Chapter 7

Questionnaire design and data collection

This chapter is a guide through the design process of the methodological tool at the heart of all WQs: the questionnaire. It introduces the steps towards a custom-built WQ that is tailored to a particular situation, language variety or dialect. The applications of WQs are manifold: they range from large-scale projects that survey regional linguistic variation over large stretches of geographical space, down to microlinguistic studies of only one syntactic phenomenon or even small undergraduate class assignments.

The method is largely language-independent. As long as the respondents are sufficiently literate it is worthwhile considering. While levels of literacy obviously vary between societies globally, in places where there have been low rates of illiteracy over a person's lifetime, the method will generally be useful. The success of any survey depends a lot on the way a question is phrased. Questionnaire design is far from "anything goes", which will be shown in this chapter, though it is also sometimes difficult to formulate precise dos and don'ts. By the end of the chapter, you should have a fairly good idea of:

- what you can and you cannot ask in WQs
- how to ask your questions
- the pros and cons of different survey formats (e.g. length, presentation, paper or online)

We will first reflect on the planning stages of questionnaire design. The research question at the base of the questionnaire defines its purpose and therefore needs to be clarified (Section 7.1). The structure of the questionnaire will be considered next (7.2), before ample discussion will be dedicated to question types (7.3), including those used in newer WQ approaches, and useful strategies to mitigate against socially sanctioned answer behaviour. After these core questions, we will address some issues about population sampling (7.4).
7.1 Planning the questionnaire: Purpose & research question

The most obvious question to ask is about the purpose of the questionnaire. In reality, this process begins with the formulation of a research question. The research question should be of a broader spectrum and not as specific as a particular questionnaire item. At this early stage of the research, literature searches on the general research question are particularly informative.

For Canadian English, this book includes a lot of reference points. For other areas, the researcher would gather the precise knowledge elsewhere. Examples of research questions, drawn from the previous chapters and relating to Canadian English, include:

(7.1) a. Is Canadian English a homogeneous variety?
b. Is the autonomy of Canadian English in danger?
c. Do women in Vancouver use more standard variants than men?
d. Are traditional Canadian lexical variables on the decline?
e. Do Chinese-Canadians in Vancouver use different lexical items than Anglo-Irish Canadians?

Boberg’s (2010) assessment on lexical autonomy in Canadian English, for instance, is of immediate interest to research questions (7.1a, b, & d) and will directly inform the design of the survey in some respects. Research question (7.1e) would require theorizing on the concept of ethnicity, for which Hoffman and Walker’s (2010) Ethnic Orientation Index (Section 8.2.3) or Nagy et al.’s (2014) indices will be useful. Without a proper literature search one is bound to repeat some mistakes or one will not be able to find that particular angle that is missing in existing work. Equally importantly, without a thorough literature review one will be missing the studies that will help to contextualize one’s own findings.

It is important that a research question is **empirically testable**, which is best reached with a set of questions that, taken together, address the research question. Any of the research questions in (7.1) will need to be fleshed out further with follow-up questions, such as the following:

- How do you intend to study the phenomenon, which variables will you study?
- From whom do you intend to collect data? Location (city, province, state, country etc.), social group(s)?
- What are comparisons based on? Is there a control group or data that you can use to the effect or will a control group also need to be polled?

Looking at Example (7.1b), we see that the research question is fairly general and in need of specification, which includes the following issues. First, how does one intend to measure “linguistic autonomy”? Second, what does “endangerment” refer to (in this context)?

A very generalization of the question (7.1b) is as follows:

1. How do you define the autonomy of English?
2. What is the condition of the “endangered” variety?
3. How is the autonomy of the variety measured?
4. In what context is the variety used?
5. How is the variety used?
6. What are the implications of the variety’s use?
7. What are the benefits of the variety’s use?

Before proceeding, the researcher should have a clear idea of what the variety is and how it is used. Considering these questions will help to refine the research question.
Chapter 7. Questionnaire design and data collection

In reality, our questionnaire will have as such question items as a telephone contact item, a questionnaire item, and so forth. Each question should have its own purpose.

In other areas, the main purpose of research is to keep the researcher or researcher interested, and include:

7.2 Structure of the questionnaire

A very basic but important design feature of questionnaires is its structural organization. The distinction is made between an introduction that includes respondent INSTRUCTIONS, a survey part that arranges the questions, preferably in an intuitive order, and a conclusion, that expresses thanks to the respondent and provides contact details. Schleef (2013) lists seven essential parts of introductions to WQs:

1. Title of the questionnaire
2. Brief explanation of its purpose
3. Polite request to fill in the questionnaire fully and honestly
4. A short outline of what the questionnaire will cover (including a time estimate)
5. The promise of anonymity
6. The researcher’s name, institution and contact details
7. An expression of thanks

Before respondents start on the survey part, they need to give their consent to participate. Often times, the WQ is preceded by an information sheet, which may be framed as a letter of informed consent. Some details from Schleef’s list may be included in a CONSENT FORM or Information Sheet that is handed out prior to the distribution of the survey or survey link, such as the following shown in Illustration 7.1:

this context obviously the decrease or even loss of linguistic autonomy compared to a previous point in time? We often need a benchmark to compare our present-day results with previous results that makes such assessments possible. Such results usually come from previous studies that can be used as real-time comparisons. For Canadian English, Dialect Topography data from various points from the 1990s and early 2000s are one such benchmark, the Survey of Canadian English or Boberg’s NARVS are other reference points. Often times, however, it means to collect two data sets, which requires double the effort for data collection alone.

Not all comparative approaches require the collection of two data sets. Example (7.1a), for instance, can be answered without a real-time component, but still requires a comparative component. To assess homogeneity, one needs to define the concept for the terms of the study and show that homogeneity is either greater or less (or equal) in CanE when compared with another variety of English, most obviously American English. In one way, though, this question can be addressed with only one data set. If it can be shown that most traditional linguistic variables behave similarly across the country (within a given, pre-defined range), one would be able to postulate homogeneity. Ideally, though, one would include a comparison with, e.g. American English varieties.
In the event of filling out a 30-minute online questionnaire on English language features for a UBC class project, in the questionnaire you will be asked how you pronounce certain words, what you call certain items, and what constructions you use.

The questionnaire consists of two parts. In the first part, we would collect some information on your social background, such as your age, birthplace, or your parents’ birthplace. You may be asked to self-identify your ethnicity. We need this information in order to spot trends in the language. The second part includes the linguistic questions. In most cases, you will be offered a number of choices, so the process is easy and quick. Please be advised that your interview will be stored on a server in the European Union.

All data will be used for statistical analyses only and will not (and cannot) be linked to you personally. We will not ask for your name or your contact details. Your questionnaire would be one in a pool of around 500 questionnaires. If you decide to participate in the survey, please feel free to discontinues the interview at any time without providing any reasons.

If you are interested, please connect to the link to start the survey: ENTER LINK.

If you have any questions or concerns, you can ask me or you can contact my office.

Dr. Brian Dollinger
Assistant Professor
Department of English
University of British Columbia
#307-1533 East Mall
Vancouver, BC. V6T 1Z1
(604) 822 4055

You will not directly benefit from this study, but your answers will assist us in spotting linguistic changes. If you are interested in the results, please contact me or my instructor in April 2013. Thanks for your cooperation.

STUDENT FIRST NAME

If you have any concerns about your rights as a research subject and/or your experiences while participating in this study, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8686 or by phone at 1-877-822-6597.

Illustration 7.1 Informed Consent Form (UBC 2013)

This Consent Form was approved by UBC’s Behavioural Ethics Review Board – a process which is now compulsory in most university settings (at least at the M.A. and Ph.D. levels and beyond). This paper version was accompanied by a text-only version for social media advertising (without the UBC crest and formatting).

Once an invitee agrees to this anonymous Consent Form, the link to the survey brings her to a site such as the one shown in Illustration 7.2. In terms of Informed Consent, it is important aspect for WQs is what may be called “Implied Agreement”: after being informed about the study and its implications, the invitee has the choice to follow up with a link to a survey or to leave it. If she follows the link, she agrees to the study.
The survey in Illustration 7.2 works in conjunction with the Informed Consent Form shown in Illustration 7.1. A title is given (Washington State English Survey), a purpose is given that should be accurate but not too general ("to spot trends in the language" - Illustration 7.1), a request to fill in the questionnaire is offered, a time estimate is given (15 or 20 min) and the researcher's name and contact details are provided, as well as an expression of thanks.

Washington State English Survey
Answers marked with * are required.

Hello! Thanks for taking the time to fill out the questionnaire. It should take no longer than 20 minutes, but it is usually completed in less than 10 minutes. All questions appear on one page, so you always see how much is left. There are 54 questions total, 8 of which are optional.

A tick next to * after a question means that you are required to answer it. Whenever you see "Optional" at the beginning of a question, you can skip it to the next one. These "Optional comment/number" fields are there to give you the opportunity to add what the form may otherwise not capture. You can find them every few questions through the questionnaire. You can use them to write in any question (please use the question number).

If you have questions or concerns, contact Dr. Stefan Dollinger, Department of English, University of British Columbia, Vancouver, Canada, at stefan.dollinger@ubc.ca.

Thanks a lot for helping with this and enjoy the survey!

1. Your gender: *
   - male
   - female

2. Your age cohort: *
   - 14-19
   - 20-29
   - 30-39
   - 40-49
   - 50-59
   - 60-69
   - 70-79
   - 80 and over

3. Your education (highest completed or in progress): 

Illustration 7.2 Beginning of Washington State English Survey (2011)

There are a number of ways to skin a cat, and this applies to the order of the background information questions in a questionnaire as well. One difference between Schleef's basic layout, which recommends the listing of the background questions at the end of the survey, and the Washington Survey is that the latter lists these questions upfront. In addition, in the Washington Survey detailed instructions on how to deal with the linguistic questions are only added after social background questions were asked. The assumption here is that people are generally familiar with social background questions. Once age, gender, education, residence history and the like are asked in the beginning of the questionnaire (and out of the way), one can focus the attention of the respondent on the all-important linguistic part with a statement such as the following:
Thank you! We will now start the questions on your use of language.

Please read through the following short questions and mark the option that is most natural to you.

We are interested in the language you use when you are among friends, and not in what somebody else might consider "better". Usually, gut reactions work best here. If you use a word that we do not list, please make sure to write it in one of the "Optional commentary/word" text boxes with the question number.

PLEASE DO NOT DO BACK AND CHANGE ANY ANSWERS (we are interested in your gut reactions!)

Illustration 7.3 Instructions for the linguistic part (Washington Survey)

The respondents are encouraged not to answer what might be prescriptively desirable but instead to reveal their more informal language use, "among friends". This may or may not work, as we will see in Section 7.3.6, but most surveys include such instructions. In this survey of 54 questions, including background questions and 8 optional questions, no reminders were included about the style of language that was aimed to be elicited, but in longer surveys, they will be very useful.

Longer surveys, such as the 96-item Dialect Topography Survey benefit from some more elaborate form and structure. Following the established model of the FI questionnaires (such as LAUSC or SED), the linguistic section is arranged into groups of questions that are semantically related. There are five parts in Dialect Topography, for instance:

I. Around the House
II. Food and Drink
III. Outdoors
IV. Neighbours
V. Potpourri

In the first part – Around the House – typical questions are q5 “What do you call the knob you turn to get water outdoors or in the garden?” or q6 “What do you call the small cloth you use for washing your face?” Part II elicits, among others, answers to q20 “What do you call food eaten between meals or before going to bed?” or q27 “Which do you say? (a) He has drank three glasses of milk. Or (b) He has drunk three glasses of milk.” Part III asks respondents questions outside the home, e.g. q37, “In ASPHALT, the PH sounds like f. Does the S sound like sh?” or whether one says, in q40, “The cat wants to go out. Or The cat wants out,” fitting with the outdoor theme. The idea behind such ordering is to keep respondents engaged with a subject. It is believed that thematically arranged questions, with changes of topic, are more entertaining than a random mix or mélange of questions, and might increase the quality of the data.
7.2.1 Questionnaire length

Length is perhaps one of the most crucial categories when designing a survey. On the one hand, researchers want to elicit as much information as possible, on the other hand, they often forget the negative effects of long questionnaires: not only is there great danger that respondents would discontinue or not even start a survey, which would influence the representativity of the survey sample, but there is good evidence to suggest that the longer the survey the less reliable the results.

In a 2008 survey in Vancouver the explicit comparison was made between WQ and FI data. The survey included 32 linguistic questions following 18 social background questions and was therefore fairly short (Dollinger 2012b). Figure 7.1 shows the deviations in frequencies between the WQ and FI data that are greater than 10%, grouped by responses from the first and the second half of the survey.

