LOTs of THOUGHTs on the Endangered PALMs of New York

Michael Newman*

1 Background

1.1 Low-Back Mergers in the Northeastern U.S.

Johnson (2010) describes the ongoing simplification of the low-back vowel subsystem in the northeastern US as a two-step process summarized in Table 1. First, an initial state Johnson calls *3-D* with distinct LOT, PALM, and THOUGHT, using Wells's (1982) keyword notation, is reduced to two vowels, although which two merge at this stage varies. In (North)eastern New England in a system Johnson calls *ENE,* LOT and THOUGHT merge with PALM kept distinct. Elsewhere in what he calls *MAIN*, PALM and LOT merge, with THOUGHT distinct. Finally, in either case the last vowel joins creating what he calls *3-M*, although of course that full merger is not necessarily foreordained:

	(North)eastern New England	**Elsewhere**
Initial State	*3-D*: LOT vs. PALM vs. THOUGHT	
Intermediate State	*ENE*: LOT-THOUGHT vs. PALM	*MAIN*: LOT-PALM vs. THOUGHT
Final State	*3-M*: LOT-PALM-THOUGHT	

Table 1: Low back merger trajectories.

Most of Johnson's participants had either ENE or MAIN, with some of the youngest having 3-M and a few of the oldest retaining 3-D. His research was concentrated near the Rhode Island-Massachusetts border, but he included a small substudy in New York City, where he assumes a MAIN system, as do such other researchers on New York City English (NYCE) vowels including Becker and Coggshall (2009), Becker (2010, 2014a, 2014b), Wong (2010, 2012), and Olivo (2013). Those researchers along with Wong and Hall-Lew (2014) coincide in finding THOUGHT lowering over time towards the (assumed) combined LOT-PALM. Becker (2010, 2014b) and Coggshall and Becker (2010) find lowering to be greatest among Whites, although robust among Asians and Latinos, and Wong (2010, 2012) and Wong and Hall-Lew (2014) confirm this tendency among Asians. Evidently, if lowering continues, THOUGHT tokens will begin to overlap with LOT-PALM and, barring other differences such as duration or rounding, the vowels will eventually become indistinguishable. As such, the lowering suggests a possible future 3-M through what Trudgill and Foxcroft (1978) call "merger by approximation"; see also Johnson 2010. Johnson actually finds a small percentage of the New York teenagers that he surveyed with a minimal pair test already had merged the vowels perceptually.

1.2 3-D systems in NYCE

Nevertheless, other researchers find continued 3-D systems in NYCE including Labov, Yaeger and Steiner (1972), Labov (2006 [1966]), Labov, Ash and Boberg (2006), Kaye (2012), and Newman (2014). In ENE, PALM at [ɑ] is lower and fronter than merged LOT-THOUGHT at [ɔ], but in NYCE the relative positions are reversed; PALM may be [ɑ̝] but it can approach [ɔ], whereas LOT is [ɑ]. As an illustration, the well-known Boston dialect performance phrase highlighting PALM (along with non-rhoticity) "I parked my car in Harvard Yard" could actually come out quite differently if said by a 3-D r-less New York visitor:

Bostonian rendition: [aˈpʰɑktmaˈkʰɑʔɪnhɑvədjɑd]
New Yorker rendition: [aˈpʰɔktmaˈkʰɔʔɪnhɔvədjɔd]

---

*I would like to thank my former undergraduate student E. Brian Kelly who gathered much of the data and did the legwork on the initial analyses. He was co-author on the presentation.

There is also a difference in PALM composition. As described by Kaye (2012), NYCE PALM has three sources, the first two shared with ENE, but the third limited to NYCE:

1) Middle English long-A words before later-elided /r/ such as *bar, star, cart*, etc., and also later-elided /l/ such as *palm, calm*, etc. and sporadic exceptions such as *father*.
2) Borrowings such as *pajama, garage*, and *spa*.
3) Middle English short-O words in lengthening contexts, such as when the vowel precedes voiced consonants, such as *god, doll*, and *job* whereas words like *got* and *stop* with following voiceless codas remain LOT. The process is independent of subsequent morphology; *plodded* and *plotted* are not homonyms despite sharing the flap (Kaye 2012).

