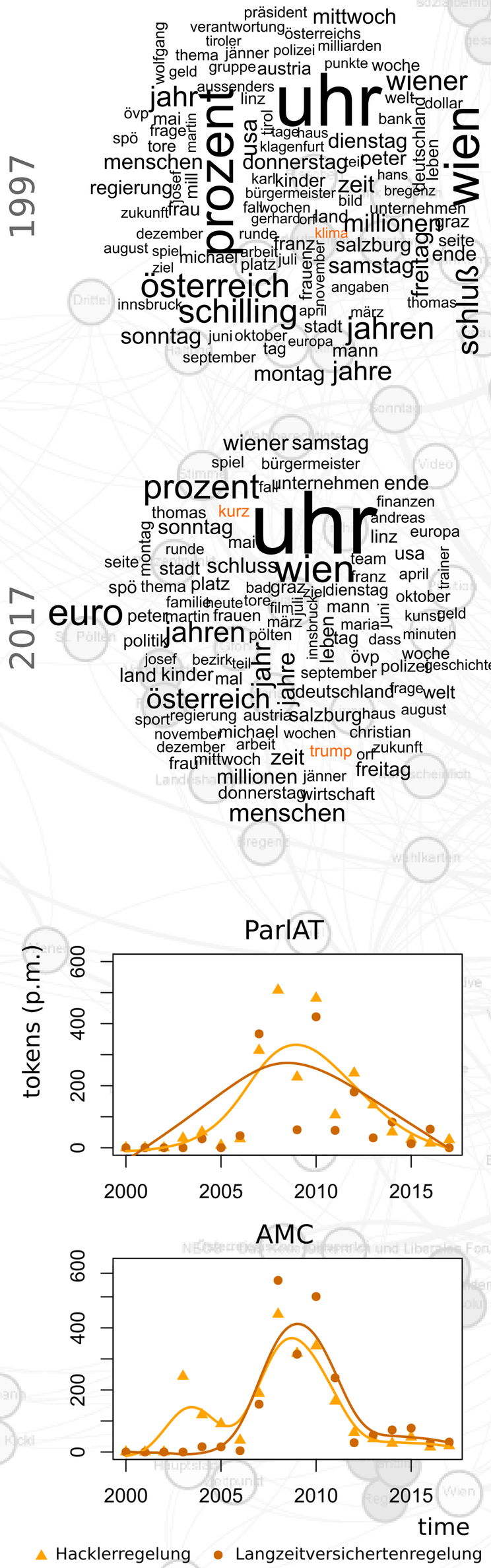
Language, Data & Knowledge 2019 | 20-23 May 2019, Leipzig, Germany

Diachronic dynamics of lexical networks: A roadmap

Andreas Baumann[1,4], Julia Neidhardt[2], Tanja Wissik[3]



Languages change

...but to what extent is lexical change driven by individuals such as politicians?

We analyze their effect on the propagation of lexical innovations.

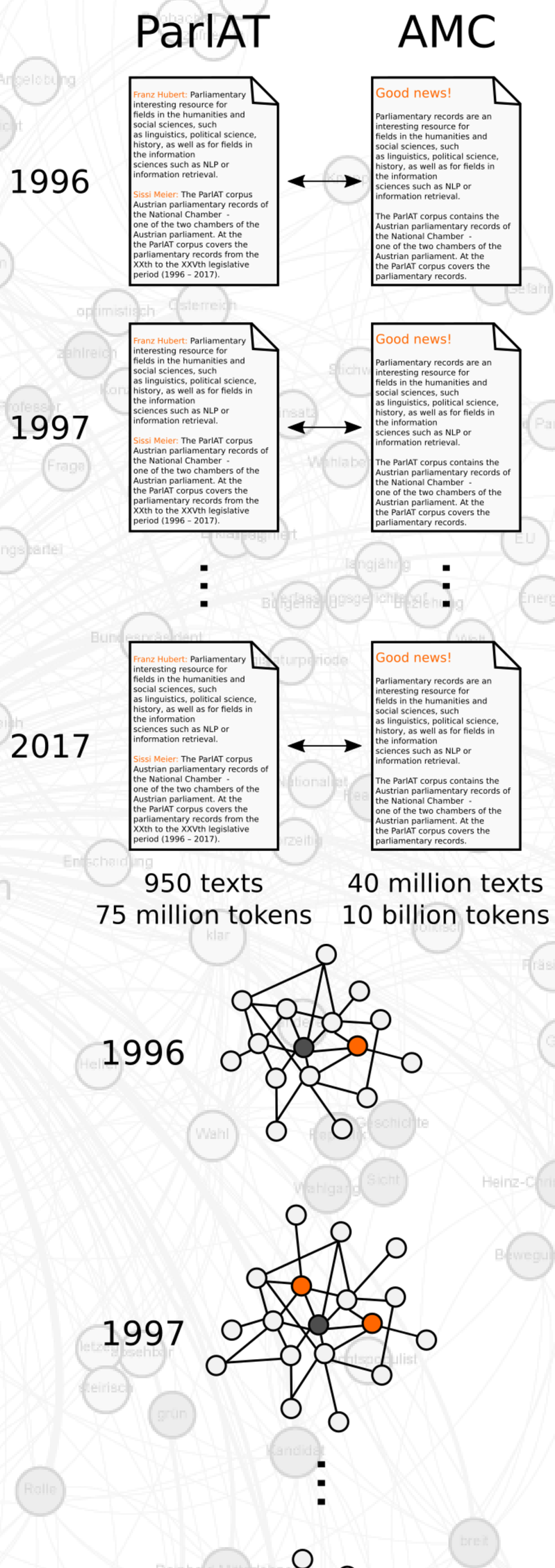
More questions:

- What determines the successful spread of lexical innovations?
- Can we disentangle social factors from cognitive factors (e.g. frequency) in language change?

