



10-1-2015

# Ethnic Orientation without Quantification: How Life “On the Hyphen” Affects Sociolinguistic Variation

Luiza Newlin-Lukowicz

---

# Ethnic Orientation without Quantification: How Life “On the Hyphen” Affects Sociolinguistic Variation

## **Abstract**

The effect of ethnicity on sociolinguistic variation has been studied quantitatively from multiple angles. The existing methodological range has uncovered that ethnicity is reflected in various aspects of social life. However, the diversity of existing measures has i) prevented a direct comparison between studies and ii) relied on subjective quantification of social data. This paper introduces a novel methodology that unites previous measures and removes the need for subjective quantification. Specifically, I apply hierarchical cluster analysis to social data collected from an ethnographically-informed Polish-involvement survey to assess the impact of multiple social factors on regional variation for Polish New Yorkers. The analysis identifies the maintenance of transnational ties as the strongest predictor of linguistic variation and reveals speakers’ “hyphenated” (Polish-American) identities.

# Ethnic Orientation without Quantification: How Life “On the Hyphen” Affects Sociolinguistic Variation

Luiza Newlin-Lukowicz\*

## 1 Introduction

Ethnicity has been increasingly recognized as an important force driving sociolinguistic variation for minority groups. Yet because of its inherent fluidity and complexity, ethnic identity has been elusive to operationalize, which is evident in the range of approaches to the study of ethnolinguistic variation. This range includes qualitative, ethnographic accounts centered on interaction (e.g., Mendoza-Denton 2008), as well as quantitative analyses with a focus on social networks (Dubois and Horvath 2000, Sharma 2011), ethnic orientation (Hoffman and Walker 2010, Nagy, Chocie and Hoffman 2012), ethnic lifestyle (Wong 2007), and transnational ties (Sharma 2014). The vast array of methods is necessary to capture the fact that ethnicity may be reflected in various aspects of social life. Indeed, depending on the community, some aspects of ethnic identity may take precedence over others. However, the diversity of existing measures has prevented sociolinguists from establishing generalizable principles for cross-community and cross-linguistic study of ethnicity (Nagy et al. 2012). This issue is confounded by the common reliance on subjective quantification of social data in assessing (and effectively “measuring”) ethnic identity.

Much of the work that has investigated ethnolinguistic variation from a quantitative perspective has relied on survey data (e.g., Dubois and Horvath 2000, Hoffman and Walker 2010, Nagy et al. 2012, Newlin-Lukowicz 2013, Sharma 2014, Wong 2007). In each of these studies, the survey was tailored to the needs of a specific community, and thus necessarily involved a different set of questions, a different answer format, and finally, a different method for the quantification of results. The quantification issue thus subsumes subjective decisions about all of the following: Should scores be added or averaged? Should the survey be treated as a whole, or should questions be divided into subsets? And, at a very basic level, how can we determine whether a question is at all informative about speakers’ ethnic identity?

This paper investigates whether ethnic identity is reflected in the English of Polish New Yorkers by analyzing speakers’ answers to a survey not unlike the ones mentioned above. The analysis will introduce a novel methodology that unites previous measures and removes some of the need for subjective quantification. Specifically, I apply hierarchical cluster analysis (HCA) to survey data to assess the impact of multiple social factors on regional variation for Polish New Yorkers. HCA groups speakers based on the *similarity* of their answers to survey questions, rather than on some raw score. In so doing, HCA simultaneously analyzes multiple social factors and groups speakers into subsets defined by similar answers to core questions. To establish subsets, HCA assesses which questions are informative for clustering and ignores the ones that are not.

The analysis presents social and linguistic data. The social information is drawn from an ethnographically-informed Polish-involvement survey that probed multiple aspects of speakers’ lives: i) Polish lifestyle, ii) Polish community involvement, and iii) transnational ties. The linguistic data involves three regional features of New York City English, with tokens derived from sociolinguistic interviews and word lists, and analyzed acoustically. The results suggest that Polish New Yorkers do not form a homogenous group with respect to ethnic identity. The maintenance of transnational ties polarizes the group into two camps: those with strong and weak ties to Poland. Speakers with strong ties to the motherland further differ with respect to their degree of involvement in the Polish New York City community. These social distinctions are found to account for the co-occurrence of regional variables of New York City English: speakers with strong transnational ties manifest high TH-stopping rates and a lowered /ɔ/, while speakers who are additionally involved in the Polish New York City community exhibit a robust nasal split for /æ/. I argue that these results reflect speakers’ “hyphenated” identities as Polish-Americans and underscore the individual aspect of ethnic identity.

---

\*This work was supported by an NSF grant to Luiza Newlin-Lukowicz and Gregory Guy (BCS-1324916).

