

**Title: A large-scale online study of dialect variation in the US Northeast:  
Crowdsourcing with Amazon Mechanical Turk**

**Authors:** Chaeyoon Kim (Dartmouth), Sravana Reddy (Wellesley), Ezra Wyszogrod (Columbia), James Stanford (Dartmouth)

Due to the Founder Effect and the early English colonies (Chambers & Trudgill 1998), the US Northeast has some of the smallest dialect sub-regions in North America (Labov, Ash & Boberg 2006, Kurath 1939, Carver 1987, Nagy 2001, Nagy & Roberts 2004, Roberts 2007, Nagy & Irwin 2010, Wood 2011, Johnson 2011). Can such fine-grained distinctions be observed with an online crowd-sourced survey? What is the current status of New England's sub-regions? As for theory, Carver predicts that new features emerge along the same lines as previous generations (1987:32). Is that happening here?

Hypotheses: (a) The sub-regions of the Northeast can be quickly and clearly observed with a large-scale crowd-sourced survey. (b) New features are emerging along the same geographic lines as earlier generations.

Methods: We used Amazon Mechanical Turk for two online crowdsourcing tasks: a self-reporting survey and an audio recording task. The self-reporting survey involved 535 respondents raised in New England (MA, NH, ME, VT, RI, CT) and also neighboring states (NY, PA, NJ, MD, DE) as a contrast. Our survey had 35 lexical items and 7 phonological features (rhoticity, intrusive-r, PALM/LOT, LOT/THOUGHT, MERRY/MARRY/MARY, NORTH/FORCE, "broad-a" BATH). Example: "When you say *father* and *bother*, do they rhyme?" These traditional Eastern New England (ENE) features were based on prior work listed above, recent new fieldwork, and an online pilot-study. We specifically targeted ENE and its features, unlike prior online surveys (Vaux & Golder 2003; Wood et al. 2015; Bowerman 2010).

Results: The results support our hypotheses. Labov et al. (2006:171) report PALM-fronting in ENE and backing/raising in NYC, i.e., distinct from LOT in both locations. We find the same patterns in our mapped data. Labov et al. (2006) show the low-back merger in ENE and Western Pennsylvania, and our results are similar. /r/-lessness and other features are also consistent with Labov et al. Statistically, our Rbrul logistic regressions (Johnson 2009) on age, sex, and home state show that many well-known ENE features are decreasing in apparent-time, including r-lessness, intrusive-r, PALM/LOT distinction. Many old lexical items have vanished (*rock maple* 'maple-sugar tree' Cassidy 1985[1960s data]), but new regional items are emerging, including *jimmies* 'ice-cream sprinkles' and *wicked-pissah* 'really awesome'.

Examples:

/r/-lessness: Decreasing in apparent-time, log-odds -0.026 per year younger, p=0.027, sex: n.s. Top 4 states for /r/-lessness (Rbrul weights): MA(.944), ME(.936), RI(.907), NH(.868).

*Rotary* for 'traffic circle/roundabout': Decreasing in apparent-time, log-odds -0.037 per year younger, p=0.011, sex: n.s. Top 4 states for *rotary*: MA(.997), RI(.993), ME(.990), VT(.973).

*Jimmies* for 'ice-cream sprinkles': Stable in apparent time. Top 4 states for *jimmies*: MA(.988), NH(.987), ME(.968), RI(.929).

Conclusion from the self-reporting survey: (a) Online crowd-sourcing can quickly and accurately differentiate fine-grained sub-regions. (b) Carver correctly predicted that new generations would

follow the same lines with new features. For newer lexical items, our relatively young “computer-savvy” online respondents still pattern along the old colonial-era dialect lines.

Online audio recordings: We also ran an online task asking respondents to record themselves reading 12 sentences that we designed for automatic vowel analysis with FAVE (Rosenfelder et al. 2011) and DARLA (Reddy & Stanford 2015). Several hundred New Englanders have completed these online recordings. The acoustic sociophonetic results from this second online task will be included in the conference paper.

### **Selected References**

- Boberg, Charles (2001). The phonological status of western New England. *American Speech* 76:3–29.
- Bowern, Claire (2010). North American English dialect survey. Yale University.
- Carver, Craig (1987). American regional dialects. Ann Arbor: University of Michigan.
- Cassidy, Frederick (ed.) (1985). *Dictionary of American regional English*, volume 1. Cambridge: Harvard University Press/Belknap.
- Chambers, J.K. & Peter Trudgill (1998). *Dialectology* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Johnson, Daniel Ezra (2009). Getting off the GoldVarb standard: Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1):359–383.
- Johnson, Daniel Ezra (2010). *Stability and change along a dialect boundary: The low vowels of southeastern New England*. PADS No. 95. Durham, North Carolina: Duke University.
- Kurath, Hans (1939). *Handbook of the linguistic geography of New England*. Providence: Brown.
- Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William (2012). What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics* 10(2):265-93.
- Labov, William, Sharon Ash & Charles Boberg (2006). *The Atlas of North American English*. Berlin: Mouton.
- Nagy, Naomi (2001). “Live Free or Die” as a linguistic principle. *American Speech* 76.1:30-41.
- Nagy, Naomi & Patricia Irwin (2010). Boston (r): Neighbo(r)s nea(r) and fa(r). *Language Variation and Change* 22:241-78.
- Nagy, Naomi & Julie Roberts (2004). New England: Phonology. In Edgar Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds), *Handbook of varieties of English: The Americas and the Caribbean*. Berlin/New York: Mouton de Gruyter.
- Preston, Dennis (1993). The uses of folk linguistics. *International Journal of Applied Linguistics* 3:181-259.
- Preston, Dennis (ed) (1999). *Handbook of perceptual dialectology, vol. 1*. Amsterdam:Benjamins.
- Ravindranath, Maya (2011). A wicked good reason to study intensifiers in New Hampshire. *New Ways of Analyzing Variation* 40, Georgetown University, October 27-30.
- Reddy, Sravana & James Stanford (2015). Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard*.
- Roberts, Julie (2007). Vermont lowering? Raising some questions about /ai/ and /au/ south of the Canadian border. *Language Variation and Change* 19:181-197.
- Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Jiahong Yuan (2011). FAVE (Forced Alignment and Vowel Extraction) program suite.
- Vaux, Bert & Scott Golder (2003). The Harvard Dialect Survey. Cambridge, MA.
- Wood, Jim (2011). Short-a in northern New England. *Journal of English Linguistics* 39(2):135-65.
- Wood, Jim, Laurence Horn, Raffaella Zanuttini & Luke Lindemann (2015). The Southern Dative Presentative meets Mechanical Turk. *American Speech* 90:291-320.