4. Sampling and field methods

4.1. The pilot project

The Telephone Survey "Telsur" and Atlas project "ANAE" began as a pilot study of dialect differentiation in North American English, conducted from November, 1991 to April, 1993.[1] The area chosen for study consisted of all or parts of six states: Illinois, Wisconsin, Minnesota, Iowa, Nebraska, and South Dakota. This project aimed to contribute both to the specific geography of American dialects and to the study of the principles of sound change. The specific area to be studied was chosen because it included major regional boundaries and new phenomena that had not previously been mapped. Sampling in communities with a range of sizes was undertaken in order to represent both the dimensions of geographic dispersion and population density. First, seven focal places were targeted: Chicago, IL; Milwaukee, WI; Duluth, MN/Superior, WI; Minneapolis/St. Paul, MN; Des Moines, IA; Sioux Falls, SD; and Omaha, NE. Four of these are dominant metropolises with a 1990 population over 300,000: Chicago, Milwaukee, Minneapolis/St. Paul, and Omaha. The remaining three were selected to provide geographical coverage; they all have a population over 100,000, and they provide points 150 miles or more from the four larger cities.

The sample design for the pilot project entailed the selection of places within this 150-mile radius of each of the focal cities. In each area, eight cities were to be selected, two in each of four ranges of population:

50,000 to 200,000
10,000 to 50,000
2,000 to 10,000
under 2,000

Cities were selected within a 150-mile radius of the largest cities first. Where the territory of focal cities overlapped, sampling was frequently reduced because not enough cities of the requisite size existed.

Each of the focal cities was to be represented by two subjects, with the exception of Chicago, which was to be represented by four speakers. The smaller towns within the 150-mile radius of the focal cities were each represented by one speaker. In the course of the pilot project, 52 speakers were interviewed in 41 communities ranging in population from 2,605 (Lena, IL) to 6,793,132 in the urbanized area of Chicago, IL in 1992 and 1993.

4.2. Expansion of the project

The acoustic analyses of the first set of speakers showed a clear differentiation of the dialect regions of the Inland North, the North Central region, and the Midland, generally in accordance with the dialect boundaries established by other researchers, but showing a level of detail, precision, and consistency not previously achieved. The next phase of the project extended the territory to a fifteen-state region, from Ohio to the Continental Divide and from the Canadian border to the Ohio River, with Missouri, Kansas, and Colorado forming the southern tier of states west of the Mississippi River (1993–1994).[2] In the third phase, coverage was extended to the entirety of English-speaking North America (1994–1998).[3] As that sample approached completion, more detailed investigation of a set of cities in the Midland region was undertaken (1998–2000) to try to account for the extensive variation found among them. In addition, in 1997 and 1998, interviews of a sample of 41 African-American speakers were conducted in 15 cities with a high proportion of African-Americans.[4]

The sampling strategy for the Telsur/Atlas project was designed with the goal of representing the largest possible population, with special attention to those speakers who are expected to be the most advanced in processes of linguistic change. It has been established that most sound changes are initiated in urban centers (Trudgill 1974; Callary 1975; Bailey et al. 1991); thus the first tier of communities to be sampled consisted of places with the greatest concentration of population. Each community was selected as the focal point of an area, and the areas were determined so as to cover all the territory of English-speaking North America. Three defining terms are involved: Zones of Influence (ZI), Central Cities, and Urbanized Areas (UA). The selection of places to be sampled involves intersecting characteristics of the three levels, as will be explained below. The terms will first be defined, and then the selection criteria that produced the overall sampling plan will be described.

Zone of Influence

A Zone of Influence (ZI) is a set of counties. It is derived from the 1992 County Penetration Reports of the Audit Bureau of Circulations (ABC). ABC audits data from member organizations on the circulation of newspapers and other publications. For every county with at least 100 households, the County Penetration Report lists the name of each member newspaper, gives its circulation, and indicates whether it is a daily or weekly and morning or evening publication. A ZI, defined for the Telsur/Atlas project, is determined by Central Cities (see below). A county belongs to the ZI of a given Central City if, in that county, the circulation of the newspaper(s) from that city is greater than the circulation of the newspaper(s) from any other city that has been designated a Central City for the purposes of the research project.

Once the Central Cities have been selected, it is in theory possible to assign every county to a ZI. In practice this is not true, because some counties have fewer than 100 households and so are not listed in the County Penetration Reports. In most cases, such counties can confidently be assigned to a ZI on the basis of the ZI assignment of surrounding counties. In a few cases, the assignment of a given county could arguably be made to either of two ZIs. In those instances,

1 This project was supported by NSF under grant BNS91-11637, "A Survey of Sound Change in Progress".
2 The second phase of the Atlas was supported by NSF under grant SBR 92-22458, "A Telephone Survey of Sound Change in Progress".
3 The third phrase of the Atlas was supported by grant RT-21599-94 from NEH, "A Phonological Atlas of North American English".
4 The study of the Midland was supported by grant SBR 98-11487 from NSF, "Linguistic Diversity in the North American Midland".

the assignment was made on the basis of considerations such as proximity to the Central City.

Central City

This term is used in two senses. First, it is used as a synonym for the U.S. Census Bureau's definition of a Central Place as the defining feature of larger census units, including the Standard Metropolitan Statistical Area (SMSA) and the Urbanized Area (see below). The second sense is defined for the Telsur/Atlas project: a Central City is the central place of a Zone of Influence. As in the Census Bureau definition, a Central City may actually consist of more than one city: examples are Minneapolis/St. Paul, MN and the Quad Cities on the Mississippi River (Moline and Rock Island in Illinois and Davenport and Bettendorf in Iowa). The basic criterion for the selection of a Central City of a ZI is that it is a place for which the Urbanized Area (see below) has a population of at least 200,000 according to the 1990 census. Due to low populations in some areas, it was necessary to designate a number of cities smaller than this limit as Central Cities, such as Burlington, VT, Roanoke, VA, and Boise, ID. Three of the Central Cities are even smaller than the threshold of 50,000 which is used by the Census Bureau as a criterion for status as the Central Place of an Urbanized Area; they were assigned the designation of Central Cities for the same reason as the other Central Cities with a population under 200,000: to provide well-motivated geographic coverage. The status of such towns as regional centers is demonstrated by the existence of a local newspaper that has wide circulation in the area. The three Central Cities which are not UAs are Minot, ND, Aberdeen, SD, and Rutland, VT. Thus a Central City serves as the defining place of a Zone of Influence, and at the same time it is the Central Place of an Urbanized Area.

Urbanized Area

This term is defined by the U.S. Census Bureau in order to provide a better separation of urban and rural population than is given by the SMSA, which takes the county as its building block. It consists of a central city or cities and the surrounding densely settled territory. By definition, it has a population of at least 50,000. The densely settled surrounding area consists of contiguous incorporated or census designated places having either a population of 2,500 or more, a population density of 1,000 persons per square mile, a closely settled area containing a minimum of 50 percent of the population, or a cluster of at least 100 housing units. Further details on the definition of an Urbanized Area may be found in the Census Reports. The composition of each Urbanized Area is shown on maps in the series of census reports *1990 CPH-2: Population and Housing Unit Counts*.

In the design of the Telsur/Atlas sample, the Urbanized Area is taken to be a conservative estimate of the territory of the speech community of the corresponding Central City. If a speaker is a native of any place within the Urbanized Area of a Central City, he or she is taken to be linguistically representative of the Central City's speech community. The areal extent of the UAs as mapped by the Census Bureau is quite restricted, which allows us to be confident that this is a valid sampling decision.

The Central Cities selected to define ZIs are further divided into four types by population of the corresponding UA and by area of the ZI, as follows:

p1 UA population > 1 million;
p2 UA population > 200,000, non-restricted (area > 5,000 square miles);
p3 UA population > 200,000, restricted (area < 5,000 square miles);
p4 UA population < 200,000.

These four levels are used to differentiate the amount of sampling to be done in smaller cities within each ZI. At the level of the Central Cities, the only difference in sampling is between the p1 cities and all others: in p1 cities, at least four speakers were to be interviewed, while in all others, at least two were to be interviewed. Furthermore, in every city, an effort was made to insure that at least one speaker would be a woman between the ages of 20 and 40.

Appendix 4.1 lists the 145 Central Cities that were selected for sampling and gives the corresponding ZI and UA populations. The figures show that 54 percent of the population of the United States lives in the 145 Urbanized Areas (or smaller cities) that were selected for sampling. Thirty-three of the UAs have a population over one million, and 112 have a UA population under one million. Thus the total minimum number of speakers to be represented in the completed national sample of the United States would be 356 speakers. A similar sample, consisting of about 40 speakers, was designed for Canada. A sizable number of speakers from smaller towns were interviewed in the course of the pilot project, and it occasionally happened that a speaker in one place was actually a good representative of a different speech community – small or large – and had moved to her/his present community recently. Thus many "extra" speakers were interviewed: the Telsur sample of North America consists of 762 speakers. The additional speakers add greatly to the depth and richness of the data, and they provide further confirmation of the validity of the methods employed through the consistency of the findings that they yield. Figure 4.2 shows the speakers interviewed for the Atlas; each speaker is located by a symbol that corresponds to the size of the speech community represented. The cities designated as "Primary" in the legend are Central Cities of Zones of Influence.

4.3. Selection and recruitment of speakers

Once a place was selected, the next step was to locate representative speakers. This was accomplished by searching local telephone directories for names marked by the most prominent national ancestry groups. In most of the pilot project area, the largest group of Euro-Americans is of German ancestry. English and Irish ancestry are also reported widely, Scandinavian ancestry is frequent in the northern region, and Polish ancestry is prevalent in the industrial centers. To maximize the likelihood of reaching speakers who are native to their places of residence, names were selected that occurred in clusters. Ideally, names were chosen that were listed as "Jr". when the senior with the same name was also listed. The initial interchange with a person who answered the telephone was the identification of the interviewer by name, giving the affiliation with the University of Pennsylvania; the explanation that a study of communication among people from different parts of the country was being conducted; and the question of whether the speaker had grown up in the town where he or she was located. If the answer was affirmative, permission to conduct the interview was requested. If the speaker agreed, permission to record the interview was requested. The complete script of this introduction, as well as the entire interview schedule, is given in Appendix 4.3. The make-up of the interview schedule will be discussed below.

