LINGUISTIC STABILITY & MALLEABILITY IN LONG-TERM DIALECT CONTACT: THE COT/CAUGHT CONTRAST AMONG MOBILE SPEAKERS

Jennifer Nycz

Georgetown University
jennifer.nycz@georgetown.edu

ABSTRACT

This paper examines the low back vowel systems of two mobile English-speaking groups: 40 people from Toronto, Ontario now living in New York City and 30 native New Yorkers living in Toronto. Vowel measurements of words in the COT and CAUGHT lexical sets were drawn from minimal pair tasks and conversational speech to determine whether speakers in either sample show evidence of acquiring the system of their new ambient dialect. Results indicate that natively COT/CAUGHT-merging Torontonians exhibit small yet significant phonetic distinctions between word classes after years living in New York, though this seems to be the result of changes in individual lexical items, not a category-level split resulting in new phonemic structure. Similarly, natively COT/CAUGHT-distinguishing New Yorkers in Toronto show evidence of phonetically approximating these vowels, but their underlying contrast remains robust. That is, both groups exhibit structural stability yet phonetic malleability as a result of mobility-induced contact.

Keywords: dialect acquisition, accent change and mobility, sociophonetics, vowel merger and split

1. INTRODUCTION

People often change the way they speak after moving to and living in a new region. This phenomenon is typically referred to as *second dialect acquisition* (SDA), but speakers do not change every (or even most) of their dialect features to match those of their new community. Studies of dialect change among mobile individuals indicate that specific features are altered depending on a range of linguistic, social, and developmental factors (see [1, 2] for overviews). Focusing on the linguistic domain, several scholars have argued that the likelihood of adopting particular features depends on the linguistic complexity of the change involved [3, 4, 5]: changes which seem to require learning a complex set of phonological, grammatical, or lexical conditioning factors are less frequently attested than those which are less constrained. The most "complex" (and thus most difficult-to-acquire) type of feature in this view is a non-native phonemic contrast. To acquire a new phonemic distinction, a learner would have to not only create new contrastive categories but also assign every relevant lexical item to the correct new category. Mobile speakers faced with a non-native merger, however, need only implement a relatively simple change (a single process of neutralization), to accommodate to new dialect forms [6].

Previous studies have investigated whether mobile speakers exposed to a new contrast show signs of acquiring that contrast. In a study of six Canadian children living in England, Chambers [4] found that most showed little evidence of distinguishing words like COT and CAUGHT (which have been merged for several generations in most varieties of Canadian English [7]) in word list production. Through acoustic analysis of conversational speech, however, Nycz [8] found that a majority of seventeen Canadian adults who had been living in the New York City area showed gradient shift towards the local COT/CAUGHT contrast, producing a significant (though small) phonetic distinction between these word classes. Neither study, however, can address the relative difficulty or likelihood of acquiring a merger versus a distinction, as they do not examine the reciprocal case, in which individuals from a low-back-vowel-distinguishing region move to a place characterized by a low back merger. This paper fills this empirical gap, by examining production and judgments in both types of contact scenarios: Torontonians in New York (TO-in-NYC) encountering a local dialect in which two vowel categories map onto their single low back vowel, and New Yorkers in Toronto (NYC-in-TO) facing the opposite situation.

2. METHODS

2.1. Participants

The data analyzed here are drawn from interactions with 71 participants: 40 people born-and-bred in Toronto, Ontario, Canada who had moved as adults to New York City, New York US, and 31 natives of New York City who had been living in Toronto.[1] All participants had been living in their current city for at least 4 years at time of interview, though time in that city as well as the age at which speakers moved there varied within and across the samples (Figure 1). The two speaker groups have a similar mean "Age of Arrival" (mid-20s), though on average the Toronto sample is 12 years older than the New York sample, and thus (again, on average), has spent longer in their current region. Both samples vary internally with respect to self-reported ethnicity and social-attitudinal factors which will not be discussed here.