An interesting pattern can be seen. The younger age cohorts, teens to thirties, show more deviations in the first half of the questionnaire. The 30-year-olds are in the lead with two deviations in the second half but six in the first. The 40-year-olds show five big deviations in each half. Then, the 50- and 60-year-olds show an increase in deviations in the second half: the longer they work on the questionnaire, the more the WQ answers diverge from the FI scores. The 70- and 80+-year-olds are in a category of their own: regardless of the section, they show a steady deviation with ten or eleven and thirteen or eleven errors in the first and second halves, respectively. These figures indicate that the two older age cohorts may indeed be more unreliable respondents.

![Figure 7.1 Position of questions and error rate by age of respondent (Dollinger 2012b: 90)
Figure 7.1 clusters into two levels of deviations: one level is around five, the other around twelve deviations, as represented by the dotted lines. The 60-year-olds provide a clue to what is happening. Their deviations in the second half skyrocket from five to twelve— that is, from one level to another. Whereas sampling size affects the younger age cohorts, it is the 60-year-olds, whose sample size is in between the younger and the older ones, who appear to show some fatigue factor in the second half. In terms of sample sizes, the cohorts seventy and older have fewest respondents, while the 50-year-olds and younger have around fifty respondents and more. The 60-year-olds are in the middle, with twenty-four respondents. Since their deviation score is low in the first half, their deviations are not an effect of sample size but appear to point toward fatigue in the 60-year-olds and over.

Length, therefore, does seem to be a critical issue. What has been long known in the opinion polling literature, the discipline that pioneered written surveys, is true also for linguistic questionnaires. As Moser and Kalton (1971:309) put it in their classic account: “The temptation is always to cover too much, to ask everything that might turn out to be interesting.” They continue with dear words of advice: “This must be resisted.” The question is then: how long should a questionnaire be? A definite answer would be difficult to give, as questions vary in processing time— consider the difference between a question for “place of residence” on the one hand and “offer your residential history from birth to present” on the other hand.

However, one hundred questions (including background questions) appear to be an upper limit that key surveys abide by (Davis 1948 listed 100 linguistic questions), but in most cases the number of response items should be lower than that. Seiler (2010:520) suggests a maximum of one hour as the time to be spent on a questionnaire, which translates well into a 100 questions for the slower respondents. The DT of Canada questionnaire, for instance, is about 30 minutes to an hour in length, depending on a number of factors, such as literacy, age, and current fatigue levels of the respondent. As such, it is much shorter than previous linguistic questionnaires, including Allen’s “check list” for the Linguistic Atlas of the Upper Midwest. It is most important that questionnaire length is balanced with the anticipated sample size: more data points may possibly skew results (due to fatigue), especially in later questions, while fewer questions will provide more reliable results and, most importantly, higher response rates. If one wishes to collect substantial numbers of responses, a shorter questionnaire would aid tremendously. A high response rate is one of the universally acknowledged advantages of WQs and long WQs undercut this design advantage, which must be avoided.

In some fields, such as psychology, it is customary to pay respondents and there are some linguistic projects, e.g. the Roswell Voices project (Kretzschmar et al. 2007), which as a part of their policy pay the informants. Such practice is also common in fieldwork with aboriginal and First Nations communities, which has in these contexts the added
blessing of expressly breaking with a historical and unreflected practice of exploiting aboriginal peoples. It is very important to ensure that respondents are in no power relationship with the researchers, e.g. students who do not want to fill out the form but do not dare to decline for fear of losing favours in an ongoing course. These are some of the issues that Ethical Review Boards will look out for and, at least in the North American (research) university context, apply very high standards to prevent any kind of coercion.

However, there is also a lower limit of the number of questions asked which is based on respondent expectations: if one is asked 15 or 18 background questions, a respondent might reasonably expect roughly an equal number of linguistic questions in order to “make it worth” the respondent’s time. Respondents clearly understand that their social data is just the backdrop for the linguistic questions one is interested in. It is important to meet the respondents’ expectations, which means that unless one plans to elicit very few social categories there is a lower limit of questions under which one should not go. The shortest questionnaire I have ever used included only four linguistic questions and 8 background questions, bringing the total number up to 12 questions. This worked, because the linguistic questions were time-intensive. I announced it as a “5-minute-or-less survey”, which helped increase the number of responses. In fact, I was primarily interested in only two of the four linguistic questions (which were on take up #9, as discussed in Chapter 4), while the others were used as filler and practice questions that also helped to contextualize the meanings of the variable in question (e.g. some of the more widely used meanings of phrasal verb take up, e.g. take up where you left ‘continue’).

7.2.2 Choice of medium: Paper or online?

A very important aspect is the choice of medium: will the survey be printed on paper or should it be an online survey, or a combination of both? Until recently, the paper method was the most frequently used method and most of the literature refers to this type still (e.g. Dillman 2000; Brown 2001; Dörnyei 2003; Schlee 2013).

Structural features are different in paper and online environments. For instance, in an online questionnaire that required respondents to click “Next” to go to page 2, only slightly more than 50% actually did so. This resulted in a small sample for the second half of questions. What does not pose problems on paper (Please turn over), was overlooked by half of the respondents online despite a big “NEXT” button, which illustrates that both media have structural requirements that need to be considered. The best way to discover what works and what does not is to pilot the WQ (Section 7.3.7).

Internet surveys have proven immensely useful and follow the societal trend of self-administration (Dillman 2000: 7). It needs to be kept in mind, however, that even in the Western World not all layers of society can be successfully and equally reached with screen-based collection methods (be it on a personal computer or on a handheld
device). The same problem applies to a much lesser degree to written surveys as such, since pen-and-paper have been available much longer.

Considering this drawback, which will decrease with time as IT technologies reach even the most disadvantaged societal layers and regions, there are considerable advantages when collecting data digitally. While, for the computer savvy, constructing one's own online database and collection form is a task that can be done with open source software such as php/MySQL quite easily, the more practical approach is to use a service such as limesurvey.org. Lime Survey (2012) is one of the most interesting and versatile options and free of charge: its open software concept and its impressive features give you every flexibility, though it would take some time to get familiar with its a-typical interface. Other suites include surveymonkey.com and esurveyspro.com, fluidsurveys.com or qualtrics.com, though all of these charge a fee for the packages you would need for most WQs. The output of survey suites can be imported into a spreadsheet program, such as Excel (to which Chapter 8 serves as an introduction).

7.3 Question design

In English linguistics, the tradition of asking linguistic questions is deeply entrenched in the FL method. As most attention has been given to this method (as shown in Chapters 1 & 2), it is only logical that FL questions are the more advanced genre. English dialect geography has consistently emphasized the importance of indirect questions with the maxim to never ask an interviewee directly for a linguistic item. One should refrain from asking an interviewee how he pronounces "barn", since the fieldworker's pronunciation will influence the interviewee's answers (as discussed in the previous chapter's section on koinéization theory), and one better asks indirectly, e.g. 'what do you call the structure in which you keep the cows', as it will produce more varied lexical items.

What is a perfectly logical elicitation requirement in phonetics, however, has also been applied to non-phonetic domains in a blanket fashion. For instance, if one is interested in the use of a particular lexical item, it might just be better to ask for it directly and have the interviewee elaborate on the item to verify whether she really knows the word and its meaning, as argued in Pratt (1983). As a side effect, one also saves a lot of time and is able to conduct the interview more naturally rather than skirting around a target that is hard to reach.

Today, a clear trend towards more varied and less dogmatic approaches to question design can be seen that includes face-to-face interviewing. In plans for a new dialect atlas of the UK and Ireland, for instance, Kerswill et al. (1999) elaborate on the idea of a WQ component as part of an interview setting. The WQ is given to interviewees a few days prior to the appointment and is to be filled out on their own: their WQs, which are "Sense-Relation-Networks" (SNRs) ask direct linguistic questions, with "standard
Chapter 7. Questionnaire design and data collection

...notion words” acting as linguistic cues for dialect words. This direct method is not just meant to decrease the time needed to elicit words, but also to influence the interview setting in a positive way, as the authors write

with an indirect question, the interaction may feel more like an interview or a test of some sort, rather than a conversation, and this may have the [effect of increasing the formality of the speech of the informant.]

(Kerswill et al. 1999: 262)

While the follow-up interview is an essential part of this method, the use of a WQ component clearly shows that more recent FI methods have begun to embrace direct questioning that is traditionally used in WQs and has been shunned in FIs.

Self-reporting and community-reporting

An important aspect of all types of questionnaire studies is the kinds of responses which are elicited: are the respondents asked to report on their own use, assessments, attitudes or report on somebody else’s? The term self-reporting is used to refer to reports of one’s own use, attitudes and the like and is distinguished from community-reporting, which is the reporting of features and usage in a second-hand fashion as used by others.

This distinction is important as we are dealing with two different types of questions when for instance, respondents are informed (in the DT questionnaire, for instance) that

We are only interested in what you say when you are among friends – not what you think you should say, and certainly not what you think other people think you should say.

Such questions are exclusively aimed at the respondent’s own language use or attitudinal assessment and therefore self-reporting in nature. The social context is defined as being “among friends”, implying that informal language use is what is aimed to be elicited here.

Self-reporting questions are different from questions that elicit an assessment of community norms or behaviour, such as the request in Krug and associates’ questionnaire which asks respondents to assess whether a feature could

a. be said in their home country in an informal conversation or
b. be written in their home country in an email to a former teacher by =

= everyone, = most, = many, = some, = few or no one (Krug & Sell 2013: 81)

This type of community reporting is more complex than self-reporting, in particular as the social circumstances of language use vary and are difficult to be defined in a way that is clear to everyone, though contexts, such as in Krug and Sell’s example, are useful. Depending on the personality type of the respondent and the reach of the respondents’ social networks, however, one would expect quite different answers from respondents of otherwise similar social backgrounds. Community-reporting questions are likely to produce a more heterogeneous data set than self-reporting questions.
7.3.1 Types of questions

There is an array of question types to be used in WQs. Different classifications have been proposed for basic types of question content – Dillman (1978) and Patton (2001) list five types that overlap to a great degree. The recurring types of content questions are

- Behaviour and Experience
- Knowledge
- Beliefs and Opinions
- Attitudes and Feelings

Whatever the classification, it is important to design questions that target one of these areas. When transferring raw questions into questionnaire items one may want to ensure that one uses the correct basic type of question – if we’re interested in behaviour – e.g. what word do you use for X – we must ensure to phrase the question accordingly.

The main distinction for question format is between open questions and closed questions: open questions (or open-response items) invite answers with an open text field that the respondents are asked to fill in, which naturally produces variation in spelling, framing, verb forms (if the answer is more elaborate) and the like. Closed questions (or closed-response items), by contrast, offer a limited set of answer categories in one form or another. The latter is the type that is more common in linguistic WQs, but closed questions require both more thorough knowledge of the linguistic variable than most open answer questions and necessitate the researcher to have a very definite idea of the answer options that a question will trigger. In return, if a closed question is designed in the right way, the analysis is much easier than with open answer questions, as the answers are already classified.

In the social science questionnaire tradition (such as opinion polling), closed questions are the norm and a great deal of literature has been produced on their design in sociology, psychology and market research (e.g. De Vaus 1991; Holm 1998; Dillman 2000, Dörnyei 2003, Brown 2001 and Groom and Littlemore 2011: 98–105) reflect on questionnaires in the context of applied linguistics, which is, overall, close to traditional social science methodology and of immediate relevance to linguistic attitude and perception studies, which we will address in Section 7.3.4 below.

Apart from linguistic attitudes and perception studies, questions in social dialectology generally aim to elicit information on the regional and social uses of linguistic variables and represent an interesting sub-type of a somewhat atypical nature. Compared to the usual social science opinion-poll type of questions, language use questions target only forms of behaviour and not knowledge, attitudes, beliefs or
opinions. For instance, one feature of the opinion polling of sociological and psychological study is that questionnaire items “rarely take the form of actual questions that end with a question mark” (Dörnyei 2003: 28). In linguistic questionnaires, however, it is almost always the case that respondents are directly asked whether they say variant A or variant B or what they call an item that is used for X. As we will see below, for some aspects linguistic questions show differences when compared to social or psychological question design.

Closed-response items

Closed-response items are the preferred type of question in the social sciences for their ease of analysis. There are many different types of closed items, though only some of which have been used in social dialectology and sociolinguistics. These will be presented below.

Checklists

Checklists are a question format that was among the earliest closed-item response types to be used in linguistic WQs, as we recall from Davis’ (1948) study. Checklists deliver categorical answers offered in a closed list of possible answer options. The respondents are asked to select (circle, click) all that apply. For instance:

(7.2) a. What do you call the wheeled conveyance you put your groceries in while shopping: shopping basket / shopping buggy / shopping cart / shopping trolley

b. How do you spell the following word:

colour  color

With checklists, multiple selections are usually encouraged: if a person uses both shopping buggy and shopping cart, both choices would be selected. Likewise, if a person uses both spellings, colour and color would be marked. The advantage of this method is its speed (circling or ticking off options is fast), its downside is that it offers no information on the frequency of the variants: is colour used more frequently or color? Generally, respondents would not be in the position to give adequate answers as to frequency, beyond stating their primary variant. Checklists do not allow for such assessment. For perception and attitude studies, however, checklists can be used quite effectively, as in the following example from Schleef (2013: 45):

(7.3) The following is a list of cities in England. In which of these do you think some locals speak a widely recognizable local dialect? Please circle.

