

Crowd-sourcing Variation in Minority Languages: Illustrated with Frisian

Nanna Haug Hilton

University of Groningen

Adrian Leemann

Lancaster University

Charlotte Gooskens

University of Groningen

Sociolinguistics as a discipline has predominantly focused on large language communities that are monolingual, yet it is generally agreed that theory development depends on insights from other contexts, including lesser-used varieties in multilingual settings (Meyerhoff & Nagy 2008). However, the lack of usable recordings of speech data presents scholars wanting to do variationist sociolinguistic research on such varieties with obvious challenges. *In this paper we present a new methodological tool for crowd-sourcing speech and perception data in endangered languages, and discuss its benefits and limitations on the basis of the example of Frisian spoken in the Netherlands.*

A number of language documentation efforts using smartphone technology have come on the market in recent years (e.g. Bird et al. 2014), relying on crowd-sourcing of speech recordings and users' ability to translate in writing. At the same time, crowd-sourcing of data over the internet has become a popular methodology for dialectological and folk-linguistic research. Ventures employing dialect quizzes in smartphone applications have shown how studies of phonetic and phonological variation can benefit from the addition of crowd-sourced material (Leemann et al. 2016).

Our talk presents the smartphone application 'Stimmen fan Fryslân' (Voices of Fryslân) in its prototype state (full release at the end of 2017). The app combines the approaches above. It contains a perceptual dialectology task (here for Frisian and Dutch language areas), and a picture-naming task (usable for any language) that documents phonological and phonetic variation.

Our presentation also includes a discussion of the caveats of crowd-sourcing variation in a threatened language. We discuss the extent of linguistic variation that can be expected in minority language communities, and the types of extra-linguistic variables that are important to monitor in situations of intense language contact.