Chapter 5

Summary

The research presented in the preceding chapters applies a variety of quantitative tools to English and American corpus data to yield insights into synchronic and diachronic aspects of English phonetics and phonology. The tools reveal elements of phonetic structure in English dialects while underscoring the largely random character of much variation; they point to historical sources of American dialect variation, and they help recover historical processes of new dialect development in the American colonies. In the process, the research illustrates the usefulness of quantitative tools in linguistic research, the relative importance of different tools, and ways in which various tools can be used in concert to advance linguistic research.

5.1 The Usefulness of Quantitative Tools

The quantitative tools used in the previous chapters achieve several important goals of linguistic analysis:

- They yield reasonable measures of variation in phonetic usage among speakers within and among localities.
- They illustrate the largely continuous geographical nature of that variation, particularly with respect to vocalic features.
- They identify a number of major and minor dialect regions (or groups of speakers) as clusters of speakers or localities with relatively similar patterns of usage, distinguishing regions of relative uniformity from transitional zones with substantially greater variation.
In some cases, they isolate regionally coherent groups of variants or features that can be said, with appropriate caveats, to distinguish the dialect regions.

The techniques reinforce each other to provide a consistent picture of the dialects to which they have been applied in these studies, tending to confirm the applicability of computational techniques to the study of dialect variation. Applied to data from traditional dialect speakers in England, as discussed in the next section in greater detail, the techniques largely corroborate standard characterizations in the literature. However, the results differ in important ways from those of previous studies in the placement of several important dialect boundaries, in the association of features with dialect regions, and in placing those systematic characteristics against a background of largely continuous variation. Applied to dialect speakers in both England and the United States, the techniques allow one to quantify differences in variation in the two countries, provide insight into the development of American speech varieties, and reveal regional English origins of sociolinguistically relevant American dialect differences.

All of the techniques appear useful, but they need not all be used simultaneously. The results suggest that even a small but judiciously chosen feature-based dataset can yield a great deal of insight, and that in the absence of feature detail, variant-based data also can be used quite effectively. Clustering techniques are useful for delineating dialect regions, but they do not necessarily distinguish between regions of greater uniformity and greater diversity—a task for which distance regressions, variant-area regressions, and barrier analysis are more suitable, and any one of which is likely to be sufficient. Finally, principal component analysis appears to be useful in at least some cases in isolating distinctive features or variants that are more or less characteristic of dialect regions, and that in some cases are structural in nature.

In general, unrotated principal components are comparatively difficult to interpret linguistically in the research presented here, but varimax rotations consistently yield components that have comparatively straightforward interpretations. ¹ The technique may be equally applicable to the study of linguistic variation across class, gender, and other dimensions as well. In some circumstances, feature-based principal component analysis may not only capture structural features such as upgliding or centering diphthongization, but also provide measures of their strength, while variant-based PCA may isolate groups of variants that reflect structural shifts. In this respect, feature-based

¹Although not discussed in the preceding chapters, oblimin and equamax rotations yield generally unsatisfactory results when applied to the data used in this thesis, but quartimax and promax rotations yield results that are largely indistinguishable from those derived using varimax rotation.
and variant-based approaches may be substitutes rather than complements. Nevertheless, PCA does not invariably identify linguistic patterns that are clearly characteristic of particular traditional English dialect regions. That failure arises from the fact that PCA isolates only groups of variables that occur together, whereas the variation of most phonemes in the English dialects is relatively distinct. As a result, any phoneme pattern that might be identified as characteristic of a particular region generally includes phonemes that are not unique to that region or that are unique to the region but only by a subset of speakers, and that are therefore not isolated as part of a regionally distinct pattern.

5.2 The Traditional English Dialects

The application of the quantitative techniques described above to phonetic data from the Survey of English Dialects yields a variety of insights into the structure of traditional English dialects in the mid-twentieth century that, to my knowledge, have not been previously been presented together as a coherent whole. Many of the results are standard insights long since arrived at by dialectologists using traditional approaches, but here they result from statistical analysis of large numbers of features. Close examination of the simple statistics of phonetic features and variants reveals three sources of variation:

- Some variation involves distinct and mainly binary splits of consonants—the most extensive being the voicing of fricatives in much of the Southwest, contrasted with unvoiced fricatives elsewhere. In the more complex case of rhotics, five separate realizations can be distinguished on the basis of place and manner of articulation.

