

Escaping the TRAP: Losing the Northern Cities Shift in Real Time

ANJA THIEL
University of Bern

AARON J. DINKIN
San Diego State University

ABSTRACT

We examine the loss of the Northern Cities Shift raising of TRAP in Ogdensburg, a small city in rural northern New York. Although data from 2008 showed robust TRAP-raising among young people in Ogdensburg, in data collected in 2016 no speakers clear the 700-Hz threshold for NCS participation in F1 of TRAP—a seemingly very rapid real-time change. We find apparent-time change in style-shifting: although older people raise TRAP more in wordlist reading than in spontaneous speech, younger people do the opposite. We infer that increasing negative evaluation of the feature led Ogdensburg speakers to collectively abandon raising TRAP between 2008 and 2016. This indicates a role for communal change in the transition of a dialect feature from an indicator to a marker.

The presence of the Northern Cities Shift (NCS) formed the basis on which the *Atlas of North American English* (Labov, Ash, & Boberg, 2006; henceforward ANAE) defined the dialect region of the Inland North, focused on the urban areas south of the Great Lakes, from central New York to Wisconsin. The NCS involves a rotation of the short vowels, including the raising and fronting of TRAP,¹ the fronting of LOT, the backing and/or lowering of DRESS, and other changes. The most distinctive feature of the NCS is the raising of TRAP, which is the focus of this paper.

The NCS has received a great deal of attention in sociolinguistic research (e.g., Durian, 2014; Eckert, 1988; Labov, 1994; Labov, Yaeger, & Steiner, 1972; McCarthy, 2010). Although most of these studies focused on urban environments (e.g., Chicago, Buffalo, Rochester), and the NCS has generally been described as an urban phenomenon, research into small-town communities has shown that the shift has also been adopted by rural speech communities. For example, Dinkin's (2009, 2013) study of several smaller cities and rural communities in central and northern New York found moderate NCS participation in some of these, which he termed Inland North fringe communities. A study of the shift in small-town Michigan (Gordon, 2000) led to similar conclusions. The results from all these studies suggested that, in both rural and urban communities, raised TRAP was a stable feature.

More recent research, however, has found that NCS communities are starting to lose this vowel shift. Apparent-time backing of NCS-fronted LOT was observed first in Chicago (McCarthy, 2011) and upstate New York (Dinkin, 2011). Retreat from raised TRAP has been observed in Chicago (D’Onofrio & Benheim, 2020; Durian & Cameron, 2018), Buffalo (Milholland, 2018), Rochester (Kapner, 2019; King, 2017), Syracuse (Driscoll, 2016; Driscoll & Lape, 2015), and Lansing (Nesbitt, 2018, 2019; Wagner, Mason, Nesbitt, Pevan, & Savage, 2016). In Lansing, Wagner et al. (2016) and Nesbitt (2018, 2019) reported the development of an allophonic alternation: prenasal TRAP remains in raised position while preoral TRAP retracts, producing what ANAE (p. 174) terms the nasal short-*a* system; King (2017) and Kapner (2019) both reported the same in Rochester. Driscoll and Lape (2015) found that this retraction seems to be progressing faster among speakers in the rural environs of Syracuse than among urban speakers, which suggests that urban and rural communities might be treating TRAP differently in terms of its reversal.

The force behind the retraction of TRAP may be related to an increased level of social salience. Sociolinguists have frequently described the NCS as transparent to its speakers and, therefore, apparently unavailable for negative social evaluation. For example, Gordon (2000:178) wrote that “For the most part, speakers... do not seem to be aware of the variation”; and Preston (1998) has reported that residents of Michigan regard their state, despite the presence of the NCS there, as one where unmarked standard English is spoken. Labov (2010:194) argued that “NCS variables appear to be indicators rather than markers, with little style-shifting associated with their social distribution and with no evidence of conscious awareness in the Inland North.” If style-shifting was observed in the usage of NCS features at all, their incidence seems to have increased in more careful speech styles: Labov (2010:59) cited Ash (1999) as finding that “the comparison of word lists and spontaneous speech in ANAE interviews showed that the raising of short *a* in the Northern Cities Shift was more advanced in word lists.”

Recent research, however, has found that at least some NCS variables have become subject to social evaluation. In Lansing, attitude studies have shown that the fronted NCS variant of LOT is evaluated negatively, while the retracted NCS variant of DRESS is evaluated more favorably than its unshifted counterpart (Nesbitt & Mason, 2016; Savage, Mason, Nesbitt, Pevan, & Wagner, 2016). For TRAP, the evidence of social evaluation is less substantial. Accounts from focus-group interviews and anecdotal evidence in Lansing and Syracuse show that raised realizations of TRAP have attracted overt comment (Savage et al., 2016) and suggest an increasing stigma around raised TRAP (Driscoll & Lape, 2015). However, since none of the above studies took style-shifting in speech production into account, the question of how potential negative evaluations might affect the treatment of TRAP in production remains unanswered.

The goal of this study is to address two of the questions posed above. Is raised TRAP retreating in a remote rural area, as it has been found to be in urban areas and their environs?² And how does suspected negative evaluation of TRAP affect its

production? To answer these questions, we will track the production of TRAP in different speech styles in apparent and real time in the small, rural community of Ogdensburg in northern New York. So far, few studies have looked at speech patterns in this rural part of the state, and none have focused in depth on a single speech community; Dinkin (2009, 2013) grouped Ogdensburg with the Inland North fringe based on data collected in 2008. This paper reports on new data from the same community, collected in 2016. We observe interaction between year of birth and speech style for preoral TRAP, and an unexpected effect of year of interview; we argue that the NCS raising of TRAP has virtually disappeared from the community. This retraction appears to be driven by a growing stigma around raised TRAP, as we find speakers avoiding raising especially in careful speech.

PAST RESEARCH AND MOTIVATION

Ogdensburg is a small community of about 11,000 people in northern New York, located in sparsely populated St. Lawrence County. Despite its small size, Ogdensburg is one of the most populous settlements in the county and the only one with the legal status of a city. It is situated on the St. Lawrence River, the national border between the US and Canada, and is home to the Ogdensburg–Prescott International Bridge that connects the city to Ontario. As shown in [Figure 1](#), Ogdensburg is relatively remote from American population centers; the nearest cities with populations greater than 50,000 are Ottawa and Kingston, both in Ontario.

Ogdensburg was once a vibrant industrial city. Benefiting from its position on the St. Lawrence River, which connects it to Lake Ontario, it was an important shipping and railroad hub. The city entered a sharp economic decline in the second half of the twentieth century and never fully recovered; today, Ogdensburg performs below average in terms of socioeconomic success compared to the rest of St. Lawrence County and New York State. As of 2016, only half of Ogdensburg's residents over the age of sixteen were in the city's labor force, while an estimated 9.2% were unemployed.³ St. Lawrence County has a relatively low median household income (\$46,000), among the ten lowest of the sixty-two counties in the state of New York. Even by these low standards, Ogdensburg's median household income is well below average: at \$36,832, it was third lowest in the county in 2016. Although there are several institutions of higher education in the vicinity,⁴ only about 16% of Ogdensburg's population had attained a college degree in 2016, while the average in New York State was more than twice as high (U.S. Census Bureau, [n.d.](#)).

Dinkin (2009, 2013) found NCS vowel realizations, including raised TRAP, in Ogdensburg, but not in the nearby village of Canton less than twenty miles to the southeast, leading him to describe Ogdensburg as the northeasternmost limit of the NCS. This description was corroborated by a later study (Dinkin, 2020) sampling more villages in the vicinity. Ogdensburg's only geographical

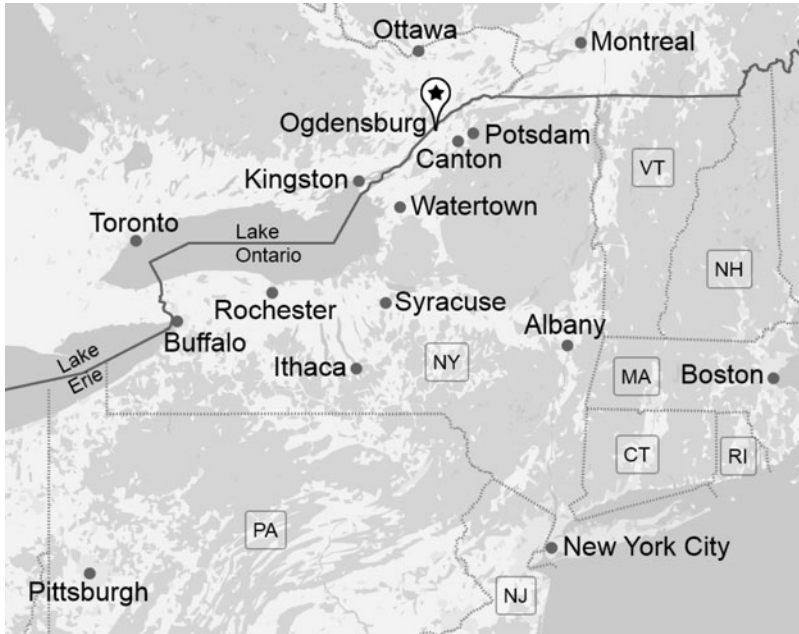


FIGURE 1. Ogdensburg's location in New York State.

connection to other known NCS communities lies to the southwest, toward the city of Watertown. To the northwest of Ogdensburg lies Canada, and to the east is the dialect region Dinkin (2009) termed the North Country; in both of these neighboring regions, the NCS is absent (Dinkin, 2009, 2013; Labov et al., 2006).