Bias in Telephone Listings

By using published telephone listings to locate prospective speakers, we introduce the possibility of bias from the exclusion of those with unpublished telephone numbers. Labov (2001) reports that in the Philadelphia neighborhood study of sound change in progress conducted from 1973 to 1977 (LCV), a strong negative correlation was found between social class and the rate of unlisted telephone numbers, as follows:

Social class	% unlisted telephone numbers
Lower working class	80
Upper working class	56
Lower middle class	44
Middle middle class	31
Upper middle class	0
Upper class	0

This finding was understood by the fieldworkers to stem from the varying degree to which the different groups felt the need or wish to be available to the outside world. The effect of this bias on the study of sound change in progress was tested in the LCV study. Telephone listing for the subjects in the neighborhood study was entered as a variable in the regression analysis of the first and second formants for all the vowels under investigation, and this was compared with the results of a complementary survey of sound change in progress carried out by telephone (Hindle 1980). If telephone listing biased speakers towards either greater or less advancement of sound change, it would appear as a significant effect on the normalized vowel formant value. No such effect was found for any vowel.

While we can therefore assume with reasonable confidence that we are not likely to be misled as to the direction of sound change in the present study by relying on telephone listings to locate speakers, we must recognize that the pool of accessible speakers is reduced as we descend the social scale. This may not alter our finding as to the progress of sound change, but it is likely to affect the speaker sample's representation of the population as a whole. In this work, we have employed the Socio-economic Index (described below) developed by Duncan (1961) and updated most recently by Nakao and Treas (1992) to rank speakers on the social scale. Indeed, the distribution of the Telsur/Atlas speakers by Socio-economic Index appears to be weighted towards those who are higher on the social scale. Table 1 compares the social class distribution of the population in a selection of cities of varying sizes and locations with the social class distribution of the Telsur/Atlas sample as a whole.[5]

Table 4.1. Population by social class in selected cities

	Population	Upper middle	Middle middle	Lower middle	Upper working	Middle working	Lower working
New York City CMSA	8,716,770	16	25	13	30	11	5
San Francisco CMSA	3,239,687	17	26	12	29	11	5
Dallas CMSA	2,010,378	14	24	13	32	12	6
Miami CMSA	1,500,947	13	23	12	33	13	6
Minneapolis/St. Paul, MN	1,329,371	15	24	13	31	12	5
Cleveland, OH	1,266,993	13	22	12	33	13	7
St. Louis, MO-IL	1,154,922	14	23	13	32	13	6
Denver, CO	975,817	16	26	13	30	11	4
Kansas City, MO	777,523	14	23	13	32	12	6
Montgomery, AL	128,656	13	23	13	33	12	7
Muskegon, MI	65,424	10	18	11	36	15	10
Monroe, LA	58,100	13	23	12	33	14	6
All (N)	21,214,588						
Percent		15	24	13	31	12	5
Telsur speakers (N)	633	98	250	65	114	57	49
Percent		15	39	10	18	9	8

Table 4.1 shows that the Middle Middle Class is over-represented in the Telsur/Atlas sample as compared to the general population, and the upper working class is under-represented. The skewing found here is much less than the skewing of telephone listings by social class, however, and, most importantly, all the social classes are well represented. While it is thus evident that the speakers interviewed for the Telsur project do not precisely reflect the social class distribution of the population at large, this does not interfere with the investigation or analysis. The aim of ANAE is to determine those structural patterns that differentiate communities rather than those that differentiate speakers within the community. Various tables throughout the Atlas will take advantage of the distribution of social parameters throughout a dialect or regional area to establish their influence on the progress of a change. In these multivariate analyses, regression coefficients for education and occupation are generally much lower and less significant than those registering the effects of age, gender, and city size

Sociolinguistic studies of large cities show that centrally located social groups – lower middle and upper working class speakers – are the initiators of those sound changes internal to the system, which operate below the level of consciousness. Though these changes eventually affect the entire community, these centrally located speakers are more advanced in ongoing sound changes than are speakers at the extremities of the social scale. With two-thirds of the Telsur speakers falling into the upper working, lower middle, and middle middle classes, we can have some confidence that newly emerging sound changes will be represented in the data. As a further brake on any bias of the sample towards higher-class speakers, in the last stages of interviewing to complete the sample, special techniques were developed to locate speakers who satisfied the strictest criteria of nativity and social class. These will be detailed below.

4.3.1. National ancestry

The methods described in Section 4.4 are appropriate for a study of the central tendencies of speech communities, but not for a detailed examination of social differentiation within a community. Over the past two hundred years, large numbers of immigrants have entered most of the cities studied here; the great majority of them have become speakers of the current local dialect in the second and following generations. Even when a majority of the population consists of groups of foreign stock, the doctrine of First Effective Settlement applies: the new groups assume the cultural patterns of the smaller groups who preceded them (Zelinsky 1992; Mufwene 1996). In order to maximize the chances of recruiting local speakers, the Telsur method tended to focus on the majority ethnic groups in each area.

Table 4.2 gives the overall distribution of the major ethnic groups in the sample by the regions established in Chapter 11. In response to the question, *What's your own family's national ancestry?* (Appendix A), 79 responded "White", "American", "European", or some other non-specific information. These are summed up as "White" in Table 4.2. The other figures show only the first identification given.

Most of the subjects named more than one nationality in response to this question (418 out of 762). Table 4.2 shows only the first response given; the overall proportions of national ancestral groups are similar for second, third, and fourth items given. The bold figures show the mode for each region. The right-hand column gives the percentages of each group in the 2000 U.S. Census for all Americans.

5 The data in Table 4.1 are based on figures given for occupation of employed persons 16 years old and over in Table 18, "Labor force and disability characteristics of persons: 1990" from the census volume series CPH-3. The calculation of the social class distribution from the data on occupation is described in Appendix 2.

In general, the proportions of national ancestral groups are ordered similarly to the census. The largest single identification is German. In the Telsur sample, the German group is by far the largest in the Midland, the North, and the West. There is a much more even distribution of ethnic groups in the South, with a heavier representation of English and Scots-Irish. The Scots-Irish are the modal group in Canada. The Mid-Atlantic region (which includes New York City, Philadelphia, Wilmington, and Baltimore) is the only region in which Italians are the predominant ancestral group.[6]

The emphasis of the Telsur method on the predominant ethnic group is seen most clearly in the high numbers of subjects of German background; the proportion is about twice as high (28%) as in the Census (15%). So far, German nationality has not been associated with the greater or lesser development of the phonology of the Midland and the North, but this bias in the population must be borne in mind. The Telsur method has not led to the elimination of smaller ethnic groups. Lithuanian, Finnish, Welsh, and Lebanese are represented in the 22 speakers in the "Other" category. Considering all responses, 14 of the Telsur subjects identify Jewish ethnicity in their background. A much larger number mention some Native American group. In terms of primary identification, the greatest number of Native Americans are found in the South.

Table 4.2. National ancestral groups identified in first response to Telsur questionnaire. Bold figures indicate largest group in a region.

	Canada	Midland	Mid-Atlantic	North	South	West	Transitional	Total	2000 Census %
English	5	11	2	19	23	9	3	72	8.7
Scots-Irish	**11**	7	2	7	21	2	0	50	1.5
Irish	3	17	5	14	16	1	2	58	10.8
German	4	**80**	7	**67**	**29**	**29**	1	**217**	**15.2**
Dutch	1	5	1	5	1	2	0	15	1.6
Scandinavian	0	4	0	20	2	10	0	36	3.5
French	4	5	1	12	5	4	0	31	3.0
Canadian French	0	0	0	4	2	1	0	7	0.8
Italian	0	7	**13**	16	12	2	1	51	5.6
Jewish	1	2	0	2	2	1	0	8	
Polish	1	6	2	11	2	2	0	24	3.2
Other Slavic	5	8	2	9	4	0	0	28	.25
Other	1	4	2	3	3	0	0	13	
"White"	1	17	0	23	27	11	1	80	
Hispanic	0	0	0	2	5	6	0	13	12.5
African-American	0	2	5	7	27	4	0	45	12.9
Native American	1	4	0	1	7	1	0	14	1.5
Total	38	179	42	222	188	85	8	762	

4.3.2. Race

Although thirteen subjects gave some Hispanic or Latino identification in response to the question on ethnicity, the Telsur survey did not focus on the 12.5 percent of the U.S. population that is Hispanic. The studies of Latino/a English that have been carried out in the last several decades indicate that there are some common features of the second generation dialect that differentiate it from others (Santa Ana 1992; Bayley 1994). Detailed sociolinguistic studies have found that Latino speakers are subject to several competing influences: traditional Spanish, AAVE, and the local white dialects (Wolfram 1974; Poplack 1978; Fought 2003). A thorough and accurate study of geographic differences in the English of Latinos from the Caribbean and various countries of Central and South America is beyond the scope of the current work. It is not likely that the Telsur interview would be able to trace the many variable tendencies in these English dialects, where consistent dialect patterns are still in the process of formation.

The study of geographic differentiation among African-American speakers raises a different set of questions. Studies of AAVE have shown a remarkable geographic uniformity in those grammatical and phonological features that are distinctive to this dialect (NYC: Labov et al. 1968, Labov 1972; Detroit: Wolfram 1969, Edwards 1992; Washington DC: Fasold 1972; Mississippi: Wolfram 1974, Loman 1967; North Carolina: Anshen 1969; Los Angeles: Baugh 1983; San Francisco: Mitchell-Kernan 1969). In general, African-American speakers do not participate in the regional sound changes that are the main focus of ANAE (Labov and Harris 1986; Veatch 1992; Labov 2001: 506–508; Thomas 2001). Thomas finds a remarkable uniformity of vowel systems among African-Americans throughout the U.S. (p. 165), even in the South (p. 170).[7] At the same time, there are consistent differences between African-Americans and whites in the South, even in the earliest records.

