

Identifying speakers' dialectal background in Finnish heritage communities of North and South America: a comparative approach

Ilmari Ivaska, Mirva Johnson, Tommi Kurki, and Eeva Sippola

Previous research suggests the Finnish dialectal background of an American Finnish speaker, their parents, or their grandparents is often identifiable in a speaker's speech, despite evidence of contact with English and other dialects of Finnish (Jönsson-Korhola 2003:408–9). We examine these observations with a comparative case study of two communities founded by Finnish migrants in the early 1900s. We focus on results of a supervised machine-learning model's identification of Finnish dialect features present in the Finnish spoken by each community's last generation of heritage speakers, and how the feature distribution relates to Finnish regional variation.

The communities under study are Oulu, Wisconsin, the United States, and Colonia Finlandesa in Misiones, Argentina. Oulu was established in 1904 by mostly Finnish-speaking migrants. The community was rural and had its largest population in 1920 with 1077 residents. Colonia Finlandesa was established in Misiones in 1906 and had its highest population in the 1930s with around 500 people (Sippola 2021:5). Data from each community come from interviews conducted with the community's last generation of active Finnish speakers. Interviews were conducted in Oulu, Wisconsin in 2016 and 2017 and we compare previous analysis of four interviews (from [Author 1, 2, 3]) to recordings from Colonia Finlandesa. Data from Colonia Finlandesa were selected from oral history interviews conducted in the community in 1972 and recently transcribed.

We examine the Finnish dialectal background of these speakers using two supervised machine-learning models we built that 1) predict the probability of an individual to belong to different dialect groups and 2) situate an individual on a map of Finland based on transcripts of their speech. These models were tested using data from the Finnish Dialect Syntax Archive (LAX), whose material was collected in the 1960s from all Finnish dialect regions (see Ikola (1985) for more on the LAX). The model uses a distributive approach to dialect identification, defining usage of a feature (e.g. the 1st sg. pronoun) and the degree to which individuals use different possible variants of that feature (e.g. *mä/mie/miä/minä*). In previous work ([Author 1, 2, 3]) the model successfully identified the dialectal distribution of five second- and third-generation American Finnish speakers from northern Wisconsin (see Figure 1, highlighting the dialectal distribution of a second-generation speaker). We extend usage of the previously vetted model to a group of heritage speakers in a different majority language context.

Work on moribund heritage languages often connects regions from which settlers emigrated and the dialect common to the region at the time (Bousquette & Putnam 2020:202). Some work has been done examining specific dialect features that remain in the speech of heritage speakers of German, Norwegian, and Spanish (ex. Litty, Evans, Salmons 2015). Our preliminary comparison of language contact settings between Finnish and different majority languages also offers a way to estimate the dialectal input a speaker may have received and gives another perspective on determining reasonable baselines for comparison—a longstanding question of interest to the field (e.g. Polinsky & Scontras 2019:4-5).

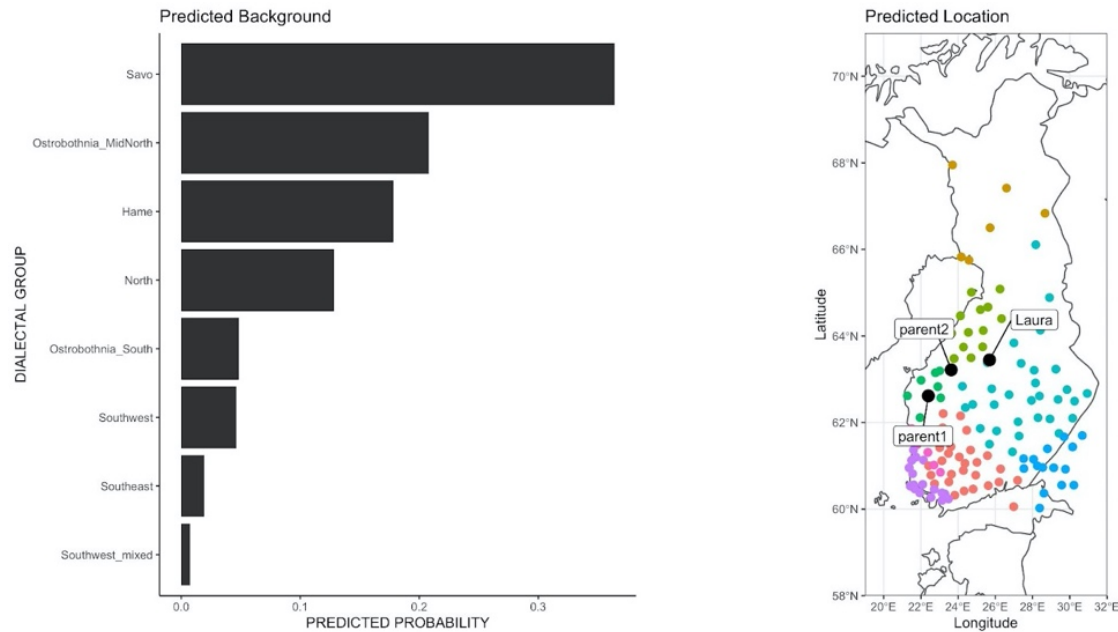


Figure 1: Distribution of Laura (pseudonym)'s dialectal background and predicted geographic location

References:

[Author 1, 2, 3]. Identifying the dialectal background of American Finnish speakers using a supervised machine-learning model. (in prep).

Bousquette, Joshua, & Putnam, Michael T. (2020). Redefining Language Death: Evidence from moribund grammars. *Language Learning*, 70(S1), 188–225. <https://doi.org/10.1111/lang.12362>

Ikola, Osmo (ed.) (1985). *Lauseopin arkiston opas* [Dialect syntax archive guide]. Turku: Lauseopin arkiston julkaisuja.

Jönsson-Korhola, Hannele & Lindgren, Anna-Riitta (eds.). (2003). *Monena suomi maailmalla. Suomalaisperäisiä kielivähemmistöjä* [Diverse Finnish around the globe: Language minorities of Finnish origin] (Tietolipas 190). Helsinki: Suomalaisen Kirjallisuuden Seura.

Litty, Samantha, Christine Evans & Joseph Salmons. (2015). Gray zones: The fluidity of Wisconsin German language and identification. In Peter Rosenberg, Konstanze Jungbluth, Dagna Zinkhahn Rhobodes (eds.), *Linguistic construction of ethnic borders*. Frankfurt: Peter Lang Editions, 183–205.

Polinsky, Maria, & Scontras, Gregory. (2019). Understanding heritage languages. *Bilingualism: Language and Cognition*. 23(1), 4–20. <https://doi.org/10.1017/S1366728919000245>

Sippola, Eeva. (2021). Spanish contact influence in a Finnish heritage community in Misiones, Argentina. *Nordic Journal of Linguistics*. <https://doi.org/10.1017/S033258652100025>