London  Leicester  Liverpool  Leeds  Sheffield  Nottingham
Manchester  Birmingham  Northampton  Carlisle  Norwich  Plymouth
Multiple-Choice items: Binary (nominal) & categorical

Checklists deliver categorical answers. A special case of categorical answers are binary answer choices (also known as True-False items in the wider context). In this case, only two answer options are offered and the respondent may only choose one. In other cases, more than two options are offered and only one may be chosen. In Multiple Choice, a number of options are offered. Instructions may be permitted to either only choose one or more. Examples include:

(7.4) a. What do you call a small house in the countryside, often by a lake, where people go on summer weekends: cabin / camp / chalet / (summer) cottage / summer house / summer place / the lake / vacation home (NARVS, q45)

b. What do you call a multilevel building for parking cars: car park / indoor parking / parkade / (parking) garage / parking lot / parking ramp (NARVS, q48)

c. What do you call the sweet hard substance that covers some cakes?
   frosting
   icing

   (SCE, q34)

d. What do you call milk with more than 2% fat content?
   whole milk
   homo
   homo milk
   other

   (Vancouver Survey, q27)

It is important with categorical answers to add an option of “Do not use” or “Other”, to avoid respondents getting stumped by answer variants they do not use. While binary options are not favoured in social science research, “the more options an item contains, the more accurate evaluation it yields” (Dörnyei 2003: 42), there are good reasons to use them for linguistic features. For instance, it would make little sense to ask respondents how long their glide is in news. For these questions, one might ask:

(7.5) Does the ū in STUDENT sound like the oo in too, or the ū in use? (DT, q52)

A binary answer option – possibly with an option “Do not know” – is as much detail as self-reporting WQs can obtain on phonetic behaviour, for instance.

Rating scales

Rating scales are a standard answer feature in social science research. There are a number of different types: Likert scales and Semantic Differential Scales. The original Likert scales were developed in the 1930s by the American psychologist Renesis Likert. Respondents are asked to indicate the degree of their agreement or disagreement to a set statement, e.g.
For Likert scales to work, they need to be worded in a favourable or unfavourable manner. “Neutral” wording does not work well, e.g. Austrian German is alright, as it does not offer an incentive to use the extreme ends of the scale.

Semantic differential scales are an adaptation of Likert scales. Instead of responding with degrees of agreement or disagreement, a set number of answer choices is offered in between two logical extremes, e.g.

(7.7) Canadian English is

| Widely appreciated | X | : | : | : | Not appreciated |
| Well studied       |   | X | : | : | : | Unstudied |

Rating scales allow more advanced statistical processing than is possible on categorical answers, which is one of the reasons for their popularity. In sociolinguistics, attitude and perception studies make frequent use of them, as we have seen in Jenkins’ (2007) study in Chapter 5 or as will be shown in Section 7.3.4. The issue of whether to use an even or odd number of categories (which results in the creation of a “neutral” category in the latter case), is a matter that is highly disputed among practitioners. While 20% of respondents seem to select a “neutral” category if one is offered to avoid making a choice, the relative responses, the ratio of answers in both extremes, are not affected by an even or odd numbered set of options (Dörnyei 2003: 38). In general I have used even-numbered categories, thus forcing the respondents to take a stance.

These are the most popular types of closed-response questions in social dialectology, though they only represent a fraction of choices as there are many more answer options and presentation modes than have been applied in linguistic contexts (see, e.g. De Vaux 1991; Holm 1998: Vol. 1). Likert scales have been used for decades to assess language attitudes, starting with the work of Lambert and Giles. With the advancement of digital technology, sound files have been used more frequently, with Campbell-Kibler’s (e.g. 2007) and associates’ (Wanjema et al. 2013) work being a good case in point for sound delivery online. Analog tape recorders, however, have been put to use for a long time so that a decisive factor here is the ease of delivery with digital formats with internet technology. To give only one recent example, Watson & Clark (2014), for instance, combine time-tested methods in a digital environment, using Likert scales to have respondents rate speech samples for attributes such as pleasantness, while presenting material in the form of sound files on the internet in their WQ.
Multiple-items scales in social dialectology

An interesting difference between social science and linguistic questionnaires is the fact that one of the most important tools in social science questionnaires, which is a must-have feature in that discipline, is very difficult to incorporate in a social dialect WQ. This feature is called **multi-item scaling** and refers to the measuring of one attribute with multiple questions. For instance, if a psychologist considers measuring the level of extraversion in a person, the questionnaire would include a number of inter-related questions and scales, not just one question on the topic. Then, the aggregate score of the questions targeting one feature is calculated, which yields much more reliable results than just one question on the issue. Multi-item scales are important, as it has been shown that

the actual wording of the questions assumes an unexpected importance; minor differences in how the question is formulated or framed can produce radically different levels of agreement or disagreement. (Dörnyei 2003:33)

Multi-item scaling is the antidote to this problem. The score averaged over a number of differently worded questions on the same issue is more likely to balance out individual weaknesses of individual questions.

This key feature of social science questionnaires is, unfortunately, only of very limited use in social dialectology. This is because of the different types of questions dialectologists and sociolinguists need to ask. In social behaviour the issues to be polled are generally relating to beliefs, opinions and attitudes, which are quite easily asked in five different ways (such as attitudes towards free speech, sympathy scores of politicians or personality traits such as honesty). In dialectology, by contrast, quite a number of linguistic items have unique conditioning factors that make it very difficult to link them with other items. Even if we consider related questions, such as yod-dropping in student, news and avenue, we see that all three contexts and lexical items show their individual behaviours and express different social connotations so that they are usually not collapsed into one score.

It is the same with lexical variables: whether someone uses *tap*, parkade and *toque* are three isolated cases of linguistic variables that will hardly be grouped together in one kind of measurement. While in these three variants a combined score would indicate the use of Western Canadian variants over American variants, devising questions that target the same linguistic item and do not seem repetitive is challenging.

Where multi-item scaling would be possible, it simply has – to my knowledge – not been done. For grammatical patterns, one could offer different contexts for the same grammatical rule (e.g. case following prepositions: *my* house is different from/ than/to yours, asked again in different contexts, e.g. *his* car is different from/ than/to hers). Researchers probably consider such questions as repetitive and as too obvious for a respondent not to align the answers – after all, in Western societies consistency in linguistic conformity is a compulsion.

In case of the self-report form or similar forms for linguistic phenomena in dialectology and society, it may be very sure to have a number of questions to this approach, which will not be repeated at each item. This means there is not much room for the multi-item scaling.

Inter-relatedness

Inter-relatedness is another way to employ this form of knowing. In closed-response forms, the self-report form, for such cases a positive statement is constructed, contexts such as a rating scale, in positive statements it means the respondent is not sure.
in linguistic forms has been taught to the population as a whole since the onset of compulsory schooling.

In cases where parallel forms of asking questions is possible without sounding too obvious and repetitive, however, multi-item scaling would be an area of improvement for linguistic WQs. As we said, multiple-item scaling do not work well in social dialectology due to the nature of the variables, which means that researchers need to be very sure that their questions are as unbiased as possible. The closest one can easily get to this approach is by asking inter-related questions, i.e. questions on the same type of variable in a number of contexts. An example of these is discussed below.

**Inter-related items**

Inter-related questions are useful for variables that are formally the same (e.g. word form or syntactic construction), but require special situational contexts. They are a way to ensure that the respondent needs to offer more than one cue correctly for closed-response items to demonstrate that the variable is part of her grammar. One such case is the lexicogrammatical variable *positive anymore*, i.e. *any more* used in positive sentences. Standard English, generally, requires *any* instead of *some* in negative contexts only, e.g. *There are no cigarettes anymore*. Positive anymore, however, occurs in positive sentences with a different meaning: e.g. *John smokes a lot anymore*, which means that *John smokes a lot nowadays* (Chambers 2007).

The semantics is complicated because of the presence of *negative anymore* in any standard and many non-standard dialects of English (e.g. *John no longer smokes anymore*), a presence which clouds the picture considerably. The existence of two similar constructions, *negative anymore* as a standard feature and *positive anymore* as a non-standard feature may lead respondents to report on their use of the Standard English variety that was taught (or at least reinforced) in school instead on their knowledge of the non-standard construction.

With *positive anymore*, the questions are necessarily more ambiguous than for yod-dropping. This should not come as a surprise as we are trying to elicit relatively fine semantic nuances. The DT questions 48, 50, 51 and 54 are shown below in (7.8):

(7.8) a. 48. Someone said, *John smokes a lot any more*. Does this mean that
John hasn't been able to cut down, let alone stop?
or John wasn't smoking much for a while but now he is?
or John has almost quit?

b. 50. What does *any more* mean in *John smokes a lot any more*?
still
or nowadays
or negative
(7.8) c. 51. That sentence, *John smokes a lot any more*, does it sound like something you might say under the right circumstances? or others might say though you wouldn’t? or no one you know would say?

d. 54. Someone said *Harry likes rock music any more*. Does this mean that Harry’s turned off rock? or Harry’s finally seen the light? or Harry’s always been a great rock fan?

Note that the questions do not all immediately follow one another. Example (7.8a) gives three paraphrases of the entire sentence – all written in informal style. The chance that a respondent is going to guess the answer right is 1 in 3 (right answer: the second option). We therefore need more safeguards to avoid data that would be unacceptably skewed (note also that respondents are not given the option to say: “I don’t know”, which would present a different scenario). Next, in (7.8b), we do not translate the meaning of the entire sentence, but we merely elicit the meaning of “any more” in this context: again, the chances are 1 in 3 to guess it right (which many people would rather do than not answer at all). The combined likelihood that all three interpretations would be guessed correctly (7.8a, 7.8b, 7.8d) would be \(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}\) and thus only a miniscule chance to get all three answers right. If one is looking for a safe method to discriminate guessers from users and one does not wish to work with open-response items, this method would do the job.

In addition, question (7.8d) uses highly idiomatic language such as “turn somebody off rock”, “seeing the light” (getting to appreciate rock music) that is likely to test the certainty of the respondent. As the meanings are ambiguous, a respondent would have to be quite certain in the interpretation of *Harry likes rock music any more* to select the second option (“has seen the light”). The lack of a “Don’t know” category was apparently inspired by the idea to elicit passive knowledge rather than active knowledge: the goal was to detect all people that would passively know the correct meaning, although they might not use the construction themselves. Example (7.8c), then, directly asks the respondents whether they use the construction or are passively familiar with it. Matching all four questions offers an example of the usefulness of inter-related questions and allows the identification of passive users to a degree that would not be possible for any of these questions in isolation.
Open-response items

Open response items are very frequently used in the field and most often for lexical variables. In social science research, however, these items carry the connotation of being suboptimal questionnaire items. Their major shortcomings are that answers are more time-consuming for respondents to write and significantly more time-consuming in the analysis. Open-response items, however, do play a vital role in dialectology, as only open answer questions offer the range of variants some dialectological studies are interested in.

There are a number of ways to elicit open responses, which range from open text fields, to semi-structured answers to more innovative modes of responses. For instance, Preston’s (1989; Preston & Long 2002) method of having respondents identify areas on a map for certain linguistic attributes, is a discipline-related response mode that is a form of open response to be discussed in Section 7.3.4. The more traditional types of questions are shown below:

(7.9) a. What do you call someone who studies “too much” and tries very hard to impress the teacher?

b. What do you call the building where people park their cars when they go shopping?

(Vancouver Survey, q29)

The examples in (7.9) elicit lexical items. Note that the description of the lexical target, rather than asking “what is your word for brown-noser, keener etc.”, renders these questions indirect rather than direct. The example in (7.9b), however, is an open-response version of (7.4b). Open-answer questions trigger more lexical variation than can be listed in closed-answer questions, which, in turn, are more quickly answered than open-answer questions and help focus the respondent’s attention on the variable.

For finer semantic distinctions, open answer questions are often the most effective method. The examples below come close to a short-answer type of question, where respondents are invited to comment on a variable of interest.

(7.10) a. What is the meaning of hydro in Chris bought a few hydros down the street.

b. Three interlinked questions on take up variable

(Vancouver Survey, q21)

(Language Survey Vancouver-Washington State, 2011)
The meaning of *hydro* in (7.10a) targeted a type of marijuana joint and a mere six of 429 respondents identified this special, fringe-group meaning. Example (7.10b) is an expanded version of an open-response type of question. It first invites a paraphrase of the meaning of a sentence in open text format. It asks *who might use this sentence* and then invites further, optional comment. One would dedicate such space – three of the recommended maximum of a 100 questions – to a variable of prime interest that is semantically rather narrow, such as *TAKE UP #9*.