Our data

We work with two diachronic text corpora of Austrian German:

- ParlAT: parlamentiary records
- AMC: Austrian media texts



Preparations

We analyze frequency trajectories to identify lexical innovations, and we use named entity recognition (NER) to identify innovators.

Methods & tools:

- NER: spaCy enhanced with Wikidata
- Trajectories: time-series modeling (GAM; NLS; growth rates)

Lexical networks

We build co-occurence networks to see if innovations detach from their innovators and start to get used in other contexts as well.

Methods & tools:

- Co-occurrence networks (sentence/paragraph/text)
- igraph, Neo4j, Gephi
- Time-series modeling of network parameters

Background image | Lexical network around NE 'Alexander van der Bellen' (by Grill, 2018)

[1] Department of English and American Studies, University of Vienna; andreas.baumann@univie.ac.at
[2] Faculty of Informatics, TU Wien; julia.neidhardt@ec.tuwien.ac.at
[3] Austrian Centre for Digital Humanities, Austrian Academy of Sciences; tanja.wissik@oeaw.ac.at
[4] ITSV GmbH, Austrian Social Security

Funding: ÖAW go!digital Next Generation grant (GDNG 2018-020)

Barabási, A. 2016. Network science. Cambridge: CUP.
Bastian, M., Heymann, S. & Jacomy, M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. Association for the Advancement of Artificial Intelligence.
Bybee, J. 2010. Language, Usage and Cognition. Cambridge: CUP.
Chen, H., Chen, X. & Liu, H. 2018. How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks. PLoS ONE 13(2): e0192545.
Csardi, G. & Nepusz T. 2006. The igraph software package for complex network research, InterJournal, Complex Systems 1695. http://igraph.org.
Ellis, N., O'Donnell, M. & Römer, U. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. Cognitive Linguistics 25(1). 55–98.
Grill, G., Neidhardt, J. & Werthner, H. 2017. Network Analysis on the Austrian Media Corpus. Vienna Young Scientists Symposium, Austria. 128-129.
Hagberg, A., Swart, P. & Chult, D. 2008. Exploring Network Structure, Dynamics, and Function Using NetworkX. Proceedings of the 7th Python in Science Conference.
Hamilton, W., J. Leskovec & D. Jurafsky. 2016a. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. Proc Conf Empir Methods Nat Lang Process. 2116–2121.
Hamilton, W. L., Leskovec, J. & Jurafsky, D. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. arXiv:1605.09096.
Hilpert, M. & Perek, F. 2015. Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts. Linguistics Vanguard 2015; 1(1): 339–350.
Honnibal, M. & Montani, I. 2017. spaCy: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.
Hunter, J. 2007. Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95.
Jones E., Oliphant E., Peterson P. 2001. SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/
Karjus, A., Blythe, R., Kirby, S. & Smith, K. 2018. Identifying linguistic selection and innovation while controlling for cultural drift. Cuskley, C. et al. (eds.) Proc Evolang 12.
Kim, Y., Chiu, Y. I., Hanaki, K., Hegde, D. & Petrov, S. 2014. Temporal analysis of language through neural language models. arXiv:1405.3515.
Montero, P. & Vilar, J. 2014. TSclust: An R Package for Time Series Clustering. Journal of Statistical Software, 62(1), 1-43. URL http://www.jstatsoft.org/v62/i01.
Oliphant, T. 2006. A guide to NumPy, USA: Trelgol Publishing.
Ransmayr, J., Mörth, K. & Ďurčo, M. 2017. AMC (Austrian Media Corpus). In Rech, C. & Dressler, W. U. (eds.). Digitale Methoden der Korpusforschung in Österreich. 27-38. Wien: ÖAW.
Sagi, E., Kaufmann, S. & Clark, B. 2012. Tracing semantic change with Latent Semantic Analysis. In A. Kathryn & J. Robinson (eds.), Current methods in historical semantics, De Gruyter Mouton.
Schakel, A. & Wilson, B. 2015. Measuring word significance using distributed representations of words. arXiv:1508.02297v1.
Shvachko, K., Kuang, H., Radia, S., & Chansler, R. 2010. The hadoop distributed file system. In Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on (pp. 1-10).
Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., ... & Saha, B. 2013. Apache hadoop yarn: Yet another resource negotiator. In Proc. 4th Symposium on Cloud Computing (p. 5).
Vrandečić, D. & Krötzsch, M. 2014. Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85.
Webber, J. 2012. A programmatic introduction to neo4j. In Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity: 217-218. ACM.
Wissik, T. & Pirker, H. 2018. ParlAT beta. Corpus of Austrian Parliamentary Records. Fišer, D. et al. (eds.) Proc LREC2018 WS ParlaCLARIN. 20-23.
Wood, S. 2017. Generalized Additive Models: An Introduction with R (wnd edition). Chapman and Hall/CRC.
Yao, Z., Sun, Y., Ding, W., Rao, N. & Xiong, H. 2018. Dynamic word embeddings for evolving semantic discovery. In Proceedings of the Eleventh ACM Int. Conf. on Web Search and Data Mining: 673-681.
Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. 2010. Spark: Cluster computing with working sets. HotCloud, 10(10-10), 95.

Diachronic Dynamics of Lexical Networks YLEN
ÖAW Österreichische Akademie der Wissenschaften
TU Wien Technische Universität Wien
universität wien