6 Surveys and interviews

Natalie Schilling

1 Introduction

Intersubjective agreement is best reached by convergence of several kinds of data with complementary sources of error. (Labov 1972b)

Early in the development of modern sociolinguistics, William Labov taught us that there is no single best type of data or method for linguistic study, since all are limited. Thus, our fullest understandings are reached only through approaching our research questions from several angles, with each vantage point providing a unique perspective that offsets its necessarily limited scope.

In this chapter, I examine two important and complementary methods for collecting data on language in its social setting: surveys in which researchers more or less directly elicit information on linguistic features, patterns, and inter-relations; and interviews in which connected speech is elicited. In addition, I consider methods designed specifically to elicit information on language attitudes, though both surveys and interviews can provide valuable information on language attitudes as well as language use. The chapter takes a largely variationist sociolinguistic approach, to complement the other chapters in Part I of this volume. Data collection methods designed to glean information on language variation can usefully inform other types of linguistic study; at the same time, elicitations associated with theoretical, anthropological, psycholinguistic, acquisitional, and other approaches to language study can be invaluable for sociolinguistic research. For example, data from a relatively relaxed, conversational interview might yield information on stigmatized language features that are resistant to elicitation via grammaticality judgment tasks; conversely, sociolinguists can use grammaticality judgment tasks to distinguish accidental gaps from genuine ungrammaticality (see Section 2.3).

I discuss both designing and implementing the various techniques, as well as advantages, disadvantages, and purposes of each. At issue throughout is the question of the “authenticity” of our data, and I consider whether it is possible or even desirable to seek to remove researcher effects, in an effort to overcome Labov’s (1972b: 113) “Observer’s Paradox”: “To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed.”
2 Surveys and survey questionnaires

Since the late 1800s, linguists have been gathering data via survey questionnaires. The earliest surveys were designed to elicit a range of language forms – lexical, phonological, and grammatical – from across a wide geographic range, thereby obtaining a picture of language variation across space, with an eye toward how synchronic variation reflects historical connections and developments over time, in keeping with the then main current of linguistic study, historical/comparative linguistics. With the advent of structuralist linguistics, there came an interest in interrelations among elements in individual linguistic systems, so elicitation techniques for uncovering such relations were added to the empirical linguist’s toolkit – for example, tasks designed to elicit judgments of “same” or “different” to determine minimal pairs, and sentence permutation tasks to determine the subject-verb forms constituting person-number paradigms. Generative linguistics brought in other types of elicitation – tasks designed to elicit grammaticality judgments (or, more properly, acceptability judgments; see Chapter 3 in this volume) and information on the structural limitations of forms – that is, which forms can and cannot be used, as well as in which environments they can grammatically occur.

In the sections that follow, we consider different modes of survey administration, as well as direct vs indirect means of eliciting lexical, phonological, and grammatical features. We also briefly examine how the elicitation of listener judgments has been usefully applied to sociolinguistic study, leaving full discussion of judgments of grammaticality and “same” vs “different” to Chapters 3 and 4. We devote special attention to a type of survey in which linguistic information is elicited completely indirectly – the rapid and anonymous survey – since this clever design mitigates observer effects on “naturalness” of data, while still ensuring that desired forms are produced.

2.1 Modes of administration: from long-distance to up-close to the World Wide Web

The first dialect geographic survey, Wenker’s 1876 survey of German dialects (Mitzka 1952; cited in Chambers and Trudgill 1980: 15–36), was administered via postal questionnaire, since again, the aim of such early studies was to obtain information on language variation across wide geographic areas. However, researchers soon decided that it would be advantageous to send out trained fieldworkers to administer questionnaires in person; so starting with studies conducted in the late nineteenth century, most dialect geographic studies up until the past few decades have relied on face-to-face rather than long-distance methods. Among the earliest such studies are Gilliéron’s 1896 linguistic survey of France (Gilliéron 1902–10; cited in Chambers and Trudgill 1980: 15–36) and Grierson’s (1905) Linguistic Survey of India, conducted between 1894 and 1928,
and resulting in the publication of an impressive eleven-volume description of hundreds of dialects of Indian languages, collected via elicitations and gramophone recordings. Important twentieth-century surveys conducted via in-person fieldwork include the various projects associated with the Linguistic Atlas of the United States and Canada, such as the *Linguistic Atlas of the Middle and South Atlantic States* (McDavid and O’Cain 1980), the *Linguistic Atlas of the Gulf States* (Pederson et al. 1986), the nationwide survey of US dialect regions conducted in connection with the *Dictionary of American Regional English* (DARE) project from 1965–70 (Cassidy 1985; Hall and Cassidy 1991, 1996; Hall 2002, 2012); and Orton et al.’s Survey of English Dialects (1950–1961; e.g., Orton, Sanderson, and Widdowson 1978).

Face-to-face surveys were long preferred over long-distance for a number of reasons. For example, they allow more control over who the respondents are, since they provide at least a degree of verification of participants’ claimed demographic characteristics (e.g., age, sex, race). In addition, administering surveys in person allows researchers to record responses themselves (through on-the-spot phonetic transcription or audio recording), to record multiple responses indicative of variable usage, and to exchange clarifications. In addition, because in-person surveys are usually administered orally, respondents may be more relaxed than when filling out a written form, and so yield more naturalistic data.