Dinkin's corpus includes nine native Ogdensburg speakers, interviewed in 2008. These nine speakers include five born in the 1980s, one each born in the 1970s, 1960s, and 1950s, and one old outlier born in 1922. Seven are female, two are male; all nine are white. Seven were interviewed in person, following the protocol for a Short Sociolinguistic Encounter (Ash, 2002) that lasted 15–25 minutes and included the reading of a wordlist and other targeted elicitation tasks in addition to free conversation. Two participants were interviewed by telephone following the ANAE methodology. Vowel formants were extracted at points chosen manually in Praat (Boersma & Weenink, 2005) and originally normalized using the same algorithm and group norm used in ANAE, based on Nearey (1977). Phoneme means were calculated following ANAE methodology, excluding tokens before sonorants, after glides, and after stop+liquid clusters. An average of 395 vowel tokens per speaker were measured, of which an average of thirty-seven were TRAP⁵: twenty-four in spontaneous speech, nine elicited through wordlist reading, and four elicited through other procedures. Full methodological details can be found in Dinkin (2009).

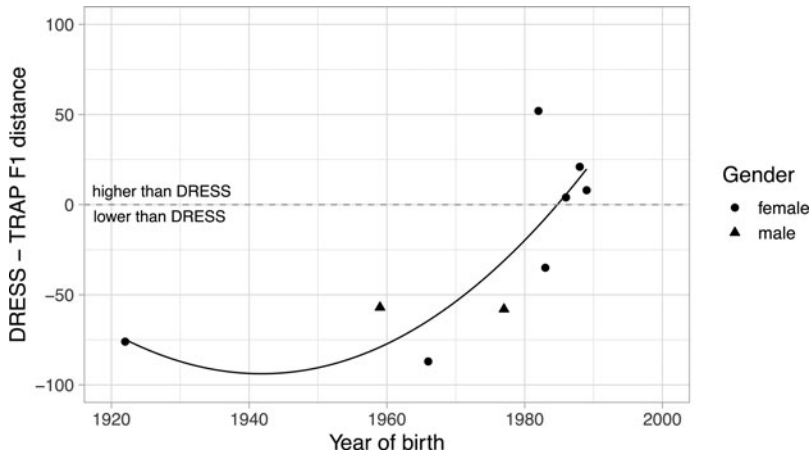


FIGURE 2. The relative height of TRAP and DRESS in 2008, according to Dinkin (2009). On this and other scatterplots, the LOESS curve shows the overall trend.

The analysis of TRAP in Dinkin (2009, 2013) was based on two criteria defined by ANAE, designated AE1 and EQ. The AE1 criterion considers a speaker to have NCS-raised TRAP if their mean normalized F1 of TRAP is less than 700 Hz. The EQ criterion defines raised TRAP as being both higher (i.e., having a lower F1) and fronter (i.e., having a higher F2) than DRESS. The AE1 and EQ criteria are both valuable for measuring participation in TRAP-raising: AE1 measures TRAP directly, but its value depends on the normalization methodology chosen and on the somewhat arbitrary value of 700 Hz; EQ is based on more objective criteria but cannot necessarily distinguish between the raising of TRAP and the lowering of DRESS.

All Ogdensburg speakers interviewed in 2008 produced a relatively high TRAP. Out of the nine speakers, three fulfilled the AE1 criterion, raising TRAP to a mean F1 less than 700 Hz; the others all had mean TRAP F1 less than 770 Hz. Four of the five youngest speakers fulfilled the EQ criterion, raising TRAP above DRESS, indicating an apparent-time trend toward the EQ criterion, as shown in Figure 2; Dinkin (2009:92–3) concluded that this trend was the result of ongoing lowering and backing of DRESS. Ogdensburg was the only community sampled by Dinkin (2009) where apparent-time trends toward NCS features could be observed. In all other sampled NCS communities, most of which had greater EQ participation than Ogdensburg, the NCS was apparently no longer in progress. Full details on their production can be found in Dinkin (2009).

Ogdensburg's status as a small city squeezed between non-NCS regions to the north and the east, but where NCS was nonetheless seemingly increasing, motivated an in-depth follow-up study of the community. There is no obvious reason why Ogdensburg should have been behind in adopting NCS features. Communities of similar sizes and with similar median household incomes and geographical distances to urban NCS communities had completed the adoption

of the NCS despite these characteristics. Gloversville, for example, is only slightly bigger, and has an even lower median household income than Ogdensburg. Glens Falls, while somewhat bigger and wealthier, is just as remote from urban NCS communities as Ogdensburg. In neither of these two communities did the NCS seem to still be in progress.

While Ogdensburg was apparently late in adopting the NCS, it showed relatively strong signs of incipient change toward the low back merger of LOT and THOUGHT, contrary to the traditional maintenance of the distinction in NCS communities. This suggested that Ogdensburg might be subject to influence from Canada and/or the North Country, neighboring dialects with more thorough low back merger.

Thus, Ogdensburg appeared to be a fruitful site for an in-depth sociolinguistic study examining how two linguistic changes that have been thought to be mutually exclusive (Labov et al., 2006) co-exist and interact in such close proximity. As will be shown below, however, Ogdensburg's status as an NCS community may have been short-lived. This paper will focus on unexpected findings relating to the raising of TRAP in Ogdensburg, which will illuminate the means by which retreat from dialect features such as the NCS takes place; analysis of the low back merger in Ogdensburg will be left for future work. In our analysis we emphasize the effect of style-shifting, which has apparently gone underexamined in other recent studies of NCS TRAP.

METHODS

New Ogdensburg data

The 2016 data was collected by Anja Thiel during a three-month fieldwork stay in Ogdensburg. Participants were recruited through social and print media, as well as by using the “snowball” technique (i.e., interviewees recruited friends or family members as participants). Interviews were scheduled in advance and lasted 1–2 hours. The interviews generally included spontaneous speech elicitation in a casual conversation that lasted 30–90 minutes, followed by wordlist and minimal-pair readings. In total, the 2016 corpus comprises thirty-nine speakers born between 1932 and 2002 (mean: 1969), including twenty-five women and fourteen men. All speakers except one⁶ are white. The data were subjected to automated vowel extraction using FAVE (Rosenfelder, Fruehwald, Evanini, & Yuan, 2011), which produced formant values normalized using the Lobanov (1971) algorithm, setting each speaker's overall mean F1 equal to 650 Hz (with a standard deviation of 150) and mean F2 equal to 1700 Hz (standard deviation 420). Phoneme means were calculated according to the ANAE methodology as described above. Formant measurements for an average of 2,016 vowel tokens per speaker were extracted, including an average of 150 tokens of TRAP: nine in wordlist style and 141 in spontaneous speech. Regression analyses for F1 of TRAP reported below exclude tokens of TRAP preceding nasals or /r/. Full details on this corpus can be found in Thiel (2019).

TABLE 1. *Comparison of methods for the 2008 and 2016 data sets*

TRAP data	2008	2016
Data collection	<ul style="list-style-type: none"> - 3 days - Spontaneous encounters - Short socioling. interviews - 15–30 mins - Spontaneous speech + wordlist + other elicitation tasks - 2 phone interviews: <ul style="list-style-type: none"> Spontaneous speech + other elicitation tasks 	<ul style="list-style-type: none"> - 3 months - Scheduled interviews - Full-length socioling. interviews - 1–2 hours - Spontaneous speech + wordlist + other elicitation tasks
Interviewer	- Native English speaker (non-NCS), male	- Nonnative English speaker, female
Corpus	<ul style="list-style-type: none"> - 9 speakers - 7 f + 2 m - Born: 1922–1989; mean: 1972 - White 	<ul style="list-style-type: none"> - 39 speakers - 25 f + 14 m - Born: 1932–2002; mean: 1969 - White (except one)
Formant extraction	<ul style="list-style-type: none"> - Hand-measured - Lobanov-(re)normalized 	<ul style="list-style-type: none"> - Automated (FAVE) - Lobanov-normalized
Mean calculation	- ANAE methodology	- ANAE methodology

In this paper, we compare the production of TRAP in this new corpus to that in Dinkin (2009)'s data collected in 2008. For comparability with the new Ogdensburg data in the present paper, the 2008 data were renormalized using the Lobanov method,⁷ with the same standardized FAVE means. As the summary of methodological characteristics of the 2008 and 2016 data in Table 1 shows, there are noteworthy differences between the two studies. For example, in 2008, subjects were interviewed for twenty minutes by a native English speaker who recruited them on the spur of the moment, whereas in 2016 they were interviewed for more than an hour by a nonnative speaker during a scheduled appointment. But for reasons to be discussed below, we do not believe these methodological differences are responsible for the differences between the treatments of TRAP in the two corpora. Potentially of more importance, formant measurements in the 2008 data were selected by choosing a measurement point by hand, while in the 2016 data they were automatically extracted. Severance, Evanini, and Dinkin (2015) contended that the differences between these two measurement methods are in general small, and recent studies such as Turton and Baranowski (2020) have productively combined hand-measured data with FAVE measurements; but to be sure, in the next section we will verify that, in our data, the two methods yield sufficiently comparable results to be includable in the same analysis.