For questions that are asked in a direct fashion, or on a binary basis (e.g. yes/no answer choices) it is recommended to *INVITE ELABORATE COMMENT*. Pratt (1983) has found that only if the respondent (or interviewee) is able to fully contextualize the linguistic item, one can consider her as actually knowing how to use the item. This method offers an alternative to inter-related questions by replacing some closed-response items with one open-response item.

*Mixed-response types*

One technique that combines some of the benefits of closed-response questions (quicker answer submission and ease of analysis) with the benefit of open-response types (greater range of variants) has been used in a number of studies and may be called major response types plus other field. The examples in (7.11) show the principle:
one combines a list of the most prevalent response types with a field "Other" and invites the addition of other, not-listed terms.

(7.11) a. Which of the following names do you use for a waterway smaller than a river?

☐ crick
☐ stream
☐ brook
☐ rill
☐ snye
☐ run
☐ creek
☐ other ____________________________

(7.11) b. Please select the word you would use most often in your everyday speech. Select more than one answer only if necessary. If the word you use is not listed, please write it in.

(7.3.2) From raw questions to questionnaire items

Now that the most common question types have been introduced, we need to address the types of linguistic variables that can be elicited, which are first treated as "raw questions". An obvious way to arrive at raw questions is by reading the literature and by looking at previous studies. For newer variables, the process is less straightforward: it is generally recommended to start with a "question repository" where one collects potential variables that may be included in the questionnaire. Raw questions are only the first step and must not be confused with final questionnaire items.

Transferring raw questions into items involves an analysis of each question with the general research question in mind, which will help decide whether to include a particular variable or not. Once this decision has been made, the next step is to check the question type and then to explore different kinds of question wording. The quality of the questions is a most important and crucial point, because if not formulated in the best possible way the results will be biased. In order to turn raw questions into questionnaire items, three principles should be kept in mind:
- **Accessibility**: respondents have to have conscious access to the given linguistic feature
- **Clarity**: the questions must be clearly and straightforwardly worded and tailored towards a diverse readership
- **Conciseness**: the items must be short and to the point

**Accessibility** of a speaker to the polled linguistic feature is the most basic requirement. It needs to be assessed for each question individually and thoroughly. The two examples below are typical WQ questions with binary answer choices and may serve to illustrate the principal type of accessibility assessment:

(7.12) a. Which do you say?
   Just between you and me, your aunt is often wrong.
   Just between you and I, your aunt is often wrong.
   (Dialect Topography, q56)

b. Does LEVER, as in 'Pull the lever', rhyme with clever or cleaver?
   (Dialect Topography, q35)

While (7.12a) elicits responses to syntax (or phraseology) about *between you and I/me*, (7.12b) polls respondents on a pronunciation variable in the form of a lexically conditioned phonetic realization of the stressed vowel in the word *lever*. The binary answer choices above in (7.12b) are designed to tackle phonemic variation in a given lexical item.

It is important to assess carefully whether respondents have conscious access to this kind of information, which will involve trial runs for new variables. In (7.12a) respondents can be expected to give a reliable answer, as prescriptive grammar questions such as this one have traditionally been at the forefront of grammar classes in compulsory schooling in western countries. With this question the challenge (and the risk) is to get the respondents to report what they actually say rather than what they were told by their English teachers in school. In other words, we need to assess whether there is a potential social stigma attached to one variant or another and if there is, we need to clarify whether there are sufficient indicators that the variant is not too highly stigmatized to lead respondents to report on what they "ought to be saying". If so, we would be accessing language attitudes and not the reported behaviour. In the Canadian context, we have seen in Chapter 4 that the formerly stigmatized construction between *you and I* is now increasingly heard in the Canadian media landscape. It seems warranted to include this variable in a self-report survey, but one would need to keep this possible bias for underreporting between *You and I* in mind in any interpretation.

In (7.12b) the accessibility problem is different. For phonemic questions, the major pitfall is to elicit information that is phonetic, not phonemic in nature. For instance, if we have respondents say in the word *ban* and *bat* which they would get more statements that are trustworthy, then we get a higher token (ban a higher token (in ban) (in bat) (Bolinger 1989) and *ban* is higher token (ban) (in ban). Respondents may have phonemic values, that is, a counter example, related to a lexical network, as the contextual (or situational) network is related to the situation.

**Clarity** may have been brought up in Gold (2001: 46) for the following purposes (in a slightly different double meaning):

(7.13) **Do you**

Example: We presented the question for clarity whether the respondents with whom we were speaking would need to accommodate their language use (yes, no, or don’t know) to accommodate the language use of Gold.

As a result, Gold decided to include the question in the form **Do you**
instance, if we were to ask whether our Canadian respondents pronounce the vowel in the word ban with a lower, higher or equal tongue position than the vowel in bat, we would get nonsensical results. We would get some results, but they would not be trustworthy, because this variation is phonetic and generally not accessible to the language user. It requires spectrograms and a large sample of pronunciations of each token (ban and bat in certain contexts) to show that Canadians raise the vowel <a> (a higher tongue position) before nasals (in ban) but not before voiceless obstruents (in bat) (Boberg 2010: 144). Some respondents with a good ear might notice that ban is higher in vowel height than bat, though generally, people would be confused. Respondents cannot reliably report on phonetic variation, but they can report phonemic values, such as whether they pronounce leaver as rhyming with clever or clever. With phonemic questions it is crucial to give clear reference points that are stable in pronunciation across age groups, social groups and dialect regions. In the current example, reference points either and never rather than clever and cleaver would not work, as the pronunciation of either varies between [i] and [a]. For variables where situational variation plays a role and speakers might use both variants depending on the situation, WQs are not a good method of data collection.

Clarity of the item is another important aspect and the researchers themselves may have problems assessing this aspect objectively without outside help. Brown (2001:46) spots the problem in the researcher’s familiarity with questions and their purposes that the researchers themselves “may not be able to spot ambiguities and double meanings”. One such example is shown in (7.13).

(7.13) Do you use the expression So that’s what he thinks, eh?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>sometimes</td>
</tr>
</tbody>
</table>

Example (7.13), from the 1972 Survey of Canadian English (SCE), is a bad example for clarity: it asks for two issues to be assessed (a “double-barrelled question”). First, whether the construction with eh is used at all (yes/no) and second the frequency with which it is used (in answer C). It would have been clearer to just ask for general use (yes, no) and to possibly offer the qualifier “do you use it in the right situations” – to accommodate for the social stigma that some uses of eh carry (e.g. Avis 1972; Gold 2008).

As an early study, the SCE made more than one suboptimal choice and most problem questions can be attributed to lack of clarity. The variable chesterfield can serve to illustrate a number of potential problems in questions design. In SCE it was offered in the format shown in (7.14a):
(7.14) a. What do you call a piece of furniture that seats two or three people in a row and has upholstered arms and back?
   A. sofa
   B. chesterfield
   C. davenport
   D. by any other name  (SCE, q29)

   b. What do you call the upholstered piece of furniture that 3 or 4 people sit on in the living room?  (DT, q2)

SCE offers answer options, yet couch is not listed as an incoming variant – a focus group interview with younger speakers would have produced that variant, which was first reported around the time, e.g. Gregg (1973). It is paramount to offer a description that is maximally transparent for as many people as possible, ideally all participants. However, the description cannot be too precise or too cumbersome. Instead, one must work with some degree of "common sense". As this feature is culturally sensitive, in multicultural societies we face a bigger issue with clarity than first meets the eye. A pilot study that examines the question format is therefore absolutely necessary. One efficient way to do so is to have people read the questionnaire and comment, in a think-aloud protocol, on the questions themselves: are they confused? If so, where is the problem? Are they clear on what is being asked of them?

For chesterfield, Dialect Topography (DT) takes a different route, as shown in (7.14b) by offering a slightly adapted question that seems to capture the essential features of the item in question: 3 or 4 people are seated (two avoid confusion with the variant love seat, which generally seats two people), adding the typical location in a house (living room) yet removing "clutter" by not specifying which parts are upholstered. The DT question is shorter, which is generally also good.

Conciseness. As Brown (2001:45) states: "As a rule of thumb, short questions are good questions". As long as the target form is clearly identified, which applies better to (7.14b) than the longer question in (7.14a), the question is considered as detailed enough. One goal should be to write the shortest question possible that identifies the item for as many people as possible. This is easier said than done and will not always be accomplished. But the goal must be to come as close as possible to the shortest question that unambiguously identifies the item and sounds "natural".

Pilot testing with an array of speakers of different backgrounds (that are found in the target group) is an essential step towards finding the right balance between conciseness and detail (Section 7.3.7). Testing is a crucial part of all WQ design and it is often only at this stage that difficulties with and differences between question phrasing are discovered.

7.3.3 Selecting response options

The question format and the version of a response option are important considerations in elicitation and questionnaires (SCE, q30). Example (7.14a) is a standard, everyday, own use ("by anyone") option. The OE equivalent is "in the house".

Because of the subjective and cultural nature of dialect phenomena, there are no linguistic labels and it is often necessary to use linguistic techniques from sociolinguistics such as the translation of language proficiency levels or Estimation (SCE, q31).

Inventories. As we have noted above, one aim of elicitation is to capture the forms as they are used in everyday life, and their status within a sociolinguistic framework. The elicitation tools are designed to capture the forms as they are used. The next stage requires a careful analysis of the data.

The choice of forms to include in one's inventory is crucial and depends on the purpose of the inventory (SCE, q36). For example, the form of a verb in a specific tense or a specific location in a sentence may be of interest. The inventory must be designed to capture the forms in the context in which they are used.
7.3.3 Self-reporting linguistic behaviour

The questions discussed so far in this chapter have all been designed for self-reporting of a respondent’s typical linguistic behaviour; they are all traditional self-report questions of linguistic behaviour. There have only been two exceptions to this: Example (7.3), which asks the respondent to report on her perceptions, not on her own use (see Section 7.3.4), and (7.8c), which includes questions about community use ("others might say but you wouldn’t") (see Section 7.3.5).

Because the reader will thus be fairly familiar with traditional linguistic behaviour reporting, we only need to highlight possible limitations of the self-reporting of linguistic behaviour. We will first address the issue of linguistic inventories and social correlations elicited in WQs before we will look at two more recent elicitation techniques from the social study of syntax that are not found in traditional WQs. These are translation and reformulation tasks and grammaticality judgements with Magnitude Estimation Tasks.

Inventories vs. social correlations

Some findings suggest that WQs are not the ideal method for data collection when one aims to study the social correlates of stigmatized linguistic phenomena. Cornips (2002) shows in a study on Dutch verbal complementizers om and voor that both forms are reported in the questionnaires, yet are distributed differently when compared to natural speech. While inventories of variants may be established with WQs, their social distribution may be skewed. It is essential to note that the prestige variant is om and that voor is categorically avoided by educated respondents. This means that the elicitation of socially stigmatized variables (in this case of non-educated speech) requires some fact-checking.

The relationship of reported use to actual use is the matter of some debate. On the one hand, there are those who treat reported data as representative of the evaluation of linguistic features. On the other hand, reported use can be treated as indicative of some form of use that is only indirectly linked to actual use. It seems that Seiler’s statement is closest to the nature of the data:

> It is impossible to compare the informant’s linguistic behaviour in the interview with his/her ordinary speech. [...] It seems that the indirect method yields reliable results as far as the geographical distribution of linguistic variables per se is concerned. However, its limitations are to be found in the exploration of the exact functional properties of the variants [...].

In other words: WQs generally produce very reliable inventories of variants, but they seem to offer only limited cues towards the linguistic behaviour of the respondent in concrete functional contexts. After all, one does not ask respondents in the questions presented so far to “rate” the social acceptability of particular constructions. Such
questions would yield different results from the type of “What do you call/Which do you say” type we have used so far. WQs, however, do not always yield the full social picture of variant distribution and any WQ study must be very aware of this limitation.

Socio-syntactic reformulations

Traditionally the domain of generative linguistics, grammaticality judgements are now used in more sophisticated versions in the new field of “social syntax” or socio-syntax. These question types increase the versatility of WQs for the study of syntactic and morphological phenomena beyond the traditional types and questions. There are two basic question types. The first is a translation or “reformulation” task where a test sentence is given with the instruction to “translate it into the variety under study.” The test sentence can either be given in the standard variety with the request to translate it into the target variety, as applied in *Syntaktischer Atlas der deutschen Schweiz* (*Syntactic Atlas of German Switzerland*, e.g., Glaser 2000). With this method, we run into problems somewhat reminiscent of Wenker’s missing transcription system, because respondents are required to render dialectal forms in writing that is usually only performed in the standard variety. The results, however, appear more promising because their focus is, unlike Wenker, on syntax and not phonetics:

(7.15) GerG Ich sprach schon
I spoke already

AutG I hob scho g’redt (*reden* [g’redt] & *sprechen* [g’sprechen] are synonyms)
I HAVE-AUX already spoken-PASTPART

Example (7.15) illustrates the principle by using a Standard German test sentence and its Austrian German translation – one can see that even in the absence of a normed writing system the difference in the tense (preterit *spoke* vs. perfect *have spoken*) can be clearly rendered and it is immaterial what letter to indicate vowel quality is used.

