

New methods for old languages: the comparability of data

W13, ICHL26, Heidelberg

Alessia Cassarà (Universität Stuttgart); Lena Kaltenbach (Universität Mannheim);
Mariapaola Piccione (Universität Stuttgart); Tara Struik (Universität Mannheim)

Corresponding convenor: tara.struik@uni-mannheim.de

While historical linguistics is traditionally known to suffer from a “bad data problem” (Labov 1994: 11), the field has seen a surge in the development of (annotated) data collections and computational tools to trace quantitative changes throughout the history of languages, allowing researchers to get more out of the (often sparse) data than ever before. This availability of data opens up many new avenues for research, in particular in explaining the cognitive mechanisms behind language change. In this workshop we want to bring together researchers working in different disciplines to discuss novel empirical methods that allow us to investigate the relation between the structural changes we observe in historical texts and the factors which arguably led to these changes. We aim to do this by focusing on a) how historical corpus data can be related to **models of language learning**, b) **contemporary psycholinguistic models** and c) how we can deal with the **heterogeneity of historical data** in relation to these models.

Historical linguists have discussed the link between historical change and changes in the input and have proposed models that make use of psycholinguistic explanations, especially in terms of language acquisition (e.g. Lightfoot 1999, 2017). However, a challenge for the study of the role of language acquisition in language change is that there is no direct access to the input for past stages of languages. Approximating the input by using corpora of child-directed speech (e.g. CHILDES) for contemporary languages has resulted in the development of learning models, which may also be informative for the historical stages. For instance, Yang’s (2016) Tolerance Principle has been shown to work effectively with small amounts of data, making it very attractive for historical work (Kodner 2020, 2022; Drescher and Lahiri 2022, Ringe and Yang 2022, Trips and Rainsford 2022). However, the application of such models on corpus data requires careful consideration of how the data obtained from corpora can be compared to the input a child received (cf. Trips and Rainsford 2022 for discussion). One potential solution is to compare the frequencies of the most common verbs in a corpus to the most common verbs in a sample of child-directed speech, as Kodner (2019) demonstrates that there is a substantial overlap.

From both a psycholinguistic and historical linguistic perspective the relationship between language change and mechanisms of language processing has only rarely been explicitly addressed (for exceptions, see Jäger & Rosenbach 2008; de Smet & de Velde 2017; see also the contributions in Hundt et al. 2017 and the ongoing work by the the DFG Research Unit SILPAC (FOR 5157)). Notably, some authors have recently pointed to the importance of cross-linguistic and within-language structural priming and syntactic adaptation for studies

of (contact-induced) language change (e.g. Pickering & Garrod 2017; Kaan & Chun 2018; Kootstra & Şahin 2018; Kootstra & Muysken 2019). Effects of priming may be observable in historical corpora in the form of persistence of linguistic forms (see Ecay and Tamminga 2017; also Gries 2005; Szmrecsanyi 2006). From a Uniformitarian perspective (see Bergs 2012, Walkden 2019 for discussion), it follows that psycholinguistic processes active in language change should not differ fundamentally across languages or language stages. Methodologically, changes observed in diachrony could in principle also be elicited in psycholinguistic experiments and the results and methods of psycholinguistic experiments could inform historical corpus analyses.

Applying psycholinguistic methods and learning models to historical data also requires us to think critically about the nature of our data and how informative they are about the actual linguistic environment in which language acquisition and change takes place. Historical corpora may be heterogeneous in nature, consisting of many different genres (e.g. legal prose, narrative verse, etc.), which may not all be equally representative of a language user's input. Some types of text, e.g. theatrical texts, conversation manuals, direct speech in verse narratives, etc. have been argued to be particularly close to spoken language in the past (Ernst 1980, Ayres-Bennett 2000); also, it has been shown that language change does not proceed at the same rate in all text genres (Whitt 2018). However, it is not clear whether a restrictive approach to selecting corpus texts is preferable to one which instead draws on as much data as possible, using statistical techniques to evaluate the effect of genre. A further open question is the extent to which the writers of historical texts are themselves influenced by mechanisms such as priming, whether it is self-priming within a single text, between the two writers in private correspondence or even between two languages in translations. Similarly, it is not always clear what the impact of the linguistic background of individual authors is on the output – are they, for instance, monolinguals, early bilinguals, or possibly late bilinguals writing in their first language or late bilinguals writing in their second language?

In this workshop, we aim to compare different types of historical corpus data not only with each other, but also with the input to language acquirers and with data elicited in psycholinguistic experiments in order to develop novel methodologies bringing the fields of historical linguistics, psycholinguistics and language acquisition closer together. We invite contributions which answer or relate to the following research questions and topics:

- How can models of learnability be applied to historical data?
- What are the psycholinguistic processes behind historical language change?
- Which insights does historical linguistics provide for the study of these psycholinguistic processes?
- Which methods and resources are the best to use if we want to relate historical data to language learner input and which are best for researching the relationship between experimental data and historical data?
- Which additional data types/methodologies can contribute to bridging the gap between the disciplines of historical linguistics, acquisition studies and psycholinguistics, e.g. artificial language studies, longitudinal studies, computational models of language change, etc.?