Even in those Northern cities in which African-Americans form the majority (e.g. Detroit), African-Americans do not appear to have had any influence on the evolution of the white vernacular, either in the city or the surrounding suburbs. For this reason, the Telsur survey did not specifically search for African-American speakers in the North, the Midland or the West. In those areas, 22 subjects identified themselves as having African-American ethnicity, in whole or in part.

The procedure in the South was the opposite. Using the special methods for locating speakers of a given background discussed in Section 4.7 below, African-American subjects were targeted in five major cities: New Orleans, Jackson, Birmingham, Atlanta, and Durham. Chapter 22 reports on the phonological inventories and phonetic patterns of these speakers, comparing them with the white subjects in the same cities. The chapter includes a summary of the phonetic analyses of rural and small-town African-Americans by Erik Thomas.

4.4. Methods of recruitment

Understandably, many speakers are wary of an unsolicited telephone caller who begins speaking from a prepared script. Telsur interviewers were trained to initiate the interview in a slow speech style to achieve maximum clarity in explaining the purpose of the call. The overt purpose of the interview was explained in the following initial script:

> *Hi, my name is _____. I'm calling from the University Pennsylvania in Philadelphia. We're doing research on communication between people from different parts of the country, so we're looking for people who grew up in one place to help us by telling us a little about how people say things in your area. Did you grow up in _____? If yes: Can you take a few minutes now to answer some questions?*

6 There are 11 subjects in New York City: three are Italian and three are Irish, one German, one Scots-Irish, one Dutch and two African-American

7 In the North, some recent studies show partial movements of African Americans in the direction of the white regional pattern (Thomas 1989/93 in Ohio, Deser 1990 in Detroit, Henderson 2001 in Philadelphia). Studies of African-American English in Northern cities show stylistic variation in the vocalization of /r/ and monophthongization of /ay/ (Myhill 1988). In the South, African-Americans show vowel systems that are related to general Southern patterns, though the earliest records show consistent differences between African-American and white speech. Many of the older black speakers show monophthongal [e:] and [o:] for the vowel classes of long *e* and *o*, and /aw/ is consistently further back than in white speech (Labov, Graff, and Harris 1986).

If the respondent asked to know more about the purpose of the interview, the interviewer proceeded as follows:

> *People across the country are talking to each other more and more, and at the same time we know that local accents are getting more different, in spite of the fact that we all watch the same TV programs. We want to find out how people talk in each region of the country and whether local ways of talking are changing in any way.*

Since North Americans have a general interest in the existence of dialect differences within American speech, refusal rates were low by comparison with other telephone surveys (see Table 4.3).

PERMISSION TO RECORD. The following routine was followed closely in securing permission to record over the telephone.

> *In order to be able to keep track of everything you can tell us, I need to be able to make a tape recording of this conversation. Is that all right with you? (If informant is hesitant: I can assure you that this information is used only by our research group for our reports about general trends in American English, and no information identifying individuals is ever released.* If still hesitant: *If we come to a question you don't think you want to answer, just tell me and we'll skip it. I don't think you'll have a problem with any of the questions I'm going to ask you.)* If permission is given, turn tape recorder on and tell informant you have done so.

In the small number of cases where the person did not agree to be recorded (7 to 16%), the interviewer was instructed to thank the person for their time and terminate the interview.

Though the Telsur interview did not as a rule reach the levels of intimacy and rapport characteristic of the best sociolinguistic interviews, a large part of it was designed to replicate friendly conversation. The interviewer was trained to call upon all of his or her knowledge and experience of the place where the speaker lived. With each successive interview in a given place, the interviewer was better informed about that place and could converse more effectively with people local to the place. The interviewer was trained to be sensitive to the level of interest shown by the subject in order to maximize the flow of spontaneous speech.

Sensitivity to questioning was most likely to arise in the section on demographic data, which was positioned at the end of the interview. It includes the speaker's age and occupation and also the speaker's parents' occupations. Speakers occasionally declined to give some of this information, but the refusal rate was low. Most speakers had already talked about their own occupations by the time the interviewer reached this section, so the question was a matter of filling in details.

4.5. Records of calls required for successful interviews

The Telsur project kept detailed records of all telephone calls made, in order to trace regional differences in the difficulty of locating local speakers and rates of refusal and acceptance. The ease or difficulty of achieving a successful interview varied greatly. The first phone call of the Atlas was made to Sioux Falls, SD, at 3:30 in the afternoon on February 24, 1992. A woman answered the phone and listened politely to the investigator's request for an interview. She explained that she had a day care center in her home, so she was not free to talk during the day. The interviewer thanked her and dialed a second number in Sioux Falls. This call was answered by a man who agreed to be interviewed after asking, "It doesn't cost anything, does it?" The ensuing tape was labeled TS 1. The last interview, TS 835, was conducted by the same interviewer on November 14, 2001, in San Diego, CA. This interview, with a roommate of the college student in whose name the phone was listed, was achieved after dialing the telephone 142 times. The outcomes of these calls to San Diego were as follows:

Frequency	Result
9	No answer (6%)
54	Answering machine (38%)
7	Busy signal (5%)
12	Phone disconnected (9%)
5	Call screening, fax machine, modem (4%)
42	Respondent not local (30%)
10	Interview refused – not interested, busy, refused recording, etc. (7%)
2	Respondent asked interviewer to call back later (1%)
1	Successsful interview (< 1%)
142	Total calls

These two interviews, the first and the last of the Telsur project, represent the extremes of the task of garnering a successful interview. (There were also occasional instances of getting a good interview on the first phone call to a city or town.) In general, the most difficulty was encountered in places where there was a high proportion of non-local residents. City size was not necessarily a problem. In Chicago, for instance, the following record was made in February, 1993, without any special screening for census districts:

Frequency	Result
1	No answer (6%)
5	Answering machine (29%)
3	Interview refused – not interested, busy, refused recording, etc. (18%)
4	Respondent asked interviewer to call back later (24%)
4	Successful interview (24%)
17	Total calls

In a sampling of cities in the Midwest, another investigator made recordings between January, 1993 and April, 1994 in Wisconsin (Hayward, Steven's Point, and Oconto), Minnesota (Chisholm, St. James, and Minneapolis), Iowa (Grinnell and Denison), South Dakota (Redfield), Nebraska (Wayne and Falls City), Illinois (Lena and Fairbury), and Ohio (Cleveland and Cincinnati). These are mostly small towns, but a number of large cities are included as well. The results were as follows:

Frequency	Result
12	No answer (14%)
8	Answering machine (9%)
3	Busy signal (3%)
7	Phone disconnected (8%)
13	Respondent not local (15%)
14	Interview refused – not interested, busy, refused recording, etc. (16%)
5	Respondent asked interviewer to call back later (6%)
4	No adults at home (5%)
20	Successsful interview (23%)
86	Total calls

In another part of the Telsur region, the state of Texas, three interviewers working together made the following record between June, 1996 and January, 1997, in calls to Austin, Amarillo, Houston, and Dallas:

Frequency	Result
34	No answer (14%)
66	Answering machine (27%)
6	Busy signal (2%)
19	Phone disconnected (8%)
19	Fax machine, business, etc. (8%)
62	Respondent not local (26%)
22	Interview refused – not interested, busy, refused recording, etc. (9%)
3	Respondent asked interviewer to call back later (1%)
1	No adults at home (< 1%)
9	Successful interview (4%)
241	Total calls

In yet another region, the following record was made by two interviewers during April and May, 1995. These calls were made in New York State (Syracuse, Albany, Rochester, and Buffalo) and Pennsylvania (State College, Harrisburg, Pittsburgh, Erie, and Scranton):

Frequency	Result
19	No answer (14%)
50	Answering machine (36%)
4	Busy signal (3%)
9	Phone disconnected (7%)
2	Fax machine, business, etc. (1%)
26	Respondent not local (19%)
8	Interview refused – not interested, busy, refused recording, etc. (6%)
3	Respondent asked interviewer to call back later (2%)
17	Successful interview (12%)
138	Total calls

These records are summarized for purposes of comparison in the following table of percentages of outcomes of each dialing of a telephone number.

Table 4.3. Percentage of outcomes of dialing the telephone in five cities or regions

	San Diego	Chicago	Midwest	Texas	NY & PA
No answer	6	6	14	14	14
Answering machine	38	29	9	27	36
Busy signal	5	0	3	2	3
Phone disconnected	8	0	8	8	7
Not a residence	4	0	0	8	1
Respondent not local	30	0	15	26	19
Interview refused	7	18	16	9	6
Call back later	1	24	6	1	2
No adults at home	0	0	5	<1	0
Successful interview	1	24	23	4	12
Total number of calls	142	171	86	241	138

Overall, the table reflects the relative difficulty of accomplishing a successful interview in terms of the number of times it is necessary to dial the telephone. There is partial comparability among the different places defined here, but there are also differences, as was stated above. The table registers two general types of outcome, which can be considered separately: the first five lines are outcomes in which the phone is not answered by a live person, and the last five lines are outcomes in which the interviewer speaks to a potential interviewee.

Table 4.4 summarizes the frequencies of outcomes in which the interviewer reached a person, in order to assess the rate of actual refusal and success. The case of no adults being at home – when a child under the age of 18 answers the phone – is not included, since those are cases of the interviewer not reaching a potential participant.

It must be kept in mind that speakers were screened as quickly as possible for locality status, in order to weed out non-local speakers with a minimum investment of time and effort. Respondents were told, "We're looking for speakers who grew up in one place to help us by telling us a little about how people say things in each area. Did you grow up in _____?"