An interesting variant of the reformulation task is suggested in Buchstaller and Corrigan (2011:32–3). By offering the test sentence in the vernacular and asking respondents to perform a linguistic task, some of the issues of prescriptive influence can be by-passed. Example (7.16) shows their instructions and a test sentence.

(7.16) You will hear and then see a question, and you will be asked to turn it into the equivalent statement that sounds natural to you.

Training session

Question: Was John’s friend Ian at the party?
Statement: John’s friend Ian was at the party.

Now please do the same for the following sentences:

Question: Will it be Susie what presents the cheque?
Statement: ____________________

(Buchstaller & Corrigan 2011:33)
It is reasonable to assume that respondents who are familiar with relativizer *what*, which is used in the Northern English & Scottish locales that Buchstaller and Corrigan are studying, will be more likely to produce it in the positive sentence than if a standard English prompt was provided (e.g. *Will it be Susie who presents the cheque?*).

**Magnitude Estimation Tasks and grammaticality judgements**

Socio-syntacticians also use alternative methods in grammaticality judgements by employing a methodology known as the MAGNITUDE ESTIMATION TASK (MET) (Bard et al. 1996), which has been applied over the last decade or so. MET is a departure of traditional assessments of grammaticality on a binary (nominal) scale (yes or no). Inspired by psycholinguistic methods of measurement, METs offer respondents a reference sentence, which they are required to rate: it is up to the individuals to rate the sentence 1 or 5 or 100, as the value in isolation is immaterial. Only when a second sentence is rated and a relationship between the example sentence score and a test sentence is established, the numbers gain significance. An example is given in (7.17):

(7.17) **Reference sentence:**
I'm going home and got an umbrella.

**Your rating:**

10

Now, please rate all sentences below in relation to the sentence above:

**Sentence**

<table>
<thead>
<tr>
<th>Your rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The man put his coat on a hanger.</td>
</tr>
<tr>
<td>2. That's what I hate, is that she's always late</td>
</tr>
<tr>
<td>3. I'm not going to eat nothing hot no more</td>
</tr>
</tbody>
</table>

(Buchstaller & Corrigan 2011:36)

An advantage of this method is that it allows respondents to express subjective feelings of correctness. Sentence #3 was rated lower than the reference sentence because of its multiple negation though the person could rank it as low as 1 or even negative scores, which could be easily considered in the analysis, which deals with relative scores only.

A version of the MET is to offer ratings in a visual form. This adaptation offers a visual line that represents the scales of acceptability and respondents are asked to mark their assessment on the line, as shown in (7.18):

(7.18) **Please use an 'X' to rate the acceptability of this sentence:**

I'm going home and got an umbrella.

Now please do the same for the following sentences using 'X' again – this time to represent whether you think these are better or worse than the sentence above in bold:

I really wants to buy those red shoes.
Sometimes the girls thinks it's boring.
The mark on the lines will then be measured for its distance from the edges (in millimeters, Buchstaller & Corrigan 2011:38). A powerful argument for this question format is that some respondents may have problems handling the abstract (and open-scale) ranking system shown in (7.17). Consequently, the respondents "with low numeracy skills" who would otherwise fail the original MET are allowed to partake with the visual format (ibid: 37).

These methods represent some of the new approaches to the study of dialect syntax for which WQs play an important role. They allow a reassessment of traditional WQ methodology, by introducing gradient scales. Thus questions such as John smokes a lot any more (is this something you would say/someone else/no one would) would not necessarily need to be answered in a binary fashion. Combined with a tendency of indirect reporting via the reporting of community behaviour rather than personal use, one would defray prescriptive pressures on the respondents (see Section 7.3.6).

7.3.4 Self-reporting language attitudes and perceptions

Language attitude studies and perception studies make explicit the underlying beliefs of respondents' conceptions of language. WQs have been used extensively to assess people's beliefs, so much so, that they can be considered the field's primary data collection tool. We have seen, in Chapter 5 for instance, Jenkins' assessment of teacher attitudes towards English as Lingua Franca. The prominent role of WQs in this area has been long in the making. Mather and Speitel (1975:25, fn 80) anticipated the use of WQs for "evaluation procedures, i.e. specific questions in which the informant is called upon to comment on his linguistic responses".

Attitude and perception studies tap into the beliefs and attitudes of respondents, which is different from the reporting of linguistic behaviour or the knowledge about linguistic forms found in a community. For instance, in the Canadian context, data on language attitudes towards Canadian English has traditionally been hard to come by. Those studies that do exist, however, have successfully employed WQs (Owens & Baker 1984; Gulden [Halford] 1979; Warkentyne 1983).

Language attitudes

Language attitude studies aim to elicit the sociolinguistic evaluation of a feature, not its use. For instance, we might ask someone to "rate this construction for pleasantness" or the like and offer a scale for assessment from "very" to "not at all", whereas in usage-based self-reporting we use the typical questions discussed in Section 7.3.3. The study of language awareness is a central part of the WQ tradition and central to sociolinguistics, since social evaluations of a linguistic feature are an intrinsic part of an explanation of linguistic change (Weinreich, Labov & Herzog 1968). William Labov, among others, has been very clear in making this point:
The speech community is not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms; these norms may be observed in overt types of evaluative behaviour, and by the uniformity of abstract patterns of variation which are invariant in respect to particular levels of usage.

(Labov 1972: 120–1)

Put plainly, membership in the speech community is not defined by the simple notion that people speak the same, but by the more abstract notion that they evaluate communal linguistic variations similarly, and by shared linguistic patterns. It is here where attitude studies provide much valuable insight. One attitude result, from a Vancouver sample of 429 respondents (2009) is shown in Figure 7.2. The data was collected with question (7.19):

(7.19) Is there a Canadian way of speaking?

- Strongly agree
- Agree
- Somewhat agree
- Somewhat disagree
- Disagree
- Strongly disagree

The question's vagueness is intended, as we did not want to specify the details, but primarily gauge the 'feelings, wishes, or attitudes' of the respondents as expressed in a Likert Scale rating.

Figure 7.2 shows the influence of post-secondary education, with the university educated respondents' answers on the bars on the left and the non-university-educated on the right. It can be seen that the strongest believers in Canadian traits are found in the university educated (columns 1 “strongly agree” and 2 “agree”). In column 3 “somewhat agree”, the non-university educated take over and also lead the sceptics in columns 4 “somewhat disagree”, 5 “disagree” and 6 “strongly disagree”.

![Figure 7.2 Answers to “Is there a Canadian way of speaking?” by education](left: university education, right: up to community/trade college)
Perceptions

As mentioned in Chapter 6, Wallace Lambert's (1960) study is generally considered as the starting point of speaker-assessment studies. How linguistic varieties are perceived can be assessed in a number of ways (see Giles and Billings 2004 for a succinct genesis of the field). The most famous method is Lambert and associates’ matched-guise technique, in which bilingual speakers, one time speaking French, the other time speaking English, provided the audio cues that listeners rated by a number of character traits of the person behind the voice they just heard, such as good looks, self-confidence, kindness, ambition, likability and even body height. The method was soon extended to varieties of the same language, from which a strong research tradition developed.

A more recent development and extension of speaker (in this case hearer)-based assessment was pioneered in Preston (1989) and following publications with "perceptual dialectology", which has developed a methodology for WQs with ratings and map-based response types to elicit perceptions of dialects. This methodology puts the regional perspective of language front and centre but does not deal with "factual" distributions of dialects and dialect regions, yet with their perceived constructions. Illustration 7.1, for instance, shows the information gathered via map-delimitation and labelling tasks, in this case from a young southern Michigan student.

Illustration 7.1 Raw material from Preston’s map task (18-year old southern Michigan female student) (Preston 2005)

One can see from the region delimitations but more so from the region labels that the method grants access to "the ordinary speaker’s understandings of language variation" (Preston 1989:2). This ethnolinguistic (macrolinguistic) approach to linguistic
perception can probe into popular taxonomical aspects of a language or variety, the social characteristics that are "overly regarded by a speaker as supporting linguistic differences" and an ordinary speaker's belief in geographical language differences (ibid).

All of the above is present in the WQ answers in Illustration 7.1 – from "Hillbillies: Hick talk" to Canadian slang "Ey" and the "society freaks" on the East Coast: there is a mix of perceptions of people and their varieties. When analyzed in a consistent way over a large sample, one arrives at very coherent pictures of perceived dialects, as we will see in Illustration 7.2.

Preston's approach has been applied to Canada in McKinnie and Dailey-O'Cain (2002), whose goal was to document young Ontarians' and Albertans' "perceptions of the cultural and linguistic make-up of the country, as well as their attitudes towards English speakers in other geographical areas in Canada" (p. 277). Following Preston's (1989) methodology, respondents were asked to judge each province or territory's variety for perceived 'correctness', 'pleasantness', and similarities to their own variety on a scale from 1 (most) to 5 (least). A possible attitude elicitation question is shown in (7.20):

(7.20) Rate the English of Canada's provinces and territories on a scale
from 1 "most correct" to 5 "least correct":

<table>
<thead>
<tr>
<th>Province</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Columbia</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Alberta</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Manitoba</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

In part two of their study, the authors provided respondents with a map of Canada with only provincial and territorial boundaries and asked to "demarcate areas in which they believe people speak the same as each other but differently from people in other areas of the country" (p. 279). The study yielded three types of data: quantitative data on correctness, pleasantness and 'sameness', quantitative data on regions identified in the map task (for which all regions were considered that were identified by at least 10% of the respondents) and qualitative data via the labelling of the dialect regions.

The results are highly interesting and show some striking consistencies. For instance, young Albertans view their province's English as very pleasant. As Table 7.3 shows, with a mean rating of 2.05, Albertan English is considered as pleasant. BC English, however, is considered even more pleasant, while Ontario English is rated comparatively low. This is surprising, as Ontario is the traditional economic powerhouse of the country and such assessment is counter-intuitive to ad hoc assessments.
Table 7.3 Perceived pleasantness of dialects (McKinnie & Dailey-O’Cain 2002:280)

<table>
<thead>
<tr>
<th>Albertan respondents’ perceptions of “pleasantness”</th>
<th>Mean</th>
<th>s.d.</th>
<th>Ontario respondents’ perceptions of “pleasantness”</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>1.95</td>
<td>1.09</td>
<td>ON</td>
<td>2.17</td>
<td>1.20</td>
</tr>
<tr>
<td>AB</td>
<td>2.05</td>
<td>1.09</td>
<td>BC</td>
<td>2.22</td>
<td>1.12</td>
</tr>
<tr>
<td>MB</td>
<td>2.33</td>
<td>1.07</td>
<td>AB</td>
<td>2.57</td>
<td>1.05</td>
</tr>
<tr>
<td>SK</td>
<td>2.39</td>
<td>1.09</td>
<td>PE</td>
<td>2.67</td>
<td>1.11</td>
</tr>
<tr>
<td>NB</td>
<td>2.59</td>
<td>1.07</td>
<td>SK</td>
<td>2.68</td>
<td>0.93</td>
</tr>
<tr>
<td>PE</td>
<td>2.60</td>
<td>1.08</td>
<td>MB</td>
<td>2.69</td>
<td>0.95</td>
</tr>
<tr>
<td>ON</td>
<td>2.63</td>
<td>1.19</td>
<td>NB</td>
<td>2.76</td>
<td>1.06</td>
</tr>
<tr>
<td>NS</td>
<td>2.64</td>
<td>1.11</td>
<td>NS</td>
<td>2.81</td>
<td>1.06</td>
</tr>
<tr>
<td>YT</td>
<td>2.68</td>
<td>1.14</td>
<td>NL</td>
<td>2.94</td>
<td>1.35</td>
</tr>
<tr>
<td>NT</td>
<td>2.77</td>
<td>1.12</td>
<td>YT</td>
<td>2.95</td>
<td>1.02</td>
</tr>
<tr>
<td>NL</td>
<td>2.89</td>
<td>1.29</td>
<td>NT</td>
<td>3.02</td>
<td>1.07</td>
</tr>
<tr>
<td>QC</td>
<td>3.06</td>
<td>1.22</td>
<td>QC</td>
<td>3.07</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Ontarians, by contrast, rate their English as the most pleasant (2.17), followed by BC (2.22) and they are more generous with Albertan English (2.57) than vice versa. However, they share the low opinion of Quebec English with Albertans, but assign generally worse scores than the Albertans.

In terms of perceived “correctness”, BC is considered the most correct English by both Albertans and Ontarians (ibid: 283), followed by their own province’s variety in second place. In third place are Ontario, respectively Alberta. At the other end of the spectrum, Newfoundland and Quebec are considered the second least correct and the least correct varieties. McKinnie and Dailey-O’Cain (2002:284) remark that the ratings of Newfoundland and Quebec English “are only slightly higher than three, on a scale where three is the mid-point, which indicates that [Canadians] find these varieties neither particularly ‘correct’ nor particularly ‘incorrect’.”