However, long-distance surveys have their own advantages: they require much less time, effort, and money than in-person surveys, and so allow for broader population coverage in a shorter amount of time. In addition, one can argue that respondents may actually feel more rather than less comfortable with long-distance surveys, since the presence of a fieldworker who is either an advanced student or a professor may be intimidating. Some relatively recent sociolinguistic studies using written questionnaires include Chambers’ (1994, 1998a, 1998b) study of lexical, phonological, and morphological variation in the “Golden Horseshoe” region of Canada (along the western tip of Lake Ontario, from Oshawa to Niagara Falls, and encompassing Toronto); Boberg, Roberts, and Nagy’s McGill-New Hampshire-Vermont Dialect Survey (which investigates attitudes, lexical items, and pronunciations; see, e.g., Nagy 2001); and Gordon’s investigation of the Northern Cities Vowel Shift in the American Midland (Gordon 2006).

In recent decades, technological advances have led researchers to long-distance survey methods that allow for both breadth of coverage and audio recording. Thus, for example, Labov, Ash, and Boberg’s (2006) *Atlas of North American English* is based on a sweeping telephone survey (TELSUR) of larger population centers in the US. Interestingly, whereas traditional large-scale dialectological surveys were interested in obtaining information on historical forms and so focused on the most conservative speakers in the populations of study (i.e., non-mobile, older, rural males, or NORMs), TELSUR turns the tables on this traditional bias by purposely selecting young female respondents from urbanized areas, in order to record the usages of the most innovative speakers, and hence
Researchers increasingly have been turning to the convenience of internet-based surveys, which afford maximal geographic (and social) coverage with minimal time, effort, and expense. The latter also have the advantage of allowing for not only detailed audio recording of speech data, but also recording such measures as response time and eye movement. In addition, computer-based surveys enable one to relatively easily include computer-manipulated stimuli, in order to test subtle facets of linguistic perception. For example, Plichta, Preston, and Rakerd (2005) conducted an internet survey of people’s perceptions of vowels associated with the Northern Cities Vowel Shift (currently taking place in much of the inland northern US), in which participants from both within and outside the Northern Cities region listened to computer-synthesized words with various degrees of vowel shift (e.g., words ranging along a continuum from sod to sad) in sentences offering no contextual clues (e.g., “Did you say, ‘sod’ or ‘sad’?”), and then selected the word they thought they had heard.

Of course, with the return of long-distance surveys with no fieldworker presence during administration comes the return of the same disadvantages that plagued the earliest postal questionnaires, including inability to (at least partially) verify participants’ self-characterizations and lack of immediate availability to either ask or answer clarification questions. A further caution in moving from the field to the World Wide Web is control over data access; while most internet surveys do not elicit specific identifying information (but only general information such as participant demographic characteristics), great care must be taken to guard potential identifying information from public access (e.g., audio responses to an internet survey that could be used to identify participants by voice).

2.2 Direct and indirect elicitation of linguistic features

Linguistic surveys can be composed of different types of elicitation frames, designed to yield different types of information. Almost all will elicit basic demographic information, often at the end of the survey, in order to reduce speaker self-consciousness, for example, about age or socioeconomic status. However, such placement is risky if the survey is long (as were traditional dialectological surveys, which sometimes could take an entire day to complete), since participants (and researchers) may tire and perhaps not finish the survey. Also often included in surveys are questions designed to elicit information on attitudes and orientations, to see if such matters correlate with patterns of language use – for example, use or non-use of local dialect forms, maintenance or attrition of an endangered language. (See Section 3 on surveys designed specifically to elicit language attitudes.) The bulk of the linguistic survey, however, involves eliciting information on linguistic features, structures, and systems.

In order to elicit particular linguistic features, researchers can use either direct elicitations (self-reports) or indirect ones. For example, the telephone portion of
the Survey of Oklahoma Dialects (SOD-T), conducted in the mid 1990s (Bailey, Tillery, and Wikle 1997), included both types, including, for example, a number of direct elicitations of the type in example (1).

(1) Have you ever heard the term “SNAP BEANS” used for the bean that you break in half to cook?
   a. yes
      {IF YES} How often would you use that term: all of the time, some of the time, not very often or never?
      (1) all  (2) some  (3) not often  (4) never
   b. no
      {IF NO} What term would you use?

Indirect elicitations can take several different forms (e.g., questions of the form, “What do you call . . . ?” or fill in the blanks), and they may be used to elicit a full range of linguistic forms. For example, indirect elicitations of lexical items in SOD-T were formulated so that respondents could give more than one synonymous answer – for example, “What do you call those little bugs that get on you in the grass and make you itch?” (looking for redbugs or chiggers), and “What do you call the enclosed place where hogs are kept?” (looking for pig pen or sty). The Linguistic Atlas of the United States and Canada projects used questions that were a bit more indirect – for example, “Where did you keep your hogs and pigs?” “The thing you put in your mouth and work back and forth and blow on it. Do you remember any other names for it?” (looking for harp, breath harp, mouth organ, harmonica, etc.).