Comparing automated and hand measurements

We compare hand measurements of a selection of the 2016 data to the automated measurements of the same. We randomly selected fifteen speakers from the

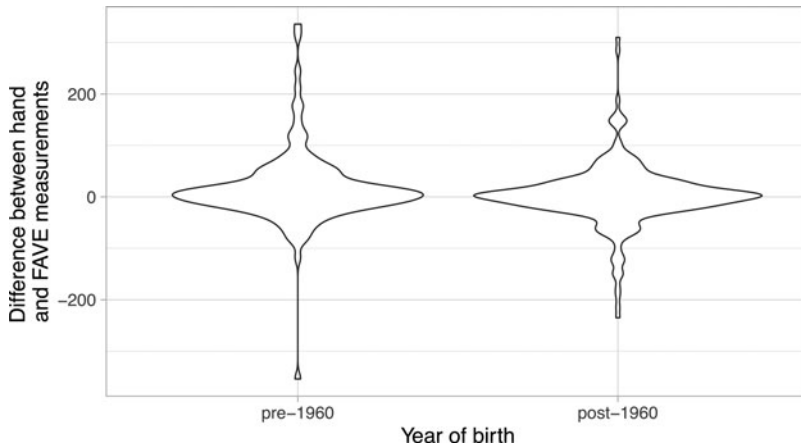


FIGURE 3. The difference between F1 measured by hand and by FAVE-extract for 393 randomly selected TRAP tokens from the 2016 corpus. Positive values mean that hand measurement produced a larger F1 value.

thirty-nine interviewed in 2016; and from these speakers we randomly selected 393 preoral TRAP tokens. F1 of these 393 tokens was measured in Praat 6.1 (Boersma & Weenink, 2020), at measurement points chosen by hand according to the same methodology used for the 2008 data. These F1 measurements were then compared to the unnormalized FAVE measurements for the same set of tokens.

For the majority of tokens, the two measurements differ by less than 25 Hz; three quarters differ by less than 50 Hz, although there are outliers ranging beyond 300 Hz in either direction (see Figure 3). The mean absolute magnitude of the difference between the two measurements is 39 Hz. There is a very slight tendency for hand measurements to produce larger F1 values than FAVE measurements ($p \approx 0.02$ in a paired t -test); the mean difference between the two measurements is 7.6 Hz, which we believe is small enough to consider negligible on the scale of the F1 differences that will be relevant in this paper.

We may also calculate the mean difference between hand and FAVE measurements of TRAP F1 for each speaker (see Figure 4). The absolute value of this mean difference is less than 15 Hz for the majority of speakers, and less than 30 Hz for all but one. The mean of the speakers' mean differences is 7.8 Hz.

Table 2 shows the results of a mixed-effects regression model investigating whether any factors systematically bias FAVE to produce results higher or lower than hand measurements on this subsample of 393 tokens. We find that hand measurements tend to be slightly more extreme than FAVE measurements: for large F1 values,⁸ the hand measurements tend slightly larger than FAVE, and for small F1 values, the hand measurements tend slightly smaller. There also appears to be an age correlation: hand measurements are more likely to be larger than FAVE measurements for older speakers than for younger speakers. However, the sizes of the effects are fairly small. For tokens with F1 between

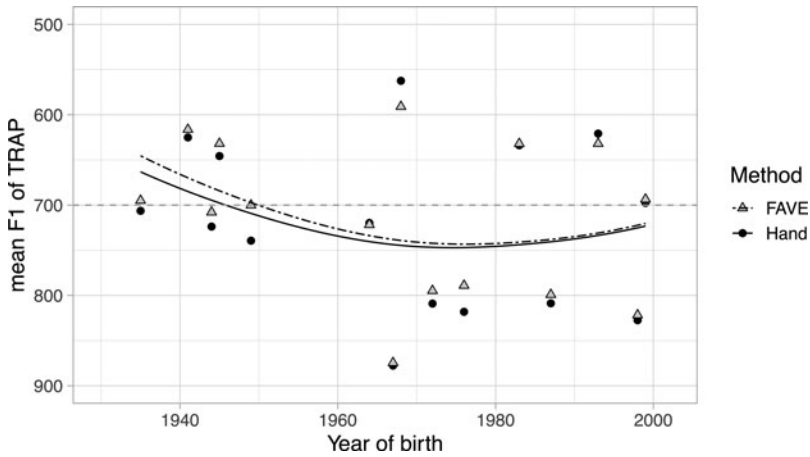


FIGURE 4. Unnormalized mean FAVE and hand measurements of TRAP F1 for fifteen randomly selected speakers interviewed in 2016.

TABLE 2. A linear mixed-effects regression model for the F1 difference between hand measurements and FAVE-extract measurements of a subsample of TRAP tokens

Predictor		Coefficient	Mean	n	p
(Intercept)		-65.2 Hz	7.6 Hz	393	
<i>Estimated F1</i>		+0.087		393	0.005
<i>Coda manner</i>	lateral	+58.4 Hz	55.2 Hz	12	0.01
	stop	reference level	9.4 Hz	214	
	affricate	-15.3 Hz	-5.5 Hz	31	
	fricative	-17.6 Hz	3.7 Hz	136	
<i>Coda voicing</i>	voiceless	reference level	12.2 Hz	269	0.03
	voiced	-26.2 Hz	-2.4 Hz	134	
<i>Age</i>		+0.377 Hz/year		393	0.03

Note: “Estimated F1” denotes the mean of the two measurements of F1; “age” denotes age in 2016. Positive coefficients favor hand measurements being larger than FAVE measurements. Model calculated by a step-down procedure.⁹ Not significant: speaker gender, speech style, coda place of articulation, onset. Random factors: speaker, word.

500 Hz and 900 Hz (within which the vast majority of tokens lie), produced by a speaker born in 1986, the model predicts that hand and FAVE measurements will differ on average by less than 25 Hz; and as we will see below it is among speakers born in the 1980s that the differences between the 2008 data (measured by hand) and the 2016 data (measured by FAVE) are of greatest interest. Even for 75-year-old speakers, the model predicts that hand-measured F1 values will exceed FAVE values within that range by an average of less than 42 Hz. We infer from this that F1 values in the hand-measured data from 2008 and the FAVE-measured data from 2016 appear sufficiently comparable to be included in the same analysis.

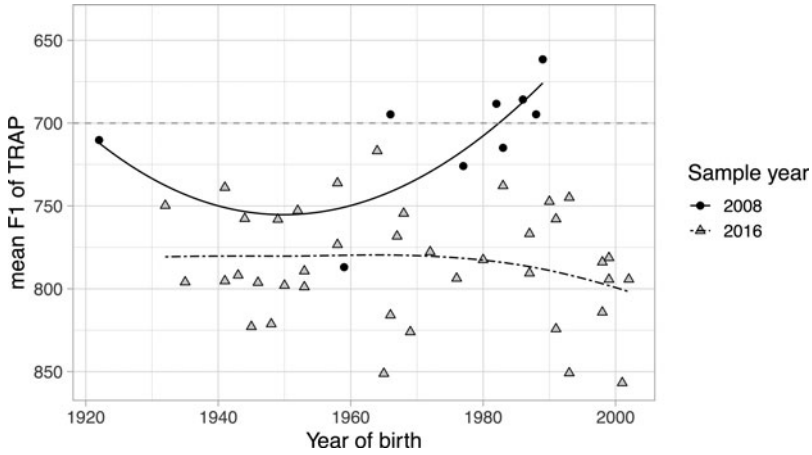


FIGURE 5. Overall F1 TRAP means and AE1 participation in 2008 and 2016.

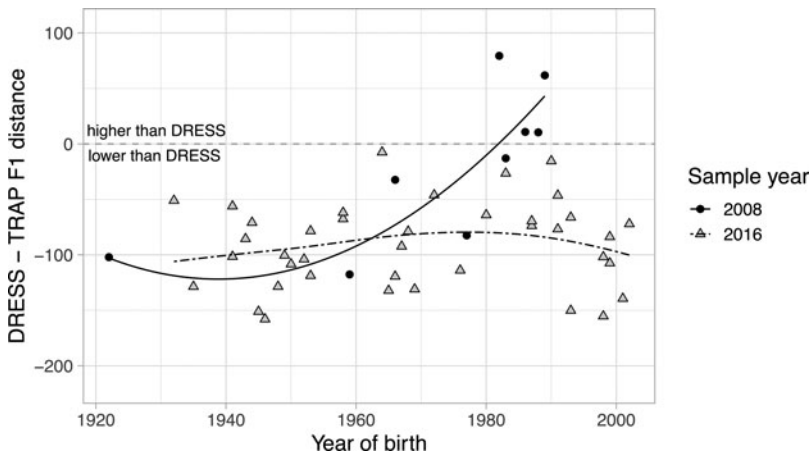


FIGURE 6. Overall EQ participation in 2008 and 2016.

RESULTS

Overall results

Unlike in the 2008 data, speakers in the 2016 sample fail to meet the two criteria for NCS TRAP. Based on overall means, none of the speakers interviewed in 2016 raise TRAP to a (normalized) F1 less than the AE1 benchmark of 700 Hz.¹⁰ In the 2008 data, five speakers satisfy AE1, four of them born after 1980. Figure 5 shows the comparison between the 2008 and 2016 interviews in this regard. Similarly, as shown in Figure 6, none of the 2016 speakers raise TRAP above DRESS to fulfill EQ. In the 2008 data, on the other hand, four of the youngest speakers fulfill EQ.

Because of the patterns in TRAP production in the 2008 data, the lack of AE1 and EQ participation in 2016 is highly unexpected. To find potential explanations for the sudden absence of raised TRAP in the community, more detailed insight into the data is necessary. To conduct this more in-depth analysis, we separate the data by style, and examine F1 of TRAP in spontaneous speech and the more careful wordlist style individually. We begin with a distributional inspection of each combination of style and sample year, and then calculate a multiple linear regression on the combined data set.

