

# Is a heritage variety just a regional variety spoken outside the national boundary?

Naomi Nagy (University of Toronto)

## Abstract

Comparative variationist sociolinguistics examines variable aspects of language, or “different ways of saying the same thing,” by simultaneously considering multiple probabilistic factors that impact speakers’ choices. These include inter-speaker (social) and intra-speaker (linguistic context) predictors, and account for community members’ shared grammars, as well as how we use stochastic information to recognize speakers’ group memberships (cf. Labov et al., 2011). This approach has been applied to many languages, but most ensuing generalizations are based on studies of large well-documented varieties. In this talk, we will navigate the boundary between (analysis of) standard homeland varieties and less-standardized heritage varieties.

For every variable element of a language, a *Probability Matrix* must be acquired by speakers, containing probabilistic information about when each form is (more) appropriate. Sociolinguists model such matrices through multivariate regression analyses that reveal significant predictors (and levels within each predictor). One way to understand differences between the language varieties used by homeland and heritage<sup>1</sup> speakers is to ask how a Probability Matrix compares between Heritage and Homeland speakers. (How) can these fairly be compared? Although there have been some proposals tendered, the variationist field lacks a robust comparative methodology to determine how/if varieties differ. In this talk, I focus on one weakness: different-sized samples are often compared, as it can be harder to find/build a large corpus of a small, under-documented language. This difference in sample size implicates different levels of statistical significance even when the two populations’ patterns are identical.

I illustrate one solution through comparison of variable patterns in Heritage and Homeland Cantonese. Revising analyses conducted previously of two morphosyntactic variables: *prodrop* and *classifiers* (Nagy, 2015; Nagy & Lo, 2019) and applying a bootstrap procedure to mitigate issues associated with unequal-sized datasets frequent in studies of minoritized varieties, I offer a reproducible comparison method (excerpting from Nagy & Gadanidis *fc*). From these analyses, we learn that heritage and homeland grammars’ degrees of complexity are similar: their Probability Matrices are the same size. This approach allows us to consider the complexity of the decision-making process the speakers apply in selecting among forms. As one might expect, heritage and homeland speakers are capable of equally complex processes. This adds another report from the [Heritage Language Variation and Change in Toronto Project](#) that finds little difference between Homeland and Heritage varieties of 10 languages spoken in Toronto, when applying corpus-based rather than experimental methods.

## References

- Labov, W., S. Ash, M. Ravindranath, T. Weldon, M. Baranowski & N. Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15.4:431-63.  
<https://doi.org/10.1111/j.1467-9841.2011.00504.x>

<sup>1</sup> A heritage language is one that is not the broader community’s majority language; a homeland language is.

- Nagy, N. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164:309-27. <https://doi.org/10.1016/j.lingua.2014.04.012>
- Nagy, N. & T. Gadanidis. *fc*. Heritage language variation and change – How complex is it? *Heritage Language Journal*.
- Nagy, N., & Lo, S. 2019. Classifier use in Heritage and Hong Kong Cantonese. *Asia-Pacific Language Variation* 5(1):84-108. <https://doi.org/10.1075/aplv.17001.nag>