The Linguistic Atlas survey also included fill-in-the-blanks to elicit pronunciations and grammatical features. For example, variant pronunciations of the word *yolk* were elicited with the question, “What do you call the two parts of an egg? One is the white; the other is _____,” while variants of the past tense of *drive* (e.g., drove, drove, driv) were elicited with “I wanted to hang something out in the barn, so I just took a nail and _____.” To study Danish vowel mergers, Ejstrup and Hansen (2004) devised a task in which respondents spoke informally about a set of common objects, presented in picture form and selected to elicit target vowels.

Other elicitations can be even less direct: Thus, for example, Bailey, Tillery, and Wikle (1997) used the following frames to elicit phonological features, even though they seemed to be lexical elicitations: “When are you most likely to hear an owl hoot?” (looking for the pronunciation of /æ/ as either [aɪ] or [aː], not the lexical item night), and “Now what about those large birds that sit on telephone poles and swoop down to kill mice and other small animals, what do you call those?” (looking for the pronunciation of hawk as either [hɔk] or [hɑk] to elicit information on the /s/-/ɑ/ merger taking place in much of the US in word pairs like hawk/hock).

Finally, researchers can use direct and indirect methods to elicit forms from entire semantic domains. For example, Teresa Labov (1992) elicited information on slang terms in various domains by having US students select terms from lists she provided (direct elicitations) of slang words for people, approval/disapproval,
and being “under the influence,” and by and by having them fill in blanks she left open in each category (indirect elicitations).

2.3 Direct and indirect elicitation of listener judgments

In addition to eliciting data on which linguistic features can be used in particular languages and language varieties, linguists need to know which forms and structures cannot be used. Further, if we are interested in language variation, we need to know which environments favor, disfavor and prohibit variable usages (see also Chapter 20). Because non-use in speech or straightforward elicitation may be due to accidental gaps rather than ungrammaticality, sociolinguistics can usefully apply sentence permutation and judgment tasks of the types discussed in Chapters 3 and 4.

As one example, in our study of ethnicity-based language variation in Robeson County, in Southeastern North Carolina, composed of residents of Lumbee Indian, African American and White ethnicity (Wolfram 1995; Wolfram and Dannenberg 1999), we observed in everyday conversation and in the sociolinguistic interviews from which we obtained the bulk of our data (see Section 4 below) that I’m could be used as a perfective, as in I’ve been there a long time for “I’ve been there a long time,” in Lumbee English. We also noted that the Lumbee seemed to be able to use the I’m form as part of simple past constructions, but we wanted to make sure. Hence, Wolfram (1995) devised a sentence permutation task designed to yield tense-marked auxiliaries that would provide more direct evidence of the tense-aspect status of I’m than we were able to obtain via observation or elicitation of conversational speech. The task involved having speakers make three types of changes to a number of stimulus sentences, some past tense (e.g., I forgot to do it yesterday) and some perfective (e.g., I’s seen the toten [i.e., presage of a fateful event]): (1) change positive to negative, (2) provide an elliptical version of the sentence via VP deletion, and (3) change declarative sentences to questions. Wolfram hypothesized that for the simple past sentences, do forms would surface (e.g., I didn’t forget to do it yesterday rather than I haven’t forgot(ten) to do it yesterday), while for the perfectives, we would get have (e.g., I’m seen the toten, I know I have rather than I know I did). For the most part, the results conformed to expectations, though we did get a few have forms with the past tense sentences and one or two isolated cases of am with perfectives (e.g., Am I seen the toten?). Hence, we were able to confirm that I’m could indeed be used to indicate both past and perfective meanings and that in such constructions the underlying verb is do or have, not be.

Sociolinguists also use judgment tasks to help determine structural limitations and linguistic systematicity. One important line of inquiry in variationist sociolinguistics has been the study of vowel mergers in progress, a task for which judgments of “same” or “different” can add invaluable information to that gleaned from conversational data. Some important mergers in progress that have been subject to sociolinguistic inquiry include the /ɔ/-/ɑ/ merger in US English...
conditioned mergers such as /i/-/ɪ/ before /l/ in Southwestern US varieties (as in filled/field; e.g., Di Paolo 1988; Labov, Ash, and Boberg 2006), and /ɪr/-/ɛr/ (as in beer/bear) in Norwich, England (Trudgill 1988) and in New Zealand (Hay, Warren, and Drager 2006), and the seeming merger of /ay/ and /oy/ in Essex, England (as in live/loin; e.g., Labov 1994: 377–84).

The same/different task can take a number of forms, ranging from the very direct (e.g., playing the respondent a pre-recorded list of word pairs and having them state whether they are the same or different), to playing elements of the pairs interspersed with other items and asking for definitions, to constructing and reading (or playing) to respondents elaborate stories whose interpretation depends on the listener’s understanding of one or more word pairs involved in ongoing merger (e.g., Labov 1994: 403–6).