Labov et al. (2006) nevertheless observe some seemingly idiosyncratic variation involving LOT and PALM, with a number of speakers showing somewhat different lexical distributions. The formulation by Kaye would only seem carve out the maximum set. In Newman 2014, in a small illustrative sample of sixteen New Yorkers, I confirm this variation among three of five White participants, with the remaining two having MAIN systems. By contrast, the four Latino participants and four Black participants (of various backgrounds) preserve intact PALM classes. Of the three Asian participants only the oldest has 3-D. Variation of words between classes may be just that, sociolinguistic variation; but there are two consequences, one methodological and the other empirical, that need accounting for.

Methodologically, the variation creates an obstacle to determining the position of the affected vowels because of the impossibility of making a reliable prior assignment of words into their appropriate classes, which is, after all, a prerequisite for plotting. The only obvious workaround would be to determine class membership by position after measurement, but to do so is problematic particularly when the relevant classes are adjacent as in this case. For example, a token may be assigned to, say, LOT because it appears to be in LOT vowel space, but what the researcher is trying to do is to determine where LOT vowel space is. Empirically, the existence of variation raises the question of whether a change in progress is occurring as words shift classes in what Trudgill and Foxcroft (1978) call "merger by transfer." Both issues are explored in this study by examining the low back systems of eleven White New Yorkers, the racial group that showed the most variation between MAIN systems and 3-D systems in Newman 2014.

2 Methods

Besides variation in word class membership, another obstacle to determining the condition of the NYCE low-back vowel system is the relatively low number of potential PALM words and the fact that few are particularly frequent. Consequently, sociolinguistic interview data, as in Newman 2014, may simply not provide sufficient tokens, which is particularly the case for rhotic speakers due to the fact that following [ɹ] effectively removes the vowel from the low-back system (Labov 2006). Consequently, I devised a short passage with an abundance of PALM words for a read-aloud task. Although there is no evidence for social evaluation of low-back word class membership in New York, it seemed prudent to follow Labov's (2006) suggestion that efforts be made to mitigate any formalizing effects of the reading task. In this case, similar to Labov's original strategy, the passage was in the voice of a child narrator recounting a humorous anecdote.

In the reproduction of the passage "God Dolls" below, there are a total of 37 target words. Expected PALM words following Kaye's (2012) criteria are double underlined; LOT are single underlined, and THOUGHT are dot underlined. Only consistently fully realized vowels were used, so for example, *always* was not measured, and *cloths* was sometimes eliminated due to misreading as *clothes*. The passage also targets PRICE/PRIZE, although these diphthongs are not analyzed here.

God Dolls

> Our parents are always buying educational toys. Like even though my sister Marge is only five, they got her all these Greek god dolls. Not normal Barbie dolls: gods and goddesses in these skimpy cloths with palm fronds everywhere. I guess so she'd learn about those times.

> Well, they lasted like two weeks. See, she's always fighting with our middle brother Bob. So, last week, after she tried to bite him on the arm, he got seriously mad. I told him to calm down since she's only five, but no, he bided his time until everyone was out of sight. Then he climbed right up to her room, grabbed the dolls and chopped them into pieces. Then he threw them off the dock and into the pond out back of the house. I don't know why he thought he wouldn't have to answer for his crime or maybe he was so full of spite he didn't care. Anyway, Mom found them, and Marge came to me sobbing: "Josh! "Bobby broke my dollies and threw them in the pond" I went to see and there they were: plastic Greek god arms, legs, and other body parts all bobbing up and down in the water. So sad.

The participants included eleven native White New Yorkers from Queens and Nassau County (an inner suburb entirely within the New York City dialect region) distributed among three families: two generations each of the Hugheses and the Baracio-Ledwiths and three of the Granger-Happelmaths. All included various ethnic combinations including Jewish, Irish, German, and Italian. Recordings were made on a Zoom H-4 recorder with its built-in microphone and analyzed with the Penn Forced Alignment and Vowel Extraction suite. Normalized formant extraction was taken at 35% of vowel duration to reduce diphthongizing effects. Plots were made with Graphpad Prism.

With regards to the methodological question, the study can be seen as a test of hierarchical cluster analysis (HCA), using Wessa's (2015) R-based online statistical tool. Cluster analyses have been used for some sociolinguistic purposes such as grouping social variables (see Moisl 2010). However, they have not been used as far as I am aware for determining which tokens belong to a given linguistic category.