The analysis of the map-drawing and labelling task affords a more nuanced view on the perception of dialect regions in Canada that may be compared and contrasted with the regional dialect zones derived from linguistic data in Table 6.10 (p. 218). Illustration 7.2 shows the accumulated responses with a minimal overlap of 10% of respondents. Together with the labels assigned to these regions, one arrives at very interesting insights into perceptions of Canadian English.

The most frequently identified regions are Quebec and Newfoundland, which are “simply the most distinct forms of English in Canada” (ibid: 292), whereas in the case of Quebec the L2 English of francophones is commented on with labels like “Franglais” or the extremely pejorative label “Bad French Bad English”. Ontario is singled out next, as are the Maritimes (New Brunswick, PEI and Nova Scotia) and the Territories. The
latter are perceived collectively because their English appears to be "Influenced by First Nations", as one respondent put it.

BC and Alberta are discriminated by 10–25% of respondents, while the same percentage recognizes a Southern BC variety and a Southern Ontario variety. While Southern BC was rated as "Canada's most correct English" or "more proper" and "normal", the Southern Ontario region received less favourable attributes, such as "American style" or "Industrial Northern States type accent" (ibid: 290–1).

It is obvious that these perceptions are of relevance. Not just to the extent that they inform decisions about language in real-world contexts, but as part of the description of a sociolinguistics that places social aspects and interactional issues at the centre. For instance, from traditional WQ data, one would have no means to gauge the radically different assessments of the Southern BC and Southern Ontario varieties. In addition, McKinnie and Dailey-O'Cain point out that the Englishes of the three economically most powerful Canadian regions – BC, Alberta and Ontario – were "consistently rated most highly in terms of 'correctness', which is an important sociolinguistic finding that is matched in varieties, though is perhaps less obvious in the Canadian context.

**Illustration 7.2** Perceived dialect regions by 100 Ontarians and 100 Albertans
(McKinnie & Dailey-O'Cain 2002:291)

- 10–25%
- 26–45%
- 46–65%
- 66–85%
- 86+

To be continued...
7.3.5 Community-reporting of linguistic behaviour

While self-reporting questions have been used most frequently in the Canadian context, the reporting of general community-wide linguistic behaviour has a long tradition since Stadler’s early 19th-century dialectological approaches and, of course, Wenker’s method of sentence transcriptions by school teachers. So far we have seen community-reporting questions interspersed with self-reporting questions that have been worded in ways such as the following:

Do you use the construction under the right circumstances?
Do you know of others who use it?

No one you know uses the construction.

In contrast to self-reporting, community-reporting questions do not target the linguistic behaviour but the knowledge of the respondent: does s/he know anyone who uses the variable? They are thus knowledge questions, not behavioural questions in Dillman’s (1978) typology and would need to be framed accordingly.

Most recently, Buchstaller et al. (2013) rediscovered community-reporting as they chose to frame questions on non-standard grammatical items in a community-reporting way, in order to circumvent the social stigmata of some variables. In other words, rather than asking whether a person uses the construction They wants more cookies, they asked whether people in the community use the construction on a scale of grammatical acceptability, as shown in (7.21):

(7.21) 1: This type of sentence would never be used here – it seems very odd.
2: This type of sentence is not very common here but it doesn’t seem too odd.
3: I have heard this type of sentence locally but it’s not that common.
4: People around here use this type of sentence a lot.

Please rate these sentences as described above.
The local supermarket got robbed and the police were looking for a witness.
They were asking a group of children whether they had seen anything.
Suze pointed at a little girl. She said ‘That’s the girl seen it’.

1______________2______________3______________4

(Buchstaller et al. 2013: 95)

Using WQs, Buchstaller et al. (2013) assess both morphosyntactic and phonological features, which offer them detailed data that had been unavailable on their phenomena. What seems important is that they approach syntactic and morphological non-standard phenomena in a more indirect way than previously. Rather than asking whether someone uses an non-standard construction (or someone respondents know), they ask about behavioural norms in a community. This more indirect approach seems poised to offer new avenues for the successful employment of WQs in social dialectology.
As a positive side-effect, providing non-standard cues is one of the most effective and simple means available to mitigate standard language interference. The use of written language is, at least until the advent of digital and social media communication, generally associated with more standard-like behaviour, which is an effect that would cause respondents to answer with more formal responses. With non-standard cues, this kind of standard-interference is reduced as much as possible for the WQ method. There are more methods to reduce prescriptive influences in WQs, to which we turn next.

7.3.6 Mitigating prescriptive influence: Framing the questions

Standard sociolinguistic practice has long favoured the study of observed linguistic behaviour whose (elusive) goal is to observe, usually via audio recordings, natural speech situations with the aim to capture unmonitored, vernacular style. Observation is only possible with an observer in place which increases the likelihood of linguistic monitoring. With a written, self-reporting questionnaire, we do not have the same kind of problem, albeit the respondents who fill out the questionnaire know full well that somebody is going to read their answers at some point and he or she may even assume that the person handing them the questionnaire might do that. Still, while we have a kind of implied Observer’s Paradox, the fact that WQs do not ask for names, addresses or phone numbers – and are thus anonymous – would surely instil some level of confidence. If the questionnaire is returned by mail, the respondent is not asked to provide a return address, in an online format, the IP addresses would not be saved or exported into the analysis sheet. In this section we explore some verbal and non-verbal measures that help mitigate prescriptive influence.

The role of instructions

WQs have decidedly different methodological challenges to address than interviews. As respondents read linguistic questions, they interpret them from their various backgrounds and provide answers from among a set of choices (yes or no questions, a set of answer choices), and sometimes fill in their own response(s) to open-answer items. The written medium and its association with schooling is a complicating factor as some respondents would inevitably select the more formal choices that researchers generally do not want to elicit. To alleviate such behaviour to the extent possible, the respondents are informed about the purpose of the questionnaire, at which point a statement against prescriptive influence is warranted.

Dialect Topography, as seen in its questionnaire on the webpage, informs respondents about the general purpose of this study – words and their pronunciation in

Canada and bordering U.S. regions, but, more importantly, respondents are asked to report on what they 'say when among friends' and not on what they think is 'proper' linguistic etiquette. With these instructions, one aims to circumvent the effect of schooling ("good grammar" vs. "bad grammar") and the avoidance of linguistic behaviour that some speakers consider as not preferable. DT also asks respondents not to ponder over their answers and that their gut reactions usually work the best (third paragraph). The idea is that snap judgements produce more authentic assessments than well-crafted and reasoned responses.

The B.C. Linguistic Questionnaire as developed by Polson included in a later version quite elaborate and informal instructions, which are quoted in Illustration 7.3 from Stevenson (1976:84–5). The informality of style and directness of Robert Gregg's instructions (Stevenson's & Polson's supervisor and director of the BC Survey) could hardly be more extreme. It clearly shows that the "city boys" (in 1970s generic masculine wording) and "girls" were aiming to reduce prescriptive influence. In today's WQ one would probably aim to word a bit more concisely, but the general tone would still work.

**INSTRUCTIONS FOR THE B.C. LINGUISTIC QUESTIONNAIRE**

Please read these instructions before trying the questionnaire.

This is NOT a test. We do NOT presume to judge people's speech habits; we merely record them.

So when you answer the questions, 

**PLEASE put down what you actually say.**

Do NOT put down what you think you SHOULD say.

Do NOT put down what your friends and relations think you should say.

We picked YOU as our informant and we want YOUR answers. Don't let anyone else tell you what to put down. If someone else wants to add his two-bits worth and you think the information might be interesting, add a note to your answer telling who the other person is and where he comes from. But PLEASE don't let the other person influence your answers.

It isn't very likely that you'll be able to answer all the questions. Don't worry about it. The fact that you don't use a particular word can be just as important as the fact that you do. [...] Furthermore, we're city boys. If you would assume that we're shockingly ignorant about most things, and try to set us straight, or give explanatory notes, or add information, it would be helpful (and don't worry about your writing - we're not snobs). [...] [examples of questions follow]

**Illustration 7.3 Instructions from Gregg's BC Linguistic Survey (Stevenson 1976:84)**

While all WQs use some sort of instructions, their effect seems to be less clear. Instructions to linguistic surveys have been found to offer only a limited kind of guarantee that the respondent would observe them. Schütze (1996:133) reports of a study on grammaticality judgements and reasons, somewhat disappointingly, that "the exact
contents of those instructions might not matter a great deal. One set of instructions “invoked English professors marking term papers, the other emphasized the absence of right or wrong answers and appealed to personal reactions”, yet the different instruction sets “turned out to have almost no effect on the pattern of responses” (ibid).

Such results are mirrored in the social science literature, where it is widely acknowledged that “many respondents do not read the entire content of questionnaires in a thoughtful way” (Dillman 2000:81). While we would still want to offer instructions, one might want to consider using indirect verbal cues and visual aids rather than instructions alone to elicit the desired low-level formality response that is anticipated.

**Using informal language in the questionnaire**

Since instructions were shown to only offer a small effect on the results, we may think of other aspects of the questionnaire to elicit non-standard responses. One of the most effective ways to do so is by using informal language in all texts offered, in the consent email, in the instructions as well as in the questions themselves. In the Canadian context, Polson’s (1969) study is credited for establishing a model for informal question tone. Examples include:

(7.22) a. We aim to discover → we’d like to find out
b. Dear participant → Hi there (depending on the context and target group)
c. My residence is different from yours → My house is different than yours

In short, the goal is to mirror the informality of the spoken language in a given region and considering the target group, in as much as the written medium allows. As (7.22a) shows, lexical choice tends towards the informal (use of phrasal verb *find out* instead of Latinate verb *discover*) and the use of contractions (’d is common. Terms of address are more informal as in (7.22b), in as much as the social conventions allow. The examples in the questionnaire items should also remain within a common core vocabulary (house instead of residence) (7.22c). Another way to approach the issue of linguistic formality of the WQ text as such is to write the questionnaire with the reading skills of a Grade 9 student in mind.

**Reliability**

Some studies examine respondent consistency when presenting the same item twice. When given the same sentence to be judged again, some respondents show somewhat erratic behaviour. Carden (1973) found reliability measures between 87 and 92 percent for responses within one point, on a scale of five or six, of the previous rating, which falls to 67% if any changes are included and rises to 97% if only “radical” changes are considered. Greenbaum and Quirk’s (1970:43) test of individual consistency in grammatical performance and evaluation tests found a “very high level of consistency” at 84 and 82 percent of matching results.
The overall message seems to be that ratings are reliable within a given range, but not for absolute categories. This result will facilitate the interpretation of gradient data and offer a motivation to collapse answer classes, e.g. on a six-tiered scale, results that are reported in three groups (the two most extreme classes each combined with a neutral class) or even two groups would offer more reliable results overall. To increase this match further, one approach would be to opt for binary variables rather than gradient variables or to reduce gradient answers to basic two or three categories, however. This "safe" choice comes with a downside, however, as it would mean that a number of statistical tests, which are common in social evaluation studies, would not be applicable. It is a choice that is probably best made based on a particular set of variables and a number of variant options in a pilot study (see Carden 1976: 104), since at this stage little work has been carried out on that important matter, as, in general, different methodological approaches to the same variables and social contexts are the exception. Such comparative studies, however, would be indispensable for good judgement calls.

Some tentative insights: How to ask and how better not

Pilot testing of single, draft questions (raw questions) is a very effective way to learn about the needs of the target population. The following findings are based on 20 respondents from quite homogeneous sub groups of the target group, e.g. university students, retirees and the like and were solicited in the ENGLISH 489 majors seminar in the fall term of the 2014–15 academic year. The question was handed to the respondents on a slip of paper with the instruction to “answer this question”, such as in the examples in (7.23):

(7.23) a. Do Marry and Mary sound the same?
   Yes
   No

b. What do you call the toilet facilities in public places, such as airports, restaurants or shopping malls?
   bathroom
   restroom
   washroom
   lavatory
   other (please specify)

c. Complete the sentence: He didn’t mean to bump into you, he did it ____
   accident.
   on
   by

(Tawnie Chambers 2014)
d. Please read the following sentence, then select the statement that is most applicable.

*It is stressful not to know what to do.*

This type of sentence would never be used in my community – it seems odd. This type of sentence is not very common in my community, but it doesn't seem odd.

I've heard this type of sentence in my community, but it is not that common. People in my community use this type of sentence a lot. Hirota (2014)

The examples in (7.23) represent different stages in the development of questionnaire items from a first hunch or "raw question". We will address their strengths and weaknesses, as identified in the pilot test phase, one by one. Example (7.23a) has, in one form or another, been asked for a long time in the Canadian and American contexts. Everybody asked understood the text (clarity), but some asked the administrator of the question whether they were supposed to sound the same (Jeff Ashkinasi). Since it is one of three angles on the merger of the front (first) vowel in *merry*, *mary* and *Mary* in CanE, it would need to be followed up, preferably after some other questions and the instruction not to go back and "align" or correct previous answers, with questions on *Mary* and *merry* and *mary* and *Mary*. Interestingly, once the overall purpose of the question was clear – being part of a set of questions – the lay respondents suggested to group the questions together. This, however, is not to be recommended as it would entice the respondent to draw comparisons that seem to be more based on spelling, causing respondents to assign different phonetic values to different spellings, a common phenomenon in highly literate societies.