2.4 Limitations of elicitation tasks

Despite the utility of surveys composed of linguistic elicitations for obtaining large-scale information on the use and patterning of particular features of interest, they do have their limitations. Even the best designed elicitation tasks are removed from how people use (and think about) language in everyday life, and people’s reports of their linguistic usage may or may not match up with what they actually do. For example, direct elicitations such as “Have you ever heard the term ‘snap beans?’” can induce respondents to claim knowledge and use of features they have never heard prior to the research situation, while less direct elicitations can be very difficult to devise and/or yield a wide range of responses falling far from the targeted item. Hence, for example, Bailey, Wikle, and Tillery (1997) report on their frustrating attempts to indirectly elicit double modal constructions from Southern US speakers, since there is virtually no single indirect frame that is guaranteed to yield a double vs a single modal. (Indeed, the elicitation frame for might could from the Linguistic Atlas of the Gulf States is cringingly unwieldy: “When you get something done that was hard work all by yourself and your friend was standing around without helping, you say ____.”) In addition to inadvertently priming or failing to prime respondents to produce desired forms, elicitations can also induce people to purposely over- or under-report linguistic usages. In one well-known case, Trudgill (1972) noted both phenomena in a single city, Norwich, England, with women over-reporting their use of standard English features in direct elicitations compared with their actual usages in conversational interviews, and men under-reporting their use of standard variants. Most likely, these mismatches are due to the association, in many communities, of vernacularity with masculinity and what are often considered to be its component qualities, such as being tough or hardworking, and having “street smarts.”

Issues regarding the social valuation of linguistic features also affect acceptability judgments and sentence permutation tasks, and studies must be designed quite carefully to steer respondents away from their tendency to confound
linguistic grammaticality with social acceptability or “correctness,” and corresponding rejection of well-formed constructions that do not correspond to standard forms. Further, non-linguists tend to focus on language content, not subtleties of linguistic form, and so base their responses on semantic sense, pragmatic felicity, and/or social appropriateness. For example, in our studies of Smith Island, Maryland, we were interested in learning about possible subject–verb agreement patterns with existential it (e.g., It’s a lot of crabs in the bay for “There are a lot of crabs in the bay”; see, e.g., Parrott 2002), and so administered a series of judgment tasks to respondents selected for their interest in our dialect studies. However, even the most linguistically curious tended to focus on content rather than form, and stimuli such as “Can you say ‘It’s many politicians elected by Smith Islanders?’” were met with responses such as, “Actually, nobody would say that because it’s not true.” Further, when they did pay attention to form, they focused on prescriptive standards rather than grammatical dialectal usages, and so rejected forms because they were “illiterate” or “improper” rather than unnatural in the Smith Island dialect. (See Buchstaller and Corrigan 2011 on strategies for investigating non-standard linguistic grammatical usages; Chapters 3 and 4 in this volume also discuss the issue of subjective interpretations of judgment tasks.)

Finally, there are issues related more to the testing situation than the test items themselves – for example, ordering effects, participants’ possible discomfort with the test-like nature of the elicitation task, and their resulting desire either to do “well” on the test by providing the answers the “teacher” (i.e., researcher) expects of them, or to get the test over with as quickly as possible, perhaps by giving the exact same response to every prompt. Such issues can be mitigated by including control frames and varying the order of frames, as well as framing elicitation tasks as “games” rather than test-like research tasks. For example, as part of his study of African American Vernacular English in Washington, DC, Fasold (1972) included a set of “Word Games” designed to uncover sources of non-conjugated be in African American Vernacular English – for example, distributive be vs be derived from deletion of will or would, as in “He be working” (regularly vs soon). However, we should always remain aware of possible mismatches between what people do and what they say they do; and ideally, elicitations will never form our sole source of data on language use. (See Chapters 3 and 7 for more on test effects and how to compensate for them.)

2.5 Rapid and anonymous surveys

One ingenious method of overcoming the unnaturalness of direct elicitations is the rapid and anonymous survey, in which linguistic elicitations are disguised as ordinary non-linguistic questions, and research participants respond exactly as they would in everyday life, in non-research contexts. The classic example is Labov’s (1972a) study of r-pronunciation vs r-lessness in the speech of personnel in three New York City department stores, each catering to a different social class group. The elicitation frame was exceedingly clever in its
simplicity: in each store, Labov asked workers the location of a particular item he had already determined to be on the fourth floor, thereby eliciting the desired variable in two word and phrase positions. Further, he obtained information on stylistic variation by asking each respondent to repeat the utterance in a more careful/emphatic style, by following his initial inquiry with “Excuse me?” Labov recorded each response immediately afterwards, by stepping out of view of the respondent and noting in writing whether each potential case of [r] was realized or not.

As predicted, the personnel in the store catering to the highest social class group, Saks, had the most r-pronunciation, in keeping with the fact that this variant was an incoming prestige form, and employees could be expected to accommodate to the speech of their prestigious clientele. The mid-level store, Macy’s, had middle levels of /r/ usage and the greatest amount of “self-correction” in repetition, while workers in the bargain store, S. Klein, showed the lowest levels of r-pronunciation.

