CHAPTER 1

Introduction

The goal of this book is to map regional linguistic variation in written American English. To investigate this topic, a large corpus of modern American letters to the editor was collected from hundreds of cities from across the United States. This corpus was then used to map hundreds of measurements of grammatical variation in American English for the first time. Statistical analyses of these maps found that regional grammatical variation exists in written American English and that most grammatical variables follow only one of a few basic regional patterns. In addition, five modern American dialect regions were identified: the Northeast, the Southeast, the Midwest, the South Central States, and the West. These results challenge standard theories of American dialect regions and show that regional linguistic variation is far more complex than is generally assumed. This chapter situates this study by reviewing previous research in American dialectology and by presenting an outline for the rest of this book.

1.1 American dialectology

The first large-scale survey of regional dialect variation in American English was the *Linguistic Atlas of the United States and Canada*. As recounted by the director of the *Atlas* Hans Kurath (Kurath et al., 1939), the project was first proposed in December 1928 by members of the Modern Language Association, inspired by the national dialect surveys being conducted across Europe at the turn of the century. A committee chaired by Charles C. Fries and including Kurath was formed to consider the feasibility of such a project. In January 1929, unaware that this committee had been formed, E. H. Sturtevant at Yale proposed a similar project to the American Council of Learned Societies. The two groups were united at a meeting in February 1929 that was organized by Fries, where a formal proposal for the project was drafted. Sturtevant then presented this proposal in March 1929 to the Executive Committee of the American Council for Learned Societies, who

agreed to fund a conference to further discuss the project. This conference took place that summer at Yale and resulted in the appointment of a new committee – chaired by Hans Kurath and including Leonard Bloomfield among other top American linguists of the time – that was charged with presenting a proposal and a budget for the *Linguistic Atlas of the United States and Canada* to the Executive Committee. The plan was approved by the Council in January 1930, although they recommended that the committee first conduct a survey of New England before a continental survey was begun.

The *Linguistic Atlas of New England* began in 1931, with Kurath as the director and Miles L. Hanley as associate director. Data collection for the survey was completed in 1933, with 416 informants in 213 communities from across New England, as well as New Brunswick, having been interviewed by 9 fieldworkers, including Guy Lowman, the primary fieldworker, and Kurath, as well as noted linguists Bernard Bloch and Martin Joos. The fieldworkers gathered data using a standardized questionnaire designed by Kurath to elicit upwards of 800 different items, especially words used to discuss common subjects and regional activities, such as geography, weather, time, flora, fauna, farming, mining, and forestry. Grammatical data on a limited number of function word alternations (e.g. *whom/whom*, *ran across/into*) and morphological alternations (e.g. *dived* vs. *dove*) were also collected. In addition, responses were phonetically transcribed by the fieldworkers so that phonological features could be analyzed. In most communities only two informants were interviewed – an elderly informant from an old, local family and a middle-aged and more well-educated informant from a local family. Informants with university educations were also interviewed in larger urban areas. This approach to selecting informants was taken because it was only possible to interview a small number informants at each location, making it necessary to focus on informants who were most likely to use regional forms. In addition, because Kurath was specifically interested in identifying historical patterns of regional variation, non-mobile, older, rural male informants were generally preferred.

Following data collection, maps were produced showing the distribution of each of the linguistic forms. Because the maps were often quite unclear, with different variants dispersed across New England, Kurath also manually plotted linguistic borders known as *isoglosses* to divide the region into sub-regions where the different forms predominated. This allowed for Kurath to make sense of the complex data he was faced with and focus his analysis on the underlying patterns of regional variation in these maps.

In addition, the maps for various linguistic variables were compared in order to identify *bundles of isoglosses* – isoglosses for multiple variables that follow similar paths. In this way, common patterns of regional linguistic variation were identified and used to locate dialect regions. The major finding of this survey was that there were two principal dialect regions in New England: eastern New England and western New England, with the border between these regions running through Connecticut, Massachusetts and Vermont. For example, the survey found that post-vocalic /*r*/ deletion, the pronunciation of *library* with two syllables, and the use of the term *comforter* rather than *quilt* were all features found primarily in eastern New England. These patterns were explained by appealing to historical settlement patterns, as the eastern region had been settled by colonists originating from the Atlantic coast whereas the western region had been settled by colonists originating from the Lower Connecticut River Valley and the Long Island Sound. The methods and results of this survey were published in three volumes beginning in 1939 (Kurath et al., 1939–1943) along with a handbook (Kurath et al., 1939).