Non-local respondents are not candidates to be a Telsur speaker. However, they still have the opportunity to refuse to be interviewed, without divulging their locality status, by cutting off the phone call before the interviewer is able to determine that they are non-local. (Some respondents simply hung up the phone during or immediately after the interviewer's request for participation. Others had reactions such as "Heaven's sakes!" or "We can't help you. Bye", before hanging up.) The number of flat refusals of the total number of adults reached by phone, including non-locals, is given first, as the minimum refusal rate. In another sense, the refusal rate is the number of refusals out of those who either refused after the request for participation was made or who terminated the interaction before responding to the interviewer at all; this calculation is given on the second line of refusal rates. The "true" refusal rate must be somewhere in between.

The success rate may also be judged by several criteria. The most realistic measure from the standpoint of the interviewer is the rate of successfully completed interviews in relation to the number of live people contacted; this is the proportion given as success rates in the last line of the table.

Table 4.4. Percentages of refusal and success in obtaining interviews

	San Diego	Chicago	Midwest	Texas	NY & PA
Respondent not local	42		13	62	26
Interview refused	10	3	14	22	8
Call back later	2	4	5	3	3
Successful interview	1	4	20	9	17
Refusal rate, incl. Non-locals	18	27	27	23	15
Refusal rate, excl. Non-locals	77	27	36	65	29
Success rate, incl. Non-locals	2	36	38	9	31

These variations in refusal and success rates are related to differences in regional histories and population mobility. The greatest differences between regions are in the proportions of non-locals, reflecting the well-known migration patterns in the U.S. towards the sun belt. Chapters 11 and 20 will show that the defining features of the West as a dialect area are more complex and less consistent than for other areas, and the high proportion of non-locals in San Diego is correlated with this situation. Large-scale inmigration to the largest Texas cities is reflected as well in the variable realization of Southern features in that state (Chapter 18).

Finally, we must confront the fundamental question of any sampling procedure: to what extent does the sample represent the population of local speakers? Are the local speakers who refused the interview different linguistically from those who agreed to be interviewed? The early study of New York City included a method of sampling those who refused face-to-face interviews by means of a telephone interview, and found no such bias (Labov 1966, Appendix D), but there is no practical way of re-sampling those who refused the telephone interview. It

is possible that persons with greater linguistic insecurity are more likely to refuse the Telsur interview, or that leaders of linguistic change are more likely to accept it. We have no way to estimate such biases. The major way of assessing the representativeness of the sample is through the regularity of the results, in the form of homogeneity and consistency of isoglosses (Chapter 6).

4.6. Contacting speakers: pinpointing the ideal speaker

As we approached the end of the interviewing, we found that we needed a few more speakers in places where experience had demonstrated that it was difficult to locate speakers of the traditional vernacular. One example is New York City, where the status of preconsonantal /r/ is a crucial issue. /r/-vocalization is waning fast among upper middle class speakers, and we needed to determine its status in the working and lower middle class, where vocalization historically has been very high. Yet finding a white, native, working or lower middle class New Yorker in a city of seven million people by choosing names from a telephone directory is difficult. In a borough where such speakers are most likely to be found, such as Queens, 28 percent of the residents are not native-born Americans, 22 percent are African-American, and 20 percent are Hispanic. In several sociolinguistic studies, it has been found that African-American and Hispanic speakers do not participate in the major sound changes in progress that are the focus of ANAE. The problem of ethnicity can largely be circumvented by selecting names from the telephone directory that are marked for national ancestry as Irish, Italian, German, Slavic, Jewish, or other European nationalities that are well represented in the area of interest. We exclude English names, as those are prevalent among African-Americans, and Spanish names. But the problem of locating a native-born speaker from centrally located social classes remains a difficult one.

The same problem arises in Sunbelt cities such as Atlanta, GA and Dallas, TX. These places are populated largely by native-born Americans, but the rate of inmigration from the North and from the surrounding regions raises a serious obstacle to locating natives of the respective cities. Furthermore, under the pressure of so much outside influence, it becomes even more important to interview speakers who participate in relatively closed social networks and thus are less subject to the leveling influence of imported dialects; these speakers, again, are those from the interior social classes.

It may seem paradoxical that it is difficult to locate speakers with the desired characteristics when the goal is to represent the speech patterns of the community as a whole. But it is not uncommon to find that the main stream of vernacular tradition is obscured by the presence of large numbers of recent arrivals in the adult population. Studies of the formation of new communities (Payne 1980; Kerswill and Williams 1994) have shown that the children of these recent migrants adopt the local vernacular with great regularity, confirming the Doctrine of First Effective Settlement (Zelinsky 1992). The future course of any speech community cannot be traced from the diverse patterns of adults whose children reject their non-local dialect. Thus the original study of New York City was based on 81 of the 700 subjects interviewed in the primary social survey (Labov 1966).

Two strategies for reaching speakers who satisfy these criteria present themselves. One is to make many phone calls and to be very particular about which respondents are interviewed. However, the years of interviewing had demonstrated that it is frustrating to the interviewer to have to make an enormous number of phone calls in order to obtain a satisfactory interview. It is also wasteful, in that each telephone call incurs an expense. Most importantly, we do not systematically elicit the information necessary for classification by social class, occupation and education, until the end of the interview. Asking a respondent for this personal information as part of a screening process would likely produce an intolerably high refusal rate.

An alternative method is to identify neighborhoods in the city where the desired speakers live and to restrict calls to those neighborhoods. The 1990 census reports contain a wealth of detailed information on social characteristics of the population, which is listed by census tract in the series *1990 CPH-3: Population and Housing Characteristics for Census Tracts and Block Numbering Areas*. A census tract is a rather small area, usually having a population of 2,500 to 8,000 and averaging about 4,000. If the interviewer can identify census tracts in which a high proportion of the residents satisfy the necessary criteria, it is likely that a much higher success rate can be attained in reaching the desired speakers. In the CPH-3 set of census reports, the most useful tables for this purpose are Tables 8, 16, and 20, dealing with race, ancestry, and social and labor-force characteristics.

The order in which the tables are consulted depends on the nature of the area under consideration. To locate speakers in New York, the county of Queens was selected. A list was made of all the census tract numbers which satisfied the criterion of 10 percent or less foreign born white persons (Table 20). From that list, those who did not satisfy the criterion that two-thirds of the population should be white (Table 8) were eliminated. Table 20 contains data on only about 225 of the approximately 670 census tracts that are listed in Table 8, so many tracts that would otherwise be candidates for consideration were not reviewed. Returning to Table 20, the census tracts still on the list were examined for the percentage of the population holding a bachelor's degree or higher; those in which the rate was greater than about 20 percent were eliminated. Finally, Table 16 was consulted for the predominant national ancestries of the targeted census tracts. There were eight census tracts that satisfied the criteria well, and ten more that were somewhat marginal.

Obtaining telephone listings for the targeted areas requires further steps. The atlas of the census tracts is consulted to locate the boundaries of the tracts. A further resource is the Census Tract Street Locator on the website of the Census Bureau,[8] which locates streets by census tract and gives the corresponding zip code, as well as other information. From commercially available databases of telephone listings, phone numbers are easily searched by zip code.

Using this extensive preparation, telephone listings of a number of Jewish and Irish names were printed for parts of Queens, NY. In three sessions, the phone was dialed 19 times. In eleven cases, no one answered the phone. Of the eight people contacted, four refused to be interviewed and two were not native New Yorkers. Two highly successful interviews were completed with women having precisely the desired social histories. In addition, an arrangement was made to interview the daughter of one of the women a few days later. This and subsequent applications of the method proved to offer a very high rate of return for the time invested.

4.7. Age and gender distribution of the sample

The sampling methods discussed above produced a range of subjects from age 12 to 89. It is not important for the goals of ANAE that all ages be equally represented; as noted above, emphasis was put on the early adult years. It is important that the age range be roughly equivalent for all geographic regions. If not, a

8 The web address for this utility is http://tier2.census.gov/ctsl/ctsl.htm. This and other Census Bureau databases are listed at http://tier2.census.gov/dbappweb.htm.

constant difference in age-grading in the population might appear as a regional difference. Table 4.5 shows the age distribution of the Telsur sample in decades for seven regions.[9] A graphic comparison of the five major regions appears in Figure 4.1. All regions show a heavy concentration in the young adult period, 20 to 40 years. The major differences that appear are differences between the South – with more older subjects. and a modal range of 40 to 49 – and the West, with a modal range of 20 to 29. The three other regions are intermediate, with modes in the 30 to 39 range.

Table 4.6 and Figure 4.2 show the distribution of the sample by gender. The excess of women over men is apparent, and is also the parallel distribution across decades of age. The ratio of women to men is 1.7:1. The chief departure from this is in the concentration of women in the 20 to 29 age range as against the relatively high proportion of men in the decade from 30 to 39. For the decade from 20 to 29, the ratio of women to men is 1.8:1; for age 30 to 39, it is 1.2:1.

