

Topics in Dutch Minority Languages on Twitter

Anna Katrine Jørgensen

Meertens Instituut*

anna.jorgensen@meertens.knaw.nl

Lysbeth Jongbloed-Faber

Fryske Akademy*

ljongbloed@fryske-akademy.nl

Jolie van Loo

Meertens Instituut*

j.j.e.vanloo@gmail.com

In this paper, we focus on the use of two Dutch minority languages, Frisian and Limburgian, on Twitter and the effect of topics on choosing language. Almost nine out of ten Dutch people are active on social media, with Twitter claiming 3.5 million tweeters from the Netherlands [2]. Furthermore, over 10% of tweeters tweet in more than one language [3]. Frisian is the mother tongue of approximately half of the Frisian population. The majority of the 650,000 inhabitants of Fryslân can understand the language (very) well (85%), 64% of the population can speak it (very) well, while only 12% indicate that they can write it (very) well [4]. Limburgian mainly functions as a spoken language and is highly variational. 70-75% of the population in Limburg can speak the local dialect and speakers from all socio-economic levels do so in a variety of settings [5].

Interestingly, both languages are used on Twitter, despite their mainly spoken quality. A Frisian or Limburgian bi- or multilingual's choice of language and variety from her linguistic repertoire relies on a number of factors, including the audience [6], offline language use with peers [7], attitude [7], writing skills [7], the perspective, the tone of the tweet and the topic [1,6].

This paper presents an interdisciplinary study of the use of topics by Frisian and Limburgian Twitter users. We present a large-scale, rule-based classification of topics in these minority languages, and discourse about the promises and pitfalls of automatised topic classification on Twitter in this setting.

Furthermore we describe the influence of topic on the (de-)selection of a language variety from a tweeters repertoire. The reasons for choosing a minority language for a certain topic are debated and reasons hereof provided. We discuss the differences between the Frisian and the Limburgian distributions of topics, both in their use of Dutch and in the differences between the distributions in the two minority languages.

This interdisciplinary approach of data-driven topic modelling and sociolinguistics provides a detailed overview of the impact of topic on the use of minority languages on Twitter. The strength of the approach is two-fold: the sociolinguistic study is generalisable and relies on the analysis of large data sets, while the output of the rule-based topic classifier is subject to thorough investigation and is applied insights from sociolinguistic theory.

* This paper is the result of the interdisciplinary project 'Twidentity' which involves Meertens Instituut, Universiteit van Twente, University Maastricht and Fryske Akademy. The Twidentity team consists of Anna Katrine Jørgensen, Jolie van Loo, Lysbeth Jongbloed-Faber, Leonie Cornips, Theo Meder, Dong Nguyen and Dolf Trieschnigg.

References

- [1] Androutsopoulos, J. Code-switching in computer-mediated communication. Pragmatics of Computer-mediated Communication. De Gruyter Mouton. 2013.
- [2] Boekee, Steven et al., National Social Media Survey 2014. Newcom Research & Consultancy B.V., Amsterdam 2014.
- [3] Hale, S.A., Global connectivity and multilinguals in the Twitternetwork. Proceedings of CHI 2014.
- [4] Province of Fryslân. Friese Taalatlas: <http://www.fryslan.frl/taalatlas>, 2011.
- [5] The Limburgish Academy Foundation. Limbugs, Modern Usage 2009
<http://www.limburgs.org/en/limburgish/modern-usage>
- [6] Nguyen, Dong et al. Audience and the Use of Minority Languages on Twitter Forthcoming, 2015.
- [7] Jongbloed-Faber, Lysbeth et al., The Impact of Social Media on Language Vitality: Online Practices of Bilingual Teenagers in Fryslân. Forthcoming, 2015.