Spontaneous Speech

The results for TRAP in spontaneous speech resemble the overall patterns described above very closely; this is unsurprising, since the majority of the data is spontaneous speech. While the 2008 data show moderately advanced TRAP-raising among younger speakers, such raising is absent in the 2016 data, indicating a real-time change away from NCS TRAP.

As can be seen in Figure 7, six of the nine speakers in the 2008 sample satisfy the AE1 criterion in spontaneous speech: the six women other than the old outlier. For four of the five women born in the 1980s, TRAP is higher than DRESS in this style. The fact that the two male speakers have lower TRAP than all seven women suggests that the NCS may have been a female-led change in Ogdensburg, though the number of speakers in the data is too small to enable a firm conclusion to be drawn in this regard. Similarly, the slope of the trendline in Figure 7 suggests an apparent-time change toward TRAP-raising, though we must exercise caution in drawing conclusions from this small sample. In any event, with the majority of speakers in this sample satisfying the AE1 criterion, we might expect to find substantial raising of TRAP in the 2016 data as well, at least among younger women, and perhaps an apparent-time trend toward further raising.

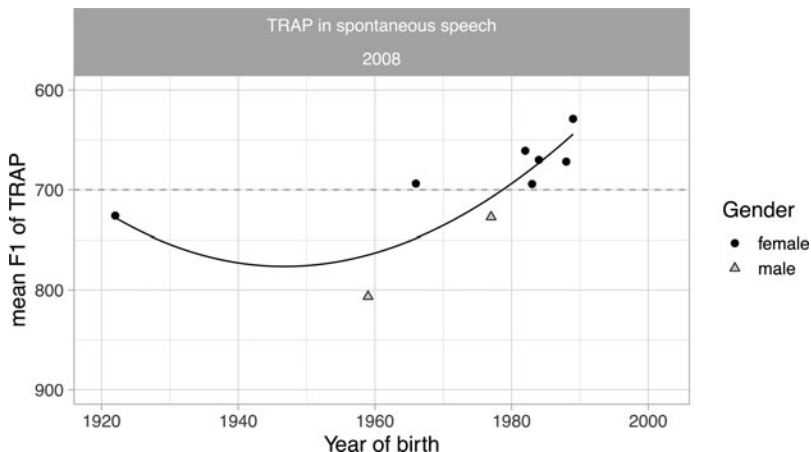


FIGURE 7. F1 TRAP means and AE1 participation in spontaneous speech in 2008. $n = 136$.

In fact, no speakers in the 2016 data satisfy the AE1 criterion in spontaneous speech, and there is no hint of gender patterning or apparent-time trend in F1 of TRAP (see Figure 8). Instead, the 2016 data suggests that TRAP F1 has remained between 720 and 850 Hz over the last seventy years, constituting a stable non-NCS TRAP system.

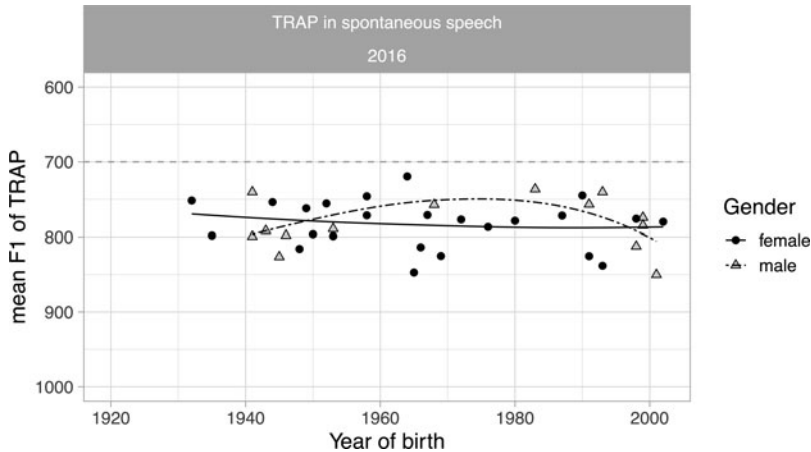


FIGURE 8. F1 TRAP means and AE1 participation in spontaneous speech in 2016. $n = 3704$.

Some of the speakers in 2016 do produce relatively high TRAP in spontaneous speech, but none of them produce it high enough to meet the 700-Hz mark, nor do any of them raise TRAP above DRESS in spontaneous speech. Indeed, all but one speaker in 2016 has mean spontaneous TRAP lower than all but one speaker in 2008. Thus, despite the robust presence of TRAP-raising in Ogdensburg in 2008, in 2016 we find TRAP substantially less raised, with no visible apparent-time trend in either direction.

Wordlist style

In wordlist style, apparent-time trends indicate a shift away from NCS TRAP, which is corroborated by real-time differences between 2008 and 2016, as 2016 speakers have lower TRAP in wordlist style than 2008 speakers. Five of the speakers interviewed did not participate in wordlist reading: the two interviewed by telephone in 2008 (including the oldest individual in the combined corpus) and three in 2016.

In the 2008 wordlist data, it is only in the two relatively *older* speakers that mean F1 of TRAP less than 700 Hz is found, satisfying the AE1 criterion (see Figure 9). This differs from spontaneous speech, in which it was the youngest speakers who produced the most raised TRAP. In the 2016 data, two speakers satisfy AE1 in wordlist style (see Figure 10), even though none satisfied it in spontaneous

speech. TRAP appears to be lowering in apparent time in this style, continuing the trend suggested by the 2008 data; the two speakers who satisfy AE1 are in the older half of the sample.

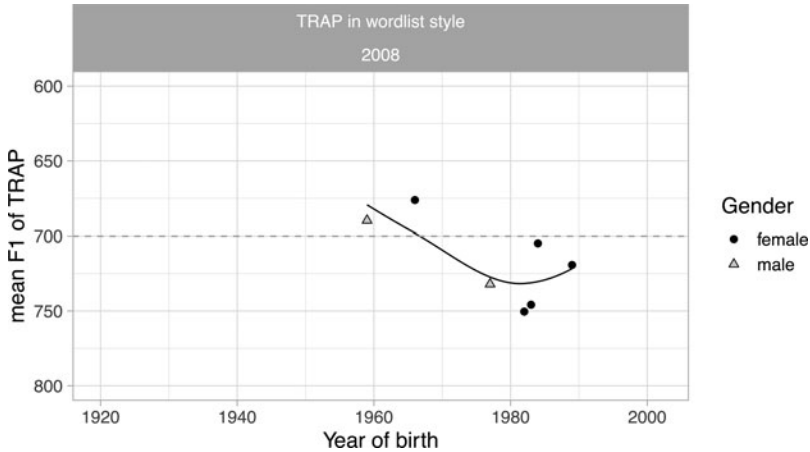


FIGURE 9. F1 TRAP means and AE1 participation in wordlist style in 2008. $n = 56$.

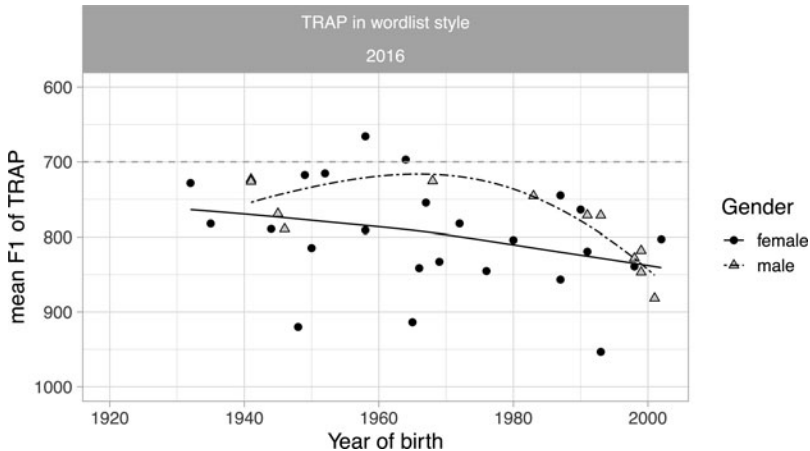


FIGURE 10. F1 TRAP means and AE1 participation in wordlist style in 2016. $n = 344$.

A real-time difference can be observed in wordlist style between the 2008 and 2016 datasets. As shown in Figure 11, speakers interviewed in 2008 produced TRAP in wordlist style on average about 85 Hz higher than 2016 speakers. Both data sets seem to show TRAP lowering in this speech style in apparent time, and the four speakers from both datasets who do raise TRAP to F1 less than 700 Hz in wordlist style were all born between 1958 and 1966.

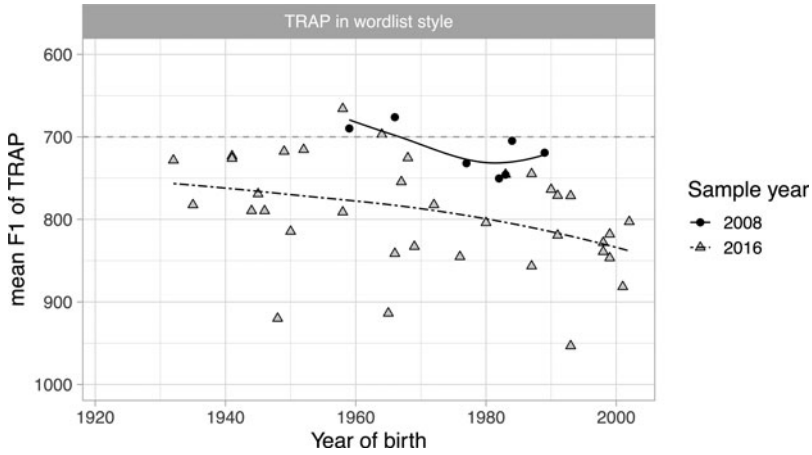


FIGURE 11. F1 TRAP means and AEI participation in wordlist style in 2008 and 2016. $n = 400$.