Clearly, a rapid and anonymous survey is advantageous in that it yields unself-conscious (and so presumably “natural”) data on targeted features from a broad population in a short amount of time. At the same time, there are a number of limitations: the researcher can only obtain information on a very limited amount of linguistic data, since recording is done only in writing (since surreptitious audio recordings are widely considered to be unethical in linguistics; see Chapter 2), and social information is limited as well, since we can only guess at respondent demographics rather than directly elicit this information. Hence, Labov discusses the patterning of the (r) variable not only by department store (and specific occupation within each store), but also by gender, ethnicity, and age, but had to admit that his demographic categorizations could not be completely accurate, especially with regard to age. Finally, this method requires a very good ear, and its practitioners need to be very well trained in discerning phonetic detail without the computational aid to which so many linguists have become accustomed.

Because of the limitations of the rapid and anonymous survey, Labov stresses that they should only be used as preliminary or supplementary sources of data, never as the basis for an entire study. Nevertheless, they can be an invaluable complement to more direct elicitations and less pointed observations, since, as Labov says, “They represent a form of nonreactive experiment in which we avoid the bias of the experimental context and the irregular interference of prestige norms but still control the behavior of subjects” (Labov 1972a: 69).

### 3 Eliciting language attitudes

In addition to eliciting and observing information on language production, linguists are interested in obtaining information on listener perceptions, including their attitudes toward particular features, varieties, and variants, as well as those who use them. Whereas language attitudes can be directly elicited (e.g., in interviews or opinion surveys), typically they are investigated partially indirectly,
via elicitation of attitudes toward speakers (not language per se), and sometimes elicitation of extralinguistic behaviors indicative of language attitudes. Probably the most widely used method of eliciting language attitudes is the Matched Guise Technique devised by Lambert and his colleagues in the 1960s (Lambert et al. 1960). The classic version involves assembling a stimulus consisting of passages spoken (or read) in different languages or language varieties produced by a single speaker. In between-subjects designs, listeners are separated into randomized subgroups (see Chapter 7), each of which hears a subset of recordings. If listeners hear more than one recording, they are typically told that the different guises are produced by different individuals and are then asked to rate the speakers on various measures, usually along a so-called semantic differential scale, involving Likert scales for various semantic categories, often along status-stressing and solidarity-stressing scales (e.g., intelligent-unintelligent, friendly-unfriendly). The method is thus partially indirect, in that it involves direct elicitation of participants’ attitudes toward speakers and only indirect elicitation of attitudes toward language.

Sometimes the adjectives chosen are based on previous studies, with the dimensions of status and solidarity again underpinning the choices. In such designs, several items (either semantic scales or attitudes questions) ideally measure the same underlying construct, such as status or solidarity, in order to counteract potentially unreliable effects caused by individual items. Each group of items for a single construct can be tested for internal reliability, and less reliable items can be removed from the analysis (Garrett 2010). The researcher may also choose to elicit potentially relevant dimensions via open-ended questions in a pilot study, perhaps with focus groups. For example, Williams’ studies in the mid 1970s of language attitudes and education used semantic differential scales derived from pilot studies with small groups of teachers who evaluated children’s speech using their own terms, which were then found to cluster along the dimensions of confidence-eagerness and ethnicity-nonstandardness (e.g., Williams 1974: 23; cited in Fasold 1984: 171–3).

The choice of speaker to produce the different guises is also a matter to be approached with care. Practically speaking, it is very difficult to find speakers who are fluent in the various languages or dialects the researcher wishes to study; hence the matched guise in its strictest sense is often replaced by the so-called verbal guise (Campbell-Kibler 2010: 378), which uses different speakers matched as closely as possible in terms of factors other than language or dialect per se that could influence listener judgments (e.g., voice quality, pitch and intonation, and speech rate). (Indeed, Williams’ studies use different speakers, matched for ethnicity and social status, rather than a matched guise, per se.)

Further, whether one is using a matched guise or voice guise, one must also control for content. This can be done using reading passages rather than conversational speech; the trade-off, though, is that reading passages sound less natural than talk. Campbell-Kibler (2010: 380–1) suggests controlling for content without resorting to readings by using passages from spontaneous speech about similar
topics (e.g., childhood games). If however, the exact same content must be used, then the unnaturalness can be alleviated by giving listeners a reason why the speakers are reading rather than talking – for example, perhaps they are allegedly auditioning for a job as a news anchor (Labov et al. 2006; cited in Campbell-Kibler 2010: 381) or for a radio commercial. A further method of circumventing the unnaturalness of having one group of speakers listen to the same passage as read by different speakers is to use a between-subjects design, in which the different guises are judgment by different, but rather large, groups of listeners (see Chapter 7).