## 2 Clustering of Social Information

The social and linguistic data for this study are drawn from a larger project on linguistic variation in two generations of Polish New Yorkers (Newlin-Lukowicz 2015). The project involved extensive ethnography followed by quantitative analysis of acoustic information. The sample reported on here includes thirty-five speakers, all Polish-English bilinguals whose parents immigrated to New York City. Those classified as generation one (N=15; 8 women, 7 men) moved to New York City between the ages of eight and fifteen. Those classified as generation two (N=20; 10 women, 10 men) were born in the City or immigrated before age five. All speakers were close in age (19–35; one speaker was 46), were raised in working-class homes and neighborhoods, and held college degrees or were enrolled in college at the time of the study.

### 2.1 Polish-involvement Survey

Based on ethnographic work (for details, see Newlin-Lukowicz 2015), the following three aspects of social life have emerged as reflecting ethnic identity for Poles in New York: i) Polish lifestyle, ii) involvement in the Polish community, and iii) transnational ties. An investigation of speakers' patterning along these three dimensions has the potential to reveal the complexity of ethnic identifications for Polish New Yorkers. Therefore, I devised a Polish-involvement survey (for details, see Newlin-Lukowicz 2015) that was divided into three components. Responses were measured on a 4-point Likert scale, indicating frequency/quantity, and were assigned 0–3 points.

The first component assessed speakers' Polish lifestyle through six questions probing the frequency with which Polish New Yorkers engaged in daily activities associated with this lifestyle, such as shopping at Polish delis, using Polish-owned businesses, and seeing Polish doctors. The second component included eleven questions which quantified speakers' involvement in the Polish New York City community. Unlike the first component, this one aimed to capture the frequency of activities that required an involvement with the community that went beyond a business exchange. This component included questions about speakers' participation in cultural organizations and events, participation in extra-curricular activities that revealed contact with Polish-speaking teachers, and use of Polish-American media. In addition, speakers reported what proportion of their friends were Polish-American.

Lastly, the third component included six questions that centered around transnational ties. Speakers were asked to report whether they had Polish friends in Poland and whether they used Polish-based media. This component additionally included questions about speakers' knowledge of Polish culture in an attempt to capture ties that could be more affective. Participants were asked to name Polish bands and movies (one point for each band/movie mentioned, up to 3 points), and to sing the beginning verses of the Polish anthem (3 points if they could, 0 if they could not).

### 2.2 Hierarchical Cluster Analysis (HCA)

Each speaker's answers to questions within one of the three survey components were added together, and the resulting three factors were subjected to HCA in R (using Ward's method and Euclidean distances). Figure 1 provides the results of the analysis in a schematic form. Overall, transnational ties were identified as the most robust factor for clustering survey data, and divided speakers into those with *weak* and *strong transnational ties*. Speakers' involvement in the Polish New York City community was found to be meaningful only for speakers with strong transnational ties, and further sub-divided these speakers into those characterized by *high* and *low community involvement*. Speakers' Polish lifestyle was not found to be informative for clustering.<sup>1</sup> As illustrated in Figure 1, HCA identified three main social clusters.

The social clusters obtained can be conceived of as representing different levels on a scale of ethnic orientations, as illustrated in Figure 2. The first group could be characterized as having an American orientation. These speakers, who are mostly second generation, establish rather weak

---

<sup>1</sup>The absence of an effect of Polish lifestyle may suggest that this lifestyle reflects whether or not a speaker grew up in a Polish neighborhood, rather than the type of ethnic identity the speaker has developed.

ties with Poles in Poland and those in New York City. The second group could be seen as oriented toward the Polish community in New York City. This group comprises a mix of first and second generation speakers who maintain strong ties to Poland, and are additionally involved in the life of the New York City Polish community. The last group is oriented toward Poland. Speakers falling within this cluster are mostly first generation Poles who remain connected to Poland. However, unlike the previous group, these speakers do not foster a relationship with New York City Poles. For convenience sake, I will refer to the three clusters by these simplified ethnic orientations.

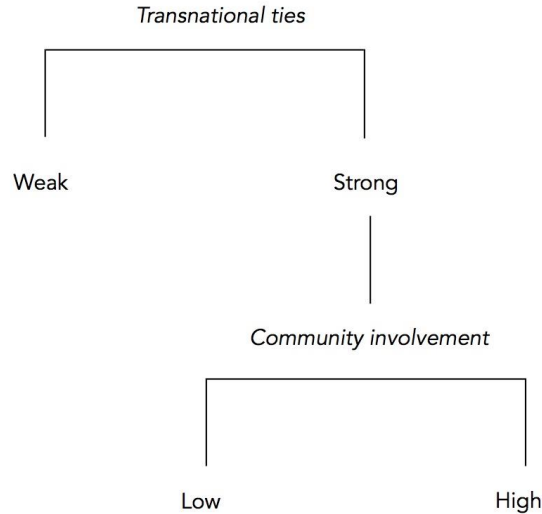


Figure 1: Schematic output of the cluster analysis based on responses to the Polish-involvement survey.