Style-shifting

Figure 12 displays each speaker’s range of style-shifting: that is, the difference between each speaker’s wordlist and spontaneous TRAP means. It shows that most older speakers shift *toward* raised TRAP in wordlist style, whereas younger speakers shift *away* from it. This pattern is visible in both the 2008 and the 2016 data. This indicates an apparent-time change in progress affecting the *direction* of style-shifting: the community is changing from treating raised TRAP as a target of more careful speech, as Ash (1999) reportedly found, to avoiding it in more careful speech.

Furthermore, a gender difference in style-shifting is visible: although all older male speakers shift toward raised TRAP in wordlist style, a handful of older

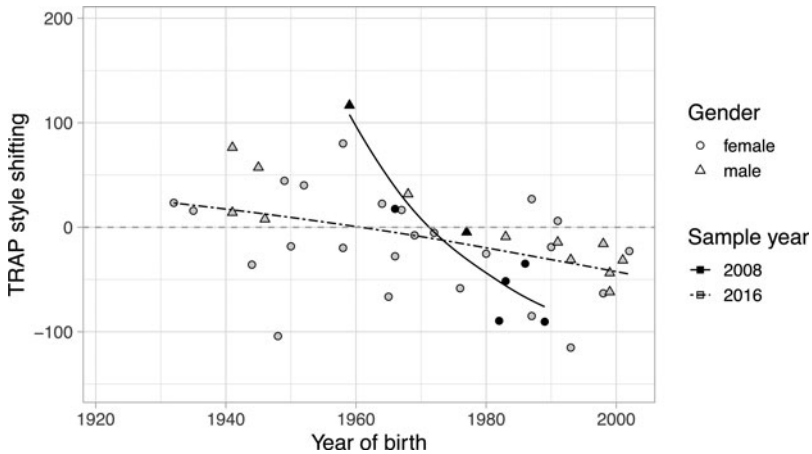


FIGURE 12. Difference between spontaneous and wordlist TRAP F1 means. A positive value means the vowel is more raised in wordlist style than spontaneous speech.

TABLE 3. *Social factors in a linear mixed-effects regression model for the height of TRAP*

Predictor		Coefficient	Mean	<i>n</i>	<i>p</i>
(Intercept)		805 Hz	785 Hz	4240	
<i>Style × age</i>	spontaneous	reference level			<10–4
	wordlist	–1.23 Hz/year			
<i>Style</i>	wordlist	+68.0 Hz	783 Hz	400	
	spontaneous	reference level	785 Hz	3840	
<i>Sample year</i>	2016	reference level	789 Hz	4048	<10–4
	2008	–76.2 Hz	707 Hz	192	
<i>Style × gender</i>	wordlist × female	reference level	787 Hz	270	0.0007
	spontaneous × male	reference level	787 Hz	1447	
	spontaneous × female	reference level	784 Hz	2393	
	wordlist × male	–28.9 Hz	775 Hz	130	
<i>Age</i>		+0.0894 Hz/year			
<i>Gender</i>	male	+8.18 Hz	786 Hz	1577	
	female	reference level	784 Hz	2663	

Note: Following consonant is included as a factor in the model; details are shown in Appendix B. Tokens of TRAP in styles other than wordlist or spontaneous are omitted from the model. Reference levels: female, spontaneous speech, 2016 interview. “Age” denotes age in 2016, even for speakers interviewed in 2008. Random factors: speaker, word. We omit *p*-values for main effects of predictors whose significant interactions are included in the regression.

female speakers do the opposite, showing the style-shifting pattern of younger speakers. This indicates that the change in style-shifting is a female-led change: some older women anticipate the direction of the change in progress, while older men uniformly have the more conservative pattern.

A mixed-effects linear regression model for the height of TRAP is shown in Table 3. The model includes terms for the four predictors whose effects on TRAP have been discussed so far—speaker age and gender, wordlist versus spontaneous speech style, and year of interview—as well as significant pairwise interactions between them.¹¹ Following consonant is also included as a predictor in the model; for the sake of conciseness, the coefficients for this predictor are omitted here but can be found in Appendix B. The coefficients in Table 3 support most of the key observations above.

- The effect of sample year indicates that speakers interviewed in 2008 have TRAP about 76 Hz higher (i.e., with smaller F1) than those in 2016.
- The main effect of style confirms that young people have TRAP lower in wordlist style than in spontaneous speech, while the interaction between style and age indicates that older speakers have TRAP *higher* in wordlist style.
- The interaction between style and gender confirms the significance of the gender difference in style-shifting noted above: women are more likely than men to style-shift away from raised TRAP in wordlist style (i.e., are leading the change toward this direction of style-shifting).

The negligibly small coefficient for the main effect of age indicates that there is no evident apparent-time change affecting spontaneous speech.

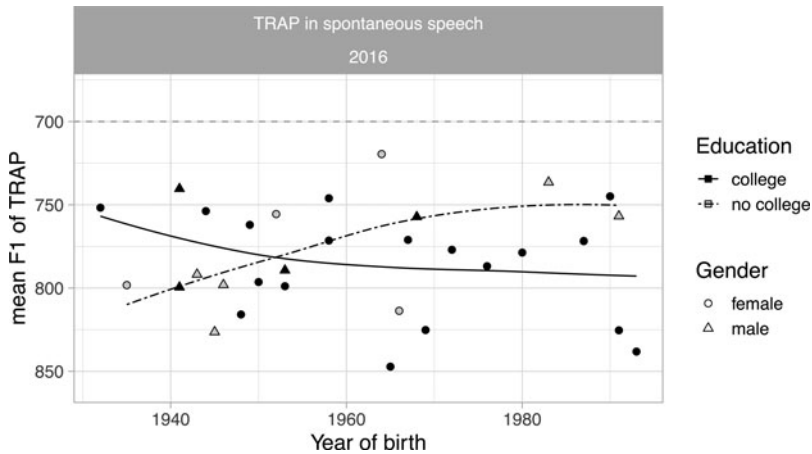


FIGURE 13. F1 of TRAP in spontaneous speech in 2016, by education level.

Education

Durian and Cameron (2018) cautioned that studies on the retreat of NCS must take care not to attribute to the community as a whole a retreat that may only be taking place in the middle class. Both Durian and Cameron (2018) and Nesbitt (2018, 2019) found indices of TRAP-raising persisting longer in speakers with blue-collar occupations; Milholland (2018) found the same in speakers with less education. Similar patterns have been reported in the reversal of the Southern Shift (Dodsworth & Kohn, 2012). It is, therefore, important to test whether the apparent loss of NCS in Ogdensburg is conditioned by social class, for which education is the most accessible indicator available in our data. However, since several of the youngest individuals in the corpus were high-school students when interviewed, their eventual ultimate education level is unknowable; for this reason, education level was not included in the regression model shown above in Table 3. Excluding students, the 2016 data include twenty-three individuals with a college education and nine without.

The trendlines on Figure 13 appear to suggest that the overall lack of age correlation in the data may actually be the combination of two contrary trends: TRAP in spontaneous speech is lowering in apparent time among speakers with a college education but is still raising among speakers without one. The difference between these two slopes, as estimated by the linear mixed-effects regression model in Table 4, does not quite reach the level of statistical significance ($p \approx 0.07$), but, if authentic, it would fit the expected profile of speakers with more access to social prestige being the first to retreat from the NCS. This may mean that our discussion above overstates the degree to which TRAP-raising has retreated in Ogdensburg overall, since speakers with a college education are overrepresented in our sample relative to their actual proportion in Ogdensburg. However, even the non-college-educated individuals in the 2016 sample do not

TABLE 4. *Social factors in a linear mixed-effects regression model for F1 of TRAP, including education, in the 2016 data*

Predictor		Coefficient	Mean	<i>n</i>	<i>p</i>
(Intercept)		838 Hz	787 Hz	3523	
<i>Style × age</i>	spontaneous	<i>reference level</i>		3249	0.0002
	wordlist	−0.94 Hz/year		274	
<i>Style</i>	wordlist	+56.4 Hz	786 Hz	274	
	spontaneous	<i>reference level</i>	787 Hz	3249	
<i>Style × gender</i>	wordlist × female	<i>reference level</i>	797 Hz	210	0.01
	spontaneous × male	<i>reference level</i>	785 Hz	1059	
	spontaneous × female	<i>reference level</i>	788 Hz	2190	
	wordlist × male	−28.5 Hz	750 Hz	64	
<i>Education × age</i>	no college	+1.48 Hz/year		967	0.07
	college	<i>reference level</i>		2556	
<i>Education</i>	college	<i>reference level</i>	789 Hz	2556	
	no college	−89.5 Hz	782 Hz	967	
<i>Gender</i>	male	+0.01 Hz	783 Hz	1123	
	female	<i>reference level</i>	789 Hz	2400	
<i>Age</i>		−0.43 Hz/year			

Note: Following consonant is included as a predictor in the model but omitted here for conciseness. Tokens of TRAP in styles other than wordlist or spontaneous are omitted from the model, as are speakers who were students when interviewed. Random factors: speaker, word.

satisfy the AE1 or EQ criteria, which the younger speakers in 2008 almost uniformly do, regardless of education level. Thus, even if we assume that less-educated individuals retain the NCS to a greater extent than more-educated speakers, they do so to a much lower degree than appeared to be the case in 2008.