After data collection was completed for the *Linguistic Atlas of New England*, Hans Kurath prepared to survey the rest of the Atlantic Coast for the *Linguistic Atlas of the Middle and South Atlantic States*. However, the Great Depression and the lack of funding and interest outside New England only allowed Kurath to send Guy Lowman into the field (Kretzschmar et al., 1993). From 1933 to 1938, Lowman traveled the eastern seaboard conducting interviews in communities from Delaware to northern Florida (McDavid & O'Cain, 1979). For these investigations, Lowman used the same basic procedure and questionnaire that was used in New England, although Kurath had modified the questionnaire, adding and removing certain forms. Due to a lack of local funds and interest, Kurath put the South Atlantic survey on hold in 1939 and sent Lowman to begin a survey of the Middle Atlantic States. Over the next two years Lowman collected data from Pennsylvania, West Virginia, New Jersey, eastern Ohio, and New York City (Kretzschmar et al., 1993). Tragically, in the summer of 1941 Lowman died in car accident while collecting data around the Finger Lakes in Upstate New York (Kretzschmar et al., 1993). Following Lowman's death, Kurath selected Raven I. McDavid, who had been recruited by Bloch at the 1937 Linguistic Institute, to complete data collection for the Middle and South Atlantic States (Kretzschmar et al., 1993). Data collection was put on hold in 1942, when McDavid joined the United States Army's Intensive Language Program (Kretzschmar et al., 1993), but McDavid returned to the field in 1945 and by 1949 over 1,200 informants had been interviewed

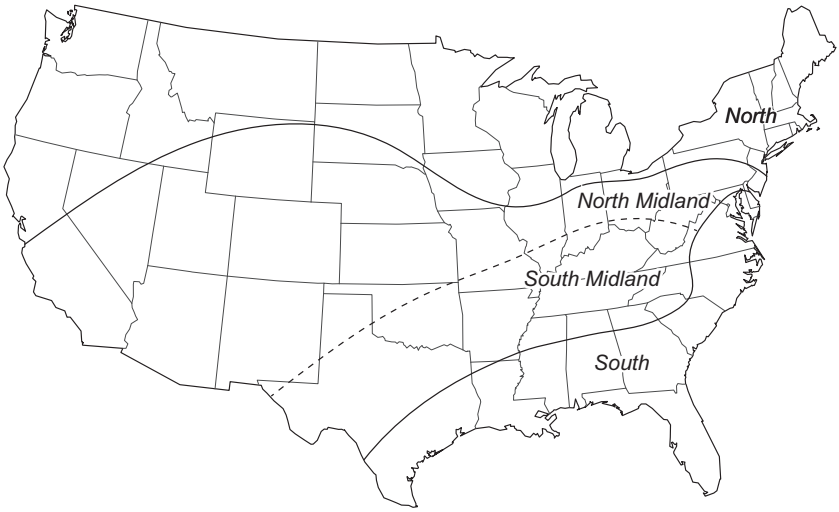

Figure 1.1 American dialect regions: *Linguistic Atlas of the United States and Canada*

from across the Middle and South Atlantic States (Kurath & McDavid, 1961).

The first major study to analyze the data from the Middle and South Atlantic States, as well as the data from New England, was Kurath's *Word Geography of the Eastern United States*, published in 1949, which mapped lexical variation from New England to South Carolina. Kurath identified three major dialect regions in the Eastern United States by plotting and comparing isoglosses, a method that he had extended since the survey of New England and that was becoming the standard approach in American dialectology. These dialect regions are mapped in Figure 1.1 and consist of the North (where words such as *pail* and *brook* are more common), the Midland (where words such as *skillet* and *snake feeder* are more common), and the South (where words such as *snap bean* and *turn of wood* are more common). In addition, Kurath also identified internal divisions within these three regions, including a distinction between the Northern and Southern Midland. Kurath considered the identification of a distinct Midland region as the main descriptive contribution of the study.

Once again, Kurath explained these dialect patterns based on historical settlement patterns. He argued that the Northern dialect region corresponds to the area settled by British colonists originating in New England, who moved through New York and into northern Pennsylvania

and Ohio, that the Midland dialect region corresponds to the region settled by British, Scotch-Irish, and German colonists originating in Philadelphia, who moved through southern Pennsylvania into western Virginia and the Lower Midwest, and that the Southern dialect region corresponds to the region settled by British colonists originating in Virginia and the Carolinas, who moved into the Deep South. Because these three groups of settlers had different linguistic and cultural backgrounds and were largely independent of each other, over time they developed distinct forms of speech, which were the foundation for the contemporary dialect regions that Kurath observed. This settlement theory of American dialect regions has dominated the field ever since.