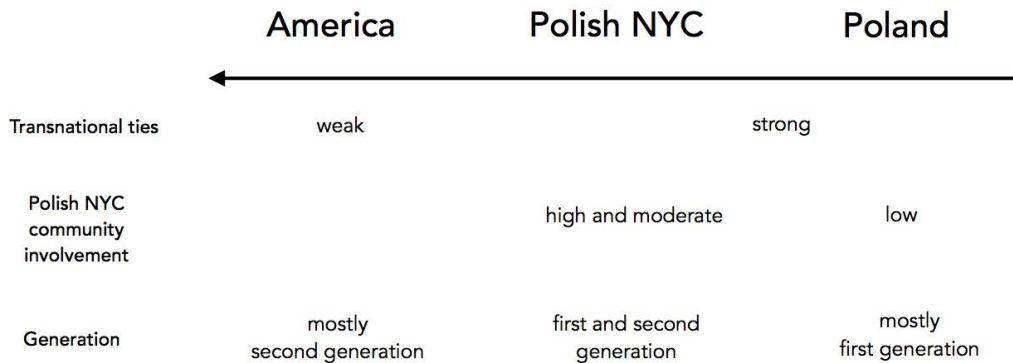


Figure 2: Ethnic orientations of Polish New Yorkers.

### 3 Linguistic Variables

The analysis of survey data underscores the social heterogeneity within the Polish community in New York City. Although speakers belonging to each generation have a similar social background, they can be seen as displaying different ethnic orientations that undoubtedly reflect their individual experiences. This section will investigate whether the three social clusters account for the linguistic variation exhibited by Polish New Yorkers. I will focus on speakers' production of three variables reported for New York City English. One of these variables, TH-stopping, has been argued to represent an ethnic marker for Polish New Yorkers. Of the remaining two variables, raised /ɔ/ has been declining from some segments of the population, but remains robust for non-white New Yorkers. The last variable, the nasal split for /æ/, has been recently reported to be replacing the traditional short-*a* split, without any obvious differences along ethnic lines.

### 3.1 TH-stopping

TH-stopping refers to the realization of stops for underlying interdental fricatives, as in [d]is for *this* and [t]in for *thin*. In the United States, TH-stopping has been reported in various immigrant communities, where its origins are often attributed to L1 interference, e.g., from Cajun French in Louisiana (Dubois and Horvath 1998), from Spanish in Latino communities (e.g., Mendoza-Denton 2008), and from German in Wisconsin (Rose 2006). In New York City, TH-stopping is associated with social meanings at the intersection of ethnicity, gender, and class. It has been a longstanding feature of the dialect, linked to Italian men (Labov 1966) and uneducated, lower class speech (Babbitt 1896, Frank 1948). Previous research has attributed TH-stopping to a substrate effect produced by the first immigrant groups, such as the Irish, Italians and Poles.

In contemporary New York City, TH-stopping has been argued to serve as an ethnic marker for second generation Polish New Yorkers (Newlin-Lukowicz 2013, 2014, 2015). TH-stopping rates were found to decline across two generations of Polish New Yorkers, which suggests that the production of stops represents L1 interference for first generation Poles. However, the second generation produced markedly higher rates of stopping than non-Polish white New Yorkers and realized TH-stopping with Polish-like voice onset time (i.e., the voiced stop in [d]en for *then* was pre-voiced, and the voiceless stop in [t]in for *thin* lacked aspiration), which distinguished them from both first generation Poles and non-Polish New Yorkers (Newlin-Lukowicz 2014).

### 3.2 Raised /ɔ/

Raised and in-gliding /ɔ/ has been a hallmark feature of New York City English for quite some time (Babbitt 1896). The distribution of this variable is not limited to New York City, as raised /ɔ/ can be heard in other areas of the East Coast (Labov, Ash and Boberg 2006). Despite its wider presence, raised /ɔ/ seems to be recognized by outsiders as symbolically New York, which is evidenced by media representations of New York City English that typically highlight this feature. In contemporary New York City, raised /ɔ/ seems to carry associations with older white New Yorkers, and with negative characteristics, such as meanness and aloofness (Becker 2014).

Production studies have shown that raised /ɔ/ has been lowering in apparent time in New York City (Becker 2010) and elsewhere (Labov, Rosenfelder and Fruehwald 2013). In New York City, lowering is affected by ethnicity and is being led by white New Yorkers in Manhattan, with other ethno-racial groups, such as African Americans (Coggshall and Becker 2010, Becker 2010), Caribbean Americans (Blake, Shousterman and Newlin-Lukowicz 2015), and Latinos (Slomanson and Newman 2004), as well as white residents in other boroughs (Newman 2014), still using the raised variant. Lowering of /ɔ/ is accompanied by the closing of the phonemic distance between /ɔ/ and /ɑ/ (Becker 2010): the more raised /ɔ/ is, the larger the distance.