The traditional method of eliciting language attitudes is plagued by the same problems as elicitations of speech production. The tasks are unnatural, and there is no guarantee that the results are reflective of listeners’ genuine attitudes (which can be conceptualized as a mental “state of readiness; an intervening variable between a stimulus affecting a person and that person’s response” [Fasold 1984: 147, citing Agheyisi and Fishman 1970: 138 and Cooper and Fishman 1974: 7]). This may be because listeners do not have free access to their attitudes or the ability to accurately convey them, or because they do not wish to express negative attitudes they might really hold. Thus, researchers have sought methods other than the matched guise/voice guise technique – for example, so-called “commitment measures,” whereby language attitudes are assessed by observing behaviors (e.g., Fasold 1984: 153–8). For example, Bourhis and Giles (1976; described in Fasold 1984: 155–8) devised a task designed to measure people’s attitudes toward four language varieties in Wales, by testing for correlations between the language variety in which an announcement was made to several audiences of theatregoers, and the extent to which each audience complied with the request to fill out a short questionnaire based on the content of the announcement. Among the findings were that Anglo-Welsh listeners were least responsive to broad Welsh-accented English, while Welsh listeners were most responsive to Welsh and least responsive to standard British English (RP). Clearly, designing a “commitment measure” in which actual non-linguistic behaviors are elicited is an extremely difficult task; it is also possible to indirectly measure language attitudes by gauging participants’ purported behaviors. For example, Cooper and Fishman (1974: 16–17; cited in Fasold 1984: 179) devised a task to test the hypothesis that in Israel Hebrew is considered to be better suited for scientific arguments, while Arabic is considered better suited to traditional Islamic arguments. The task involved having Muslims who were bilingual in Hebrew and Arabic listen to four passages (two per group of participants): a scientific anti-tobacco argument in Hebrew and the same argument in Arabic, and a traditional anti-liquor argument in Arabic and the same thing in Hebrew. Language attitudes were indirectly assessed by having the participants indicate whether they would or would not support higher taxes on tobacco and liquor based on the arguments they heard; and results were conclusive indeed, with the anti-tobacco tax being far more highly favored among those who heard the scientific argument in Hebrew, and the anti-liquor tax far more favored among those who heard the traditional argument in Arabic.
In addition to considering how best to elicit attitudes and perceptions, there is also the question of exactly what listeners are attending to in making their judgments. As technology becomes more advanced, it is becoming easier to devise experiments designed to tease out the effects of minute aspects of the speech signal on attitudes and identifications. Hence, nowadays, researchers can use computer manipulations to construct a series of guises differing along only a single dimension, whether binary (e.g., -in’ vs -ing endings) or scalar (e.g., the height of a particular vowel or length of a stop release burst). Similarly, the speech signal can be filtered to remove all segmental information leaving only intonational contour, or intonation can be flattened and segments can be scrambled, so that presumably all effects but evaluations/identifications of segments themselves, devoid of content or context, can be investigated (Thomas and Reaser 2004; Campbell-Kibler 2010: 381). This is not to suggest that digital manipulation of the speech signal is always easy, since some features lend themselves more readily to computer alteration than others. In addition, the quality of the stimuli may depend on the quality of the source recordings. If the source signal was recorded in a controlled environment, there will be fewer problems with manipulation, but it might sound artificial (since phonetics labs may constrain linguistic behavior). If the source signal is taken from spontaneous speech, it will sound more “natural” from an interactional perspective, but ambient noise may pose a challenge for digital manipulation (see Chapters 9 and 17).

As fascinating as studies attempting to get at perceptions of individual language components may be, the researcher would be wise to heed the caution of Auer, who points out that “the meaning of linguistic heterogeneity does not (usually) reside in individual linguistic features but rather in constellations of such features which are interpreted together ... [W]e do not interpret single variables but a gestalt-like stylistic expression” (2007: 12; cited in Soukup 2011: 350). Hence, it is not certain that ever more fine-grained analyses or syntheses of speech signals will yield ever-increasing gains in our understanding of language attitudes and speech perceptions, and experiments involving artificially manipulated speech should always be complemented with methods involving more naturalistic data, including elicitations based on listener judgments of excerpts from spontaneous speech, responses to questions about attitudes and perceptions, and, best, of all, observations of linguistic and extralinguistic behaviors reflective of language perceptions, attitudes, and ideologies, including, crucially, the quantitative patterning of variation according to social class and style that has been recognized as deeply rooted in linguistic evaluation since the inception of modern sociolinguistic study.

4 The sociolinguistic interview

One of the most important tools for collecting data on language in its social setting is the sociolinguistic interview. Like the rapid and anonymous
survey, the sociolinguistic interview seeks to mitigate observer effects. However, in this case recordings are overt, but questions are much less pointed and are designed to steer interviewees away from focusing on specific linguistic forms, toward producing connected speech about topics of interest. Because recording and elicitation are overt, researchers do not have to confine themselves to a limited amount of linguistic and demographic/social information, but can ask for the linguistic and social information they need. In addition, researchers can set up recording equipment and the interview setting to ensure optimal audio quality. Further, since the typical sociolinguistic interview involves only two participants, and the interviewee is encouraged to talk as freely as possible, it also allows us to gather a maximal quantity of naturalistic speech data in a relatively short amount of time. Given these advantages, it is little wonder that the sociolinguistic interview has long been and continues to be the primary data-gathering tool of variationist sociolinguists as well as other linguists (e.g., in second language acquisition) who need to obtain in an efficient manner high-quality naturalistic conversational data for a range of research purposes.

4.1 Structuring the sociolinguistic interview

Basically, the sociolinguistic interview is a loosely structured interview designed to yield large quantities of speech from interviewees that is as casual and natural as possible, with the most “natural,” “vernacular” speech held to occur when speakers focus their attention on what they are talking about rather than on speech itself. This focus on the vernacular, and the reasoning behind it, is captured in Labov’s Vernacular Principle, which states that “the style which is most regular in its structure and in its relation to the evolution of language is the vernacular, in which the minimum attention is paid to speech” (Labov 1972b: 112).