- Variation can also involve largely binary splits in length and rounding of vowels. That is, most localities are recorded as using either a fully rounded or fully unrounded variant but rarely a half-rounded one; similarly, localities use either a short or long variant but rarely a half-long one.

- The third and most extensive source involves essentially continuous variation in the height and backing of middling vowels. There is little variation among descendants of the Middle English low vowels or high vowels: Low vowels typically remain low in most localities, and high vowels typically remain high as well. In contrast, however, descendant of the Middle English vowels of medium height tend to have varied expressions across dialects. Thus much of the variation in the English dialects can be characterized as gradual drifts in height and backing of vowels and diphthongs.
Another important insight is that relatively little of the variation in the traditional English dialects can be characterized as structural. That is, relatively few features or variants are closely correlated with a substantial number of other features or variants, so that they all vary together in linguistically relevant ways. Only three relatively large groups of features or variants show high correlation amongst their members (note that I generally describe variations in terms of the historical developments they entail):

- One group is composed of rhotics following vowels or, in the non-rhotic dialects, their replacements. With relatively few exceptions (such as the intrusive [ɹ]) rhotic speakers use them in most contexts, whereas non-rhotic speakers drop them.

- A second is composed of fricatives, which tend all to be voiced (with voicing of medial /t/ a closely correlated feature) by Southwestern speakers but not by others.

- A third is composed of the descendants of the Middle English front and back low long vowels /æ:/ and /ɔː:/, which show regionally distinct and correlated developments. In much of the North of England, they tend to develop downglides and to merge to [iæ], [ie \sim je] or [ia \sim e]. In other parts of the North, they both tend to develop inglides [e\oe \sim eo] and [uə \sim o], respectively. In parts of the Midlands, they are simply raised to [i: \sim e:] and [u: \sim o:], and the degrees of raising of front and back vowels tend to be correlated. Finally, the vowels both develop upglides [ei \sim æi] and [ou \sim æu] in much of the South, and where they do, the heights of the initial vowels in the resulting diphthongs tend to be similar. The co-evolution of the low front and back vowels thus appears to be a genuinely structural development in most English dialects. However, the correlations between parallel developments are not invariably large, so that the structural parallels might best be interpreted as statistical tendencies rather than strict relations.

Two smaller groups of correlated features or variants are the low short vowels, which tend to split depending on the type of following consonant in much of the South, and high back vowels and second components of diphthongs, which tend to be fronted to [v] in Devonshire.

5.2.1 Linguistic Distances

Despite differences in their construction and composition, feature-based and variant-based linguistic distances between localities are closely correlated with each other and yield similar patterns of similarity, difference, and correlation
with geographic distance – the latter being a strong indication that geographic separation has played an important role in the diversification of dialects. Regression analysis indicates that distance alone accounts for about half the measured variation in the dialects. As a consequence of that geographic pattern, localities in the middle of the country tend to have more in common with localities at both ends of the country than do speakers at the ends with each other, and the comparatively few localities that are quite different from the rest of England are found mainly in the far North and Southwest.

Localities have quite diverse patterns of similarity with surrounding localities. Few localities are extremely similar to each other in their speech; close similarity appears to be the exception rather than the rule even at relatively short geographic distances. Nevertheless several broad regions appear to have comparatively uniform speech patterns reflected in large numbers of comparatively small linguistic distances – notably the Thames Valley, Leicestershire, and North Yorkshire, whereas localities in some regions, especially the West Midlands, are considerably more diverse. Localities in transition zones between regions of relative uniformity, sharing elements with both regions, can be quite distinctive. Somewhat paradoxically, a number of localities in the Midlands that have low overall average distances with localities in the rest of the country often have relatively large distances with neighboring localities because they are in or near a major transition zone between the North and South.

I also measure linguistic distances between the *Survey of English Dialects* localities and three idealized speech patterns – Middle English, to compare synchronic and diachronic distances; Received Pronunciation (RP), to seek insight into the origins of the British standard; and my own “Western Reserve” speech from Ohio, to address the origins of American speech patterns. However, I conclude that distance measures at their current state of development are only rather crude gauges that cannot provide substantial insight into those problems.