Methodological concerns

The difference in TRAP F1 between the 2008 and 2016 data sets is so striking, and the real-time separation between them so short, that we must consider alternate possible explanations for the difference before accepting the conclusion that NCS raising of TRAP simply disappeared from spontaneous speech in Ogdensburg within a span of eight years. One possible explanation might be the methodological differences that do exist between the 2008 and 2016 studies.

As noted above, the formant measurements in the 2008 data were extracted at points selected by hand in Praat by Dinkin (2009), whereas the 2016 data had its formant measurements extracted by the FAVE software package (Rosenfelder et al., 2011). When we compared hand and FAVE measurements of a subsample of the 2016 tokens, although some individual tokens had large discrepancies, the mean difference between (unnormalized) measurements was less than 10 Hz, especially among younger age cohorts; and it is in younger cohorts that the difference between raised TRAP in the 2008 data and unraised TRAP in the 2016 data appears. It therefore seems unlikely that the difference between formant measurement methodologies could produce such a large-scale bias as to lead to

a mean difference of 76 Hz. To the extent that a bias did exist, it was toward slightly larger F1 values in hand measurements; since our comparison between the 2008 data and 2016 data finds *smaller* F1 values in 2008, if anything this suggests that the difference in methodologies might slightly understate the difference between the two data sets.

The two data sets were also collected under somewhat different circumstances. In 2008, the interviews were conducted by a native speaker of (non-NCS) American English, mostly via the Short Sociolinguistic Encounter method (Ash, 2002): individuals were recruited on the spur of the moment for brief interviews, the longest of which was twenty-five minutes. In 2016, interview subjects were recruited in advance, and interviews took place at scheduled times; these interviews lasted 1–2 hours and were conducted by a native speaker of German. It is possible that the difference in interview methodologies could have elicited a reduced rate of TRAP-raising in 2016.

It is not obvious a priori whether the long interviews in 2016 would be expected to elicit more or less monitored speech than the short interviews in 2008. On the one hand, the more involved procedure of setting up an appointment could have emphasized the formal nature of the interview, thus eliciting more careful speech in 2016; on the other hand, the greater length of the interviews in 2016 may have given speakers more time to become accustomed to the conversation and the interviewer, and relax into a more casual speech style (see Chambers, 2008; Milroy & Gordon, 2003). The fact that the interviewer in 2016 was a nonnative English speaker may have prompted speakers to attempt a more standard or careful pronunciation, out of concern that she might not understand more regionally marked variants. Since young speakers usually shift away from TRAP-raising in careful style, any factor that might have produced more careful speech overall in the 2016 interviews might be responsible for the lower overall level of TRAP in that data.

However, even if we assume for these reasons that, in 2016, spontaneous speech was produced in a more careful style than in 2008, that hardly seems likely to account alone for the magnitude of the difference between the TRAP F1 measurements of these two supposed styles. As Table 5 shows, among speakers born in the 1980s or later in the 2016 data, the mean difference between wordlist and spontaneous TRAP F1 is about 33 Hz. The difference between the mean spontaneous TRAP F1 of these speakers and the mean spontaneous TRAP F1 of the speakers born in the 1980s in the 2008 data is about 118 Hz: more than three times that difference. If we compare only speakers born in the 1980s in both data sets, the ratio is even larger. In other words, if the difference between spontaneous speech in 2008 and spontaneous speech in 2016 is merely because these represent two different *spontaneous styles*, the difference between these two styles is *at least three times as wide* as the difference between the 2016 spontaneous and wordlist styles. While we cannot entirely rule out this possibility, it strikes us as improbable that the difference between two spontaneous-speech interview styles with different interviewers, even if one is a nonnative speaker, would be so much greater than the difference between spontaneous speech and wordlist style with a single interviewer. In other words,

TABLE 5. *F1 TRAP means in young age cohorts*

Sample year	2008		2016	
	1980–1990		1980–1990	1980–2002
Spontaneous	665 Hz		761 Hz	783 Hz
Wordlist	730 Hz		783 Hz	816 Hz

it seems very likely that at least some of the difference between the 2008 and 2016 TRAP F1 is the result of real-time change: that is, that the community at large actually produced TRAP less raised in 2016 than in 2008.

DISCUSSION

Overall, the results presented above suggest that raised TRAP has been abandoned in Ogdensburg fairly suddenly. This is most evident in the real-time comparison between 2008 and 2016 speakers, where 2016 speakers produce TRAP significantly lower than 2008 speakers in both spontaneous speech and wordlist reading. The suddenness of this change is indicated by the absence of any apparent-time trends away from TRAP-raising in the spontaneous-speech data, as if the entire community quit TRAP-raising abruptly sometime between the two studies.

Although the change in spontaneous speech is visible in real time but not in apparent time, apparent-time trends are visible in the lowering of TRAP in wordlist style—and even more clearly in the reversal of the difference between wordlist and spontaneous speech.¹² The difference between the two styles shows the sociolinguistic profile expected for a change in progress, with women favoring the innovative pattern and a clear age correlation. The apparent-time change in style-shifting is such that older Ogdensburgers raise TRAP more in careful style than spontaneous speech, and younger people do the opposite; though not statistically significant, the social stratification between speakers of different education levels also appears to reverse over apparent time. The apparent-time reversal of the effect of style-shifting suggests that what is changing in Ogdensburg may be people's attitudes toward raised TRAP, and the evaluation of it as a nonstandard feature: raising is coming to be interpreted as more nonstandard, leading speakers to retreat from it, especially in careful style. This is reminiscent of, for example, the recent history of THOUGHT raising in New York City, which was reversed owing to negative social perception (e.g., Becker, 2014).

Therefore, we propose that a growing negative evaluation of raised TRAP may have been one of the motivations for speakers to have abandoned raised TRAP in favor of the nonlocal standard, unraised TRAP. The change in evaluation of raised TRAP follows the textbook profile of the process of a sociolinguistic “indicator” becoming a “marker,” defined by Labov (2001:196) as stages in the lifecycle of a language change from below:

Changes from below begin as *indicators*.... At this stage, they show zero degrees of social awareness.... As they proceed to completion, such changes usually acquire social recognition as linguistic *markers*, usually in the form of social stigma, which is reflected in... a steep slope of style-shifting, and negative responses on subjective reaction tests.

As NCS TRAP-raising becomes a marker in Ogdensburg,¹³ we see the predicted steep slope of style-shifting. Thiel (2019) also found negative responses to raised TRAP on subjective-reaction tests in Ogdensburg, meaning that both of Labov's predicted indices of social stigma have emerged.

The effect of gender is also consistent with what we might expect to see in a transition from an indicator to a marker. In 2008, the two male speakers appear to have lower spontaneous TRAP than the women do. In 2016, there is no gender difference in spontaneous speech, but women lead the trend toward lowering TRAP in wordlist style. This is reminiscent of the pattern found in the reversal of several sound changes in Philadelphia by Labov, Rosenfelder, and Fruehwald (2013). Labov et al. found that changes such as the fronting of MOUTH and GOAT, which had been documented as changes from below (Labov, 2001), are in the process of reversing, under pressure of negative social evaluation. They reported that the changes in both directions—both the original fronting of MOUTH and GOAT, and the retraction once the fronted variants had become stigmatized—were led by women. The data from Ogdensburg suggests the same pattern: in 2008, when TRAP-raising in spontaneous speech was still robust, it was apparently led by women; but the lowering of TRAP in wordlist style in 2016 is led by women as well.

The absence of any visible apparent-time trend away from raised TRAP in spontaneous speech is still a puzzle, given the clear apparent-time trend in style-shifting and the seemingly rapid change visible in the real-time comparison of the 2008 and 2016 data. While wordlist style shows what appears to be generational change—a change taking place as a result of younger generations acquiring a different set of standards than their elders held—spontaneous speech appears to display what Labov (1994:83–4) termed “communal change”: change due to all members of a community simultaneously adopting a linguistic innovation. It appears that, as perception of TRAP-raising as a nonstandard feature gradually grew, its nonstandard evaluation eventually became salient enough to drive speakers to avoid it in spontaneous speech. It seemingly achieved this level of salience in the community sometime between 2008 and 2016, leading to the absence of TRAP-raising even among speakers of an age cohort that had showed raising robustly in spontaneous speech in 2008. This, again, is reminiscent of patterns found in Philadelphia. Fruehwald (2017) found some evidence of communal change (which he terms a “zeitgeist” effect) away from the fronting of GOAT, which, like TRAP in Ogdensburg, was transitioning from an indicator to a marker in Philadelphia: men, although not women, retreated from GOAT-fronting via a communal change in the 1980s. GOAT was, however, the only one of the features that was becoming a marker to show evidence of communal change; it is not clear why some changes are subject to such “zeitgeist” effects and others are not.

Labov (1994) attributed communal change chiefly to lexical and syntactic change. It is well documented that members of a community may also participate in phonetic change in progress over their lifespans; but in those cases, the change is typically still visible in apparent time (e.g., Hollett, 2006; Sankoff & Blondeau, 2007). Other studies reporting the loss of TRAP-raising in the Inland North, such as Driscoll and Lape (2015), have been apparent-time studies inferring the change in progress from differences between age groups. Why then in Ogdensburg do we see the retreat from TRAP-raising in spontaneous speech in real time but not in apparent time?