Sociolinguistic interview questions are grouped into modules focused on particular topics, and the modules can be rearranged as the interview progresses to approximate the flow of natural conversation. Questions are focused on topics believed to be of fairly universal interest, as well as matters of particular interest to each community of study, and interviewees are encouraged to talk as long as they like on any topic that particularly interests them, to tell stories or narratives, and even to go off on tangents of their own. For example, Labov has long maintained that interviewees will tell particularly animated narratives and forget about the fact that they are being recorded (and so produce truly vernacular speech) if interviewers ask them his famous “danger of death” question: “Have you ever been in a situation where you were in serious danger of being killed, where you thought to yourself, This is it . . .” (Labov 1972b: 113).

In addition to questions designed to spark naturalistic conversation, the traditional sociolinguistic interview includes a series of tasks designed to yield increasingly self-conscious, careful, and hence standard speech: a reading passage, a word list, and a list of minimal pairs — that is, words that differ by only one phoneme in standard speech, but may or may not differ in pronunciation in
vernacular varieties. For example, in Labov’s (1966) foundational study of New York City’s Lower East Side, the minimal pair task focused on \( r \)-lessness vs \( r \)-pronunciation in word pairs like *sauce/source*, *god/guard*, and so on. These tasks are administered after the conversational portion of the interview, to avoid calling attention to speech itself too soon.

In studies utilizing the full range of sociolinguistic interview tasks, the predicted patterns of stylistic variation were very often borne out, as for example in Labov’s studies of New York City in the 1960s and of Philadelphia in the 1970s (e.g., Labov 2001), with interviewees showing higher usage levels of nonstandard, vernacular variants in casual speech, and increasingly elevated levels of standard forms as they move from casual to careful to reading passage to word list to minimal pair style.

### 4.2 Limitations of the sociolinguistic interview

Despite its seeming success in capturing and pinpointing vernacular speech as well as a range of other speech styles, the sociolinguistic interview has been subjected to a number of criticisms over the decades. For example, Wolfson (1976) famously contended that the sociolinguistic interview is actually quite *unnatural*, since interviewees expect interviews to be relatively formal, and they may become disconcerted or perhaps even angry when faced with an interviewer who does not have a highly structured questionnaire and so may seem unprepared for the interview.

In addition, even though sociolinguistic researchers are supposed to do their best to relinquish control to interviewees, it has still been argued that there are insurmountable power asymmetries in the sociolinguistic interview. The interviewer usually holds the more powerful conversational role of questioner and a more powerful social role as a researcher associated with a university rather than, for example, a member of a vernacular-speaking and/or minority community (e.g., Labov 1984; Milroy and Gordon 2003: 61–3).

Further, researchers have questioned whether the various contexts in the sociolinguistic interview (e.g., conversation, reading passage, word list) really are differentiated chiefly in terms of amount of attention to speech. For example, many (most) people have a specialized “reading register” that differs in a number of ways from spoken speech, not just in terms of a slightly increased degree of carefulness (e.g., Macaulay 1977; Romaine 1978, 1980; Milroy 1987: 173–8). The reading tasks associated with the traditional sociolinguistic interview can also add to interviewees’ discomfort, thereby further increasing inherent power asymmetries, since many people are not accustomed to reading aloud; in addition, the “marginal” communities in which sociolinguists are so often interested may be characterized by low levels of literacy, rendering reading tasks even more awkward.

Finally, researchers have questioned the focus on vernacular, unselfconscious speech that underlies the Attention to Speech model, for several reasons.
Variationists increasingly are recognizing that people’s everyday speech repertoires include a variety of self-conscious as well as unselfconscious styles; and further, self-conscious, “stylized” linguistic usages are probably becoming more commonplace as people come into increasing contact with more languages, varieties, and variants (as well as their associated social meanings and norms for use) in the face of increasing mobility, globalization, and mediatization (Coupland 2007). In addition, self-conscious speech is more prevalent than we might like to think, even in sociolinguistic interviews, since interviewees can – and do – conceptualize them as occasions for dialect display, despite researchers’ best efforts to relegate speech itself to the background (e.g., Trudgill 1972; Coupland 1980; Reah 1982; Schilling-Estes 1998). Further, no matter how seemingly unselfconscious a stretch of speech may be, we can question whether there really is any such thing as an individual’s single “genuine” vernacular – a “default” style unaffected by any contextual factors – since people always shape their speech to fit the situation at hand and to suit their various purposes (e.g., Hindle 1979; Eckert 2000; Milroy and Gordon 2003: 49–51; Schilling-Estes 2008; see also Schilling 2013, Chapters 3–4, for extended discussions of the sociolinguistic interview and collecting data on stylistic variation).

4.3 Modifications of the sociolinguistic interview

Far from ignoring criticisms leveled against the sociolinguistic interview, variationists have long sought to address the seeming unnaturalness of the interview event, the power asymmetries the event entails, and the danger of interviewers exerting more control than they may realize. Thus, sociolinguists have modified the basic sociolinguistic interview technique in various ways, while still attempting to preserve its advantages in terms of amount of interviewee speech, efficiency, and sound quality.