In a sense this lack of employment is surprising. Uncertainty in variable and variant composition can easily arise in studies of mergers, splits, and hypothesized allophonic distributions. All forms of cluster analysis group items via similarity, with different models defining similarity in different ways. Since tokens of any variable are defined on at least one level as "the same" all variables can be seen as consisting of a cluster of instantiations. Conversely, tokens sorted into different variables can be seen as comprising distinct clusters. The same logic can be applied when determining variants. Moreover, in HCA similarity is understood as relative closeness in n-dimensional space, n is the number of factors fed into the algorithm. Since vowel space is usually conceived of as two-dimensional, and vowels are already understood as realized in clusters, HCA appears ideally suited for determining membership in vowel classes.

HCA joins clusters sequentially beginning with those of one, two, four items, and so on until a single cluster encompassing all the data points is formed. The output is typically expressed graphically as a tree composed of consecutive binary branchings called a dendogram. Each node on the dendogram represents a cluster, the closer to the "trunk" of the tree, the more robust the differentiation between the subsequent clusters; the closer to the "leaves" the less the distinctiveness. Different mathematical assumptions lead to different outcomes. In this case, the Ward method, which minimizes variance and is suited to continuous variables, was used. The measures fed into the algorithm were the tokens' F1s and F2s transformed into barks to avoid distorting effects of the logarithmic hertz measures.

HCA has important limitations. Perhaps the most prominent is that it is hypersensitive to structure with no way to discriminate between target and incidental factors that lead to proximity or distance. Put a different way, the forced binary branching invariably maps out more structure than is actually present in the data being clustered. Nevertheless, in the case at hand the system will work as long as irrelevant coarticulatory and random effects generate lower-level (towards the leaves) clusters, whereas the targeted phonological differences generate clusters higher up in the tree (nearer the trunk). The resulting process has to be thought of as hypothesis generation given the absence of probability checking. In sum, the hope is that HCA can aid the analyst in determining vowel categories, not that it can simply spit out those categories.

3 Data and Findings

The results of the HCA shown in Table 2 can be summarized as follows: 2 participants with intact systems 3-D systems (1 oldest generation, 1 middle generation); 3 participants with weakened 3-D systems (2 middle generation, 1 youngest generation); 6 participants with Main systems (2 middle generation, 4 youngest generation).

Consequently, there is a tendency in apparent time to increase MAIN at the expense of 3-D with an intermediate stage of weakened 3-D, suggesting albeit with caution for the small sample size a change in progress of merger via transfer. Four examples follow with participants' vowel charts compared to their respective dendograms, showing the similarities and differences between those two ways of depicting vowel systems.

Participant	**Birth**	**System Description**	**PALM -> LOT**
Brian Hughes	1954	Intact 3-D	None
Aileen Hughes	1958	Weakened 3-D	variably god, Bob , doll
Caitlin Hughes	1987	Weakened 3-D	variably god, Bob , doll
Lissette Baracio	1952	Weakened 3-D	doll, Bob, pond, variably god
Jack Baracio	1950	MAIN	all PALM words
Joey Baracio	1986	MAIN	all PALM words
Charles Ledwith	1957	MAIN	all PALM words
Kerri Ledwith	1989	MAIN	all PALM words
Jack Granger	1931	Intact 3-D	None
Lisa Happlemath	1970	MAIN	all PALM words
Karen Happlemath	2001	MAIN	all PALM words

Table 2: Participants with birth year and low back systems.

Figures 1 and 2 illustrate an Intact 3-D system using Brian Hughes's vowel chart and dendogram respectively. In Figure 1 (and all succeeding vowel charts) only the potential PALM tokens (following Kaye 2012) are labeled with words to focus on the distribution in question. The PALM and LOT tokens are adjacent, with PALM backer and higher as expected. Only one PALM token, *dollies*, which in any case presents a potential for phonotactic distortion due to the following /l/, overlaps with LOT tokens. THOUGHT is quite high as is typical of traditional forms of NYCE.

