

# Design Space for Diachronic Linguistic Visualizations in Theory and Practice

Christin Schätzle, Annette Hautli-Janisz, Michael Hund, Christian Rohrdantz,  
Miriam Butt and Daniel A. Keim

University of Konstanz, Germany

Beginning with seminal work by Collins et al. (2007), Keim and Oelke (2007), and Collins (2010), visualization methods developed within computer science have been introduced to work on a small but growing range of linguistic problems. These range, for example, from visualizing linguistic features across languages, e.g. vowel harmony (Mayer et al. 2010), syntactic trees (Culy et al. 2012), features from WALS according to their genealogy and geographical distribution (Mayer et al. 2014), pitch contours (Sacha et al. 2015, Asano et al. 2016), to discourse analysis (Gold et al. 2015a,b). The existing work has demonstrated that combining insights from Visual Analytics (Thomas and Cook 2005, Keim et al. 2010) with theoretical and computational linguistics offers the potential for ground-breaking new approaches with respect to understanding the complex, multifactorial and high dimensional data that typically underlies linguistic work.

The visualization approach taken at the University of Konstanz is primarily based on the close collaboration between domain experts from Visual Analytics (VA) and Linguistics. The effort is both iterative and collaborative, merging knowledge with data-driven modeling in order to provide the most optimal overview of and access to the data. This requires input from both domains: The linguist first has to select/generate a corpus suited for the analysis task. Then, relevant data dimensions are extracted and brought into a format which can be processed by the visualization (agreed upon with VA). Visual Analytics needs to establish design parameters that represent the data best (e.g. scatter plots, glyphs, histograms) and find statistical measures to uncover significant patterns. Those have to be tailored to the analysis task and the analyst's need. Moreover, appropriate interaction techniques granting an exploratory access to the data have to be integrated.

In the talk, our focus lies on the discussion of a generalized design space for diachronic visualizations. This design space, we claim, is valid for diachronic visualizations in general and can be used as a guideline for further research in this area, specifically with respect to how the type of data and the research questions related to it determine the design of the visual analysis system. We will provide successful examples resulting from our collaboration with respect to the design decisions inherent in this framework, ranging from the visualization of semantic change in English newspaper data to the visualization of syntactic change in Icelandic, focusing in particular on a novel approach that efficiently identifies subspaces within a larger space of data dimensions when dealing with complex historical linguistic data. In the end, Linguistics arrives at a deeper understanding of language and Visual Analytics is faced with novel interesting challenges and driven to extend existing methods or develop new approaches to solve the novel tasks.

## References

- Asano, Yuki, Gubian, Michele and Sacha, Dominik. 2016. Cutting down on manual pitch contour annotation using data modeling. In *Proceedings of the 8th International Conference on Speech Prosody, Boston, USA*.
- Collins, Christopher. 2010. *Interactive visualizations of natural language*. Ph.D.thesis, Toronto, ON: University of Toronto.
- Collins, Christopher, Carpendale, Sheelagh and Penn, Gerald. 2007. Visualization of uncertainty in lattices to support decision-making. *EuroVis07: Joint Eurographics – IEEE VGTC Symposium on Visualization* pages 51–58.

- Culy, Chris, Dima, Corina and Dima, Emanuel. 2012. Through the Looking Glass: Two Approaches to Visualizing Linguistic Syntax Trees. In *2012 16th International Conference on Information Visualisation*, pages 214–219.
- Gold, Valentin, El-Assady, Mennatallah, Bögel, Tina, Rohrdantz, Christian, Butt, Miriam, Holzinger, Katharina and Keim, Daniel. 2015a. Visual Linguistic Analysis of Political Discussions: Measuring Deliberative Quality. *Digital Scholarship in the Humanities* DOI: <http://dx.doi.org/10.1093/llc/fqv033>.
- Gold, Valentin, Rohrdantz, Christian and El-Assady, Mennatallah. 2015b. Exploratory Text Analysis using Lexical Episode Plots. In E. Bertini, J. Kennedy and E. Puppo (eds.), *Eurographics Conference on Visualization (EuroVis) - Short Papers*, The Eurographics Association.
- Keim, Daniel A., Kohlhammer, Joern, Ellis, Geoffrey and Mansmann, Florian (eds.). 2010. *Mastering The Information Age - Solving Problems with Visual Analytics*. Goslar: Eurographics.
- Keim, Daniel A. and Oelke, Daniela. 2007. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST '07)*, pages 115–122, IEEE Computer Society.
- Mayer, Thomas, Rohrdantz, Christian, Butt, Miriam, Plank, Frans and Keim, Daniel A. 2010. Visualizing Vowel Harmony. *Linguistic Issues in Language Technology* 4(Issue 2), 1–33.
- Mayer, Thomas, Wälchli, Bernhard, Hund, Michael and Rohrdantz, Christian. 2014. From the extraction of continuous features in parallel texts to visual analytics of heterogeneous areal-typological datasets. *Language Processing and Grammars. The role of functionally oriented computational models* pages 13–38.
- Sacha, Dominik, Asano, Yuki, Rohrdantz, Christian, Hamborg, Felix, Keim, Daniel A., Braun, Bettina and Butt, Miriam. 2015. Self Organizing Maps for the Visual Analysis of Pitch Contours. *Proceedings of the 20th Nordic Conference of Computational Linguistics. Vilnius, Lithuania* 23, 181–190, [www.aclweb.org/anthology/W15-1823](http://www.aclweb.org/anthology/W15-1823).
- Thomas, James J. and Cook, Kristin A. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center.