- The average linguistic distance between the “standard” or “chancery” Middle English of southeastern England and the mid-20th-century traditional dialects is about 50 percent greater than the average distance between the traditional dialects. I infer that mid-20th century speakers of traditional dialects typically differed as much in their speech as speakers in a given locality separated by roughly four centuries of time, and the difference between the dialects of the far Southwest and North are equivalent to a diachronic distance of roughly seven to nine centuries.

- Received Pronunciation displays similarities with dialects in many regions, and is most similar to traditional speech in the Home Counties and East Midlands, particularly Cambridge. However, the results are consistent both with origins of RP in those regions and with adoption
of RP characteristics by speakers in those regions and do not favor one hypothesis over the other. It appears likely that usage in many localities, as recorded in the *SED*, reflects degrees of diaglossia (adoption of features intermediate between traditional ones and the standard) and of dialect loss (adoption of features of the standard), but distance measures provide too little information to untangle such developments.\(^2\)

- Linguistic distance measures between my Western Reserve speech and the traditional dialects place it squarely among the Southern dialects. My average linguistic distance from all of the traditional dialects suggests a diachronic distance noticeably larger than the time span from even the earliest English settlements in America to the mid-20th century, though the discrepancy is considerably reduced if I compare it only to dialects of the South. The results are consistent with all the other evidence that points to an overwhelmingly Southern English in origin for most features of American dialects.

- By the feature-based measure, my Western Reserve speech is closer to Received Pronunciation than it is to all but 10 percent of the English localities, and less than 30 percent of localities have speech closer to Received Pronunciation than my speech is. Those findings are consistent with a strong influence of English norms on the development of an American “standard” but they are also consistent with a model ascribing similar processes to the development of both.

### 5.2.2 Principal Components as Dialect Structures

Application of principal component analysis to the feature-based and variant-based datasets produces differing but complementary insights. Applying PCA with varimax rotation to each dataset yields about a dozen linguistically sensible principal components that in each case account for roughly half the variation in the dataset, and whose component scores show them to be geographically continuous in various regions of England. The feature-based analysis isolates (in rough order of importance):

- Distinctively Southeastern short vowel developments and upgliding diphthongization of Middle English long low vowels, distinguished from Northern static short vowels and ingliding or downgliding diphthongization;

- Distinctive rhotics in regions of the Southwest and North, including the “Northumbrian burr” \([\mathbf{h}]\);

\(^2\)See Auer (2005) for a typology of relationships between dialects and standards in European languages.
5.2. THE TRADITIONAL ENGLISH DIALECTS

- Southwestern fricative voicing;
- Northern retention of lengthened long high vowels from Middle English, coupled with the development of inglides in the long low vowels; Southern backed and rounded onsets to upglides in descendants of Middle English /i:/;
- Southeastern merger of and backing and shortening of the onset in descendants of Middle English /e:/ and /ɛ/;
- Various East Anglian developments such as glottal stops for Middle English medial /t/;
- A region around London and along the Southeast Coast with particularly strong front vowel raising of Middle English /a/ to [ɛ] coupled with particularly strong backing and lengthening to [ɑː] after fricatives; and
- A long swath of localities from the Southwest to southern Lancashire that tend to retain long variants in the Middle English mid-back long vowels.

Applied to variants, varimax PCA identifies sets of variants that are consistent with most of the feature-based groups as well as a handful of others. The variant-based principal components even include three rather large sets of variants that could be interpreted as partial nonstandard systems of vowels unique to particular regions of England: the fronting of back vowels to [ɔ] in Devonshire, the distinctive Northern pattern of diphthongization, and Lincolnshire's unique combination of low-center onsets for the Middle English long high vowels (such as [uː] for /u:/ and centering diphthongs for the low and middle ones (e.g. as [eə] for /ɛː/).