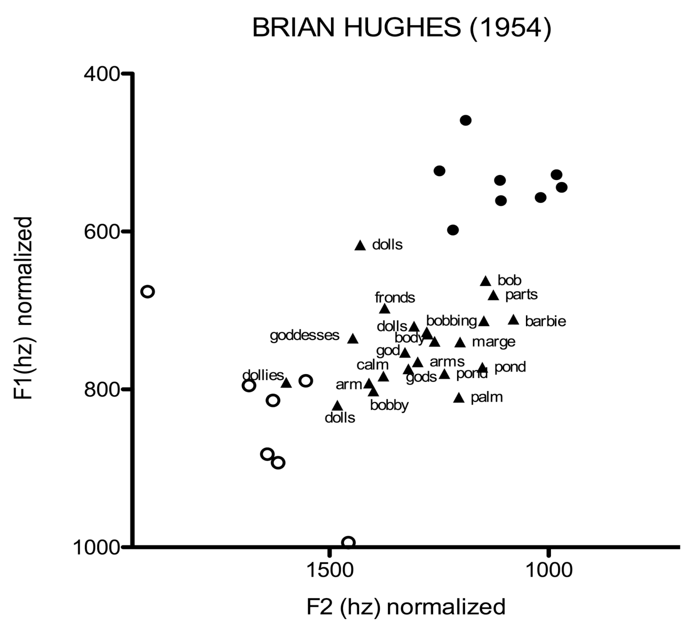

Figure 1: Brian Hughes' vowel chart. THOUGHT, closed circles; LOT, open circles; PALM, triangles.

The HCA digitizes the relative proximity of PALM and LOT compared to THOUGHT as shown in Figure 2, in the form of the highest level split first between THOUGHT and the other two classes.

The next split separates PALM and LOT, pointing to the independence of all three word classes. Since no ternary branches are allowed, one split had to take precedence. *Dollies,* as in the vowel chart, is the only PALM word clustered with LOT tokens. Even the LOT outlier *got* (visible although unlabeled as the empty circle in the Figure 1 at the left edge) is correctly clustered with the other LOT tokens.

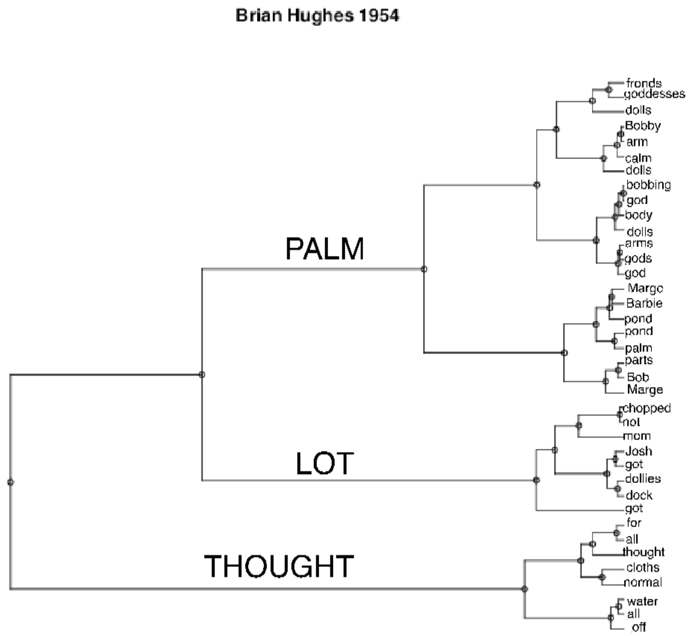

Figure 2: Brian Hughes' Hierarchical Cluster (based on Bark).

How the HCA digitalizes the clustering depicted analogically in the vowel chart is just as clearly seen in the MAIN system of the youngest participant, Karen Happlemath, shown in Figures 3 and 4.

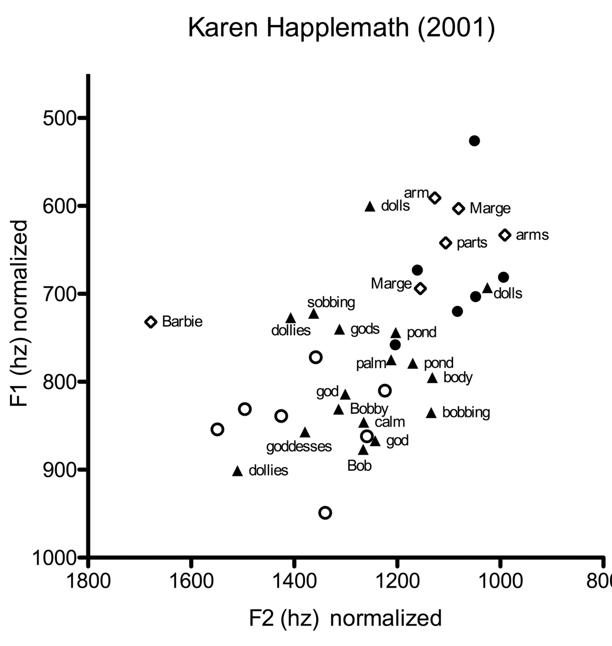

Figure 3: Karen Happlemath's vowel chart. THOUGHT, closed circles; LOT, open circles; PALM, triangles; rhotic START, diamonds.

