

Studying linguistic innovation with online data

in-person course on Monday (all day) and Tuesday morning by
Jack Grieve

Short CV: Jack Grieve is a Professor of Corpus Linguistics at the University of Birmingham. His research involves the computational analysis of large corpora of natural language to understand language variation and change, especially in the English language. He also conducts research in applied linguistics, including authorship analysis, forensic linguistics, and fake news.



Course description:

In this course, we will consider how large corpora of social media language can be used to better understand linguistic innovation. Although this is one of the basic goals of sociolinguistic inquiry, linguistic innovation has been difficult to study because innovations are necessarily rare when they first emerge and generally unknown to linguists. Working with very large and informal social media corpora offers a solution to this longstanding problem, allowing linguistic innovation to be empirically analysed in far greater detail than has previously been possible. To illustrate this approach, I will present a series of three studies that I have conducted with colleagues that examine both contemporary and historical forms of lexical and grammatical innovation in British and American English using multi-billion-word corpora of geo-located Twitter data.

First, we will discuss previously published research (see Grieve, Nini, and Guo, 2017, 2018) on how lexical innovation (e.g., *fleek*, *xans*) can be identified and mapped based on large time-stamped and geo-located corpora of natural language. Through the quantitative analysis of a large corpus of American Twitter data, I will show how we can draw generalisations about lexical innovation in American English, including identifying common word formation processes, functional motivations for the actuation and spread of innovations, and origins of lexical innovations in modern American English, arguing that these words often come from Black communities in the southern US.

Second, we will discuss research that is currently in review (with Cameron Morin) on identifying and mapping double modal constructions (e.g., *might can*, *might could*) in American English based on this same corpus. I show not only how the large-scale analysis of contemporary variation in these highly elusive forms can be pursued using social media corpora, but how the results obtained challenge many traditional theories concerning the nature of double modals in American English, including their inventory, their distribution, and, most notably, their origins. I argue that these forms appear to be innovations from Black slaves, and possibly evidence of the creole origins of African American Language, as opposed to being introduced by Scot Irish settlers, as is commonly believed.

Third, we will discuss research that is currently in review (with Christian Ilbury and Dave Hall) on the spread of lexis associated with Multicultural London English (MLE) across the UK based on a large geo-located corpus of British Twitter data (see Grieve et al. 2019 for more information on the corpus). In this case, I will challenge the growing scholarly consensus that MLE is a general urban and multicultural phenomenon and argue that MLE in fact appears to originate in Black communities in London and to have spread within England primarily through

Black social networks, especially from London through Luton and Northampton into Birmingham.

Finally, to conclude we will consider what these three studies, taken together, tell us about linguistic innovation more generally. Specifically, we will consider why all three of these studies have identified Black speakers of the English language as the primary sources of linguistic innovation. I will argue that the expression of social meaning by social groups who have been discriminated against is an important driver of the actuation and spread of linguistic innovation, at least in the English language.

References

- Jack Grieve, Chris Montgomery, Andrea Nini, Akira Murakami and Diangshen Guo. 2019. Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*. <https://www.frontiersin.org/articles/10.3389/frai.2019.00011/full>
- Jack Grieve, Andrea Nini and Diangsheng Guo. 2017. Analyzing lexical emergence in American English online. *English Language and Linguistics* 21: 99–127. <https://doi.org/10.1017/S1360674316000113>
- Jack Grieve, Andrea Nini and Diangsheng Guo. 2018 Mapping lexical innovation on American social media. *Journal of English Linguistics* 46(4): 293–319. <https://doi.org/10.1177/0075424218793191>