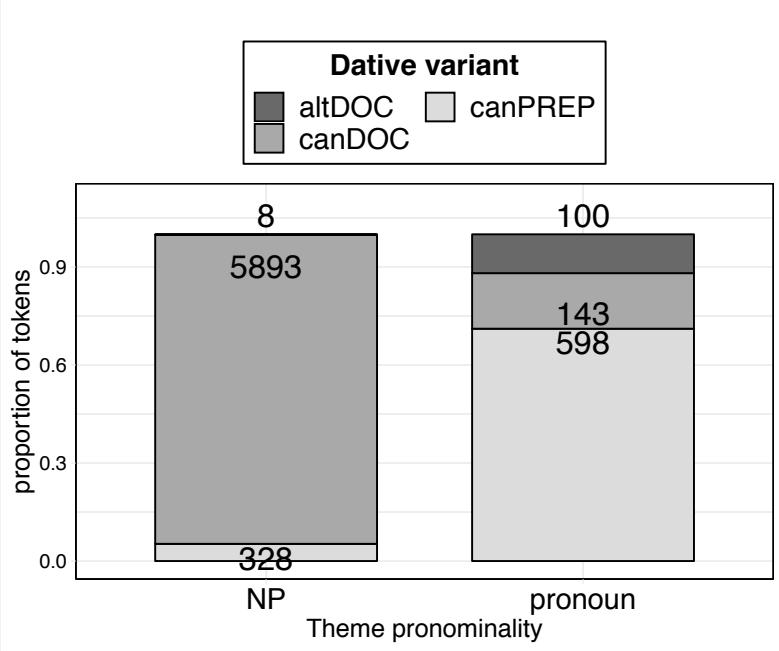


APPENDIX

Proportional distributions by recipient pronominality

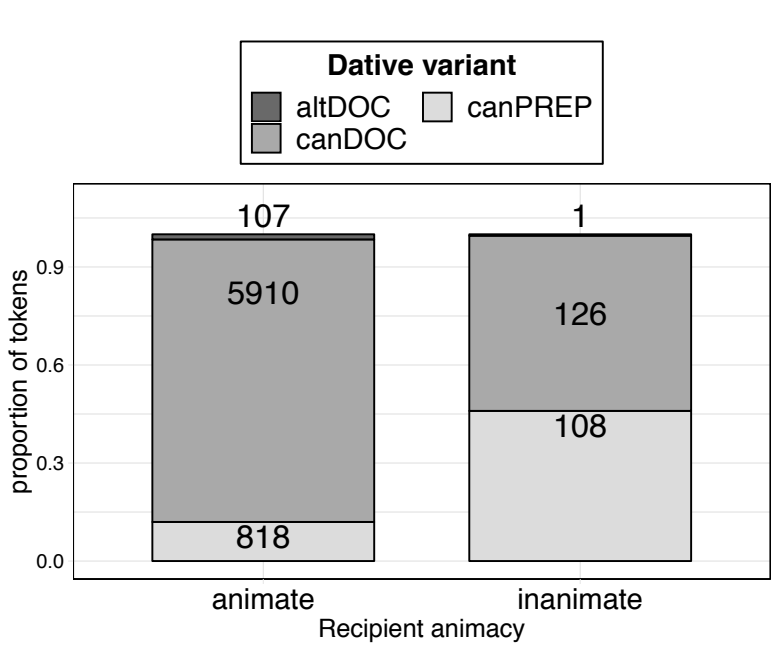


Proportional distributions by theme pronominalilty

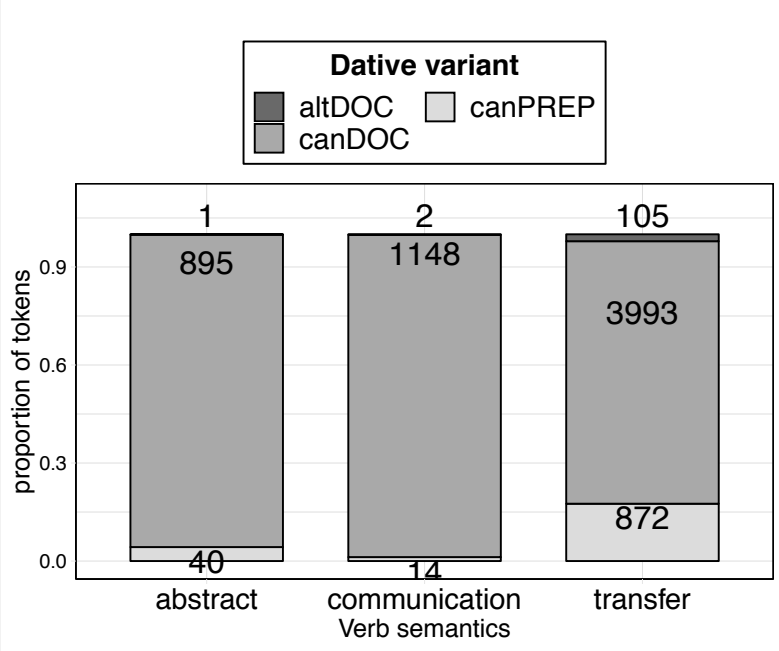


APPENDIX

Proportional distributions by recipient animacy



Proportional distributions by verb semantics



APPENDIX

Proportional distributions by weight ratio