

Sociolinguistic variation in the history of English: Productivity, neologisms and gendered styles

Dr. Tanja Säily (University of Helsinki)

In this talk I will discuss three interrelated areas of my ongoing work.

Firstly, I study sociolinguistic variation and change in derivational and inflectional productivity. My results suggest that there is similar variation in both, which provides empirical support for the hypothesis of a cline between the two (Säily 2014; Säily et al. 2018). I have co-developed a methodology for comparing productivity across subcorpora based on type frequencies and the statistical technique of permutation testing (Säily & Suomela 2017). In my ongoing study of the derivational suffix *-er*, I am experimenting with entropy-based measures of productivity using an interactive tool created by my collaborator Jukka Suomela.

My second focus is on lexical productivity and creativity. In the Academy of Finland funded project *Interfacing structured and unstructured data in sociolinguistic research on language change* (STRATAS), we are developing an open-source environment for analysing neologisms that combines the *Corpora of Early English Correspondence* with lexicographical sources and contemporary published texts (Säily et al. in press; Hämäläinen et al. 2018). In addition to providing new methods for spelling normalization and the study of lexical innovation in historical corpora, this work sheds more light on variation across social groups in the use and diffusion of new lexis and on the semantics and social meanings of neologisms in the history of English.

Thirdly, I am interested in sociolinguistic variation and change in part-of-speech frequencies, particularly noun and pronoun frequencies, in the history of English. My results support the hypothesis of gendered styles previously formulated on the basis of Present-day English data, while also challenging the notion by discovering intra-gender variation based on the specific social roles of the interactants (Säily et al. 2011; Vartiainen et al. 2013); moreover, I have been able to antedate previous results on the colloquialization process of written English (Säily et al. 2017). I have recently completed a project where we POS-tagged the 18th-century *Corpus of Early English Correspondence Extension*, and I am planning to use this corpus to continue my exploration of what POS frequencies can tell us about macro-level variation and change.

References

- Hämäläinen, Mika, Tanja Säily, Jack Rueter, Jörg Tiedemann & Eetu Mäkelä. 2018. Normalizing early English letters to Present-day English spelling. Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva, Nils Reiter & Stan Szpakowicz (eds.), *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (ACL Anthology W18-45), 87–96. Stroudsburg, PA: Association for Computational Linguistics.
<https://aclweb.org/anthology/W18-4510>
- Säily, Tanja. 2014. *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics* (Mémoires de la Société Néophilologique de Helsinki XCIV). Helsinki: Société Néophilologique.
<http://urn.fi/URN:ISBN:978-951-9040-50-9>
- Säily, Tanja, Victorina González-Díaz & Jukka Suomela. 2018. Variation in the productivity of adjective comparison in Present-day English. Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus Approaches to Contemporary British Speech: Sociolinguistic Studies of the Spoken BNC2014* (Routledge Advances in Corpus Linguistics 21), 159–184. Abingdon: Routledge.
- Säily, Tanja, Eetu Mäkelä & Mika Hämäläinen. In press. Explorations into the social contexts of neologism use in early English correspondence. *Pragmatics & Cognition*, special issue on the dynamics of lexical innovation.
- Säily, Tanja, Terttu Nevalainen & Harri Siirtola. 2011. Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing* 26(2): 167–188. <https://doi.org/10.1093/lc/fqr004>
- Säily, Tanja & Jukka Suomela. 2017. *types2*: Exploring word-frequency differences in corpora. Turo Hiltunen, Joe McVeigh & Tanja Säily (eds.), *Big and Rich Data in English Corpus Linguistics: Methods and Explorations* (Studies in Variation, Contacts and Change in English 19). Helsinki: VARIENG.
http://www.helsinki.fi/varieng/series/volumes/19/saily_suomela/
- Säily, Tanja, Turo Vartiainen & Harri Siirtola. 2017. Exploring part-of-speech frequencies in a sociohistorical corpus of English. Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds.), *Exploring Future Paths for Historical Sociolinguistics* (Advances in Historical Sociolinguistics 7), 23–52. Amsterdam: John Benjamins.
<https://doi.org/10.1075/ahs.7.02sai>
- Vartiainen, Turo, Tanja Säily & Mikko Hakala. 2013. Variation in pronoun frequencies in early English letters: Gender-based or relationship-based? Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex Philologia Lux: Essays in Honour of Leena Kahlas-Tarkka* (Mémoires de la Société Néophilologique de Helsinki XC), 233–255. Helsinki: Société Néophilologique. <http://hdl.handle.net/10138/155183>