### 3.3 Nasal Split for /æ/

The raising of /æ/ is another well-known feature of New York City English (Babbitt 1896, Labov 1966). Traditionally, New York City has exhibited a complex, phonologically-conditioned pattern of raising of /æ/ that created a predictable split between tense tokens (before voiced stops, voiceless fricatives, and front nasals /m, n/) and lax tokens (elsewhere), notwithstanding numerous lexical exceptions (Labov 2007). Recent studies have suggested a reorganization of this complex system. Specifically, white and non-white New Yorkers alike are moving toward the nasal system (Becker 2010, Becker and Wong 2010), reported for much of the U.S., where /æ/ raises before any type of nasal, e.g., *ban, bam, bang* (Labov et al. 2006).

### 3.4 Measures

The co-occurrence of the three variables will be analyzed using the following measures. TH-stopping will be represented by average stopping rates in sociolinguistic interviews. Those rates were established by contrasting speakers' realization of stops (acoustically defined by the presence of closure and burst) with their faithful realization of underlying interdental fricatives (acoustically defined by the absence of closure and burst and by the presence of frication). The mean stopping

rate took into account both voiced and voiceless stops/fricatives, and included at least 200 and at most 500 tokens per speaker, extracted from the middle thirty minutes of each interview.

Raised /ɔ/ will be discussed in terms of the acoustic distance between low back vowels. For each speaker, thirty tokens of /a/ and /ɔ/ were extracted from a word list (see Newlin-Łukowicz 2015 for details), and their onsets (25% of vowel length) were analyzed acoustically in Praat to measure formants and duration. The values of the first two formants of the vowels' onsets were normalized against four other vowels (/i/, /u/, /ε/ and /æ/) using the Vowels package in R (Kendall and Thomas 2012), and following the normalization method reported in Labov et al. 2006. For each speaker, normalized tokens of /a/ and /ɔ/ were then subjected to a Multivariate Analysis of Variance (MANOVA). MANOVA makes it possible to investigate multiple dependent variables simultaneously (i.e., F1 and F2), which makes it an attractive method for the study of mergers (Hay, Warren and Drager 2006, Hall-Lew 2009, Wong and Hall-Lew 2014). Rather than relying on average formant values, a MANOVA takes into account the entire distribution of tokens and calculates the degree of overlap between any two vowels. The output of the MANOVA that directly represents the degree of overlap is the *Pillai-Bartlett statistic*. The so-called Pillai score ranges between 0 and 1. The closer the score is to 1, the more distinct the vowels are; the closer the score is to 0, the larger the degree of overlap. For the purpose of this paper, the MANOVAs investigated the degree of distinction between /a/ and /ɔ/ in F1/F2 space. Higher Pillai scores were assigned to speakers who raised /ɔ/, and thus displayed a larger separation between /a/ and /ɔ/.

The final variable, the nasal split for /æ/, will likewise be investigated using Pillai scores from MANOVAs. The data come from a word list that included fifty-four tokens of /æ/ (see Newlin-Łukowicz 2015 for details), representing different raising and lowering environments. Tokens of /æ/ were analyzed and normalized following the same technique detailed for the low back vowels above. For each speaker, tokens of /æ/ were subjected to a MANOVA that compared the overlap of pre-nasal /æ/ with all other tokens of /æ/. Here, a larger separation of tokens implies the adoption of the nasal split. Again, this distinction is captured in terms of Pillai scores ranging from 0 (no nasal split) to 1 (robust nasal split).

#### 4 Linguistic Profile of Social Clusters

I will now investigate whether the social clusters representing different ethnic orientations (i.e., toward America, the Polish New York City community, and Poland) exhibit coherent linguistic behavior. I treat each social cluster separately, and discuss the differences observed between and within clusters. The results for speakers with a given ethnic orientation are presented in the form of density graphs. The realization of the consonantal variable is displayed separately in Figure 3, which illustrates the probability (y-axis) of TH-stopping rates (x-axis) for speakers with different ethnic orientations. Means are graphed as lines that intersect the x-axis. Figure 3 suggests that TH-stopping rates differ according to ethnic orientation: speakers oriented toward Poland occupy the upper end of the range, whereas those oriented toward America are more likely to fall on the lower end. This finding is consistent with the characterization of this variable as a Polish ethnic marker (Newlin-Łukowicz 2013, 2014, 2015).