However, the variants associated even with the largest principal components can only loosely be thought of as characterizing the dialect of a specific speaker, group of speakers, or geographical region. Even in the case of a principal component that is dominated by a large set of highly correlated variants, such as voiced fricatives or a particular type of rhotic, it is possible that no locality uses all or even most of the variants. Conversely, even if all the localities in a geographic region use a large number of variants in common — say, all the same long vowels and diphthongs — each variant is likely to have a fairly unique geographic distribution that includes other localities not in that region, with the variants happening to overlap only in the region. As a result, many of the variants may not appear in the principal component even though analysis might suggest that they are all part of a unique system of vowels common to that group of localities. Such limitations appear to be inherent in the use of principal components analysis in the analysis of linguistic variation, and they
stem from and underscore the fact that systematic variation is the exception rather than the rule in the traditional dialects.

The association of feature-based principal components with the dialects of specific speakers or regions — indeed, the ontological status of such a component — is even more problematic because vocalic features such as height may vary continuously across speakers. For example, the largest feature-based principal component scores most positively in Southeastern localities where the strongest upgliding diphthongization appears and most negatively in Northern ones with centering or downgliding diphthongs. The component scores suggest a gradual grading of diphthongization between those extremes along a north-south axis, with nearly no diphthongization in some central regions, but a rather abrupt decline in diphthongization between the Southeast and Southwest, suggesting a more distinct linguistic boundary. Such a component does not in any sense describe a distinct dialect, but it may be thought of as summarizing a set of tendencies over a large community of speakers.

5.2.3 Clusters as Dialect Regions

Various clustering algorithms, applied to either dataset, tend to yield a fairly consistent grouping of localities. Figure 5.1 illustrates the multidimensional scaling of multiple cluster analyses of both feature-based and variant-based datasets combined. The results from both datasets are remarkably similar and, taken together, can be interpreted as delineating a set of major dialect regions and minor subregions that are consistent in many respects with standard descriptions in the literature. The regions are easily characterized by elements of the principal component analyses described above:

- Three Northern regions retain the Middle English short vowels while developing upgliding or ingliding diphthongs for several of the long vowels.
  - The Far North is distinguished from the rest of the country most notably by its uvular "Northumbrian burr" [ɔ], the survival of Middle English /h/ and /hw/, and the merger, raising, and fronting of the Middle English low-back long vowels.
  - The Upper North develops centering diphthongs for most of the Middle English long vowels.
  - The Lower North combines a tendency for the high long Middle English vowels to develop into upglides with a low-center onset with a tendency for the low and middle long vowels to develop centering diphthongs. (Subregions of the Lower North can also be distinguished, for instance by use of an alveolar approximant rhotic or by fronted upglides [ɔi] ~ [ɔi] for the long low-back Middle English
Figure 5.1: The Traditional English Dialect Regions
vowels. Lincolnshire has practically no unique features yet has a sufficiently unique and uniform combination of variants that it can be clearly identified by a variant-based principal component.)

- The Central Midlands can be distinguished by the unique overlapping of northern retention of older forms for the short vowels and Southeastern innovations in the long vowels. (The East Midlands subregion, in many ways the heart of the English dialect world, has a highly uniform pattern of speech that has the lowest average linguistic distances from the rest of England. To its west, Staffordshire appears as an island of relative uniformity in an important transition zone with unusual variability.)

- The South, which generally experiences substantial shifts in the Middle English short vowels, divides roughly into three regions on the basis mainly of diphthongization, rhotics, and fricative raising.

  - The non-rhotic Southeast sees extensive shifts in the Middle English short vowels combined with systematic upgliding diphthongization of the long vowels. (Three subregions can be distinguished – a central region in which the diphthongs’ onsets are lowered and fronted, a southern coastal region which enhances the raising of short front vowels, and a North Anglian region characterized by a hodgepodge of unique and largely nonsystematic features; [l]-vocalization frequently appears in the former two regions as well.)

  - The rhotic Upper Southwest combines Southern shifts in the Middle English short vowels with relatively little movement in the long vowels other than raising and occasional centering diphthongization.

  - The rhotic Lower Southwest is characterized mainly by fricative voicing, with Devonshire distinguished from the rest of the region by its unique fronting of back vowels and the development of a low monophthong for Middle English /i:/.