Karen has some rhotic tokens. They are shown to be accountable to the data, but their r-coloring effectively removes them from the low-back system as mentioned earlier. Their mixture with THOUGHT tokens in Figure 3 is a result of the fact that preconsonantal [ɹ] raises a preceding low vowel in NYCE. More importantly for this study, Karen's LOT and original PALM tokens are also intermingled. These two mixings are also reflected in the HCA shown in Figure 4, which shows rhotic tokens, referred to here and subsequently as START marked with an asterisk. The word *thought* is an outlier as may be *dolls*. This last word may have transferred to THOUGHT or possibly had a distorting effect from the following [l], although this should have been to lower the vowel.

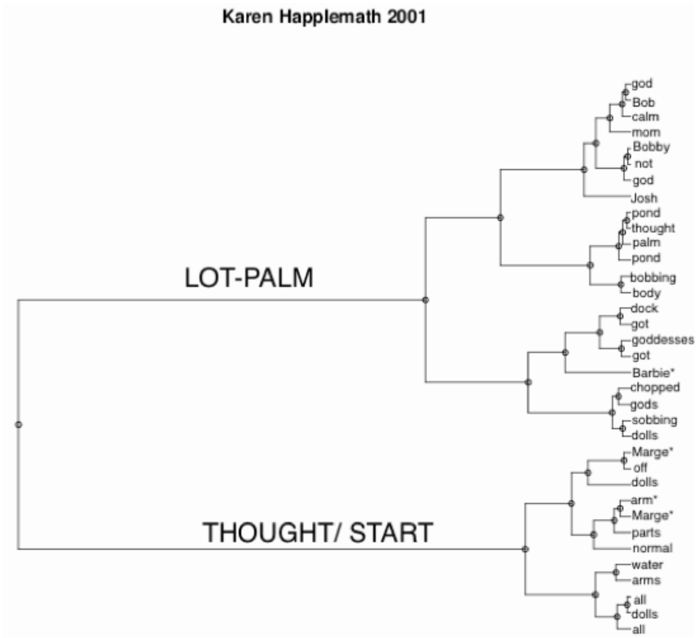

Figure 4: Karen Happlemath's Hierarchical Cluster (based on Bark).

In sum, the similarities shown between Figures 1 and 2, and between 3 and 4 reflect an underlying isomorphism between the vowel chart and dendogram. However, the binary divisions of the dendogram are more confidently interpreted for categorization purposes than the analogic distances visible in the vowel chart. That analogic nature requires that any categorizations be decided upon by the researcher, who after all may be subject various kinds of unconscious biases in addition to cognitive limitations in deciding upon them. By contrast, the HCA proposes an explicit automatized categorization.

Of course, although the HCA eliminates the subjectivity issue in analysts' judgments, it is an open question whether the HCA in any way mirrors phonological categorization. For this reason it is reassuring that the algorithm appears to successfully resolve the expected token distribution in the (different) word classes of the relatively unproblematic first two participants. This success suggests that HCA is potentially useful in more challenging cases, such as Brian Hughes's wife Aileen. Her vowel chart is presented in Figure 5. Aileen Hughes's THOUGHT and LOT tokens are quite spread out with some overlap of LOT and PALM. However, the HCA suggests more structure as shown in Figure 6 where two second level clusters separate PALM tokens from THOUGHT and LOT respectively. Both clusters are entirely composed of potential PALM words. However, the expected PALM words *god(dess)*, *Bobby, sobbing, mom*, and *doll(ie)s* are in clusters mixed with LOT tokens, which appears to indicate their transfer to LOT.