The two vocalic variables are illustrated in Figure 4, which displays their realization according to ethnic orientation. In this series of graphs, the y-axes again represent probability, whereas the x-axes capture two measures of distinction in the form of Pillai scores. The dotted lines represent the contrast between the low back vowels /a/ and /ɔ/, and the solid lines represent the contrast between pre-nasal and other tokens of /æ/. The positioning of these lines with respect to the y-axis depends on how likely speakers are to exhibit a particular Pillai score. The relevant information in these graphs is the presence of peaks, which reflect a large probability that speakers with a given orientation will exhibit a given Pillai score. For the two vocalic variables, peaks corresponding to scores of .5 and higher suggest that speakers tend to produce a large separation between the low back vowels and between tokens of /æ/, and consequently, imply the adoption of raised /ɔ/ and of the nasal split, respectively. A concentration of scores lower than .5 suggests that speakers tend not to distinguish the two contrasts, producing /ɔ/ as low and not raising pre-nasal /æ/. A cursory investigation of peaks in Figure 4 suggests that the realization of variables differs according to ethnic orientation.

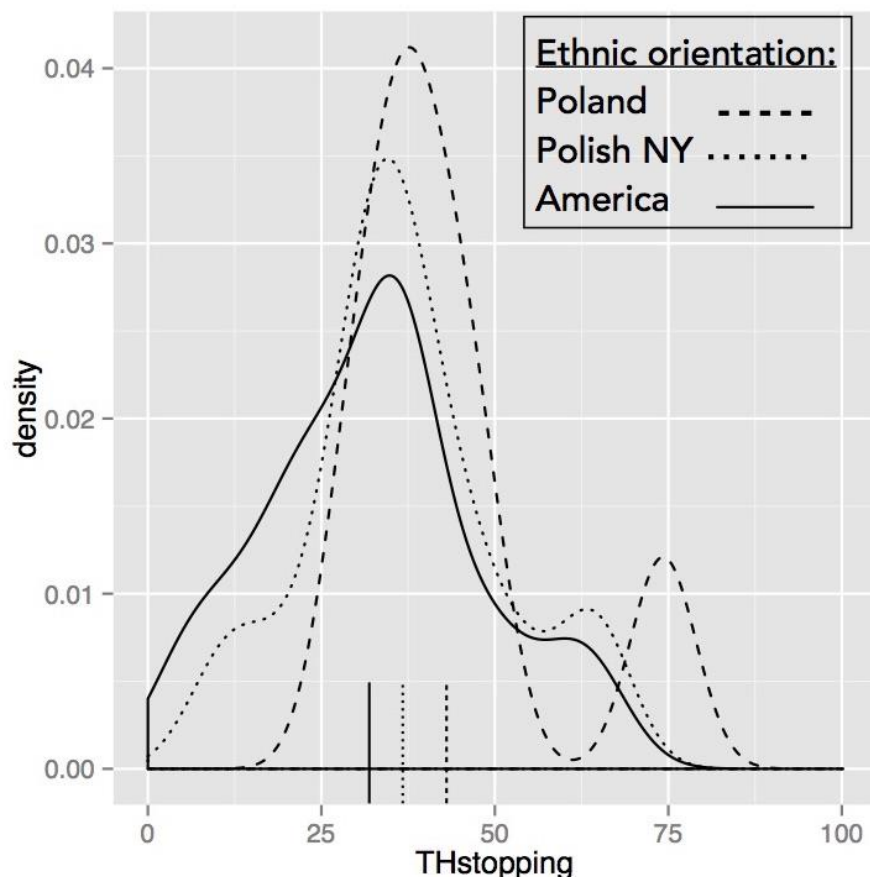


Figure 3: TH-stopping rates according to ethnic orientation. Lines intersecting the  $x$ -axis represent corresponding means.

#### 4.1 Orientation toward America

A group of ten speakers, seven of whom are second generation, were identified as oriented toward America. This group of speakers displays the lowest TH-stopping rates, with an overall mean of 32.6%, which further decreases to 25.5% for second generation speakers as compared to the first generation's 49%. Figure 3 illustrates that speakers oriented toward America tend to be concentrated on the low end of TH-stopping rates. In fact, the only two speakers who manifest TH-stopping rates at or above 50% for this group are first generation Poles.

The realization of the two vocalic variables by speakers oriented toward America is illustrated in the leftmost panel of Figure 4. This group tends to produce robust low back and short-*a* contrasts, seen in the fact that the density lines representing the two variables form prominent peaks that lean toward more distinction. In fact, most of the speakers oriented toward America produce raised /ɔ/ and a nasal split, and are assigned mid and high Pillai scores. However, the density lines in the leftmost panel in Figure 4 appear to be mildly bimodal, as a smaller peak is present on the side of less distinction for both variables. For the low back contrast, this second peak reflects the fact that the two second generation females in this group produce a low, rather than raised, /ɔ/. For the short-*a* contrast, the additional bump reflects the fact that two men, one first generation and one second generation, do not produce a nasal split. The generalization is that speakers oriented toward America tend not to engage in the Polish-identified variable (TH-stopping), but adopt the New York City-identified vocalic variables: raised /ɔ/ (with the exception of second generation women) and the nasal split.