Example (7.23b) is a question that did not trigger too many problems: 18 answered with the Canadian majority variant *washroom*, while only 1 answer each for *bathroom* and *restroom* were reported (Christopher Cheng’s report). It is crucial for this lexical variable, like for others, to offer a description that is precise enough but not too awkward or bulky. In this context, most Canadians would call a room with toilet at home a *bathroom*, whereas such facility in public places a *washroom*. It is not only the rookie researchers who is prone to making errors of that type. More generally, for lexical questions with a well-researched set of variant options, very few problems pertain. Jasmine Chen reports on a similar type of question about the name of sports class in school (phys-ed, gym, P.E. other [please specify]) 18 of 20 answered the question quickly and without raising any questions or reporting any problems when asked. In one case, the respondent wanted to know whether the current name or the name during her time in school was wanted – presumably commenting on a change in progress she noticed. The other person volunteered the information on how to make the question less redundant in one minor point. Other comments included form “Oh, this is a good
question" to “It’s definitely called [P.E.] here [in Vancouver]” These comments offer valuable insights into possible variants that are missing and offer leads.

One question that arose in a number, but by far not in all contexts, was that respondents in the Canadian context were generally interested in the linguistic question and were curious as to what it was used for. This interest can and should be harnessed to generate interest in WQs. However, the most problematic kind of feedback was that the respondents generally tried to figure out “the right answer”, which is a very important reminder that every effort must be made to remind the respondents that their choices are not any “right”, prescriptively sanctioned ones. It is therefore paramount to include proper instructions, despite their sometimes less consistent effects, as suggested by the limited research on the matter in the previous section.

Questions (7.23c) and (7.23d) explore new variables and are therefore treated separately. Undergraduate student Tawnie Chambers explored first question (7.23c), while MA student Tomoharu Hirota used WQs as one of several tools to explore split infinitives (7.23d). Tawnie’s hunch that there might be some variation concerning the preposition, where traditional by is beginning to be rivalled by on, is an interesting case. Her pilot study showed that all answers for innovative on accident were by Americans, while traditional by was chosen by all Canadians, one Brit and two L2 speakers of English. As a consequence of respondent feedback, however, the question was shortened to (7.24):

(7.24) Which do you say:

He did it by accident.
He did it on accident.
Other: ______________________

Because of the unambiguous context, it was felt that the context scenario was, in fact, not needed. If the target group are, for instance, learners at a lower level of competence, the fuller context would be helpful. These considerations, obviously, depend on the situational and social contexts and need to be taken into account and tested in good pilot trial runs.

Tomoharu Hirota’s quest into assessing the rise of the split infinitive, the erstwhile pet peeve on many self-declared grammar mavens, took him from traditional corpus study to exploring WQs to that purpose. The rationale behind not drawing attention to the split infinitive was originally inspired by avoiding drawing attention to a potentially heavily stigmatized, or at least prescriptively dispreferred, construction in some social contexts. An interesting point is how to consider respondent feedback. If taken into consideration, many respondents, after being informed of the target split infinitive, suggested visual marking of it. If this choice is taken up, however, one needs to be clear of the repercussion, which lies in a – difficult to quantify – effect of increased under-reporting.
As shown above, Buchstaller et al. (2013) deliberately chose the context of community reporting – in your community – in order to allow the respondent to report on less prestigious forms without the risk of losing face, which even in anonymous surveys is a factor. However, the pilot study revealed confusion about what the “community” constitutes: one’s place of origin or one’s campus (in the case of university students), among other things. This led to the use of the phrase “in your circle of friends” rather than “in your community”. On the one hand the group is better defined, on the other, though, the stigma-avoiding effect is lost, as one clearly is part of one’s own friendship circle. It is this kind of reasoning that needs to be made explicit and then weighed. Tomoharu decided to put (7.25) on the survey:

(7.25) Have you heard the underlined construction?  It’s fine to not know everything.

a. I’ve never heard this type of construction among my friends.
b. I’ve rarely heard this type of construction among my friends.
c. I’ve heard this type of construction among my friends sometimes.
d. I’ve heard this type of construction among my friends a lot.

Whatever the decision, there will be trade-offs: gain in one dimension, e.g. less ambiguous question contextualization, is at times lost in another one, e.g. loss of stigma reduction.

These few examples should have demonstrated the principle behind raw question testing. In addition to single question testing, it is equally important to test the entire WQ with a few people of different social groups. After having them do the questionnaire, in, possibly, two or three sections, the researcher needs to follow up with question on the “flow” of the questionnaire and on any issues that seem confusing, ambiguous or just not applicable to a particular person. In addition, it is a good idea to see if the WQ is not too onerous, perhaps entertaining and not “bothersome” in any way unless that is intended. As WQ designers we need to give the respondents as little reason as possible to discontinue our WQ, or, even worse, not to start it.

Defining evaluative categories

An issue that is more often than not disregarded is to offer brief definitions for one’s evaluative categories, or at least contextualizations. For evaluation tasks, criteria such as “acceptable” or “grammatical” are frequently and generally used. There is good evidence to suggest that the criteria should be defined, as respondents interpret them in very different terms (Carden 1970). Schütze (1996:132–3) lists a host of studies that concur with Carden’s point, yet points to a sorry state of WQ research in this respect. He reasons that “if we were to ignore all studies in which we believe the instructions to subjects were inadequate to convey the subtlety of a linguistic definition, the remaining studies could be likely counted on one hand.”
In practice, defining categories need to be little more than to paraphrase terms used in the evaluation, e.g.

"acceptable" means you would use the sentence in conversations with friends if the context arises

Or whatever context one wishes to elicit. Binary categories (yes, no questions), might represent an exception to this rule:

Do you pronounce Variable X with the vowel Y? Yes  No

It seems that further explication reveals the limits of the existing question formats, as phonetic differences have not been successfully addressed in WQs. For this reason, nominal/binary choices are best left uncommented, provided that the question is phrased clearly. There might be, however, one avenue of inquiry that promises to push the envelope of WQs and phonetic study further.

Harnessing a pedagogical phonetic alphabet for social dialectology?
As has been mentioned, the opinion that WQs are not particularly well suited for eliciting phonetic detail is, of course, correct. But rather than ruling sounds out categorically, there may be ways to exploit the written medium creatively. As was shown in Chapter 2, the basic type of questions on pronunciation, despite refinements, has not been significantly altered since Hempel's late 19th century survey, and can be summarized as using key words that are pronounced (largely) invariably to elicit information. We said at the beginning of this book that this basically means that such pronunciation questions are limited to phonemic information, while phonetic quality is generally beyond reach.

However, recent developments in speech pathology and therapy seem potentially promising to push the agenda of WQs and non-phonemic pronunciation further. Sound files, played to respondents and matched with closed answer options would allow participants to align their own pronunciation with the one sound option that fits closest. This, in itself, is an interesting proposal that will likely be exploited in the years to come. Visualization techniques of phonetic features, which are being developed in speech therapy (Ruß 2008) offer a potentially simpler method, a method that can also be used on paper. As there are some contexts and users for which paper is the preferable medium, pursuing this traditional form of delivery seems worthwhile.

The basic idea is to break down each phoneme into some of its phonetic components and articulatory features, e.g. vowel height, voicing, lip rounding, airflow, and to find intuitive visuals to represent them. In a way, it is taking the 16th century approach of early phoneticians (orthoepists) into a visual realm. The orthoepists described the production of speech sounds by observing the shapes of the articulators and contrasting them with other sounds. Their appeal was their mixture of precision with generally understandable descriptions. In its 21st century guise, the "visual description" of
Lautbilder – sound pictures – is informed by pedagogical principles and not by absolute precision. This mix is of interest for WQs. As the method has proven efficient even in young children, it is reasonable to believe that its adaptation might enrich the limited number of options for eliciting phonemic and (more limited) phonetic options.

Order of stimuli and trial items
Recently, some researchers (e.g. Buchstaller & Corrigan 2011; Buchstaller et al. 2013) have begun to reordering questions to balance any error that might stem from their presentation sequence. Greenbaum and Quirk (1970:35) show that “significantly different results” are “undoubtedly attributable to the effect of sequence” in their tasks and recommend that each test battery, or questionnaire, needs to be presented “in more than one order”. Schütze (1996: 134–5) offers an alternative explanation for such mismatch in the lack of “warm-up trials”, which are common in psycholinguistic studies. Greenbaum and Quirk (1970:32) suggest for tests in which respondents perform a linguistic operation (moving adverbs in a sentence etc.) that some questions “may yield a particular result merely on account of its occurring very early in the [test]” and that the first few items “on any occasion provide in effect practice’. These findings are based on operations that are comparable with translation or reformulation tasks but will not apply as much to traditional WQ items. In any case, starting a questionnaire with two or three linguistic “practice” questions (without informing the respondents) seems like sound practice.

Intuitive formatting & item ordering
The formatting of questionnaires may be considered a trivial matter, but it is one of the few features that researchers have full control over and that might influence the response rates as well as the reliability of findings. Readability – do the fonts work well in a digital environment, is the print version clearly legible – are key factors. Issues such as colour choices for paper, ink, margins all play a role. Most importantly, an attractive and clean design contributes directly to increased response rates in general, and at higher proportions in those social groups that are least likely to respond to WQs (Dillman 2000: 81). It is therefore important for the general validity of the sample (see 7.3.8) that all formatting tools are used. Moreover, since respondents have been shown to not necessarily read instructions carefully, the visual design features should be thought of as offering cues for those respondents as one of the most important structuring devices of the WQ. These tools include, but are not limited to, the formatting of answer choices, working with line breaks and indentations in the text and the like.

The order of the items is another issue. It is often recommended that more sensitive questions are asked at the end of the questionnaire and not at the beginning as a “respondent who has spent five or 10 minutes already answering questions is less likely to respond to an objectionable question by quitting” (Dillman 2000: 87). Objectionable
questions can be questions about one's age, income and the like. Schlee (2013) recommends in general placing the social background questions at the end of the questionnaire, though the practice may need to be weighed with other concerns. The trend, though, seems to be in line with Schlee's recommendation. Within the questionnaire, the sequence of questions should be logical from the respondents' point of view, as much as discernable, as was addressed in the section of questionnaire structure (7.2).

Another important issue concerns the first question of the questionnaire. The first question is of extreme importance. The WQ's first question

- must be applicable to everyone
- must be easy to answer
- needs to be interesting (Dillman 2000: 92)

Bad examples of a first question would be an open answer question (they are more work than a limited response option), good examples those that tap into widely-discussed linguistic questions. One might even go so far to include a question that one is not interested in, but that is widely known, e.g. pot[er]to or pot[a]to is one such variable. These questions can also be used as a "warm-up question" that is recommended by Schütze (1996), as seen above. Easier still would be a variable on lexical choices: in the Canadian context chesterfield would make a good first question that is easy to be answered. Dillman summarizes: "No single question is more crucial than the first one", in the sense to increase the response rates of the survey (2000: 92). Of course, he is right: asking for age (as has been done in some social science WQs), or even income in this place would diminish the return rate.

7.3.7 Pilotizing and revising the questionnaire

WQs should be tested twice: first, when a particular raw question is transferred into a questionnaire item, as discussed above, and, second, the completed draft WQ should be exposed to trial runs from start to finish. As with any empirical study, it is paramount to run these pilot tests. They can be done informally by the investigator, or can involve focus group interviews on a bigger scale. At first it is enough to get a sense of the strengths and weaknesses of a questionnaire item or the completed questionnaire, which is best done with the help of volunteers from a small number of different social backgrounds who provide (oral) feedback while filling out the questionnaire. Such think-aloud protocols are immensely useful in identifying questions that work and those that do not. These answers provide invaluable cues to problems that need attention and may dispel fears of the researcher about anticipated problems of a given section or question.

The pilot phase is one of the most important aspects of WQ design, as there is only one chance to get the questions right. This is not only necessary from an economical
point of view, but also from an ethical one. As most WQ surveys are anonymous in nature (unless they are longitudinal studies that need to identify respondents to match their answers, or when payment is offered, which requires at least an email address and often many more details), respondents do not need to fear that their data will be matched against their names. Anonymity also helps when dealing with Ethical Review boards and the administration of the survey: since we are not interested in the identity of the respondents beyond their group-specific background information (age cohort, sex, occupation, education and so forth), WQs are generally considered “minimum-risk” studies. If we do not elicit the identity of the respondent, safe storage and access restrictions to the data collection can be simplified. However, anonymity also means that we need to get the answers “right” the first time around, as there is no second chance to go back and complete an incomplete questionnaire. Testing or piloting the individual questions in isolation as well as the completed WQ is therefore imperative, as it offers cues for revisions that the researchers themselves may not be able to see.