The data from New England and the Middle and South Atlantic States, which by then contained data for over 1,400 informants, was also the basis for E. Bagby Atwood's *A Survey of Verb Forms in the Eastern United States*, published in 1953. This book represents the first and until now the only American dialect survey to focus on grammatical variation. Atwood analyzed variation in the expression of tense (e.g. *boiled/boilt*), the present perfect (e.g. *I have/am been*), the present participle (e.g. *singing/singin'*), the infinitive (e.g. *to tell/for to tell*), verb agreement (e.g. *you were/was*), and verb negation (e.g. *ain't, hain't*), as well as the use of certain highly marked verbal constructions such as the *might could* double modal construction (e.g. *I might could do it*) and the *belongs to be* construction (e.g. *he belongs to be careful*). In line with Kurath, Atwood found evidence for the three-way division of the Eastern United States into Northern, Midland, and Southern dialect regions. For example, *clim* as the past tense of *climb* was identified as a Northern form, *boilt* as the past tense of *boil* was identified as a Midland form, and the *belongs to be* construction identified as a Southern form. Overall, however, Atwood presents a somewhat different picture of the Midland than Kurath, noting that the Midland was characterized more by the absence of distinct forms, rather than their presence, as was the case for the North and the South. Atwood also discussed the social distribution of these non-standard forms, foreshadowing the shift toward social variation that was about to take place in dialectology, led by William Labov (1963, 1966a, 1969, 1972).

This three-way division of American dialect regions was also supported by Kurath and McDavid's analysis of phonetic and phonological variation in the Eastern data set, which at that time represented the language of over 1500 informants. Although the same basic patterns of regional variation were identified, Kurath and McDavid found that the border between the Northern Midland and the Southern Midland was stronger than the

division Kurath had identified in his lexical analysis of the same data set (see also McDavid, 1993). Furthermore, while pervasive regional patterns in pronunciation were identified, like Atwood, Kurath and McDavid also found considerable variation across social groups. The results of this study were presented in *The Pronunciation of English in the Atlantic States*, published in 1961, which was the last major study based on the data gathered for the Linguistic Atlas Project in the Eastern United States. Kurath would pass away a few years later in 1964 and McDavid would take over directorship of the project, but momentum slowed. Partial records for the Middle and South Atlantic States were finally published in 1979 (McDavid & O'Cain, 1979) and, following McDavid's death in 1984, a handbook was published in 1993, led by William Kretzschmar, who took over the directorship of the project and who maintains the records today. An *Atlas of the Middle Atlantic States* was never published.

While Kurath and his team were surveying the Eastern United States, affiliated regional surveys were being conducted elsewhere in the United States. As early as 1938, data was being collected for the *Linguistic Atlas of the North Central States*, under the directorship of Albert H. Marckwardt (Allen, 1973). Although at first the survey covered the entire Midwest, at a meeting in New York City in 1948 attended by Kurath and McDavid, Marckwardt agreed that the *Linguistic Atlas of the Upper Midwest* should be conducted as a separate survey, focusing on the states of Minnesota, Iowa, Nebraska, South Dakota and North Dakota. Directorship of the survey was awarded to Harold B. Allen, who had been trained by Kurath and Bloch at the 1939 Linguistics Institute. Marckwardt continued to collect and analyze data from the Eastern Midwest and by 1978 over 550 informants had been interviewed in Michigan, Ohio, Indiana, Kentucky, Illinois, Wisconsin, and Southern Ontario (Kurath, 1979; Labov et al., 2006). While no atlas was ever published for this region, smaller studies (e.g. Marckwardt, 1957) found that the division between the Northern and Midland dialect regions in the Eastern United States extended into the Midwest, with the border between the two regions running through the northern third of Ohio, Indiana and Illinois. These results agreed with Alva L. Davis's 1948 doctoral dissertation, *Word Atlas of the Great Lake Region*, which was based on a postal questionnaire, and were replicated in Roger Shuy's doctoral dissertation, which focused on the boundary between the Northern and Midland dialect regions in Illinois (Shuy, 1962).