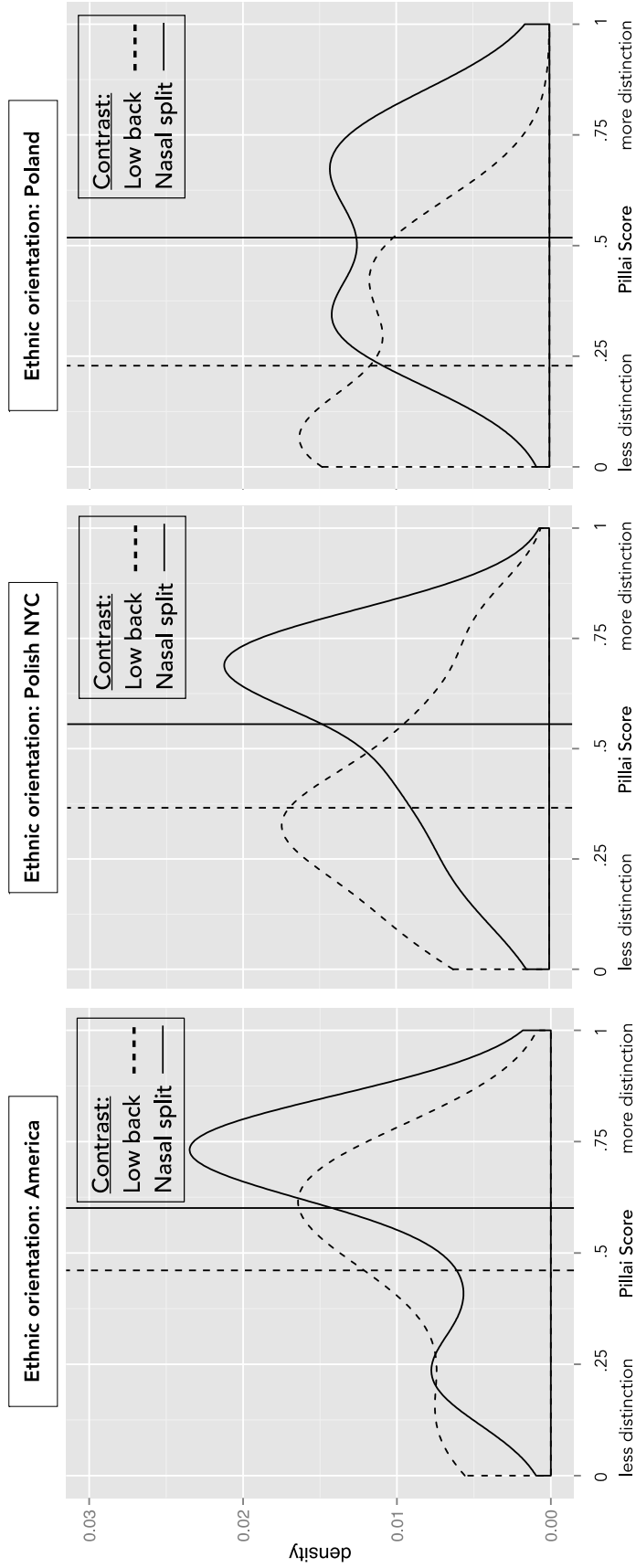


Figure 4: Robustness of low back contrast and nasal split for speakers with ethnic orientation toward America (left panel), the Polish NYC community (middle panel), and Poland (right panel). Lines intersecting the x axis represent corresponding means.

#### 4.2 Orientation toward Polish New York City

The social cluster identified as oriented toward the Polish New York City community was balanced in terms of generation, as it included ten second generation speakers and eight first generation speakers. As a whole, this group displays a different profile of variable use than speakers oriented toward America.

Unlike the group oriented toward America, speakers oriented toward Polish New York City are characterized by mid to high TH-stopping rates, with an overall mean of 37% (see Figure 3). In fact, only three speakers produce stopping rates below 25%: one first generation speaker and two second generation speakers. Further contrasting with the previous group, the speakers oriented toward the local Polish community generally produce a weak distinction of the low back vowels. Accordingly, in the middle panel in Figure 4, the dotted line representing the low back contrast has moved toward the side of less distinction. In fact, most speakers in this group do not raise /ɔ/, and instead, realize this vowel as very low, with the exception of four first generation speakers. The only similarity between this group and the speakers oriented toward America is the adoption of the nasal split, as illustrated by the fact that the solid density lines representing this variable show a qualitatively similar profile in the leftmost and middle panels in Figure 4. However, this contrast is less pronounced for first generation speakers, which likely reflects interference from Polish. As a whole, this group exhibits high rates of the ethnic marker, TH-stopping, and participates in sound changes that are ongoing in the dialect, such as /ɔ/-lowering and the nasal split.

#### 4.3 Orientation toward Poland

The group identified as oriented toward Poland includes only seven speakers. Four of them are first generation and three are second generation. These speakers further differentiate variable use, but are linguistically more similar to the group oriented toward the Polish New York City community than toward speakers oriented toward America.