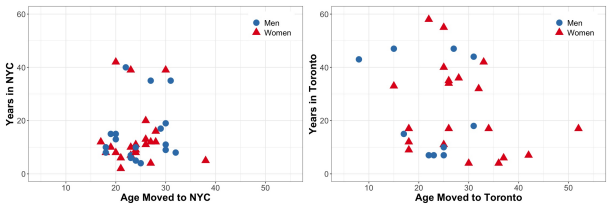

Figure 1: Study participants

2.2. Data collection and tasks

This analysis draws on data collected in New York City and Toronto in 2018-2020. Participants took part in conversational interviews focusing on growing up in Canada or the United States (as appropriate), moving to their new country and city, and impressions of the people and culture there.[2] After the conversational portion of the interview, participants completed word list readings and minimal pair tasks. In the minimal pair task, word pairs distinguished by the target low back vowels (e.g. *cot* and *caught*; *dawn* and *don*) were presented to participants on paper or via an iPad, with other potential pairs (e.g. *Mary/merry, pin/pen*) interspersed to pull focus away from the low back contrast of interest. Interviewers instructed participants to read each pair out loud and then say whether the two words sound the same or different to them. The minimal pair task thus yielded two types of data: production data as well as judgments regarding whether words in pairs contrast.

All interviews were recorded to 44.1kHz/16 bit wav files using a Zoom H4N solid state recorder and an Audio-Technica AT831b lavalier microphone. Interviews took place in the participant's home or a quiet public space (such as in a meeting room in the New York Public Library, for NYC participants). Consent was obtained and participants received an Amazon gift card worth $20 in local currency upon completion of study activities.

Data from the conversational portion of the interview and the minimal pair lists will be the focus of the analysis presented here.

2.3. Data processing and acoustic analysis

Each interview was transcribed in ELAN [9] by student research assistants. Transcripts and wav files were submitted to FAVE-Align python scripts [10] to generate time-aligned and segmented Praat textgrid files [11]. FAVE-Extract scripts were then used to extract formant measurements from all stressed vowels longer than 50 msec. Default configuration settings for FAVE-Extract were used, with maximum formant set automatically according to gender of the speaker. Vowel (i.e., word class) codes assigned by the CMU dictionary were hand-checked and corrected as needed. Raw formant measurements were Lobanov-normalized using the scale() function in R [12].

2.4. Statistical analysis

For each speaker (and in each style), a MANOVA model was fit with normalized F1 and F2 as the outcome variables and relevant phonological factors (following place, following manner, and vowel duration) and word class (COT or CAUGHT) as predictors. The Pillai statistic associated with word class was used as a measure of distinction [13]; because Pillai scores are sensitive to sample size, assessment of whether each score represented a significant phonetic distinction between COT and CAUGHT was carried out using the formula suggested in [14].

3. RESULTS

3.1. Minimal pair tasks

The TO-in-NYC and NYC-in-TO samples differed substantially in their minimal pair judgments. Of the 400 judgments produced by the 40 TO-in-NYC speakers, 335 were coded as *same*, 18 as *unsure*, and 47 as *different*, with 30 participants confidently judging at least 8 of the 10 pairs to sound the same. The 31 NYC-in-TO participants gave 309

judgments: 263 *different*, 10 *unsure*, and 36 *same*; 24 of these participants judged at least 8 pairs to sound different. The productions of these two groups are consistent with their judgments: the TO-in-NYC speakers individually produce COT and CAUGHT with very little distance between their means, and as a group show almost complete overlap between these categories, while the NYC-in-TO sample show a clear separation between word classes at the individual level and very little overlap at the group level (Fig. 2).

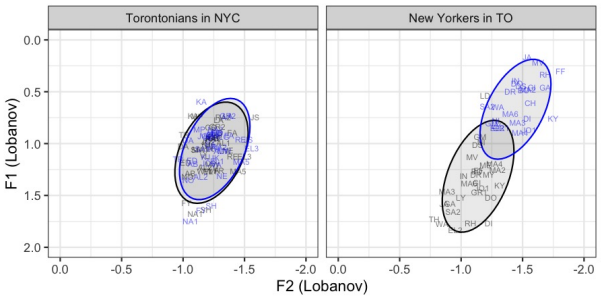

Figure 2: Productions of COT words (black) and CAUGHT words (blue) in the minimal pair task. Each label plots a speaker category mean; ellipses depict one standard deviation around the mean of the speaker means