Although the classification presented here shares many characteristics with the standard descriptions, the approach and the results differ from tradition in key respects. As discussed in Chapter 2, the dialect regions differ significantly from those of Trudgill (1999): the Central region is much more restricted than Trudgill’s (although it extends further south in the east and less so in the west) and is more closely associated with northern dialects rather than southern ones; the Southeastern region is considerably more extensive, incorporating the Southeast Coast, and the clustering of northern regions differ considerably as well. Those differences stem mainly from a basic difference in methodology. Trudgill’s classification is grounded in the distribution of variants in 8 words
drawn on the basis of his detailed understanding of the historical development of English dialect diversity. The choice of variants and the resulting classification is by no means misleading or inaccurate, but it necessarily reflects an expert's subjective judgment about the comparative importance of particular dialect characteristics. In contrast, the classification presented here is based on a more-or-less mechanical analysis of the statistical characteristics of a large amount of data. The approach still involves a great deal of subjective judgment — in the focus on phonetics to the exclusion of grammar and vocabulary, the original choice of localities and words in the *Survey of English Dialects*, the classification of variants and features, and the choice of specific quantitative techniques; nevertheless, that subjectivity is substantially moderated by the use both of large numbers of variants and features and of a variety of alternative techniques. The approach thus allows greater latitude for significant patterns of diversity to emerge from the mass of data.

Table 5.1 summarizes the main variants and features that the quantitative approach — as applied in this thesis — associates with specific English regions. Characteristics that emerge as particularly salient include: the development of inglides, downglides, and upglides in the descendants of Middle English low and middle long vowels; fronting of Middle English back vowels; lengthening of Middle English short vowels /a/ and /o/; raising of Middle English /a/; rhotics; fricative voicing, and aspirates. The table presents the broad patterns found in each region, but does not reflect the full diversity found in the data — for instance, the occasional presence of rhotics or degrees of back vowel fronting in other regions. It also excludes as comparatively unimportant a variety of other variables that also influence the regional classification, including several in Trudgill's list, the lowering and unrounding of Middle English /u/ to [ʌ], and various splits and mergers. Nevertheless, the table illustrates how statistical analysis of a broad range of linguistic phenomena can yield a relatively simple classification of dialects while giving due consideration both to the full range of variation and to the relative importance of different phenomena.

---

<table>
<thead>
<tr>
<th>Region</th>
<th>Ingles/Downglides</th>
<th>Other Features</th>
<th>Voiced Fricatives</th>
<th>Aspirates</th>
<th>ME / a/ Lengthening</th>
<th>ME / Back Vowel Raising</th>
<th>Rhonks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far North</td>
<td>U</td>
<td>O</td>
<td>U</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Upper North West</td>
<td>U</td>
<td>U</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Cumbria-N. Yorkshire</td>
<td>U</td>
<td>O</td>
<td>U</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Lower North West</td>
<td>U</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Lincolnshire</td>
<td>O</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Staffordshire</td>
<td>U</td>
<td>O</td>
<td>U</td>
<td>U</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>West Midlands</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Central South</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>North Anglia</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Central Southeast</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Central Southwest</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Devonshire</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
<tr>
<td>Cornwall</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>[h]</td>
<td>[h]</td>
<td></td>
</tr>
</tbody>
</table>

Note: U - Usual pronunciation of at least one phoneme; O - Occasional pronunciation.
5.2. THE TRADITIONAL ENGLISH DIALECTS

Students of American dialects will note that the widespread American patterns of retained rhoticity, raising of Middle English low front short vowels, splitting of back short vowels, and the development of upgliding diphthongs for the Middle English long vowels, are all characteristic of twentieth-century Southeastern coastal speech in England, and that the patterns found in the region around London can be seen largely as ancestral to the American ones.

5.2.4 Regression Analysis and Dialect Geography

Close examination using several different analytic approaches reveals that several regions – most notably Lincolnshire and Devonshire – have quite uniform speech while others – most notably Staffordshire and the Upper Southwest – are much more variable. Regression of linguistic distances on geographic ones, including regional and intraregional dummy variables, yield highly statistically significant dummy parameters, indicating that systematic differences in regional variability can account for roughly a quarter of dialect variation (in addition to roughly half that can be accounted for by geographic distance alone). In addition, barrier analysis not only identifies most of the regional boundaries identified by cluster analysis but also identifies a large number of individual localities – the great majority of them in or near the Upper Southwest – that are particularly distinct from their neighbors. Further detailed examination of variant use shows that regions and most subregions have fairly distinct overall patterns of frequently used variants, especially with respect to the descendants of Middle English short vowels, rhotics, and consonants, and that the most extensive intraregional variation occurs among descendants of Middle English long vowels and, to a lesser extent, diphthongs. The high variability of the Upper Southwest thus arises primarily from variable developments of the relatively unstable long middle vowels, which of course is also reflected in the contrasting developments of those vowels in the North and Southeast. 4