Table 4.5. Age distributions of Telsur speakers

Age	Canada	ENE	Midland	Mid-Atl	North	South	West	Total
10–	4	1	13	6	12	26	13	75
20–	10	1	30	4	26	24	24	119
30–	14	0	52	5	57	34	20	182
40–	6	2	36	8	46	47	11	157
50–	3	1	19	6	33	14	11	87
60–	1	1	20	7	17	21	8	75
70	1	2	14	4	20	15	4	60
Total	39	8	184	40	211	181	91	762
Mean	35	48	41	45	44	41	47	42

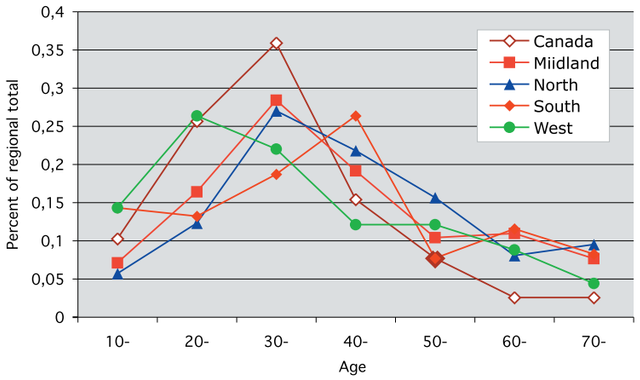

Figure 4.1. Age distribution of Telsur speakers in the five largest regions

Table 4.6. Distribution of Telsur speakers by gender and age

	Age by decade								
	10–	20–	30–	40–	50–	60–	70–	80–	Total
Female	54	80	119	87	55	48	28	9	480
Male	21	41	65	70	34	28	19	4	282
Total	75	121	184	157	89	76	47	13	762

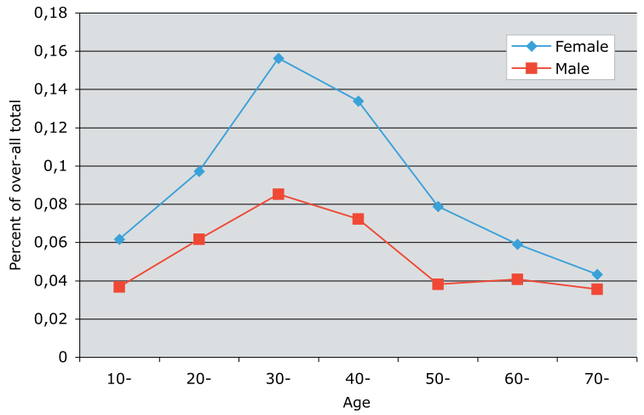

Figure 4.2. Distribution of Telsur speakers by gender and age

4.8. The Telsur interview

The original interview questionnaire was designed for the six-state pilot project area, which encompassed parts of three dialect areas and thus was written to include most of the variables that are of interest in North American English. The same form was used in the next phase of data collection in the fifteen-state area comprising the agricultural and industrial heartlands of the United States, corresponding to most people's idea of the Midwest. With the expansion of the survey to all of English-speaking North America, variants of the original inter-

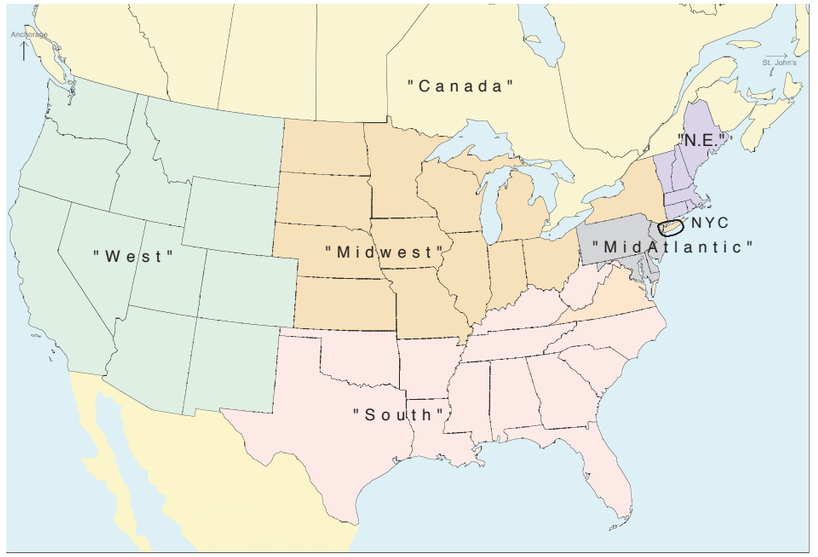

Figure 4.3. Regional variants of the Telsur interview form

9 See Chapter 11 for the distinction between dialects and regions. The region is the larger unit under which dialects are grouped.

view schedule were introduced, resulting in six forms of the questionnaire. They all share most of the same variables, but there are a number of modules which are included only in certain forms to tailor them to the different regions of North America: South, West, Mid-Atlantic, New England, Midland, and Canada. Since the dialect boundaries of Chapter 11 had not yet been established, state boundaries were used in selecting the variant forms of the interview schedule.

Following the introduction described in Section 4.5 above, which establishes that the respondent is a native of the community where she or he lives and that recording is permissible, the interview is divided into six sections.

1. *Demographic information*. Information on the native and local status of the respondent: place of birth, complete residence history, father's and mother's places of birth, and languages spoken.
2. *Spontaneous speech*. The largest portion of spontaneous speech is obtained from a discussion of recent developments in the city, the state of the downtown area, and travel outside the city. If a topic of special interest to the speaker is raised, it is pursued to the fullest extent possible. Speakers often talk about their jobs, hobbies, or other interests in this portion of the interview.
3. *Word lists*. Sequences of words that do not require reading: counting, days of the week, articles of clothing, breakfast foods, and others.
4. *Linguistic variables*. This section includes, first, *minimal pairs* in the form of judgments on rhyming (*hot*/*caught*) or "same" versus "different" (*dawn*/*Don*). In each case, the respondent is prompted to say words described but not pronounced by the interviewer (e.g. *What is the opposite of cold?* as the prompt for *hot*), then asked to give a judgment on contrast or identity of the pairs of sounds. The respondent is then asked to say the two words again. This procedure was designed to elicit two instances of production as well as a judgment of each contrast under study.

   Spontaneous pronunciations of crucial lexical items are obtained through the use of the *semantic differential* technique (Labov 1984), which uses questions about differences in meaning between two words, such as *cot* vs. *bunk* and *pond* vs. *pool*. Subjects put considerable effort into answering these questions, producing several highly stressed tokens of each word without attending to their pronunciation. Previous research shows that the use of the variables in the semantic differential approaches the values of spontaneous speech quite closely (Labov 1989).

   A series of grammatical variables was included. They were introduced with the following protocol: *I'd like to ask you to tell me what you think of a few sentences I'm going to read you. These are sentences that sound fine to people in some parts of the country but a little strange to people in other parts of the country. For each sentence I read you, I'd like you to tell me whether you think it sounds like something you could say yourself, or something you've heard around your area but you wouldn't say, or something you've never heard before*.

   Responses to grammatical features were coded on a three-point scale: 1 "could say yourself", 2 "heard but wouldn't say", and 3 "never heard".

   A small number of *regional vocabulary* items were included in the Telsur interview form. These are of the simple form, "What do you call _____?" where the interviewer gives a definition of the variable in question. For example, *couch*/*sofa* was elicited with the question, "what do you call a large piece of furniture that seats three people?".
5. *Demographic background*. More detailed information on the demographic background of the subject is gathered, including occupation, education and national ancestry.
6. *Continuation*. The final section was the request for the respondent to continue participation in the research by reading a word list, which is to be mailed to the speaker. This required that the speaker provide his or her name and address. A small number of speakers declined to give this information or refused to participate in this second part of the interview, and some asked for additional reassurance that they would not be subject to solicitations from salespeople or other unwanted callers. Most speakers readily agreed to the follow-up interview and greeted the interviewer as a familiar acquaintance when he or she called again.

   The interview form also contains suggested answers to questions that subjects often ask: "So what's this study all about again?"; "Why is this important?"; "Who is paying you to do this?"; "What are you going to do with the results?"; "Can I see some of your results?". See Appendix 4.3 for these suggested answers.

The duration of the Telsur interview averages about 30 to 45 minutes. The total volume of speech obtained proved to be more than we expected from the previous results of Hindle (1980). In the acoustic analysis of vowel systems, the mean number of vowel tokens was 306. Only 10 percent had fewer than 200 tokens.

4.9. The second interview

The second interview is designed to obtain more specific information on lexical distribution through the reading of a word list and more detailed information on contacts outside the community. Respondents are asked to read a full-page list of words, which is sent to them in the mail after the first interview. The word list is designed to cover the areas of variable contrast and variable lexical distribution in the speaker's region. A sample word list is given in Appendix 4.4. The second interview also goes more deeply into the patterns of travel, friendship, kinship, and communication that relate the respondent to other cities of interest.

4.10. Impressionistic coding

The first stage of analysis is the transcription of all demographic data, recording of lexical choices and judgments of syntactic constructions, and the coding of the speaker's pronunciation of diagnostic words in the formal part of the interview. Like the interview questionnaire, the impressionistic coding form is tailored to the speaker's region. For the phonological variables, the analyst records the speaker's judgments of "same" and "different", and then enters a fine-grained phonetic transcription of the speaker's pronunciation. Finally, the analyst codes the result in a four-cell table:

		Judged: Same	Judged: Different
Pronounced	Same	a	b
	Different	c	d

Cell (a) represents full merger, and cell (d) registers a clear distinction. Cell (b) is usually the result of the mistaking of orthographic differences for pronunciation differences. Cell (c) is the case of near-mergers, where speakers consistently make a difference between two sounds but do not judge them as different and do not use the difference for semantic interpretation (LYS; Milroy and Harris 1980; Harris 1985; Di Paolo and Faber 1990, 1995).

4.11. The socio-economic index

Occupation is widely viewed as the best single determiner of social class. Unlike other factors such as income and house value, it is an acceptable subject of inquiry and conversation between strangers. Ratings of occupational prestige, beginning with those published for 90 occupational titles by the National Opinion Research Council (NORC) in 1947, have been widely used for the ranking of occupations in terms of social standing.

In 1950 the Census Bureau began collecting data on income and education for incumbents of certain occupations, of which 270 were listed in 1950. Duncan (1961) addressed the need for a ranking of the social status of occupations by calculating a Socio-economic Index (SEI) – intended to mimic but not replicate the NORC occupational prestige score – for all 270 occupations listed by the Census. He accomplished this by performing a multiple regression of NORC prestige ratings on the income and educational levels for those occupations that were common to both the NORC and the Census listings and then extrapolating to occupational titles listed by the Census but not included in the NORC study.

