# Undergraduate-conducted surveys: Balancing learning and data collection

GERARD VAN HERK
PAUL DEDECKER
*Memorial University of Newfoundland*

JENNIFER THORBURN
*Newcastle University*

Attested pedagogical benefits of undergraduate research include gains in research skills, independence, career preparation and degree completion (Kardash 2000, Seymour et al. 2004, Lopatto 2003, Nagda et al. 1998). But when instructors attempt to scale up projects to satisfy quantitative disciplines like variationist sociolinguistics (e.g., Van Herk 2008), they must balance pedagogical needs (maximizing student involvement) and research imperatives (producing sound and robust findings).

In this paper, we describe our ongoing research project (2008-present, 15 classes) that engages neophyte sociolinguistics students in data collection and analysis, but that can also inform studies at the professional level. Students conduct professor-designed surveys with friends and family, investigating (claimed) usage of up to 41 variables, and collect demographic information about respondents. Professors collate the data; students analyze an aspect of the findings that interests them. To date, we have amassed some 5238 surveys, representing over 1% of our entire speech community (Newfoundland English).

Here, we conduct multiple multivariate analyses of collated data for five variables  interdental stopping in voiced and voiceless contexts (*dat ting for that thing*), non-standard verbal s-marking (*I loves it*), locative *to* (*Where are you to?*), and the traditional lexical item *fousty* ('musty, smelly'). Among our findings: rural respondents favour all traditional variables; the "new verbal -s" described in Childs & Van Herk (2010) clearly grows over apparent time, as does locally-associated but non-stigmatized locative *to*; both types of interdental stopping are male-associated and curvilinear, consistent with dialect revitalization; and *fousty* declines across apparent time.

Survey results reinforce quantitative studies of the same variables in actual use, but also enrich usage data by considering less frequent variables and respondents' full sociolinguistic repertoire. We conclude by discussing the strengths and weaknesses of such data, and how we try to maximize data collection *and* student learning.