The speakers oriented toward Poland manifest high TH-stopping rates, regardless of generation. Indeed, Figure 3 illustrates that the distribution of TH-stopping rates for this group occupies a higher range than for either of the remaining groups: speakers produce TH-stopping at rates no lower than 25% and up to 74%. Just like the speakers oriented toward the Polish New York City community, this group realizes /ɔ/ as low, rather than raised, but the first generation speakers exhibit slightly more differentiation of the low back vowels. Unlike the previous two groups of speakers, those oriented toward Poland do not produce a robust nasal split. The absence of a robust low back and short-*a* contrast is made evident in the rightmost panel in Figure 4, where the density lines show a clear preference for low /ɔ/ and no clear preference for the nasal split. In light of the fact that this group's English is also characterized by the highest rates of TH-stopping, the slight separation of the low back vowels and the non-adoption of the nasal split likely reflect interference from Polish, and may be perceived as a "Polish accent." Together, these findings suggest that speakers oriented toward Poland favor high rates of the Polish-identified feature (TH-stopping), disfavor the production of raised /ɔ/, and display a wider range of /æ/-raising.

### 5 Discussion

Ethnic orientation appears to have an effect on linguistic variation for Polish New Yorkers. Depending on whether Polish New Yorkers orient toward America, the Polish New York City community, or Poland, they favor a different combination of features. Speakers oriented toward America exhibit low TH-stopping rates, produce raised /ɔ/, and adopt a robust nasal split. Speakers oriented toward the Polish New York City community likewise adopt the nasal split, but produce intermediate TH-stopping rates and a low /ɔ/. Lastly, orientation toward Poland correlates with the highest TH-stopping rates, a low /ɔ/, and crucially, the absence of a robust nasal split. Hence, the three social groups are not fully differentiated in terms of variable use. Rather, they can be thought of as representing a continuum of variable use, with the conceptually intermediate group (i.e., speakers with an orientation toward Polish New York City) overlapping in their choice of variants with each of the remaining groups. Specifically, those oriented toward Polish New York City mir-

ror American-oriented speakers in their adoption of the nasal split, and pattern with Polish-oriented speakers on their realization of /ɔ/ as low.

The linguistic differences observed between speaker clusters can be accounted for if the variables' social meanings are considered. TH-stopping has been argued to constitute interference from Polish and to serve as an ethnic marker for the Polish community in New York City, when produced with Polish-like voicing (Newlin-Lukowicz 2013, 2014, 2015). Associations with a Polish accent and Polish identity may explain why TH-stopping rates are the highest for the group oriented toward Poland, and why they decrease as speakers' orientation shifts toward America. Further, the preference for low /ɔ/ for speakers oriented toward the Polish New York City community and toward Poland may reflect alignment with other New York City whites. This variable choice reflects the fact that ethnicity and race are intertwined: a strong Polish identity often goes hand in hand with a white identity. As such, low /ɔ/ reinforces the Polishness conveyed by high TH-stopping rates by aligning these speakers with other white New Yorkers. By contrast, American-oriented speakers adopt raised /ɔ/, which suggests that for these speakers a linguistic rejection of Polishness may also entail a rejection of whiteness. Note, however, that raised /ɔ/ is not produced by the two second generation females in this group, who realize /ɔ/ as low. This further differentiation underscores the fact that considerations of gender also come into play, as women are known to avoid stigmatized variables.

Lastly, the adoption of the nasal split is what sets apart speakers who are oriented toward America and the Polish New York City community from those who are oriented toward Poland. The fact that the former two groups uniformly adopt this feature may have to do with the fact that the nasal split is a recent sound change, and as such, lacks the kind of indexicalities that could be employed for ethnic identity construction, or that it participates in the construction of other types of identities. Therefore, the realization of a robust nasal split by all speakers except some of the speakers oriented toward Poland may reflect a type of place identity. As such, Polish-oriented speakers produce a low /ɔ/ because they recognize its indexical link to whiteness, but fail to adopt the nasal split because it does not aid in the construction of a "Polish" identity. Taken together, the findings for the nasal split and raised /ɔ/ imply that orientation toward Poland correlates with a general absence of regional features of New York City English, regardless of whether these features carry ethnic associations.

## 6 Conclusion

This paper set out to introduce hierarchical cluster analysis as a method in quantitative approaches to ethnolinguistic variation. Applied to a Polish-involvement survey, HCA identified three main social groupings based on the similarity of speakers' answers across all questions, rather than simply following a quantitative score. The analysis established that transnational ties and Polish community involvement divide Polish New Yorkers into socially- and linguistically-coherent clusters. HCA also determined that a Polish lifestyle did not differentiate speakers in meaningful ways. As a result of the analysis, the ethnic orientations manifested by Polish New Yorkers emerged in a post hoc manner, rather than being imposed by the researcher a priori. Therefore, HCA promises to be a fruitful method that combines quantitative and qualitative data in the analysis of sociolinguistic variation.