The Pillai statistics associated with the MANOVA models of each speaker's minimal pair productions corroborate the overall difference suggested in the previous paragraph (see the righthand portion of Fig. 3). 35 of the TO-in-NYC participants have Pillai scores in this style that are low enough to indicate no significant difference between these vowels. For the NYC-in-TO group, models for all but two participants yielded a Pillai score high enough to indicate a significant phonetic distinction between COT and CAUGHT. To summarize, in the minimal pair task, both groups as a whole provide judgments and productions which reflect the dialect of the region in which they were born and raised: the TO-in-NYC speakers, whose first dialect does not distinguish two low back vowels, largely do not produce or report a difference between relevant words in the minimal pair task, while the NYC-in-TO speakers, whose first dialect does distinguish these word classes, mostly do produce this difference in their minimal pairs and recognize these words as different.

3.2. Conversational speech

Minimal pair productions of COT and CAUGHT across the two participant groups are quite different,

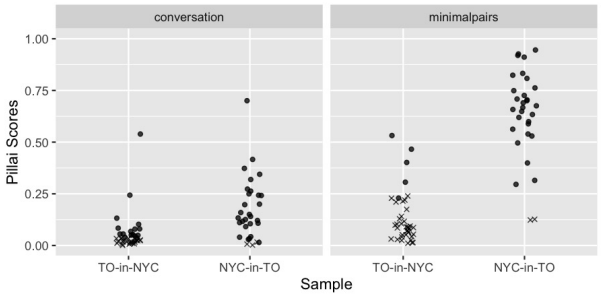

Figure 3: Pillai scores across samples and styles. Dots indicate Pillai scores high enough to suggest a distinction given token count; crosses indicate Pillais suggesting merger

reflecting a qualitative difference in the dialect that each group first acquired in their home region. The groups pattern more similarly, however, in their conversational speech: while TO-in-NYC as a group show signs of a split between these word classes, NYC-in-TO bring them closer together (Fig 4). This difference is also reflected in the Pillai scores for conversational style (see the lefthand portion of Fig. 3). For TO-in-NYC, Pillai scores still tend to be low, indicating much overlap between categories, but 17 have a Pillai that is sufficiently high given their token count to indicate a significant phonetic distinction between these categories. For NYC-in-Toronto, the majority (27 people) still have a Pillai that suggests a distinction between word classes, though these Pillais tend to be lower in conversational speech as compared to minimal pair productions.

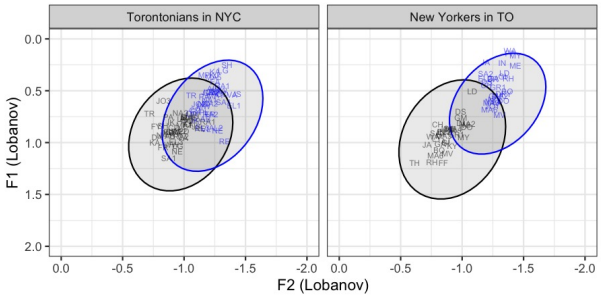

Figure 4: Productions of COT words (black) and CAUGHT words (blue) in conversational speech. Each label plots a speaker category mean; ellipses depict one sd around the mean of the speaker means

4. GENERAL DISCUSSION

In this data we observe a style difference in both samples, though in opposite directions. Native Torontonians in NYC mostly exhibit a merger in

minimal pair productions and judgments, but are more distinct in conversational productions. Native New Yorkers in TO are qualitatively more stable as a whole, showing a COT/CAUGHT distinction in both styles, though for them the magnitude of this distinction is smaller in conversational speech. The NYC-in-TO style difference could be explained in terms of a general tendency for conversational speech to be less dispersed than citation speech; the TO-in-NYC data, however, is not explainable in these terms. For both groups, the minimal pair task seems to elicit linguistic behavior which "shifts radically backward toward an older, corrected value" [15] - in these cases, productions reflecting the dialect acquired in one's region of origin. Conversational speech, meanwhile, shows more influence of the ambient second dialect: natively merged Torontonians exhibit a small but significant split which is more consistent with the New York City input they now receive, while New Yorkers whose native dialect contrasts the COT and CAUGHT sets reduce this contrast in conversation, consistent with *their* new Toronto English input. These patterns suggest that long-term convergence towards (or adoption of) a non-native *phonetic* distinction is not more difficult or unlikely than long-term convergence towards a non-native phonetic merger, contrary to claims that an acquisition asymmetry is to be expected: both groups show gradient, incomplete convergence towards their new dialect in spontaneous speech.