One possibility is that the NCS was relatively new to Ogdensburg even in 2008. In the 2008 data, there is a suggestion of an apparent-time trend toward TRAP-raising in spontaneous speech, though the details are sketchy because there are only two speakers born before 1960. It may be the case that younger speakers acquired the NCS and then abandoned it via communal change, while older speakers never had advanced TRAP-raising to begin with. This could result in a situation such that younger speakers' retreat from TRAP-raising causes them to match the less-raised TRAP of their elders, producing a flat-seeming apparent-time profile. If the same change took place in a community in which the NCS was of longer standing, the result would look like the expected apparent-time age trend: Older speakers would have had higher TRAP to begin with, and, while younger speakers retreat from TRAP-raising due to stigma, older speakers would maintain raised TRAP due to lower sensitivity to the increasing stigma.

CONCLUSION

The findings presented in this paper contribute to the growing body of research on the recession of the NCS by addressing two issues that, so far, have remained unanswered in recent NCS research.

The first concern was to add a small city in a remote rural area to the mostly urban set of communities in which recent developments in the NCS have been studied. The findings indicate that TRAP-raising is on the retreat in Ogdensburg just as it has been found to be in more urban areas. There is evidence (albeit slight) that the change toward lower TRAP is led by college-educated speakers, while speakers without a college education are more likely to continue to make use of a somewhat raised variant. Since similar findings have been reported for Chicago, Buffalo, and Lansing, it seems that the reversal of TRAP affects this rural community in much the same ways as urban centers.

The second issue concerns the extent to which negative evaluations of raised TRAP affect its production. The reversal in the direction of style-shifting suggests that the abandonment of raised TRAP might be caused by an increasing social stigma that has become attached to raised TRAP. Unraised TRAP appears to be treated as the new standard that younger speakers favor in more careful speech. Exactly what caused the negative perception of raised TRAP in Ogdensburg remains to be investigated. However, it seems plausible that, as in Lansing (Nesbitt, 2018), the increased markedness of raised TRAP is connected to

Ogdensburg's economic decline, which started in the 1960s—the same decade in which, in apparent time, we observe unraised TRAP beginning to become the favored variant in careful production (see Thiel, 2019).

The short real-time span between the data sets discussed in this paper presented an unexpected opportunity to catch a community just at the point of retreating from the NCS. Without the 2008 data, it would not have been clear that Ogdensburg had ever been an NCS community. Our results indicate a role for communal change, or change across the lifespan, in Ogdensburg's retreat from NCS. It is not just the case that younger speakers are generationally moving away from the raised TRAP of their elders; the very same age cohort whose TRAP was apparently most raised in 2008, those born in the 1980s, has TRAP no higher than anyone else in 2016. This allows us to propose that communal change can play a larger role in the transition of a linguistic indicator into a marker than Fruehwald (2017) found in Philadelphia: once a dialect feature rises to the level of consciousness and develops sufficiently strong negative social evaluation, groups who share that evaluation may collectively retreat from the feature. In Ogdensburg, since the NCS was apparently relatively new to the community in the first place, the young cohorts' retreat from TRAP raising left them about on par with the older generation, creating the apparent-time illusion that no change was in progress at all.

NOTES

1. We identify vowel phonemes of English via the lexical sets of Wells (1982).
2. Although Driscoll and Lape's research (2015; see also Driscoll, 2016) included speakers from a handful of more rural communities within twenty miles of Syracuse, the current study is, to the best of our knowledge, the first to focus in depth on the loss of TRAP raising in a rural region remote from any major cities.
3. This compares to 7.5% and 7.4% in New York State and the US, respectively.
4. SUNY (State University of New York) and Clarkson University in Potsdam, both about thirty miles from Ogdensburg; and another SUNY campus and St. Lawrence University in Canton, each less than twenty miles from Ogdensburg.
5. One anonymous referee expressed concern that the number of tokens per speaker in this sample might be too small for each speaker's mean F1 of TRAP to be a reliable estimate of their central tendency for the phoneme. To test this, we calculated the distribution density in F1 of each speaker's spontaneous-speech TRAP tokens and compared the peak density of this curve with these tokens' mean F1 (see Appendix A). The absolute value of this difference ranges from 7 Hz to 45 Hz for these nine speakers, with a mean of 25 Hz (s.d. = 12 Hz). For the speakers interviewed in 2016, introduced in the next section, the corresponding values range from 3 Hz to 63 Hz, with a mean of 22 Hz (s.d. = 14 Hz). Thus, the speakers' TRAP means in the 2008 corpus are about equally adequate as estimates of central tendency as those in the 2016 corpus, which contains roughly six times as many tokens per speaker.
6. This participant was born in China but was adopted by a white family in Ogdensburg at the age of one. Since Ogdensburg does not have a substantial Asian community, this participant was fully integrated into the predominantly white community. Production patterns of this speaker do not deviate markedly from those of peers. Therefore, ethnic background was not considered in this study.
7. An anonymous referee expressed concern that the 2008 sample may not contain enough data for all of the vowels in the vowel space for Lobanov normalization to be appropriate. Every speaker in the 2008 data has eight or more tokens of each vowel phoneme except for the relatively rare vowels FOOT, CHOICE, and NURSE; the mean number of tokens per phoneme was 24.7. (Rhotic phonemes such as NEAR and SQUARE were mostly collapsed with nearby lax equivalents such as KIT and DRESS for the sake of this calculation.) Thus, we believe the speakers' overall vowel spaces are represented thoroughly enough in the data to enable Lobanov normalization. The distribution of the data among vowel phonemes—for example, 10% of tokens measured represent the DRESS vowel, 8% represent STRUT, and so on—are largely similar between the 2008 and 2016 corpora, differing by at most 3% (GOAT constitutes 9.5% of the 2016 data and 6.5% of the 2008 data).

8. To avoid assuming that either the hand measurements or automated measurements are more accurate, this is estimated by taking the mean of the two measurements.

9. A step-up procedure found only gender and no other factors significant; an ANOVA comparison finds the step-down model significantly better. The step-up model predicts that hand measurements are more likely to be larger than FAVE measurements for women than for men, by a difference of 18 Hz.

10. Although the AE1 benchmark of $F1(\text{TRAP}) < 700$ Hz was defined based on ANAE normalization, not Lobanov normalization, Dinkin (2018) suggested that a cutoff close to 700 Hz is adequate for diagnosing NCS using FAVE-scaled Lobanov-normalized measurements. Therefore, we will continue to use 700 Hz as our cutoff for the AE1 criterion. However, changing the normalization of the 2008 data does change *which* of those speakers satisfy the AE1 criterion. Under the original ANAE normalization, three speakers in the 2008 data had mean $F1$ of TRAP less than 700 Hz; using Lobanov normalization, five of them do.

11. A significant three-way age \times year \times style interaction is excluded from the model due to the likelihood that it is due to one outlying speaker in 2008.

12. An anonymous reviewer asked, how can we infer that the age pattern in style-shifting represents apparent-time change, given that, in interpreting the spontaneous-speech data, we reject the apparent-time model and conclude that change has taken place despite a flat apparent-time curve? The comparison between the 2008 and 2016 data is relevant here; the apparent-time model depends on the assumption that each generational cohort will continue to show roughly constant behavior over time. Although this assumption is false in the case of spontaneous speech in the current study, it is true of style-shifting: with the exception of one outlier, the 2008 interview subjects show ranges of style-shifting similar or identical to those of their cohortmates interviewed in 2016. Given real-time evidence against substantial change in range of style-shifting within a cohort, we infer that the age differences in style-shifting represent generational change.

13. In some communities, there is evidence that raised TRAP has moved on to the next stage and become a stereotype attracting overt comment (Driscoll & Lape, 2015; Savage et al., 2016), though there is no direct evidence for that development in Ogdensburg.

REFERENCES

- Ash, Sharon. (1999). Word list data and the measurement of sound change. Paper presented at NWAV (New Ways of Analyzing Variation) 28, Toronto.
- . (2002). The distribution of a phonemic split in the Mid-Atlantic region: Yet more on short a. *University of Pennsylvania Working Papers in Linguistics* 8.3:1–15.
- Becker, Kara. (2014). The social motivations of reversal: Raised BOUGHT in New York City English. *Language in Society* 43:395–420.
- Boersma, Paul & Weenink, David. (2005). Praat: Doing phonetics by computer [computer program]. Version 4.3. Available at <http://www.praat.org/>
- . (2020). Praat: Doing phonetics by computer [computer program]. Version 6.1. Available at <http://www.praat.org/>
- Chambers, J. K. (2008). *Sociolinguistic theory: Linguistic variation and its social significance*. Oxford: Wiley-Blackwell.
- Dinkin, Aaron J. (2009). *Dialect boundaries and phonological change in Upstate New York*. Ph.D. dissertation, University of Pennsylvania.
- . (2011). Weakening resistance: Progress toward the low back merger in New York State. *Language Variation and Change* 23.3:315–45.
- . (2013). Settlement patterns and the eastern boundary of the Northern Cities Shift. *Journal of Linguistic Geography* 1.1:4–30.
- . (2018). Revisiting the Inland North Fringe. Paper presented at NWAV (New Ways of Analyzing Variation) 47, New York.
- . (2020). The foot of the lake: A sharp dialect boundary in rural northern New York. *American Speech* 95(3):321–55.
- Dodsworth, Robin & Kohn, Mary. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24.2:221–45.
- D’Onofrio, Annette, & Benheim Jaime. (2020). Contextualizing reversal: Local dynamics of the Northern Cities Shift in a Chicago community. *Journal of Sociolinguistics* 24:469–91.
- Driscoll, Anna. (2016). *Cold winters, flat A’s: Linguistics, geography, and the Northern Cities Shift in Syracuse*. Undergraduate honors thesis, Dartmouth College.
- Driscoll, Anna & Lape, Emma. (2015). Reversal of the Northern Cities Shift in Syracuse, New York. *University of Pennsylvania Working Papers in Linguistics* 21.2:41–7.