Allen's survey progressed independently in the Upper Midwest, using an extended version of Kurath's basic questionnaire, with a total of 208 informants interviewed and recorded between 1949 and 1957. In addition,

1,064 total informants responded to a postal questionnaire following the approach to data collection developed by Davis. The results of the survey were published by Allen as the *Linguistic Atlas of the Upper Midwest* in three volumes from 1973 to 1976 (see also Allen, 1952, 1958, 1959, 1964). Based on an analysis of lexical, phonological, and morphological features, Allen concluded that the distinction between the Northern and Midland dialect regions also extended through the Upper Midwest. Like Kurath, Allen explained these patterns based on historical settlement patterns, with settlers of the northern half of the Upper Midwest coming from New York and northern Ohio, and with settlers of the southern half of the Upper Midwest coming from the Mid-Atlantic States and southern Ohio along the Old National Trail.

Around the same time as these Midwestern surveys were being conducted, E. Bagby Atwood, who had previously analyzed verb forms in the Eastern United States, was surveying the vocabulary of Texas and the South Central States including Louisiana, Arkansas, Oklahoma and New Mexico, which was reported in *The Regional Vocabulary of Texas*, published in 1962. The data for this survey was gathered by Atwood and his students and colleagues during the 1950s, using an extended version of Kurath's questionnaire. By comparing his results to Kurath's, Atwood showed that Southern dialect words were relatively common across the South Central states, as were Midland dialect words and Spanish borrowings to a lesser extent, reflecting the mixed settlement history of this region. Based on this evidence, Atwood argued that the language spoken in Texas and the South Central States was a form of Southern English.

The last of the affiliated regional surveys to be completed was the *Linguistic Atlas of the Gulf States.* Primary fieldwork took place under the directorship of Lee Pederson between 1973 and 1979, during which 1121 informants were interviewed and recorded by 256 field investigators in 8 southern states: Tennessee, Georgia, Florida, Alabama, Mississippi, Louisiana, Arkansas, and eastern Texas. The results were published in seven volumes from 1986 to 1993 (Pederson, 1986; Pederson et al., 1986–1993). The basic finding of the survey was that there were two major dialect regions in the Gulf States – the Upland and the Lowland – with the border between these two regions running through northern Georgia, Alabama, and Mississippi. These dialect regions correspond to the Southern Midland and the South as identified by Kurath and his colleagues in the Eastern United States, showing that these Eastern dialect regions had been extended across the South through settlement, much as they had been extended across the Midwest.

Other regional surveys affiliated with the *Linguistic Atlas of the United States and Canada* were begun, but none were ever completed or resulted in major publications. Most notably, in the Far West, data collection was begun and the preliminary results were reported for two surveys. Data collection for *the Linguistic Atlas of the Pacific West* was conducted in California and Nevada between 1952 and 1959 with initial analyses showing a distinction between the language of Northern and Southern California (Reed, 1954). Similarly, data collection for the *Linguistic Atlas of the Pacific Northwest* was conducted between 1953 and 1963 (Reed, 1956, 1957, 1961), with initial results showing for example that Northern and North Midland forms were common across the region, whereas Southern terms were relatively rare. Other unfinished regional surveys included the *Linguistic Atlas of Oklahoma*, whose preliminary data was analyzed by Atwood in his study of the vocabulary the South Central States, and the *Linguistic Atlas of the Rocky Mountain States*, for which data collection reportedly began in 1988 (Labov et al., 2006). Aside from a small amount of data collected in Ontario and New Brunswick, Canada was never mapped as part of this survey.

The various regional surveys associated with the *Linguistic Atlas of the United States and Canada* mapped much of the United States, although there were several gaps in the analysis, especially in the West, and given the many years over which the surveys were completed, it is unclear how comparable these results are, or if taken together what era they could be said to represent. Nevertheless, the major patterns of regional linguistic variation identified by these surveys are combined and presented in Figure 1.1, which represents a synthesis and an interpolation of the results of these various regional surveys. The dialect regions identified in the Eastern United States are based directly on the results of the surveys described above. No data, however, is available for Missouri or parts of West Virginia, Kentucky, and Florida. In these regions the dialect borders are estimated based on the surrounding area. In the West, very little data was collected, but according to Kurath (1972) preliminary analyses on the West Coast demonstrated that the border between the North and the Midland extends to the Pacific North West, which is reflected in Figure 1.1. Although Kurath and his colleagues never produced such a national map, this map is consistent with the results of their surveys and with Kurath's view of American dialect regions. This map therefore represents a theory of what the Linguistic Atlas Project would have found had the various regional surveys been completed and combined.