- How can insights from historical sociolinguistics and philology contribute to a better understanding of the heterogeneity of historical corpus data and the linguistic background of individual authors?
- To what extent are the writers of historical texts themselves influenced by mechanisms of language processing, such as intra- and interindividual priming in monolingual and bilingual situations? How can we use notions such as persistence in historical corpora to tap into the cognitive processes behind the text production of medieval authors?

References

- Ayres-Bennett, W. 2000. Voices from the past: Sources of seventeenth-century spoken French. *Romanische Forschungen*, 112, 323–348.
- Bergs, A. 2012. The Uniformitarian Principle and the risk of anachronisms in language and social history. In J. M. Hernández-Campoy & J. C. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics* (1st ed., pp. 83–101). <https://doi.org/10.1002/9781118257227.ch5>
- De Smet, H., & Van de Velde, F. 2017. Experimenting on the past: A case study on changing analysability in English ly-adverbs. *English Language & Linguistics*, 21(2), 317–340. <https://doi.org/10.1017/S1360674317000168>
- Dresher, B. E., & Lahiri, A. 2022. The foot in the history of English: Challenges to metrical coherence. In B. Los, C. Cowie, P. Honeybone, & G. Trousdale (Eds.), *English historical linguistics: Change in structure and meaning: Papers from the XXth ICEHL* (pp. 42–59). John Benjamins Publishing Company.
- Ecay, A., & Tamminga, M. 2017. Persistence as a diagnostic of grammatical status: The case of Middle English negation. In E. Mathieu & Truswell, B. (Eds.). *Micro-change and Macro-change in diachronic syntax* (pp. 202–215). Oxford University Press. <https://doi.org/10.1093/oso/9780198747840.003.0013>
- Ernst, G. 1980. Prolegomena zu einer Geschichte des gesprochenen Französisch. In H. Stimm (Ed.), *Zur Geschichte des gesprochenen Französisch und zur Sprachlenkung im Gegenwartsfranzösischen* (pp. 1–14). Steiner.
- Gries, S. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.
- Hundt, M., Mollin, S., & Pfenninger, S. (Eds.). 2017. *The changing English language: Psycholinguistic perspectives*. Cambridge University Press.
- Jäger, G., & Rosenbach, A. 2008. Priming as a testing ground for historical linguists? - A reply to Chang, Eckardt, and Traugott. *Theoretical Linguistics*, 34(2), 143–156.
- Kaan, E., & Chun, E. 2018. Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, 21(2), 228–242. <https://doi.org/10.1017/S1366728916001231>
- Kodner, J. 2019. Estimating child linguistic experience from historical corpora. *Glossa. A Journal of General Linguistics*, 4(1), 122.
- Kodner, J. 2020. *Language acquisition in the past* [Doctoral dissertation]. University of Pennsylvania.
- Kodner, J. (2022). What learning Latin verbal morphology tells us about morphological theory. *Natural Language and Linguistic Theory*. <https://doi.org/10.1007/s11049-022-09553-2>

- Kootstra, G. J., & Muysken, P. 2019. Structural priming, levels of awareness, and agency in contact-induced language change. *Languages*, 4(3), 65. <https://doi.org/10.3390/languages4030065>
- Kootstra, G. J., & Şahin, H. 2018. Crosslinguistic structural priming as a mechanism of contact-induced language change: Evidence from Papiamentto-Dutch bilinguals in Aruba and the Netherlands. *Language*, 94(4), 902–930.
- Labov, W. 1994. *Principles of linguistic change, Volume 1: Internal factors*. Blackwell.
- Lightfoot, D. 1999. *The development of language: Acquisition, change, and evolution*. Blackwell.
- Lightfoot, D. 2017. Acquisition and learnability. In A. Ledgeway & I. Roberts (Eds.), *The Cambridge handbook of historical syntax* (pp. 381–400). Cambridge University Press.
- Pickering, M., & Garrod, S. 2017. Priming and language change. In M. Hundt, S. Mollin, & S. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 173–190). Cambridge University Press.
- Ringe, D., & Yang, C. 2022. The threshold of productivity and the ‘irregularization’ of verbs in Early Modern English. In B. Los, C. Cowie, P. Honeybone, & G. Trousdale (Eds.), *English historical linguistics: Change in structure and meaning: Papers from the XXth ICEHL* (pp. 91–112). John Benjamins Publishing Company.
- Szmrecsanyi, B. 2006. *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Mouton de Gruyter.
- Trips, C., & Rainsford, T. 2022. Tolerating subject-Experiencers? Yang’s Tolerance Principle applied to psych verbs under contact in Middle English. *Journal of Historical Syntax*, 6(12), 1–43.
- Walkden, G. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics*, 4(1), 1. <https://doi.org/10.5334/gjgl.888>
- Whitt, R. J. (Ed.). 2018. *Diachronic corpora, genre, and language change*. John Benjamins Publishing Company.
- Yang, C. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.