

The March of Data: English Linguistics across Disciplinary Borders

Conveners:

Jukka Tyrkkö (Linnaeus University)

Steven Coats (University of Oulu)

Veronika Laippala (University of Turku)

New approaches in English linguistics have emerged in recent years that blur disciplinary boundaries, facilitated by factors such as increased access to large data sets (e.g., from social media or digitization projects), the application of new methods of data analysis and visualization, the sharing of code and data on platforms such as CLARIN or GitHub, as well as by continual advances in technologies related to data storage, retrieval, and processing.

In this workshop we invite papers that focus on new developments and cross-disciplinary approaches in the landscape of contemporary English linguistics. Potential topics and thematic fields may include:

- 1) Computer-mediated communication (CMC): An ever-increasing proportion of human interaction is mediated by digital technologies. We invite papers that focus on CMC data, especially in the context of computational sociolinguistics (Nguyen et al. 2016), such as forums, blogs, newsgroups, SMS and WhatsApp messages, text chats, wiki discourse, social media platforms such as Twitter, YouTube, or LinkedIn, pseudo-anonymous “chans” such as 4chan, or streaming platforms such as Twitch or webcam sites (see, e.g., Coats 2017, Gonçalves et al. 2018, Hiiippala et al. 2019)
- 2) Digital humanities: Language data give us insight into processes and developments in specifically linguistic domains such as (e.g.) lexis, grammar, or syntax, but can also shed light on language-mediated aspects of human experience such as culture, history, politics, and economic behavior (see, e.g., Alexander 2016). We welcome papers in which English language data are utilized in order to investigate questions in the Digital Humanities.
- 3) Big data, maps, and visualization: Researchers in English corpus linguistics now often work with large data sets that are annotated with geographic or other metadata, allowing the creation of language maps (see, e.g., Grieve et al 2017), as well as many other types of data visualizations (see, e.g., Hilpert 2011). We invite papers that report on the collection and visualization of English language data, and particularly papers that discuss interactive visualization and mapping tools, as well as new innovative methods such as Virtual Reality and Augmented Reality technologies (see, e.g., Alissandrakis et al. forthcoming).
- 4) Machine learning: How can machine learning and AI be used for the study of English, for example in genre or register classification, or in the preparation and annotation of multimodal language data? What are the current best practices (see, e.g., Gries 2019) and where are we heading? We encourage the submission of papers that utilize machine learning or neural network approaches for the identification, analyses and classification of language data as text or as sound, image and video.

The conveners plan to edit a publication based on the papers presented at the workshop.

References

- Alexander, M. 2016. The metaphorical understanding of power and authority. In: Anderson, W., E. Bramwell & C. Hough (eds.), *Mapping English metaphor through time*, 191–207. Oxford University Press: Oxford.
- Alissandrakis, A., N. Reski, M. Laitinen, J. Tyrkkö & M. Levin. 2019. Visualizing rich corpus data using virtual reality. In: Parviainen H., M. Kaunisto & P. Pahta (eds.) *Corpus Approaches into World Englishes and Language Contrasts* (Studies in Variation, Contacts and Change in English XX). Helsinki: VARIENG.
- Coats, S. 2017. European language ecology and bilingualism with English on Twitter. In: Ciara Wigham and Egon Stemle (eds.), *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities*, 35–38. Bozen/Bolzano: Eurac Research.
- Gonçalves, B., L. Loureiro-Porto, J. Ramasco & D. Sánchez. 2018. Mapping the Americanization of English in space and time. *PLoS ONE* 13(5): e0197741.
- Gries, S. Th. 2020. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory* 16(3), 617–647.
- Grieve, J., A. Nini & D. Guo. 2017. Analyzing lexical emergence in American English online. *English Language and Linguistics* 21, 99–127.
- Hiippala, T., A. Hausmann, H. Tenkanen & T. Toivonen. 2019. Exploring the linguistic landscape of geotagged social media content in urban environments. *Digital Scholarship in the Humanities* 34(2), 290–309.
- Hilpert, M. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4), 435–461.
- Nguyen, D., A. Doğruöz, C. Rosé & F. de Jong. 2016. Computational sociolinguistics: A survey. *Computational Linguistics* 42(3), 537–593.