The split between the two sets of PALM tokens, nevertheless, poses a potential problem. It appears in Figure 5 in how those words are spread widely around what appears to be a horizontally flattened ovaloid. As such, it makes more sense to see the division in the dendogram as an artifact of insensitivity of the Ward algorithm to the relevance of that elongation. As such it shows the limits of this form of cluster analysis as a clarifying mechanism.

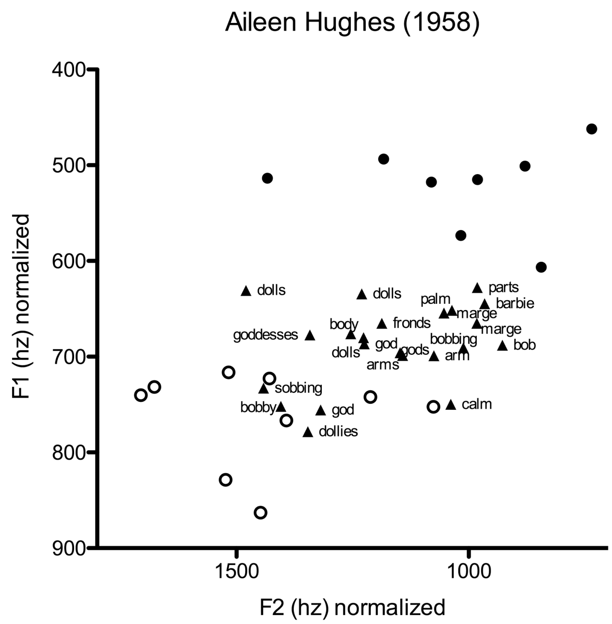

Figure 5: Aileen Hughes' vowel chart. THOUGHT, filled circles; LOT, empty circles; PALM, triangles.

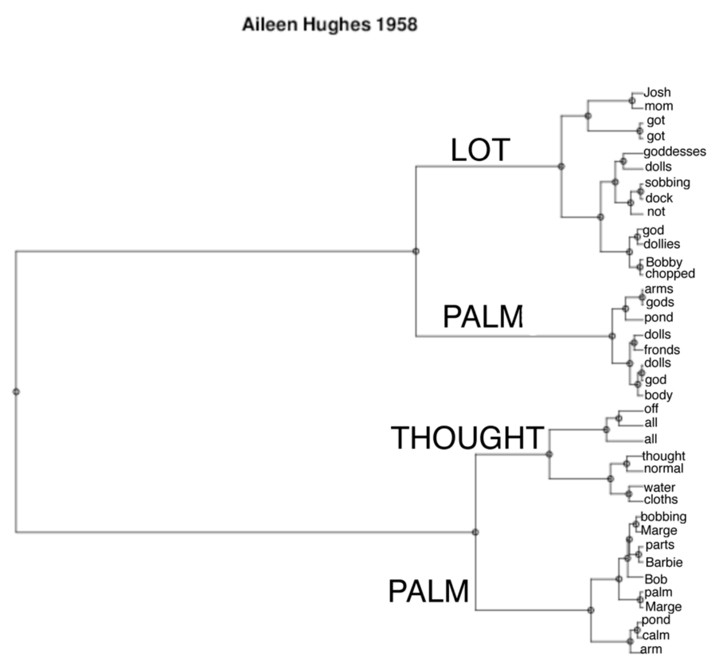

Figure 6: Aileen Hughes' Hierarchical Cluster (based on Bark).

Another weakened 3-D speaker is Lissette Baracio, whose vowel chart is shown as Figure 7. The third, Caitlin Hughes, is not shown due to space limits, but her vowel chart and dendogram closely resemble Lissette's. Lisette produced rhotic variants of the words *Marge*, *parts* and *arm,* which again effectively removes them from the low-back system. The LOT words are spread out, but given their small number, it is possible that they would have formed a denser cluster had there been more of them. Some PALM words are mixed in, but closer, on the whole, to what would appear to be the likely PALM vowel space. Again, THOUGHT forms a distinct grouping, which is raised. The corresponding dendogram from the cluster analysis appears as before with the rhotic realizations marked with an asterisk and classified as START.

The cluster analysis again joins the r-ful START tokens with the THOUGHT ones due to the r-coloring. More interesting for the purposes of this study, the dendogram shows a homogeneous PALM cluster, whereas the LOT cluster contains a good number of original PALM words, clear evidence of transfer. Even more clearly than in Aileen Hughes's case, the cluster analysis suggests an unambiguous demarcation of the two vowel clusters, one that is harder to establish confidently in the vowel chart.