7.3.8 Social background questions

In Chapter 2 we have seen the development of the demographic (social) background questions over time as part of the general history of WQs in dialectology. Today a ‘standard set’ of social variables can be offered, as shown below. Of course, depending on the study, this set needs to be adapted or expanded:

- Gender (sex)
- Age (by cohort, or continuous in years)
- Education
- Ethnicity (self-defined)
- Place of residence
- Birth place
- Residence history with approximate ages since birth
- Occupation
- Father’s Occupation
- Mother’s Occupation
- Father’s birth place
- Mother’s birth place
- Languages spoken and competence level
- Information for any indices needed (see Section 8.2).

As mentioned earlier, there is some disagreement in the literature whether to place the social background questions at the beginning or at the end of the questionnaire. On one hand, they may be used to “warm up” the respondent, as they are generally easy and straightforward. The wrong background question, however, might interfere.
with the return rate (e.g. age, education for some lesser educated respondents). On the other hand, offering the background questions early will almost certainly negatively affect the response rates as some respondents, especially if socially sensitive information is elicited, e.g. income or (in some contexts) place of residence (neighbourhood), respondents would be more likely to quit the survey. Generally, it seems that more and more surveys ask these questions at the end, though one needs to ensure that respondents do not "feel tricked", when at the end of a questionnaire and after having put in some time, they are asked social questions. It is recommended to clearly state that "social background questions (e.g. age, sex, education)" will be asked at the end. Regardless of how this is framed, some respondents will choose to discontinue the survey after the linguistic part, which means that some "orphaned" responses will need to be discarded.

### 7.4 Population sampling

By sample we mean that part of a population that can, in principle, participate in a survey and sampling is the concrete method how the participants are chosen. Ultimately, we are interested in the behaviour of a population, which is why the sample must be representative of the population. If one wishes, for instance, to study the social embedding of linguistic change in a particular setting, one needs to ensure that the sampling is in its social make-up as balanced and representative as possible (women, men, younger, older, educated, less educated and the like should be as equally represented as possible). If we aim to describe the behaviour of traditional Canadian variables in Vancouver, we need to have an idea of the Vancouver population and how we are going to sample the population, since we cannot ask every Vancouverite to partake in the survey (and even if we could, not everybody would). While balanced sampling is the goal, in reality in WQ data some groups will be over-represented. The goal must be to eliminate these differences as much as possible by actively reaching out to social groups the researchers are typically not in contact with.

#### 7.4.1 Random or judgement sampling?

Random sampling is a key component in the social sciences. The idea is that every member of a population must have the same chance to be selected in the sample. With (landline) telephone registries (assuming that every household has a phone line), one could draw samples from a publicly accessible list only 15 or 20 years ago, but reaching representativity has become a more complex issue since: which lists could be used today? One issue of big concern in random sampling are the increasing percentages of nonresponse rates: if not everyone has an equal chance to return a WQ,
because some groups tend not to do so ("nonresponse error") or are not reached in the first place ("coverage error"), the sampling will be skewed and the results not reliable (Tourangeau & Plewes 2014; Dillman 2000 for traditional solutions to this problem).

It is important to point out that most sociolinguists work under special conditions. They are either intimately familiar with a speech community or work with what may be perceived prototypical members of a group. Most sampling methods in sociolinguistics are therefore not random samples, but judgment samples. The researcher uses his or her special knowledge of the community to select the informants and once basic choices have been made, often more detailed background information is documented. In cases where the researcher is a member of the group, such documentation is often not needed since the data is already known.

Early studies in sociolinguistics, inspired by the social sciences, started with random sampling. Shuy et al.'s (1968) study recorded 702 (!) subjects, while Wolfram (1969) only had to use 36 of them to arrive at meaningful findings. A couple of years earlier, Labov applied a similar approach by selecting 122 subjects from a list of 617 provided by a social sciences survey (Labov 2006: 108; 117). Labov himself justified the smaller, selective sampling technique with the fundamental difference between data in the social sciences on the whole and linguistic data:

If the type of behavior which was being studied was similar to most forms of behavior that are investigated by social survey, the value of the study could be measured by how far it [the sample] fell short of the [...] standards [of the social science survey]. However, linguistic behavior is far more general and compelling than many social attitudes or survey responses. The primary data being gathered [...] are not subject to the informant's control in the way that answers on voting choices would be.

(Labov 2006: 114)

While there is a wide range of possibilities of someone commenting on a political leader, there are only so many alternative variants one can meaningfully identify for most linguistic variables. In these situations, an insider can make meaningful selections of informants that do not skew the data, but it is important to keep in mind that sociolinguists work under the most favourable conditions. Technically speaking, most sociolinguistic surveys have varying degrees of randomness: Gregg (2004 [1978/79]) took great pains to arrive at a randomized sampling of his subjects, quite unnecessarily so, while later studies, e.g. Tagliamonte and D'Arcy (2007), took full advantage of judgement sampling.

In WQ studies, it is easy for a person to reach 200 people and more in a week, provided one taps into the right network channels - either in person, via mail, or via the internet. The Atlas of North American English, mentioned in Chapter 1, is a ground-breaking work (Labov, Ash and Boberg 2006). Its findings for mainland Canadian English in an area stretching some 6,000 kilometres from east (Halifax,
Nova Scotia) to west (Vancouver, British Columbia) are based on the acoustic analysis of only 33 speakers (p. 220). Follow-up studies (Böberg 2008b) use data from 86 speakers across the country and are on the larger end of the spectrum of studies of acoustic phonetics, where data needs to be recorded, tokens extracted, formant measurements made and normalized to ensure the comparability across speakers. The same effort of analysis applies to other studies where speakers need to be recorded, such as variationist studies in morphosyntax. Table 7.4 gives an overview of some sample sizes in the bigger sociolinguistic studies: (for more large-scale studies, see Labov 2006; 380–403).

**Table 7.4 Large-scale sociolinguistic studies that include audio-recordings**

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolfram (1969)</td>
<td>36 analyzed (702 recorded by Shuy et al. 1968)</td>
</tr>
<tr>
<td>Trudgill (1974)</td>
<td>60 subjects</td>
</tr>
<tr>
<td>Woods (1999 [1979])</td>
<td>100 subjects analyzed</td>
</tr>
<tr>
<td>Gregg (2004 [1978/9])</td>
<td>300 subjects analyzed</td>
</tr>
<tr>
<td>Poplack (1985)</td>
<td>120 subjects</td>
</tr>
<tr>
<td>Kerswill and Williams (2000)</td>
<td>96 subjects</td>
</tr>
<tr>
<td>Eckert (2000)</td>
<td>69 (&quot;Neartown&quot;), and 60 regional sample</td>
</tr>
<tr>
<td>Labov, Ash and Boberg (2006)</td>
<td>439 acoustic analyses, 33 of which in Canada</td>
</tr>
<tr>
<td>Tagliamonte and D’Arcy (2009)</td>
<td>152 speakers (350 hours)</td>
</tr>
<tr>
<td>Walker and Torres Cacoullous (2009)</td>
<td>74 speakers</td>
</tr>
</tbody>
</table>

Some of these studies have fairly large sample sizes. What is not shown is the effort that went into the creation of these corpora. Tagliamonte and D’Arcy build on multiple years of data collection in teams, as do Labov, Ash and Boberg (2006) and Walker and Torres Cacoullous (2009); Woods (1979 [1999]) went to the upper extreme for his doctoral project by interviewing a 100 speakers in Ottawa, Canada. Poplack’s study is also one of the bigger ones with 270 recorded hours of speech in this corpus. Gregg’s study on Vancouver English is with 300 interviewees still one of the biggest data pools. The large sample size, however, came with compromises in the linguistic data analysis that did originally not go beyond the phonemic level. Wolfram’s study of 36 informants, which was carried out in the early days of sociolinguistics, shows the other extreme: a reduction of sample size and increase in linguistic detail. Most sociolinguistic studies today work with a range between 25 and 40 speakers and are thus on the lower end of the spectrum of sample sizes presented.

By contrast, the range of WQ data is different. Samples sizes in WQ studies range for Dialect Topography from 307 respondents (Quebec City) to 935 (Golden Horseshoe 1991/92). In studies in a given urban region, sample sizes of about 500 respondents that meet the selection criteria have proven to work well and allow findings for some correlations. As it is quite feasible to reach around 500 respondents today, one might...
decide to include more subjects should one’s variables and social stratification require so. The 533 responses from the Dialect Topography of Vancouver from 2004 (DT 2004) took only weeks to complete with a postal questionnaire. The Vancouver Survey 2008 (VS 2008) was administered by 43 students in an English dialectology class with students directly approaching potential respondents and handling over questionnaires and pen, over the course of two weeks. Table 7.5 shows the absolute figures in each age cohort:

<table>
<thead>
<tr>
<th></th>
<th>14–19</th>
<th>20–29</th>
<th>30–39</th>
<th>40–49</th>
<th>50–59</th>
<th>60–69</th>
<th>70–79</th>
<th>80+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT 2004</td>
<td>104</td>
<td>170</td>
<td>78</td>
<td>58</td>
<td>70</td>
<td>25</td>
<td>15</td>
<td>13</td>
<td>533</td>
</tr>
<tr>
<td>VS 2008</td>
<td>82</td>
<td>153</td>
<td>48</td>
<td>54</td>
<td>43</td>
<td>24</td>
<td>10</td>
<td>9</td>
<td>423</td>
</tr>
</tbody>
</table>

All empirical studies have to make some compromise in either linguistic level of analysis or survey size on account of the complexity of real-life scenarios. In typical sociolinguistic studies one would reduce the sample size rather than the complexity of the linguistic analysis. The rule of thumb is that WQs, as a self-reporting or community-reporting technique, do not allow for the level of linguistic detail as other sociolinguistic methods, while they excel at the relative ease with which to collect data from a large array of people.

### 7.4.2 A combined sampling method

There are a number of ways to approach sampling. Dörnyei (2003:72–72) distinguishes between four types of sampling: Convenience sampling, Snowball sampling, Quota sampling, Random (stratified random) sampling. Convenienc sampling is the easiest form of sampling for the researcher, who contacts groups that meet a criterion considered as most salient, e.g., residence in a given location. This will likely trigger a number of responses that need to be ruled out, as the respondents do not meet other criteria. It is a kind of catch-all with post-hoc control for a well-defined group, e.g., excluding visitors to a given location. Snowball sampling works with in-group contacts and depends on a chain reaction by one contact leading to another one within the same network. Quota sampling defines criteria of particular groups to be included in the sampling and aims to meet these pre-set quotas: for instance, a study on Vancouver might aim for 40% non-native speakers of English, as a reflection of this ratio in the city. Stratified Random sampling divides range within the strata every potential participant must have an equal chance to be selected, so stratified sampling is logistically more challenging and depends on the availability of lists of group members that is representative of the community, which are
often no longer accessible (like in a telephone book for an entire region). It is therefore rare that random sampling in the strict sense is applied outside of the social sciences.

For linguistic WQs, most often a combination of quota sampling and convenience sampling is used: if one wishes to compare the answers of Chinese Canadians and Anglo-Irish Canadians (those of UK or Irish descent), a quota must be set and respondents that meet the criteria (one of the ethnic groups of long-term residents in Vancouver) are invited to partake in the survey. Often, calls are placed in social media and people can be invited via email. It is important, though, to go beyond one's immediate contacts in order to ensure as broad a demographic representation as possible.

### 7.5 Chapter Summary

This chapter has offered some guidelines on question and questionnaire design. Its most important goal has been to sensitize the reader to the major issues that affect the quality and quantity of WQ data. One main concern was to offer a basic typology of WQs questions in social dialectology, which was undertaken in Section 7.3. There are aspects of WQs that are difficult to gauge, as there are many choices to be made and the literature on the topic is not plentiful. Each of these choices will have an effect on the result, so it is hoped that the recommendations offered in this chapter will keep at a minimum the remaining survey errors. In linguistic circles, it is fair to say that the number of methodological studies that compare WQs to one another are an under-studied area. Often, WQs are not foregrounded in studies, which generally focus on results that are more directly linked to linguistic theory. It is to be hoped that the methodological aspects of linguistic WQs, which this chapter has shown to differ in some crucial aspects from the more general social science WQs we have come to be familiar with, will become a focus in the empirical linguistics literature.

---

**Chapter 8**

**Word Count**

The present chapter will analyze WQs and the questionnaire and data sets, with data from the survey available on the web through the author's website, but will later be removed for the manipulation of data. We will instead explain how to assist in the questionnaire studies. Section 8.1 will allow the reader to understand the procedures.

The code for the computer with the website is available for this instance, and all technical details of this book will be presented.

### 8.1 The Procedure

The procedure that has characterized the underlying processes is that of the Canadian society and its development over its history.