A regression of the number of variants in each region on the number of localities in each region reveals a high and statistically significant degree of correlation that accounts for about two-thirds of the variation in the number of variants by region. (The residuals in that regression also indicate that regions with relatively small linguistic distances also tend to have fewer variants, and vice versa, further supporting the finding of differences in regional variability.) Given the number of localities already interviewed in a given region, an interview in an additional locality will, on average, add only a bit more

4Anderson (1987) provides an extensive discussion of the complex developments of Middle English long middle vowels. See in particular pp. 61-87 for a discussion of developments of the front vowels (including palatalization, which is probably relevant to the survival of this feature in Southern American English) and pp. 97-102 for discussion of the back vowels.
than one variant to the total population of variants identified in the region, suggesting that the *Survey of English Dialects* indeed provides a reasonably comprehensive sample.

### 5.3 English-American Speech Relationships

The linguistic data in the *Survey of English Dialects* covers only England, and the only dataset that provides comparable data for traditional English dialects in England and North America is that collected by Guy Lowman and others for the *Linguistic Atlas of New England* and *Linguistic Atlas of the Middle and Southern Atlantic*, and summarized in Kurath and McDavid (1961). Lowman's survey covered only the south of England, precluding comparison of northern English and American forms, but the detailed results described in Chapter 2 and summarized in the previous section provide compelling evidence that American dialects were overwhelmingly influenced by southern English speech during the period of settlement.

To explore the origins and diversification of American speech, I use these sources to develop a dataset of nearly a hundred words that includes nearly all phonemes in standard British and American English, recording nearly three hundred variants used by about 130 speakers in southern England, two regions of particularly early English settlement in America – southeastern Massachusetts and the Atlantic coast from Norfolk to Okracoke (henceforth in this section “Eastern Virginia”), and a region of Appalachia encompassing south-western Virginia and southern West Virginia (henceforth “Western Virginia”), an area settled by diverse groups including Germans and Scotch-Irish as well as descendants of the earliest coastal settlers. As with data from the *Survey of English Dialects*, I quantify phonetic variant usage (in this dataset, usually one variant per speaker per word) and also quantify variants in terms of phonetic features.

As with data from the *Survey*, application of the quantitative techniques described above to this dataset yields a variety of useful summary statistics, measures of similarity among English and American speakers, distinguishes clusters of speakers with similar speech patterns, and isolates groups of variants that distinguish those groups of speakers. Variant distributions reveal greater phonetic diversity in southern England than in America, indicating substantial leveling in the development of American dialects, and very few variants unique to America, indicating relatively little innovation. American speakers typically share as many variants with English speakers as the English speakers share with each other, again indicating that American speech varieties fall squarely into southern English. Speakers in Massachusetts and Eastern Virginia share roughly the same proportion of their variants with nearly every
English region despite the fact that they only share about three-quarters of their variants with each other, and tend to share more variants with southeastern English speakers than with southwestern ones. That finding suggests that both American regions apparently experienced similar degrees of influence from each English region but that leveling resulted in different resulting mixtures of variants. Moreover, Eastern Virginia speakers use considerably more distinctively southwestern English variants than do Massachusetts speakers, suggesting that settlers from that region were more influential in the development of Virginian speech patterns. Finally, Western Virginia shares most of its variants with Eastern Virginia (many more than it shares with Massachusetts) but has fewer variants altogether, suggesting further leveling processes during the westward expansion following initial colonization.

5.3.1 Linguistic Distances

Feature-based linguistic distances provide additional insights. Massachusetts speakers have smaller linguistic distances from eastern English speakers than from western ones. Distances between southern American and English speakers are relatively similar across regions, but compared with those of speakers from Massachusetts, greater in the east and smaller in the west – with the exception of the West Midlands, whose speakers have the greatest distance from and least linguistic similarity with American speakers of any of the English regions. Furthermore, the standard deviation measures are lower for the distances between southern Americans and western English speakers than eastern speakers: not only are the American southerners roughly as close to the English westerners as they are to the easterners, but their differences with westerners, by phoneme, are somewhat more uniform than their differences with easterners.