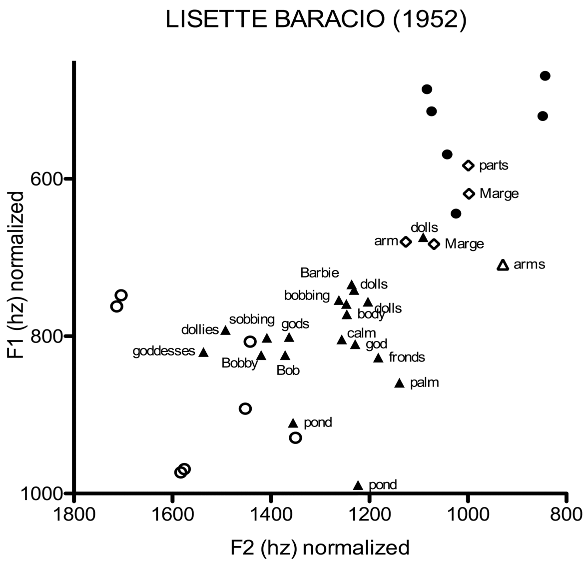

Figure 7: Lissette Baracio's vowel chart. THOUGHT, filled circles; LOT, empty circles; PALM, filled triangles; rhotic START, diamonds.

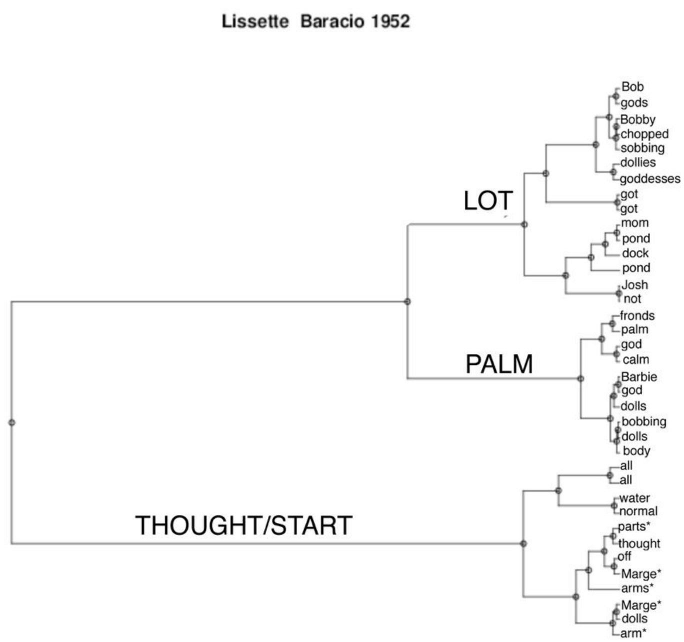

Figure 8: Lissette Baracio's Hierarchical Cluster (based on Bark).

4 Conclusions

In terms of the empirical findings, the results support first the persistence of 3-D in NYCE and second an on-going merger by transfer to MAIN. Future research with larger samples and so a fuller set of lexical items can flesh out the merger process, for example to see if transfer is conditioned by factors such as phonetic context or frequency. On that point, it seems interesting that the words that defect first beyond the loss of rhotic START tokens are those unique to NYCE PALM, i.e., lengthened short-O words.

In terms of the methodological aims, it appears that cluster analysis can play a helpful role in determining variable composition, particularly with regard to vowel classes. Traditionally, researchers do this sorting via linguistic history and native speaker intuitions. In the case of suspected merger by transfer, this technique becomes problematic because of the circularity described in the introduction. Nevertheless, the same logic applies to the apparently more frequent mergers by approximation, in which a number of statistical methods have been employed each with its strengths and weaknesses (Nycz and Hall-Lew 2014). Hierarchical cluster seems an additional useful tool for both merger types because it provides a series of objective digitalizations as hypothetical classifications. These hypotheses can then be accepted at least tentatively by the researcher as a way of proceeding with the analysis and/or bolstered with the results of other statistical procedures.

Given the status of this article as a working paper, it makes sense to suggest ways forward in this regard. In particular, the over-generation of structure has already been mentioned and is a well-known weakness of HCA. An example appears in this study in the way that the Ward algorithm grouped Aileen Hughes's PALM tokens into a two separate clusters in Figure 6. There are, however, many forms of cluster analysis including non-hierarchical ones (see papers in e.g., Hennig et al. 2015). There is good reason to explore various options to find a more effective method or combination of methods than the one used here.