The *Linguistic Atlas of the United States and Canada* was not the only attempt to map American English. Long before the possibility of a dialect atlas was discussed by the Modern Language Association and the American Council of Learned Societies, a dictionary of American English was proposed at the founding of the American Dialect Society in 1889. Although the Society published research on American regional lexicography in their journal *Dialect Notes*, which was first published in 1890, and later in the *Publications of the American Dialect Society*, data collection for a dictionary of American English was not begun in earnest until Fredric G. Cassidy was appointed as the editor of the dictionary in 1962. Fieldwork was conducted between 1965 and 1970, over which time 80 fieldworkers interviewed 2,777 informants in 1002 communities. The fieldworkers used a questionnaire developed by Cassidy that contained over 1,800 questions relating primarily to rare and archaic vocabulary items, which resulted in over 20,000 different lexical items being elicited (Carver, 1987). The results of the survey were published as the *Dictionary of American Regional English* in seven volumes between 1985 and 2013 (Cassidy & Hall, 1985, 1991; Hall & Cassidy, 1996; Hall, 2002, 2012, 2013).

The primary purpose of the *Dictionary of American Regional English* was to identify and define regional vocabulary items from across the United States, rather than to map the dialect regions of American English. However, the dictionary was the basis for Craig Carver's analysis of regional lexical variation in American English, *American Regional Dialects: A Word Geography*, published in 1987, which represents the first complete survey of regional linguistic variation in American English. In order to analyze the massive amounts of data gathered for the dictionary, Carver focused on analyzing sets of words in the aggregate. Specifically, Carver identified what he called *dialect layers*, which were defined based on sets of words that he judged to exhibit similar regional distributions. The degree to which a particular location was part of a particular dialect layer was then calculated as the percentage of the words associated with that dialect layer observed at that location. For example, Carver's New England Layer is defined based on 45 lexical items, including use of the word *grinder* for a type of sandwich and *rotary* for a roundabout. Each location was then scored based on the percentage of these 45 lexical items that had been attested at that location. Lines were then drawn around the locations with the highest percentage of those words to map that layer, with the highest concentration of New England words occurring at locations in New Hampshire, Central Massachusetts, and Western Vermont. Carver mapped a large number of

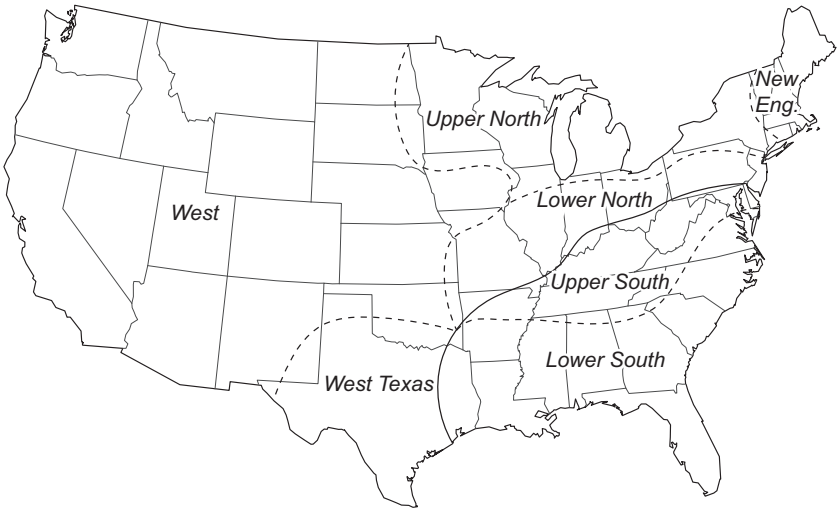

Figure 1.2 American dialect regions: *Dictionary of American Regional English*

layers in this manner and then used these results to infer the locations of American dialect regions.

Based on this approach to the analysis of regional linguistic variation, Carver identified two major dialect regions in the United States: the North and the South. In turn, Carver divided the North into three main sub-regions (the Upper North, the Lower North, and the West) and the South into two main sub-regions (the Upper South and the Lower South), as mapped in Figure 1.2, with Carver's Lower North and Upper South sub-regions corresponding roughly to Kurath's Northern Midland and Southern Midland sub-regions respectively. The identification of a Western sub-region is also notable, as this was the first time that a sufficient amount of data had been collected to allow for such a distinction to be made. Carver's two-way division of American dialect regions between the North and the South clearly differs from Kurath's three-way division between the North, the Midland, and the South, but it was not without precedent. According to Kurath (1949), before he began his surveys of the Eastern United States, it was generally assumed that the basic distinction in American English was between the North and the South. This is why Kurath considered the identification of the Midland in the *Word Geography of the Eastern United States* to be such an important discovery. The results of Carver's analysis, however, directly support this older and simpler conception of American dialect regions.