5.3.2 Clusters as Dialect Regions

Cluster analysis yields regional classifications that are sensitive to the choices of approach and distance measure, but most algorithms tend to group speakers in southern England into broadly the same regions as were identified using data from the Survey of English Dialects. English speakers near London cluster into several distinct groups rather than into a metropolitan cluster, suggesting that the extensive, protracted historical migrations from all over the British Isles to London had comparatively little effect on the speech patterns of rural speakers in surrounding regions. Under some approaches, speakers near the borders of English regions occasionally cluster with American speakers, suggesting that Americans themselves may best be thought of, like the borderers, as having characteristics of several of the English regions. American speakers tend to
CHAPTER 5. SUMMARY

Cluster with southeastern English speakers and to form two distinct groups, one northern and the other southern. American Southerners cluster so tightly that speakers from Eastern Virginia and Western Virginia are hardly distinguishable, but most folk speakers from both regions form a cluster distinct from neighbors classified as common or cultivated speakers. (A similar but weaker pattern emerges in New England as well.)

5.3.3 Regression Analysis and Dialect Geography

A regression of the number of variants on the number of localities in each region shows the same kind of high correlation as reported for the English data in the previous section, but American regions have fewer variants per speaker, again indicating substantial leveling. I also use regression analysis to test for a relationship between American speakers' degree of similarity with English speakers and the latters' geographic location from London, using dummy variables to control for the English speakers' regional affiliations. The regressions are generally all consistent with the proposition that all else being equal, American speech forms tend to be most similar to those immediately surrounding London and progressively slightly less similar in more distant regions. However, the regional affiliation of English speakers is consistently more important than distance from the metropolis in accounting for affinities with American speakers: American southerners' affinities with English speakers are more diffuse than those of the New Englanders, and they consistently show comparatively greater affinity with speakers from the English southwest and less affinity with those from the southeast.

5.3.4 Principal Components as Dialect Structures

Standard principal component analysis yields two significant principal components, accounting for about a quarter of the variance in the dataset, that are entirely consistent with the results of cluster analysis. The first principal component distinguishes a set of largely conservative characteristics found mainly found in the upper southwest of England from a set of largely southeastern English and typically more innovative characteristics – a variety of shifts, mergers, and diphthongization of Middle English phonemes identified in the preceding analysis of English dialects. The component scores indicate that the eastern innovations are far more common in American dialects than are the more conservative western variants, again demonstrating that American speech forms tend to be composed of variants found mainly in southeastern England. The second principal component also distinguishes largely southwestern English characteristics – in this case, rhoticity, palatalization particularly of low-front vowels (also noted in the Survey of English Dialects),
and various short vowel developments – from largely southeastern ones, but its component scores among Americans show a different pattern: Speakers from Massachusetts follow English southeasterners whereas speakers from the American South follow the southwesterners.

Principal component analysis of the data thus reveals two sets of oppositions involving fairly clear linguistic structural interpretations and distinct regional distributions. As represented by Lowman’s informants, the speech of southern England has a fairly strong demarcation between east and west. American speech forms appear to draw from all over the region (and possibly from others as well). However, American forms tend to be similar to eastern English ones, on the whole, but northern American forms tend to be much more so, while southern American speech reveals significant western English affinities.

A further principal component analysis restricted to southern American speakers isolates different groups of variants used by folk speakers and cultivated speakers, respectively, consistent with their separate clustering. The variants associated with folk speakers are largely those identified by the second principal component discussed above and associated with southwestern England, while those associated with cultivated speakers tend to be found in the English southeast. That finding suggests a regional English origin for some sociolinguistic distinctions surviving into the twentieth century in Southern American English.