On that point, although HCA seems particularly well-suited for determinations based on Euclidean distance such as in vowel space, it is not straightforward how to incorporate other potential factors such as nasality, duration, rounding, and rhoticity. This limitation can be seen in the present paper in the inclusion of r-colored START vowels in THOUGHT clusters. Had third formants been added, or perhaps better the difference between F3 and F2, this spurious clustering would have presumably vanished, but it is not clear whether third formants should be given equal weight to the first two. Again, more sophisticated approaches to clustering are clearly desirable.

That said, it needs to be pointed out that automated categorization measures like HCA are underused in sociophonetics, and that taking advantage of them in this study actually reveals just how primitive our traditional intuitive ways of establishing phonological categories are. Perhaps the use of such methods can clarify a number of confusing cases of category uncertainty in various languages.

References

Becker, Kara. 2010. Regional Dialect Features on the Lower East Side of New York City: Sociophonetics, Ethnicity, and Identity. Doctoral Dissertation, New York University.

Becker, Kara. 2014a. The social motivations of reversal: Raised BOUGHT in New York City English. *Language in Society* 43:395–420.

Becker, Kara. 2014b. Linguistic repertoire and ethnic identity in New York City. *Language and Communication* 35:43–54.

Coggshall, Elizabeth and Kara Becker. 2010. The vowel phonologies of white and African American New York residents. In *African American English Speakers And Their Participation In Local Sound Changes: A Comparative Study. American Speech Volume Supplement 94,* ed. M. Yaeger-Dror and E.R. Thomas, 101–128. Chapel Hill, NC: Duke University Press.

Hennig, Christian, Maria Meila, Fionn Murtagh and Roberto Rocci. 2015. *The Handbook of Cluster Analysis*. Boca Raton: Taylor and Francis.

Johnson, Daniel Ezra. 2010. *Stability and Change along a Dialect Boundary*. Publication of the American Dialect Society 95. Chapel Hill, NC: Duke University Press.

Kaye, Jonathan. 2012. Canadian raising, eh? In *Sound Structure and Sense: Studies in Memory of Edmund Gussmann*, ed. E. Cyran, H. Kardela and B. Szymanek, 321–352. Lublin: Wydawnictwo KUL.
Labov, William. 2006. *The Social Stratification of English in New York City, 2nd ed*. Cambridge: Cambridge University Press.
Labov, William, Sharon Ash and Charles Boberg. 2006. *Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton De Gruyter.
Labov, William, Malcah Yaeger and Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress*. Philadelphia: University of Pennsylvania Press.
Moisl, Hermann. 2010. Variable scaling in cluster analysis of linguistic data. *Corpus Linguistics and Linguistic Theory* 6:75–103.
Newman, Michael. 2014. *New York City English*. Berlin: Mouton DeGruyter.
Nycz, Jennifer and Lauren Hall-Lew. 2014. Best practices in measuring vowel merger. *Journal of the Acoustical Society of America 134*.
Olivo, Anne Marie. 2013. The Strong Island Sound: Sociolinguistic Evidence for Emerging American Ethnicities. Doctoral Dissertation, Rice University.
Trudgill, Peter and Tina Foxcroft. 1978. On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. In *Sociolinguistic Patterns in British English,* ed. P. Trudgill, 69-79. London: Arnold.
Wells, John C. 1982. *Accents of English I: An Introduction*. Cambridge: Cambridge University Press.
Wessa, Patrick. 2015. Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7
Wong, Amy Wing-Mei. 2010. New York City English and second generation Chinese Americans. *English Today* 26:3–11.
Wong, Amy Wing-Mei. 2012. The lowering of raised-THOUGHT and the low-back distinction in New York City: Evidence from Chinese Americans. *U. Penn Working Papers in Linguistics 18.2: Selected Papers from NWAV 40*, ed. H. Prichard, 157–166.
Wong, Amy Wing-Mei and Lauren Hall-Lew. 2014. Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language and Communication* 35:27–42.

Department of Linguistics and Communication Disorders
Queens College/CUNY
Flushing, NY 11367
michael.newman@qc.cuny.edu