Duncan's work has been updated, most recently in 1989. The NORC has reported prestige ratings (Nakao and Treas 1989) for the 503 occupational titles on which the Census Bureau gathered data in 1980, and they also report SEI assignments for those occupations (Nakao and Treas 1992), using the methods developed by Duncan, with adjustments made for current levels of educational attainment and income. In the assessment of speakers for the Atlas, it was observed that the SEI has the advantage of taking into account not only the prestige assigned to occupational titles by a sample of raters but also the objective and additional important factors of income and education associated with the respective occupations. Therefore, the calculated SEI scores are used to rank the Atlas speakers, rather than the raw Occupational Prestige scores.

Problems in carrying out the task of assigning an SEI to each speaker stem mainly from two sources: inadequate data elicited from the speaker and difficulty in matching the speaker's occupation to one of the 503 occupations in the NORC/Census list. Some speakers, queried about their occupations, give answers such as "I work for Raytheon" or "I work in an office". The interviewer did not always pursue the subject in order to determine an appropriate occupational title for the speaker. Women who report themselves as homemakers are appropriately assigned the SEI corresponding to their husbands' jobs, but often that information was not obtained. High school and junior high school students are assigned the SEI corresponding to the family's breadwinner's occupation, so the interviewer had to be careful to elicit this information. College undergraduates and graduate students are a more difficult problem: they cannot properly be assigned the SEI associated with their family's breadwinner, but it is incorrect to assign them to an occupation which they have not yet entered, associated with their field of study. When clear information on occupation is obtained, it is still often difficult to decide how the information given by the speaker best matches the occupational titles listed by the NORC survey. For all speakers where an SEI assignment is made, the Census category number is also recorded, so that the assignment can be reviewed and revised if necessary.

Appendix 4.1. Zones of influence, central cities, and UA population

	Zone	Zone pop.	UA pop. 1990	Per cent in UA	Zone abbrev.	Pop zone area sq mi
Alabama	Birmingham	2,395,674	621,703	25	Br	24,978
	Mobile	772,068	301,197	39	Mb	11,820
	Montgomery	735,752	210,060	28	Mt	12,996
Alaska	Anchorage	550,043	221,745	40	An	152,040
Arizona	Phoenix	2,754,669	2,006,568	72	Ph	91,983
	Tuscon	910,559	579,155	63	Tu	22,016
Arkansas	Little Rock	2,031,485	305,498	15	LR	47,361
California	Bakersfield	543,477	302,823	55	Bk	8,149
	Fresno	1,183,272	453,186	38	Fr	15,859
	Los Angeles	12,557,743	11,402,955	90	LA	26,142
	Modesto	597,381	231,045	38	Mo	5,767
	Riverside-SanBrndino	2,588,793	1,169,839	45	RSB	27,408
	Sacramento	2,043,240	1,097,313	53	Sa	27,520
	San Diego	2,607,319	2,348,106	90	SD	8,760
	San Francisco	5,871,470	3,629,864	61	SF	34,545
	San Jose	1,764,008	1,434,803	81	SJ	3,143
Colorado	Colorado Springs	441,755	353,026	79	CS	4,227
	Denver	3,199,682	1,517,803	47	Dn	173,268
Connecticut	Bridgeport	827,645	414,254	50	Br	665
	Hartford	1,655,252	546,074	32	Hr	3,722
	New Haven	804,219	451,486	56	NH	629
Delaware	Wilmington	737,515	450,080	61	Wl	2,424
District of Colum.	WashingtonDC	4,976,573	3,363,047	67	DC	15,522
Florida	Ft. Lauderdale	1,255,488	1,238,109	98	FL	1,220
	Jacksonville	1,420,761	738,593	51	Jc	14,673
	Miami	2,613,305	1,914,689	73	Mm	7,321
	Orlando	2,113,451	887,968	42	Or	7,630
	Pensacola	531,720	253,717	47	Pn	4,529
	Tallahassee	608,901	156,072	25	Tl	11,145
	Tampa	3,622,316	1,708,966	47	Tm	14,652
	West Palm Beach	1,177,580	795,033	67	WPB	5,955
Georgia	Atlanta	4,773,058	2,157,344	45	At	31,669
	Augusta	526,695	286,205	54	Ag	7,119
	ColumbusGA	462,445	220,651	47	CGA	6,632
	Savannah	620,623	198,609	32	Sv	9,287
Hawaii	Honolulu	1,108,229	632,498	57	Hn	6,443
Idaho	Boise	809,096	168,056	20	Bs	67,564
Illinois-Iowa	Quad Cities	556,615	264,181	47	QC	5,397
Illinois	Chicago	9,262,154	6,793,132	73	Ch	21,396
	Peoria	603,433	242,547	40	Pe	7,923
	Rockford	450,746	207,693	46	Ro	3943
Indiana	Evansville	631,670	182,908	28	Ev	9,285
	Fort Wayne	763,258	248,686	32	FW	6,190
	Indianapolis	2,893,819	914,426	31	In	19,217
	South Bend	817,583	237,481	29	SB	4,196
Iowa	Des Moines	2,364,603	293,446	12	DM	51,275
Kansas	Wichita	1,242,284	338,562	27	Wi	60,722
Kentucky	Lexington	1,277,067	221,116	17	Lx	16,186

	Zone	Zone pop.	UA pop. 1990	Per cent in UA	Zone abbrev.	Pop zone area sq mi
	Louisville	2,085,014	755,013	36	Ls	21,678
Louisiana	Baton Rouge	1,250,108	365,647	29	BR	12,136
	New Orleans	1,843,595	1,040,300	56	NO	11,328
	Shreveport	1,001,158	256,727	25	Sh	19,883
Maine	Bangor	453,541	61,374	13	Bn	24,965
	PortlandME	774,387	120,271	15	PME	8,299
Maryland	Baltimore	2,620,641	1,890,518	72	Ba	5,720
Massachusetts	Boston	4,879,886	2,774,717	56	Bo	5,943
	Springfield	812,322	532,341	65	Sp	2,853
	Worcester	709,705	315,698	44	Wr	1,581
Michigan	Ann Arbor	282,937	221,766	78	AA	725
	Detroit	6,552,441	3,697,424	56	Dt	42,232
	Flint	574,997	326,452	56	Fl	1,850
	Grand Rapids	1,024,815	436,033	42	GR	6,177
	Lansing	432,674	265,151	61	Ln	1,713
Minnesota	Duluth	389,042	122,945	31	Du	22,643
	Minneapolis	4,407,548	2,079,255	47	Mn	74,436
Mississippi	Jackson	1,524,375	289,199	18	Jk	29,231
Missouri	Kansas City	2,794,595	1,275,083	45	KC	39,830
	SpringfieldMO	590,008	159,594	27	SMO	14,637
	St. Louis	4,161,434	1,946,047	46	SL	44,618
Montana	Billings	374,142	88,206	23	Bl	87,675
	Great Falls	278,941	63,531	22	GF	56,766
	Missoula	212,007	57,006	26	Ms	24,580
Nebraska	Lincoln	309,515	192,578	62	Ln	5,976
	Omaha	1,464,098	544,273	37	Om	77,519
Nevada	Las Vegas	764,359	697,078	91	LV	40,499
	Reno	440,792	213,835	48	Rn	71,091
New Hampshire	Manchester	723,764	115,105	15	Mn	7,172
New Jersey	Trenton	325,824	298,939	91	Tr	228
New Mexico	Albuquerque	1,159,298	496,833	42	Aq	87,355
New York	Albany	1,220,151	509,196	41	Al	11,308
	Binghamton	525,354	159,059	30	Bn	6,610
	Buffalo	1,638,215	953,867	58	Bf	8,593
	New York	17,647,736	16,044,493	90	NY	11,103
	Rochester	1,238,165	620,214	50	Rc	5,486
	Syracuse	1,617,775	388,411	24	Sy	15,,638
North Carolina	Asheville	524,471	110,658	21	As	6,434
	Charlotte	2,044,904	455,386	22	Ct	11,312
	Durham	400,368	205,439	51	Dr	2,355
	Fayetteville	620,915	241,291	38	Fy	5,814
	Greensboro-/Wnstn-Salem	1,442,014	379,022	26	Gr	8,400
	Raleigh	1,846,799	305,820	16	Rl	15,555
North Dakota	Bismarck	172,140	66,607	38	Bk	26,662
	Fargo	420,712	121,351	28	Fr	28,910
	Minot	139,742	34,544	24	Mi	19,251
Ohio	Akron	791,885	527,780	66	Ak	1,908
	Canton	494,281	244,637	49	Cn	1,964
	Cincinnati	1,980,761	1,212,260	61	Ci	6,854
	Cleveland	2,104,587	1,677,554	79	Cl	3,156
	Columbus	2,410,609	944,744	39	COH	15,137
	Dayton	1,173,945	613,314	52	Dy	4,009
	Lorain-Elyria	404,145	224,007	55	LE	1,271
	Toledo	1,097,126	489,469	44	Tl	5,463
	Youngstown-Warren	697,141	361,366	51	YW	1,960
Oklahoma	Oklahoma City	2,045,951	784,367	38	OC	54,309
	Tulsa	1,232,648	475,044	38	Tu	15,328
Oregon	Portland-Vancouver	3,183,569	1,171,834	36	PV	93,817
Pennsylvania	A'town-Bthlm-Easton	1,271,505	410,244	32	ABE	3,743
	Erie	466,172	177,661	38	Er	3,427
	Harrisburg	1,394,937	293,442	21	Hr	6,736
	Philadelphia	5,802,466	4,222,377	72	Ph	6,322
	Pittsburgh	3,911,581	1,680,112	42	Pt	19,466
	SCollege-Williamsprt	320,804	118,946	37	SCW	4,397
	Scranton/Wilkes-Barre	684,514	388,610	56	SWB	3,476
Rhode Island	Providence	1,003,464	845,725	84	Pr	1,207
South Carolina	Charleston	624,369	393,302	62	CSC	5,733
	Columbia	1,266,203	328,148	25	Cl	12,743
	Greenville	1,015,409	248,525	24	Gv	5,771
South Dakota	Aberdeen	88,260	24,927	28	Ab	16,987
	Rapid City	227,134	61,077	26	RC	42,434
	Sioux Falls	430,693	100,851	23	SF	27,441
Tennessee	Chattanooga	747,891	296,882	39	Cg	7,171
	Knoxville	1,441,478	303,713	21	Kn	11,,822
	Memphis	2,190,209	825,425	37	Me	28,362
	Nashville	1,701,163	573,154	33	Nv	17,659
Texas	Amarillo-Lubbock	858,350	345,913	40	AL	52,346
	Austin	1,190,558	563,025	47	Au	11,921
	Corpus Christi	470,406	269,878	57	CC	10,617
	Dallas-Ft. Worth	6,363,453	3,198,199	50	DFW	107,873
	El Paso	897,938	571,079	63	EP	39,242
	Houston	5,358,382	2,902,449	54	Ho	42,248
	San Antonio	2,575,411	1,128,966	43	SA	44,801
Utah	Ogden	200,343	259,148	129	Og	6,970
	Provo-Orem	269,407	220,560	81	PO	5,538
	Salt Lake City	1,265,185	789,720	62	SL	69,100
Vermont	Burlington	369,128	86,873	23	Bl	6,221
	Rutland	157,785	18,230	11	Ru	2,717
Virginia	Norfolk	1,701,413	1,323,039	77	Nr	9,155
	Richmond	1,439,553	590,352	41	Rc	14,713
	Roanoke	934,433	178,384	19	Rn	11,268
Washington	Seattle	3,727,330	1,743,796	46	Se	35,857
	Spokane	1,006,349	278,939	27	Sk	50,644
West Virginia	CharlestonWV	1,063,487	393,302	36	CWV	16,337
	Huntington-Ashland	431,583	169,323	39	HA	4,405
Wisconsin	Madison	823,218	244,335	29	Md	10,747
	Milwaukee	3,627,343	1,226,060	33	Ml	33,105
Totals	Count	145				
	Sum	248,709,873				
	Average	1,715,241				