## References

- Babbitt, Eugene Howard. 1896. The language of the lower classes in New York City and vicinity. *Dialect Notes* 1:457–464.
- Becker, Kara. 2010. Social Conflict and Social Practice on the Lower East Side: A Study of Regional Dialect Features in New York City English. Doctoral dissertation, New York University.
- Becker, Kara. 2014. Linguistic repertoire and ethnic identity in New York City. In *New Perspectives on Linguistic Variation and Ethnic Identity in North America. Special issue of Language and Communication* 35, ed. L. Hall-Lew and M. Yaeger-Dror, 43-54.

- Becker, Kara, and Amy Wing-Mei Wong. 2010. The short-a system of New York City English: An update. In *U. Penn Working Papers in Linguistics 15.2: Selected Papers from NWAV 37*, ed. K. Gorman and L. MacKenzie, 10–20.
- Blake, Renée, Cara Shousterman, and Luiza Newlin-Lukowicz. 2015. African American Language in New York City. In *Oxford Handbook of African American Language*, ed. S. Lanehart, L. Green and J. Bloomquist, 280–298. Oxford: Oxford University Press.
- Cogshall, Elizabeth, and Kara Becker. 2010. The vowel phonologies of African American and white New York City residents. In *African American English Speakers and their Participation in Local Sound Changes: A Comparative Study*. Publication of the American Dialect Society (PADS) 94, ed. M. Yaeger-Dror and E. Thomas, 101–128. Duke University Press.
- Dubois, Sylvie, and Barbara Horvath. 1998. Let's tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change* 10:245–61.
- Dubois Sylvie, and Barbara Horvath. 2000. When the music changes, you change too: Gender and language change in Cajun English. *Language Variation and Change* 11: 287–313.
- Frank, Yakira Hagalili. 1948. *The Speech of New York City*. Doctoral dissertation, University of Michigan.
- Hall-Lew, Lauren. 2009. Ethnicity and Phonetic Variation in a San Francisco Neighborhood. Doctoral dissertation, Stanford University.
- Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458–484.
- Hoffman, Michol F., and James A. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22:37–67.
- Kendall, Tyler, and Erik R. Thomas. 2012. Vowels: Vowel Manipulation, Normalization, and Plotting in R. R package, version 1.1.
- Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov, William. 2007. Transmission and diffusion. *Language* 83:344–387.
- Labov, William, Sharon Ash, and Charles Boberg. 2006. *Atlas of North American English*. Berlin: Mouton de Gruyter.
- Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89: 30–65.
- Mendoza-Denton, Norma. 2008. *Homegirls: Language and Cultural Practice among Latina Youth Gangs*. Malden, MA: Blackwell Publishing.
- Nagy, Naomi, Joanna Chociej, and Michol Hoffman. 2012. Analyzing Ethnic Orientation in the quantitative sociolinguistic paradigm. In *New Perspectives on Linguistic Variation and Ethnic Identity in North America. Special issue of Language and Communication* 35, ed. L. Hall-Lew and M. Yaeger-Dror, 9–26.
- Newlin-Lukowicz, Luiza. 2013. TH-stopping in New York City: Substrate effect turned ethnic marker? In *U. Penn Working Papers in Linguistics 19.2: Selected Papers from NWAV 41*, ed. A. Freeman, 151–160.
- Newlin-Lukowicz, Luiza. 2014. From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change* 26:359–385.
- Newlin-Lukowicz, Luiza. 2015. Ethnicity, L1 Interference, and Sound Change in New York City. Doctoral dissertation, New York University.
- Newman, Michael. 2014. *New York City English*. Berlin: Mouton De Gruyter.
- Rose, Mary. 2006. Language, Place, and Identity in Later Life. Doctoral dissertation, Stanford University.
- Sharma, Devyani. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics* 15:464–492.
- Sharma, Devyani. 2014. Transnational flows, language variation, and ideology. In *English in the Indian diaspora*, ed. M. Hundt and D. Sharma, 215–242. Amsterdam: Benjamins.
- Slomanson, Peter, and Michael Newman. 2004. Peer group identification and variation in New York Latino English laterals. *English World-Wide* 25: 199–216.
- Wong, Amy Wing-Mei. 2007. Two vernacular features in the English of four American-born Chinese. *U. Penn Working Papers in Linguistics 13.2: Selected Papers from NWAV 35*, ed. S. Brody, M. Friesner, L. Mackenzie, and J. Tauberer, 217–230.
- Wong, Amy Wing-Mei, and Lauren Hall-Lew. 2014. Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication* 35:27–42.

Department of Linguistics  
 New York University  
 New York, NY 10003  
 luiza.lukowicz@nyu.edu