But what about long-term acquisition of a non-native *phonemic* contrast or merger? Here I would also argue that no asymmetry is present, because both samples show stability with respect to their underlying low back vowel system. New Yorkers may reduce the magnitude of their phonetic distinction in conversational speech, but their judgments that these words differ (and in which direction) remain robust. Similarly, while 43% of native Torontonians produce a significant measurable difference between COT and CAUGHT in conversation, their citation productions and judgments suggest that they do not recognize that this is a "real", in the sense of categorical, difference. Thus while we can say that new phonetic distinctions (or lack thereof) have been partially acquired by these mobile speakers in adulthood, there is as yet no evidence for changes at the phonemic level.

These patterns may be accounted for in a hybrid theory of lexical representation (such as current versions of Exemplar Theory [16]), which contains levels of representation corresponding to both phonemic categories and individual lexical items. The data presented here suggests that while the representations corresponding to individual lexical items accrue new exemplars over time, gradiently shifting the production targets of COT and CAUGHT words away from each other (for TO-in-NYC) or towards each other (for NYC-in-TO), the phoneme-level representations remain stable for both groups. So, while phonetic differences between word classes may emerge in certain styles (for example, conversational talk about life in one's new region, which may be more likely to activate exemplars associated with the new dialect), this does not imply that any change has occurred (or will occur) at the category level. In cases of mobility-induced dialect contact, structural stability and phonetic malleability can co-exist.

5. REFERENCES

[1] J. Siegel, *Second Dialect Acquisition*. Cambridge: Cambridge University Press, 2010.

[2] J. Nycz, "Second dialect acquisition: A sociophonetic perspective," *Lang. & Linguistics Compass*, vol. 9, pp. 469–482, 2015.

[3] A. Payne, "The acquisition of the phonological system of a second dialect," Ph.D. dissertation, University of Pennsylvania, 1976.

[4] J. K. Chambers, "Dialect acquisition," *Language*, vol. 68, no. 4, pp. 673–705, 1992.

[5] P. Kerswill, "Children, adolescents, and language change," *Lang. Variation & Change*, vol. 8, pp. 177–202, 1996.

[6] W. Labov, *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell, 1994.

[7] C. Boberg, *The English Language in Canada*. Cambridge: Cambridge University Press, 2010.

[8] J. Nycz, "New contrast acquisition: Methodological issues and theoretical implications," *English Lang. & Linguistics*, vol. 17, no. 2, pp. 325–357, 2013.

[9] ELAN, "Elan [computer software]. version 4.94." Available at: https://tla.mpi.nl/tools/tla-tools/elan/, 2018.

[10] I. Rosenfelder, J. Fruehwald, K. Evanini, and J. Yuan, "FAVE (Forced Alignment and Vowel Extraction)," http://fave.ling.upenn.edu, 2011.

[11] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," Retrieved from http://www.praat.org/.

[12] R. C. Team *et al.*, "R: A language and environment for statistical computing," 2013.

[13] J. Hay, P. Warren, and K. Drager, "Factors influencing speech perception in the context of a merger-in-progress," *J. Phon.*, vol. 34, pp. 458–484, 2006.

[14] J. Stanley and B. Sneller, "Sample size matters in calculating pillai scores," *Journal of the Acoustical Society of America*, Forthcoming.

[15] W. Labov, M. Karen, and C. Miller, "Near-mergers and the suspension of phonemic contrast," *Language Variation and Change*, vol. 3, pp. 33–74, 1991.
[16] J. Pierrehumbert, "The next toolkit," *J. Phon.*, vol. 34, pp. 516–530, 2006.

[1] This work is supported by National Science Foundation (NSF) Grant BCS-1651108.
[2] Participants in New York were interviewed by a native of Long Island, NY; participants in Toronto were interviewed by one of three graduate students, two of whom are Canadian and one of whom is a German national. All interviewers are white women in their 20s with sociolinguistics training.