Appendix 4.2. Sample interview form

A TELEPHONE SURVEY OF SOUND CHANGE
IN PROGRESS IN NORTH AMERICAN ENGLISH
Linguistics Laboratory, University of Pennsylvania

– MID-ATLANTIC VERSION –

0. Approach

Hi, my name is ____. I'm calling from the University Pennsylvania in Philadelphia. We're doing research on communication between people from different parts of the country, so we're looking for people who grew up in one place to help us by telling us a little about how people say things in your area. Did you grow up in ____? *If yes*: Can you take a few minutes now to answer some questions?

(*If speaker is hesitant* People across the country are talking to each other more and more, and at the same time we know that local accents are getting more different, in spite of the fact that we all watch the same TV programs. We want to find out how people talk in each region of the country and whether local ways of talking are changing in any way.)

In order to be able to keep track of everything you can tell us, I need to be able to make a tape recording of this conversation. Is that all right with you? (*If informant is hesitant*: I can assure you that this information is used only by our research group for our reports about general trends in American English, and no information identifying individuals is ever released. *If still hesitant*: If we come to a question you don't think you want to answer, just tell me and we'll skip it. I don't think you'll have a problem with any of the questions I'm going to ask you.)

Turn tape recorder on and tell informant you have done so.

1. Residential and language background

Confirm place of birth: Now, were you actually born in ____?
Full residence history and approximate ages in each location.
Where mother born.
Where father born.
Languages spoken in family while growing up.
Second language learning.

2. Conversation

2.1. Communication experience and travel

– Have you noticed that people in different parts of the country talk differently from yourself? What sort of differences have you noticed?
– Have you ever had a problem understanding people in other parts of the country because of their accent or because of different words they used?
– Where have you travelled?

2.2. Local color

– What's your town like? Would you say it's a nice place to live?
– What do most people do for a living in your area?
– Are there any big local industries?
– Is the economy doing OK?
– Have there been lay-offs in your area?
– Are people moving in or moving out?
– Are there lots of new houses going up?
– What do you do for fun on the weekends?
– What sports teams do you support?
– What newspapers do you read?
– What other cities do you go to for recreation or shopping?
(*Pick 2 or 3 largest cities in vicinity and explore the choice between them for different activities*.)

2.3. Downtown

– Does your city have a good downtown section?
– Are businesses moving in or out of downtown?
– Are there still some big department stores downtown?
– Are there any new buildings downtown?
– Do people hang out downtown after 5:00 on a weekday?
– Are there things to do downtown?
– Is it safe to walk around downtown at night?
– Can you find parking downtown? Is it expensive?
– Is the city doing anything to make people want to go downtown?
– Do you shop downtown or at the malls? Why?

3. Word lists

Now I'm going to ask you to say a few things for me that will help us with our study.

(a) First of all I'd like you to count for me from 1 to 10.
(b) And would you please say the days of the week?
(c) And now could you please list as many articles of clothing as you can think of.
If necessary, elicit:
– PANTS: what's another word for slacks?
– COAT: what's another word for jacket? (longer, dressier)
– HAT/CAP: what would you wear on your head?
– BOOTS: what does a construction worker or a cowboy wear on his feet?
(d) And now could please tell me what sort of things people around your area eat for breakfast, especially if they go out for a big breakfast on the weekend?
If necessary, elicit:
– EGGS: What are omelettes made of?
– BACON/SAUSAGE/HAM: What meats do people eat with eggs?
– TOAST: What do you put butter or jam on?
– COFFEE/TEA: What do people drink with breakfast?
Are there any special local foods or dishes that your area is known for?
(e) And finally could you list as many farm animals as you can think of?
If necessary, elicit:
– DUCK(S): what (other) kinds of bird might you find on a farm?

4. Formal elicitation of linguistic variables

Now I need you to say certain words, but I don't want to say them first because that might influence the way you say them. So I'll ask you questions that get you to say the words and then we'll talk about whether certain words sound the same or different to you. OK? (It's not a test or anything; it's just a way of getting you to say certain words. I'll give you as many clues as you need.)

4.1. (o-oh)

(a) If a mother deer is called a doe, what would you call a baby deer? [FAWN]
(b) What's another word for sunrise, or for the first part of the day when the sun's just coming up? [DAWN]
(c) Do those words rhyme? (Could you use them to rhyme in a poem?)
(d) Can you think of any boy's names that rhyme with those words? [DON, RON, JOHN?]

If necessary, elicit:
– DON: What's the first name of Walt Disney's famous duck?
What's short for that?

(e) Does that name sound the same as the word for *sunrise* you just said? (If someone said those two words to you over the phone, could you tell them apart?)
(f) Can you say them again for me? (*If necessary*: which one was first?)
(g) What's another boy's name that starts with D and ends with N? [DAN]

(a) What's the past tense of *catch*? (Like if today I catch the ball, yesterday I ...?) [CAUGHT]
(b) What's the opposite of *cold*? [HOT]
(c) Do those words rhyme?
(d) Can you say them for me one more time?

(a) What's the opposite of *shorter* (if you're talking about the height of people)? [TALLER]
(b) How much money do four quarters make? [DOLLAR]
(c) Do those words rhyme?
(d) Can you say them for me one more time?

(a) What's the opposite of *off*? [ON]
(b) What's the opposite of *up*? [DOWN]

4.2. Semantic differentials (1)

Now I have a few questions about the meanings of different words. Tell me, in your opinion,
(b) What's the difference between a HOME and a HOUSE?
(d) What's the difference between a DECK and a PORCH?
(e) What's the difference between to SIT and to SET?

4.3. Lexicon

(a) What's the general term you use for a carbonated beverage in your area? [POP, SODA, COKE, etc.] (*If unsure*: if you were going to buy a can of Coke or Pepsi or Sprite out of a machine, what would you call the machine?)
(b) What do you call it when you prepare meat outside over a charcoal fire in the summertime? [GRILL(ING) (OUT), BARBECUE, COOKOUT]
(c) Do grilling and barbecuing mean the same thing? *If no*: what's the difference? [SAUCE]
(d) *If not already answered:* What kinds of things would you barbecue? Grill?
(e) What do you call a large piece of furniture that seats three people? [COUCH, SOFA, etc.]
(f) What do you call the top part of a house, that keeps the rain out? [ROOF]

4.4. (i-e/_N)

(a) What would you use to sign a check with? [PEN]
(b) What would you use to fasten a cloth diaper? (A safety ...) [PIN]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(*If* pin *and* pen *are close or the same*:
(a) If you gave a book to Mary you'd say I gave it to *her*; if you gave it to John you'd say I gave it to ... [HIM]
(b) What do you call the bottom part of a dress where it's folded up and sewn in place? [HEM].
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

4.5. (tense ~ lax contrasts before /l/)

(a) What's the opposite of *empty*? [FULL]
(b) What's another word for an idiot or a stupid person? (Begins with F as in Frank). [FOOL]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(*If* full *and* fool *are close or the same*:
(a) What's a place where you go swimming in the backyard? [POOL]
(b) What's the opposite of *push*? [PULL]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

(a) What's a word for a little mountain? [HILL]
(b) What do you call the back part of the bottom of your foot? [HEEL]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(*If* hill *and* heel *are close or the same*:
(a) What's a word for the skin of an orange? [PEEL]
(b) What's the little thing you swallow when you take aspirin? [PILL]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

4.6. (oh-ow/_r)

(a) What kind of animal runs in the Kentucky Derby (what does a cowboy ride)? [HORSE]
(b) What do you call the way you feel when your throat is kind of scratchy and sore so you can't talk very well? [HOARSE]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(*If* horse *and* hoarse *are close or distinct*:
(a) What do you call the first part of the day, before noon? [MORNING]
(b) When someone is grieving because somebody close to them has just died, you say they're in ... [MOURNING].
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

4.7. (æ/_g,d) -- Semantic differentials (2)

(a) What's the difference for you in meaning between a BAG and a SACK?
(b) What's the difference between a LABEL and a TAG?
(c) What's the difference between a BAD person and an EVIL person?
(d) What's the difference between being UNHAPPY and being SAD?