Among the earliest modifications was moving from the one-on-one interview format to group interviews, with the idea that people being interviewed in peer groups would talk more with one another than the interviewer, and that the everyday interactional norms they have with peer group members would supersede any artificiality brought on by attempts to conform to the interview event or the relatively standard speech of the interviewer. Hence, Labov et al. (1968) used peer group interviews in their early studies of African American children and teenagers in Harlem, and found that interviewer effects were minimized to such an extent that the interviewer often receded into the background while the teens talked among themselves. Other researchers have enjoyed similar success in recording teens in peer group interaction, as, for example, in Hewitt’s (1982) study of London adolescents’ use of Jamaican Creole and Cheshire’s (1982) study of adolescents in Reading, England.

Another variant on the one-on-one interview was used by Wolfram and his research colleagues in their work in North Carolina beginning in the early 1990s (e.g., Wolfram, Hazen, and Schilling-Estes 1999). This involved pairs of
researchers interviewing one or more interviewees. At first glance, such a tactic seems to swing the balance of power even farther toward the interviewer; however, it was found that breaking down the one-on-one dynamic by including an extra researcher was just as effective in reshaping the “interview” into a conversation as adding extra interviewees.

Of course, researchers also move beyond the interview format entirely – for example, by recording spontaneous conversations in which the researcher plays only a minor conversational role, if any, and by having participants make self-recordings as they go about their daily activities in a variety of settings, with a range of interlocutors. Such self-recordings can yield a wider range of speech styles than would be uncovered in sociolinguistic interview data; in addition, they might reveal that people have quite different patterns of language variation in everyday life than in the interview setting. For example, in a study of British Asian English in London, Sharma (2011) demonstrated that whereas interview data indicate that women in her study shift completely away from one highly salient Punjabi-derived phonetic trait, the use of retroflex /t/, self-recordings show quite robust use of this variant in other settings – for example, in the home domain.

Although moving away from one-on-one interviews may seem to be the direction to take in seeking to overcome the Observer’s Paradox and obtaining truly “natural” data, we must remember that there are trade-offs in relinquishing the controls built into the one-on-one sociolinguistic interview. Multi-party interactions are noisier than individual interviews, and even if the researcher records each participant on a separate track or separate audio recorder, there will still be overlap between participants in the acoustic signal, rendering these parts unusable for acoustic analysis – even though they very likely constitute the most interactive portions of the interview. Further, talk may not always be the primary focus of peer group interactions. In addition, people may come and go, rendering it difficult to record necessary demographic and other social information on participants, as well as to ensure that all parties are aware that they are participating in a research study. (The reader is referred to Chapter 10 for full discussion of recording naturally occurring interactions not involving research interviews.)

Hence, despite the advantages of methods that break down the one-on-one interview structure, it really is difficult to devise a better instrument than the sociolinguistic interview in terms of efficiently obtaining large quantities of high-quality recorded speech that closely approximates everyday speech. In addition, although it can be argued that in theory the sociolinguistic interview is less than fully conducive to yielding “natural” speech due to asymmetries in social and conversational roles of interviewers and interviewees, in practice, most researchers have found that interviewer control readily falls away, as interviewees warm up to their topics and as interviewers realize that in the field they are no longer the “experts” they may be in the academy, but rather are “learners” who must cede power to their research participants, the ones who hold expert knowledge of the communities of study (see, e.g., Labov 1984: 40–1).
5 Concluding remarks: on overcoming the Observer’s Paradox

In this chapter, we have explored several important data collection methods in linguistics – surveys designed to elicit linguistic features, patterns, and interrelations; experiments/surveys designed to yield information on language attitudes; and interviews designed to provide us with connected speech. We have also seen attempts to mitigate observer effects within each methodological perspective. For example, rapid and anonymous surveys allow researchers to elicit specific linguistic forms while keeping participants completely unaware of the research context. Conversely, participants in sociolinguistic interviews are at least initially aware that they are part of a study; however, the questions they are asked are designed to yield natural, connected speech, not isolated speech forms, and also to get them so interested in what they are talking about that the research context is at least backgrounded, if not completely forgotten. And finally, with respect to eliciting information on language attitudes, researchers employ not only matched guise techniques, whose test-like nature cannot be concealed, but also commitment measures in which language attitudes are assessed, not through self-reports, but through behaviors reflective of these attitudes – just as attitudes mediate between stimuli and behaviors in non-research contexts.

In closing, though, we should note that even if the research context can be made to fade into the background, we can never truly remove observer effects on speech – nor do we necessarily want to, since speech is always being observed by some listener, in both research and non-research contexts. It is, then, arguably better to attempt to identify and account for contextual effects, including observer effects, rather than seeking to abstract them away. In other words, it may be better to dispense with the Observer’s Paradox rather than trying to overcome it, and to admit that there is no such thing as non-observed language data, and hence no such thing as one single “most important” type of language for linguistic theory – or any one “best” method for obtaining it.

References