- Durian, David. (2014). Another look at the short-a system of late 19th and early 20th century Chicago in Pederson's *PEMC* Data, *DARE*, and *LANCS*. Paper presented at NWAV 43, Chicago.
- Durian, David & Cameron, Richard. (2018). Another look at the development of the Northern Cities Shift in Chicago. Paper presented at NWAV (New Ways of Analyzing Variation) 47, New York.
- Eckert, Penelope. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society* 17.2:183–207.
- Fruehwald, Josef. (2017). Generations, lifespans, and the zeitgeist. *Language Variation and Change* 29.1:1–27.
- Gordon, Matthew J. (2000). *Small-town values, big-city vowels: A study of the Northern Cities Shift in Michigan (Publications of the American Dialect Society 84)*. Durham: Duke University Press.
- Hollett, Pauline. (2006). Investigating St. John's English: Real- and apparent-time perspectives. *Canadian Journal of Linguistics* 51.2/3:143–60.
- Kapner, Julianne. (2019). Snowy days and nasal A's: The retreat of the Northern Cities Shift in Rochester, New York. Poster presented at NWAV 48, Eugene, Oregon.
- King, Sharese. (2017). African American identity and vowel systems in Rochester, New York. Paper presented at NWAV 46, Madison, Wisconsin.
- Labov, William. (1994). *Principles of linguistic change*. Vol. 1. *Internal factors*. Oxford: Blackwell.
- . (2001). *Principles of linguistic change*. Vol. 2. *Social factors*. Oxford: Blackwell.
- . (2010). *Principles of linguistic change*. Vol. 3. *Cognitive and cultural factors*. Oxford: Wiley-Blackwell.
- Labov, William, Ash, Sharon, & Boberg, Charles. (2006). *Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
- Labov, William, Rosenfelder, Ingrid, & Fruehwald, Josef. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 89(1): 30–65.
- Labov, William, Yaeger, Malcah, & Steiner, Richard. (1972). *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.
- Lobanov, Boris M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49:606–8.
- McCarthy, Corrine. (2010). The Northern Cities Shift in real time: Evidence from Chicago. *University of Pennsylvania Working Papers in Linguistics* 15.2:101–10.
- . (2011). The Northern Cities Shift in Chicago. *Journal of English Linguistics* 39.2:166–87.
- Milholland, Agatha. (2018). Reversal of the Northern Cities Shift in Buffalo, NY. Paper presented at NWAV (New Ways of Analyzing Variation) 47, New York.
- Milroy, Lelsey, & Gordon, Matthew. (2003). *Sociolinguistics: Method and interpretation*. Oxford: Blackwell.
- Near, Terrance Michael. (1977). *Phonetic feature systems for vowels*. Bloomington, Indiana: Indiana University Linguistics Club.
- Nesbitt, Monica. (2018). Economic change and the decline of raised trap in Lansing, MI. *University of Pennsylvania Working Papers in Linguistics* 24.2:67–76.
- . (2019). *Changing their minds: The impact of internal social change on local phonology*. Ph.D. dissertation, Michigan State University.
- Nesbitt, Monica & Mason, Alexander. (2016). Evidence of the Elsewhere Shift in the Inland North. Paper presented at NWAV (New Ways of Analyzing Variation) 45, Vancouver, BC.
- Preston, Dennis R. (1998). Why we need to know what real people think about language. *The Centennial Review* 42.2:255–84.
- Rosenfelder, Ingrid, Fruehwald, Josef, Evanini, Keelan, & Yuan, Jiahong. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite. Retrieved from <http://fave.ling.upenn.edu>
- Sankoff, Gillian, & Blondeau, Hélène. (2007). Language change across the lifespan: /r/ in Montreal French. *Language* 83.3:560–88.
- Savage, Matthew, Mason, Alex, Nesbitt, Monica, Pevan, Erin, & Wagner, Suzanne Evans. (2016). Ignorant and annoying: Inland Northerners' attitudes toward Northern Cities Shift short-o. Poster presented at the American Dialect Society Annual Meeting, Washington, DC.
- Severance, Nathan, Evanini, Keelan, & Dinkin, Aaron. (2015). Examining the performance of FAVE for automated sociophonetic vowel analyses. Paper presented at NWAV (New Ways of Analyzing Variation) 44, Toronto.
- Thiel, Anja. (2019). *A northern city going elsewhere: Apparent and real-time sound change in Ogdensburg, New York*. PhD dissertation, University of Bern.
- Turton, Danielle & Baranowski, Maciej. (2020). Not quite the same: The social stratification and phonetic conditioning of the FOOT-STRUT vowels in Manchester. *Journal of Linguistics*.

U.S. Census Bureau. American FactFinder. Retrieved from <https://factfinder.census.gov/> (December 17, 2018)
 Wagner, Suzanne Evans, Mason, Alexander, Nesbitt, Monica, Pevan, Erin, & Savage, Matt. (2016). Reversal and re-organization of the Northern Cities Shift in Michigan. *Penn Working Papers in Linguistics* 22.2:171–9.
 Wells, J. C. (1982). *Accents of English I: An introduction*. Cambridge: Cambridge University Press.

APPENDIX A

Density plots of speakers’ normalized spontaneous-speech TRAP F1 measurements, illustrating the degree to which speaker-level means approximate the central tendency of token density. Speakers’ mean F1 is shown with a vertical dotted line. Tokens before sonorants and after glides and stop+obstruent clusters are excluded, following the ANAE methodology for calculating speaker means.

TABLE A1. 2008 Interviews

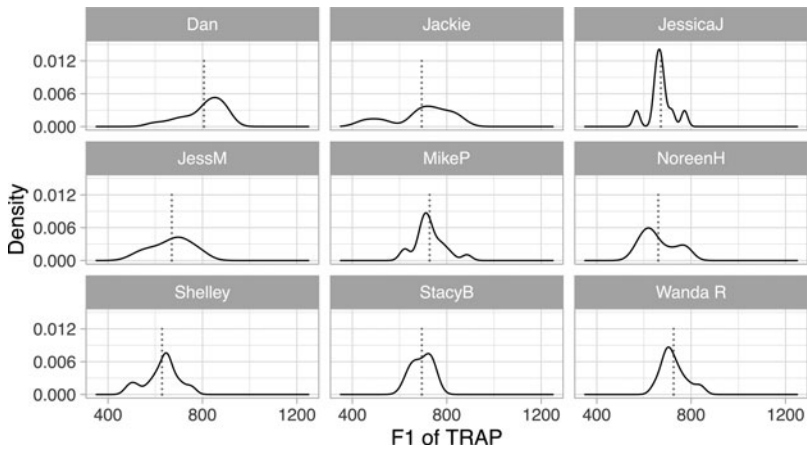
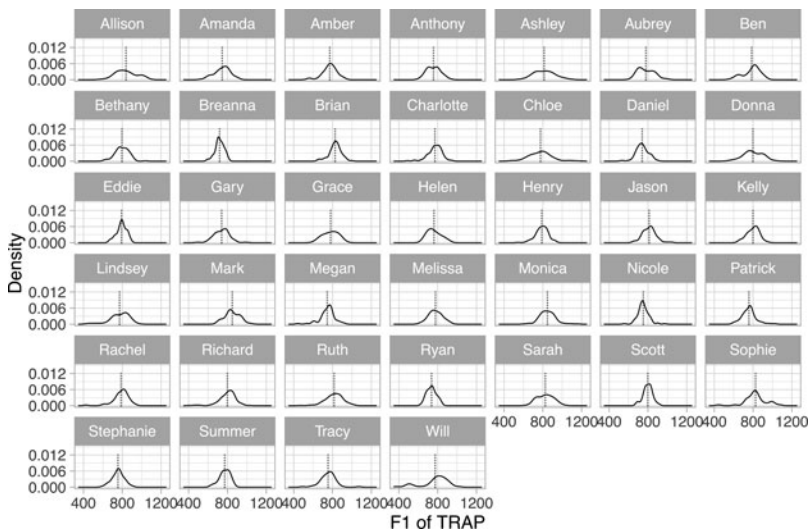


TABLE A2. 2016 Interviews



APPENDIX B

TABLE B1. *The effect of the following consonant ($p < 10^{-4}$) included in the regression model presented in Table 3*

Consonant	Coefficient	Mean	<i>n</i>
p	<i>reference level</i>	818 Hz	233
f	-2.6 Hz	802 Hz	287
θ	-5.5 Hz	805 Hz	112
s	-14.2 Hz	801 Hz	719
k	-16.3 Hz	783 Hz	1356
č	-31.9 Hz	781 Hz	81
t	-33.5 Hz	780 Hz	294
b	-35.6 Hz	778 Hz	134
g	-36.6 Hz	786 Hz	89
l	-37.5 Hz	775 Hz	141
z	-42.4 Hz	713 Hz	27
š	-43.9 Hz	773 Hz	102
v	-45.5 Hz	780 Hz	176
ð	-51.1 Hz	756 Hz	17
d	-53.0 Hz	755 Hz	336
ž	-60.2 Hz	746 Hz	5
ǰ	-68.8 Hz	742 Hz	131