4.8. Aspirated glides -- (hw, hj)

(a) What's a great big animal like a fish except it's a mammal (lives in the ocean and spouts water)? [WHALE]
(b) What do you call a sound like a siren or a baby's cry, also starts with W? [WAIL]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(a) If someone can laugh at a good joke, you say he has a good sense of ... [HUMOR]
(b) What's a word that means very, very big, or enormous, starts with H? [HUGE]

4.9. (ey-e-æ/_rV)

(a) In the nursery rhyme, who's the girl who had a little lamb? [MARY]
(b) What's a word that means happy, that people say when they greet one another at Christmas? [MERRY]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.
(e) When a man gets down on one knee and pops the question to the woman he loves, what does he say? Will you ... [MARRY]
(f) Does that sound like the word people say with Christmas?
(g) Say those two again and tell me which one's which.

4.10. (uw-juw/[+cor]_)

(a) If you're getting married, what do you say when you're asked if you take the other person to be your wife or husband? [DO]
(b) What do you call the moisture that's on the grass in the early morning? [DEW]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

4.11. Southern Shift items

(a) What's a hot drink you might put milk, sugar or lemon in? [TEA]
(b) What's a small, round green vegetable that comes in a pod? [PEA]
(c) What do 24 hours make (what are there seven of in a week)? [DAY]
(d) What's the letter in the alphabet after J? [K]

4.13. R-lessness module

(a) What's the past-tense of *fight*? [FOUGHT]
(b) What do you call a military outpost, like in the Old West, with wooden walls and towers? [FORT]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

(*If* fought *and* fort *are close or the same*:
(a) What's the sound a lion makes? [ROAR]
(b) How do you describe meat or vegetables before they've been cooked? [RAW]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

(a) What's the organ in the body that pumps blood? [HEART]
(b) What's the opposite of *cold*? [HOT]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.

(*If* heart *and* hot *are close or the same*:
(a) What's the shortest nickname for *Robert*? [BOB]
(b) What's a nickname for *Barbara*? [BARB]
(c) Do those words sound the same to you?
(d) Say them again for me and tell me which one's which.)

5. Syntactic variables

Now I just have one more section of language questions for you. In this section I'd like to ask you to tell me what you think of a few sentences I'm going to read you. These are sentences that sound fine to people in some parts of the country but a little strange to people in other parts of the country. For each sentence I read you, I'd like you to tell me whether you think it sounds like something you could say yourself, or something you've heard around your area but you wouldn't say, or something you've never heard before. OK? So here's the first one:

(a) What if there were crumbs on the kitchen floor and someone said, "The floor needs swept"?
(b) What if a mother said to her child, "Your hair needs cut"?
(c) What if you were looking at the price of a new car and someone said, "Boy, cars are sure expensive anymore!"?
(d) What if someone said, "It's real hard to find a good job anymore"?
(e) What if someone said, "I used to watch football, but anymore I watch baseball"?
(f) What if someone asked you, "I'm going to the store; d'you wanna come with?"
(g) What if someone asked, "Do you want for me to go downtown today?"
(h) What if someone asked, "Would you like for me to pick up some milk on the way home?"

6. Personal history/demographic data

Those are all the language questions I have for you. Now I just need to ask you a couple more things so that we can place you properly in our sample.

(a) What year were you born?
(b) Where did you go to high school?
(c) What were the main racial and ethnic groups in your school? (*approx. %, if appropriate*)
(d) What's your own family's background in terms of national ancestry? (→ *conversation*?)
(e) What is/was your father's occupation? Your mother's? (→ *conversation*?)
(f) Did you take any schooling beyond high school? What, where?
(g) What's your occupation? (→ *conversation*?)
– Do you enjoy your job?
– What exactly does it involve?
– So tell me, since you're an expert in this, I've always wondered . . .?
– *etc., as appropriate*.

7. Continuing contact

There's just one other thing I'd like to ask you to do. As you can tell, we try to get everybody we talk to to say certain words and the easiest and quickest way to do that is to mail out a list of words that people can read back to us over the phone, which takes about five minutes. If I mailed you a wordlist and then called you back in a couple of weeks, do you think you could spare five minutes to read me the list over the phone? *If yes*: Great, then I'll just need to get your name and address so I can send you the list. ... What would be a good time to get hold of you?

Well, once again, my name is ____, and I'm at the University of Pennsylvania in Philadelphia, and I'd like to thank you very much for the time you've taken to do this interview. You've really been a big help!

8. Answers to closing questions

Q: So what's this study all about again?
A: This is a survey of changes in the way American English is spoken across the country. We're interested in finding out what changes are going on in different regions and how fast they're progressing. For instance, one of the things I was asking you about was how you said words like *hot* and *caught*, or *sock* and *talk*. This is one of the major differences between the way people talk in different parts of the country. Most people in the West say those words the same, as do people in Canada and in a couple of other areas (Pittsburgh and Boston), whereas people in the South, the Midwest and the East mostly say them different. We want to know where the borders are between these areas and whether they are

shifting: our research suggests that the area where people say *hot* and *caught* the same may be slowly expanding.

Q: Why is this important?
A: It's important for several reasons. First, it's important to linguists who want find out more about the way language changes. (Like how did the English language evolve from Old English to the language of Shakespeare to the language of today, and why do Americans talk differently from British people?) Second, it's important to people who study dialects, because while major European countries like Britain, France, and Germany have national maps of linguistic variation the U.S. does not. Our project is the first attempt to study differences in the sounds of regional speech across the whole country. Third, it's important in developing more effective teaching methods, either in teaching English to adults or in teaching reading and spelling to children. (These strategies need to be sensitive to dialect variation, such as whether or not children will make a difference between *pin* and *pen*.) Fourth, it's important to the speech technology industry, because if computers are going to be taught how to understand human language, they have to be able to cope with different dialects. (Example: a computer at the phone company that needs to understand callers from one area who say *Don* and *Dawn* differently and callers from another area who say them the same.) We can provide some of the information that the computer designers need to create effective speech recognition technologies.

Q: Who is paying you to do this?
A: Our work is supported by a combination of public and private sector funding. We have grants from the National Science Foundation and the National Endowment for the Humanities with matching funds from a telecommunications technology company called Bell-Northern Research.

Q: What are you going to do with the results?
A: Eventually, we're working towards the publication of an atlas of American English, which will include a series of maps showing how people talk in different parts of the country. In the meantime, we'll be publishing papers on various aspects of our research in academic journals and making presentations at conferences.

Q: Can I see some of your results?
A: Certainly. I'd be happy to send you a couple of maps showing some of our results so far.

Appendix 4.3. Sample word list

The following word list is in analytic form – that is, words are grouped according to the phonemes that are being studied or the sets of phonemes or allophones under examination. The word list that is mailed to subjects is a randomized list of these words with no such structure.

The sample list in this appendix is prepared for subjects from the Mid-Atlantic dialect region. Sections modified or introducted for this geographic region are indicated with a dotted border, with words of particular interest in red.

The Mid-Atlantic word list includes an expanded list of short-*a* words for tracing the intricate pattern of the short-*a* split into /æ/ and /æh/. It also includes an extended section on contrasts before intervocalic /r/, examining the contrast between *furry* and *ferry*, *hurry* and *merry*, as well as other vowels. There is an elaborated section on contrasts between *moor* and *more*, *lure* and *lore*, which are merged for most speakers in this area. Words with /ay/ before voiced and voiceless finals are focused on, since a rapid increase in "Canadian raising" before voiceless consonants has been discovered. The lists of /aw/ words is expanded, to trace the strong fronting and raising of the nucleus of that phoneme. Back

vowels before /l/ are included, to establish the contrast between the back position of these words and the strong fronting of others. (*Pal* and *Hal* are included since in this area, these words are often homonymous with *Powell* and *howl* and, with /l/-vocalization, with *pow* and *how*.) A special list of words with two /r/s is added to trace the pattern, specific to this area, of r-vocalization in dissimilating environments, though /r/ in codas is normally constricted.

Telsur WL PA

/æ/
batch
cat
bat
mat
cap
sat
sad

cab
bad
badge
mad
bad
glad
black
bag
laugh
staff
math
bath
ask
cash
hash
man
ant
aunt
ham
camera
Janet
planet
began
thing
sing
sang
pal*
Hal*
alley
Alice
personality

/o, ah, oh/
block
bomb
calm
palm
pajama
father
pa

paw
cot
caught
Don
Dawn
e/
get
bet
bed
leg
beg
egg
step
set

aspirin
after
asterisk
alas
adze
tin can
I can
ran
swam
planning
classics
classify
Lassie
Annie
gas
math

/ay, aw, oy/
ice
sight
fight
eyes
side
tie
file

fire
time
sign
my

out
about
mountain
loud
mouse
house
down
downtown
now

Unstressed vowels
parted
rabbit
Mexico

/iy, ey, uw, ow/
bee
see
Kay
say
bay
go
hoe

do*

/uw, iw/
do*
dew
stew
goof
tooth
toot
hoop
shoot
noose

/Vg/
fish
bush
vision
measure

/_lC/
film
milk

before /l/
tool
fool
bowl
goal
cold
old
pal*

Hal*

Dissimilating
Charlie
sorcerer
forward
ordinary
quarter
extraordinary
corner
order

Distinctions
/uwr, owr/
moor
more
lure
lore
boor
bore

/owr, ohr, ahr/
four
for
far
oar
or
are

/eyr, er, ær/
Mary
merry
marry
ferry
furry
hairy
hurry

/iyr,ir/
nearer
mirror

/uw, iw/
do*
dew*

Incidence
/ohg, og/
fog
log
smog
clog
job
dog
cog

frog
bog
hog
flog
log
goggles
soggy
toggle

/or,ohr/
tomorrow
sorry
orange
horrible
forest
borrow

/wo, woh/
watch
wash
water
walrus
Washington

/uw,u/
roof
coop
route
room
root
coupon

*Appears twice in analytical table