How Nordic is Nordic tweeting?
Big Data perspectives to Online Multilingualism

In recent years, the scope of multilingualism research has been expanding from spoken to written media, and the theoretical landscape has been increasingly open to fluid and resource-centred approaches that view code-switching and translanguaging as common means of constructing and projecting metrolingual identities, rather than as signs of lacking language skills or as non-standard features (see Emi and Pennycook 2010). These practices are particularly visible in social media, which now rival superdiverse urban population centres as sites of authentic, unedited and multilingual communication (Lee 2017). Consequently, Twitter and blog data have become an important source of near-realtime linguistic data on a massive scale for studies focused on dialects, World Englishes and multilingualism (see, e.g., Grieve et al 2017; Kopaczyk and Tyrkkö 2018).

Since November 2016, the Nordic Tweet Stream (NTS) project at Linnaeus University has collected all geolocated tweets posted in the Nordic countries (Sweden, Norway, Finland, Denmark and Iceland). The current corpus comprises c. 17 million tweets, as well as diverse metadata provided by Twitter (see Laitinen et al 2017). In each country, several dozen different languages are regularly used in tweets, with non-indigenous languages like English, Spanish and Arabic making up a considerable proportion of the tweet stream. Translanguaging is likewise evident in code-switching within individual tweets and the streams of individual Twitter users.

In this paper, I will first present an overview of the NTS project, the data and metadata, current research questions and our forthcoming open-access resources. I will then discuss the major trends in Nordic multilingual tweeting in terms of regional variation, language practices within individual twitter accounts and code-switching within individual tweets applying cartographic data visualisation methods and pattern-driven approaches (Tyrkkö and Kopaczyk 2018).

References

Grieve Jack, Andrea Nini and Diansheng Guo. 2017. Analyzing lexical emergence in American English online. In *English Language and Linguistics* 21: 99-127.

Kopaczyk Joanna and Jukka Tyrkkö. 2018. Blogging around the world: Universal and localised patterns in Online Englishes". In Kopaczyk, Joanna and Jukka Tyrkkö (eds.) *Applications of pattern-driven methods in corpus linguistics*. (Studies in Corpus Linguistics 82). Amsterdam and New York: John Benjamins. 277–310.

Laitinen Mikko, Lundberg Jonas, Levin Magnus, Lakaw Alexander. 2017. Revisiting weak ties: Using present-day social media data in variationist studies. In Säily Tanja, Palander-Collin Minna, Nurmi Arja, Auer Anita (eds.) *Exploring Future Paths for Historical Sociolinguistics*. Amsterdam: John Benjamins. 303–325.

Lee, Carmen. 2017. *Multilingualism Online*. London: Routledge.

Otsuji, Emi and Alistair Pennycook. 2019. Metrolingualism: fixity, fluidity and language in flux. In *International Journal of Multilingualism*. Volume 7, Issue 3. 240–254.

Tyrkkö Jukka and Joanna Kopaczyk. 2018. Corpus-driven methods: present applications and future directions. In Kopaczyk, Joanna and Jukka Tyrkkö (eds.) *Applications of pattern-driven methods in corpus linguistics*. (Studies in Corpus Linguistics 82). Amsterdam and New York: John Benjamins. 1–12.