5.3.5 Discussion

The analysis reveals a great deal of variation as well as some interesting patterns of similarity among speakers in southern England and in American regions. To the extent that 20th-century patterns of linguistic variation in southern England and the former colonies in America reflect the patterns prevalent a few centuries ago, the analysis supports the hypothesis that American speech is a relatively uniform amalgam of variants largely brought from the southeast of England. That relative uniformity likely stems from the fact that the bulk of British emigration was from the English southeast. The bulk of the roughly 155,000 English immigrants who settled in the mainland North American colonies during the seventeenth century were indentured servants who sailed from London and came from the Thames valley, whereas other regions provided relatively small shares of the total immigration from Britain. That concentration probably has two sources: One is that the South of England held about two-thirds – and the Southeast about one-third – of the country’s population; the other is that the bulk of Atlantic travel originated in London, which was absorbing more or less all of the natural increase in population in Britain during at least part of the period of American colonization. Despite
changes in the regional patterns of emigration during the eighteenth century, the bulk of English settlers continued to come from the Southeast.

Nevertheless, clear regional differences appear between American speakers from New England and the South, differences that comparison of English and American speech varieties suggests to be attributable in significant part to differences in the speakers' affinities with speakers from English regions. The patterns of similarity are consistent with historical evidence of patterns of migration from seventeenth- and eighteenth-century Britain to North America. Massachusetts speakers tend to have greater affinities with those from eastern regions, particularly the East Midlands, consistent with the predominance of easterners among the early Puritan settlers. Southern American English speakers, in contrast, tend to have greater affinities with those from the southwestern regions, particularly from the Southwest, consistent with the West Country origins of many of the early indentured servants who emigrated to Virginia. The differences between American dialects and affinities with regional varieties of English are thus consistent with a model of new-dialect formation in the American colonies, involving competition within and selection from a pool of variants introduced by speakers from different dialect regions.

Moreover, variants used by Southern American English speakers judged as folk speakers are much more likely to be found in southwestern England than variants by speakers judged as common or cultivated, suggesting that some sociolinguistic distinctions surviving into the twentieth century within Southern American English may also be attributed to regional (and perhaps socioeconomic) British origins of the variants. One may further speculate that speech features of West Country indentured servants acquired lower prestige in some social circles in the South but also became markers of local or regional identity in others, much as similar features help establish and maintain regional identity in some forms of Southern American English today.

### 5.4 Future Directions

The insights derived from the research described here suggest several natural extensions. Feature-based analyses could be extended in Britain to include a larger number of segments from the impressionistic data from the *Survey of English Dialects*, to audio recordings from the *Survey*, and to a larger set of speakers, including speakers of differing ages, sociolinguistic status, and more regions. Those extensions could yield useful information about rates and directions of British linguistic change. Extensions to include speakers from Scotland and Ireland would be particularly useful for the study of new dialect formation in the colonies.

Feature-based analyses also could be extended by expanding the American
dataset to include more detailed impressionistic data (and from more speakers) directly from the various *Linguistic Atlases*, or by drawing on the associated audio recordings. It seems very likely that further insight into the formation of American dialects could be gained by including more speakers, particularly from the most important hearths – the Virginia Tidewater, Charleston, Philadelphia, and so forth. (I strongly suspect that a great deal may be learned by seeking the origins of Inland Northern speech in the Connecticut River valley and particularly in New Haven, from which the valley was largely settled.) In addition, the *Linguistic Atlases* provide sufficient information to further assess variation by sociological status and perhaps to seek historical sources for that variation as well.

A particularly interesting direction of research would involve the application of the quantitative toolkit described in this thesis to a dataset drawing on both the *Survey of English Dialects* and the American *Linguistic Atlases*. The corpora share roughly 170 words common, including an average of about 5 words from each class of Middle English long vowel, short vowel and diphthong, with and without rhotics, as well as the consonants for which there is any variation across the regions. Systematic analysis of such a combined dataset would almost certainly yield a variety of additional insights into both English and American dialect development.

Finally, the quantitative toolkit might be expanded by developing a variant of principal component analysis more tailored to linguistic data. Such a variant might involve a constraint that each principal component must include one variable from each of several vowel categories – perhaps using priors from categories in the standard language, or perhaps in terms of features. An alternative might involve iterations between clustering and principal component analysis, constraining each principal component to more closely resemble the set of variants or features characteristic of a cluster. To my knowledge, neither such algorithm has yet been developed and their utility remains undetermined.