CHAPTER 10

Cues of Communication Difficulty in Telephone Interviews

Frederick G. Conrad
University of Michigan, USA

Michael F. Scheber
New School for Social Research, USA

Wil Dijkstra
Vrije Universiteit van Amsterdam, The Netherlands

When people converse, they do not just send and receive messages. They also give each other ongoing visual, auditory, and textual cues about the extent to which they believe they are understanding each other—about the extent to which their utterances are "grounded," to use Clark's (1996) terminology (see also Clark and Wilkes-Gibbs, 1986; Clark and Schaeffer, 1989; Clark and Brennan, 1991, among others), as well as about their emotional reactions to what their partners are saying. Cues that addressees give to speakers—warts of the head, looks of confusion, back channels like "uh-huh" or "huh?", and requests for clarification like "what do you mean?" or reactions like "vocal"—can alter what speakers say, such that both parties can be seen as molding each other's language use simultaneously.

Such the same sort of thing goes on during survey interviews. While respondents are answering questions, they simultaneously are giving cues about their comprehension and their reactions. Respondents who simply answer a question smoothly, without delay

by training interviewers in particular techniques or by simulating telephone speech interfaces, so that causal inferences can be drawn. Thus far, the evidence from one larger scale replication in a U.S. national telephone sample (Conrad and Schoder, 2000) of an earlier laboratory study (Schoder and Conrad, 1997) suggests that the statistically reliable findings from these laboratory studies are likely to generalize to larger populations. But obviously, without full testing it is unknown whether all the results will generalize, and we instead our discussion here to be suggestive or more work exploring rather than definitive or providing immediate prescriptions for practice in larger scale telephone surveys.

10.1 CUES IN TELEPHONE CONVERSATION

To a certain extent, the grounding cues in telephone surveys reflect the constraints of grounding understanding on telephones more generally, which differ from the constraints on grounding in other media (see Clark and Brennan, 1991, for extended discussion; see also Williams, 1977; Whittaker, 2003). Telephone interviewers are audible to each other, but they do not have access to visual cues such as eye gaze, gestures, and frowns (although this is changing with the advent of video telephones). Unlike people conversing via handwritten letters, email, or instant messaging, telephone interviewers have access to multimedia paralinguistic cues that can be useful for understanding what their partner means (long delays can be a sign of trouble—see Brennan and Williams, 1995)—and other discourse features like and of and ums, sentence restarts, and other disfluencies that can give evidence about what speakers do and do not mean (e.g., Fox Tree, 1992; Clark, 1996; Brennan and Schoder, 2005). Unlike the phone connection is poor, telephone interviewers perceive each other’s signals more or less instantaneously, without the kinds of delays that can occur in handwritten letters, email, or even on a different scale) “instant” messaging. Because the channel of communication is speech, it leaves no unequivocal trace that would allow further inspection; again unlike email or chat room discussions, or messages left on telephone answering machines. And telephone conversationalists can produce and receive messages simultaneously, allowing overlapping speech (of course, within limits—simple overlaps can lead to communication breakdown), which differs from one-way forms of communication like voice mail messages or walkie-talkie radio communication.

Other facts about grounding understanding in telephone survey interviews arise from what is unique about standardized survey interaction. As various reviews have noted (see, e.g., Sebseman and Jordan, 1990; Schoder, 1991; Schaeffer, 2003; Schoder and Conrad, 2002), respondents in standardized surveys are in the unusual position of trying to ground their understanding with a partner (the interviewer) who is not the originator of their utterances. Typically, interviewers read questions scripted and presented by survey designers, and the scripts of the survey interview in predetermined and instantiated entirely by that script. The agents who are generally responsible for what a question means (and that the person with whom
on the basis of fictional scenarios. Because we had access to official definitions of survey terms, we know what the survey designers hoped would be included and excluded in respondents' answers, and could judge respondents' interpretations of the questions based on their answers. For each question, we designed two alternative scenarios that a respondent might answer; in one case, there was no likely ambiguity in how the question mapped onto the respondents' (fictional) circumstances, and in the other case there was potential for a mapping ambiguity. For example, a respondent answering a question about how many bedrooms there are in his house would likely find this a straightforward question to answer when the fictional scenario is a four-plex of a house with three rooms labeled "bedrooms." But that respondent is likely to have more trouble if one of those rooms is labeled "originally designed as a den, this room is being used as a bedroom." The hypothesis was that how interviewers handled groundings cues should be particularly important for these more complicated mappings.

Both in this laboratory study and in a subsequent field study in which interviewers telephoned a national sample of respondents (Cronin and Schorber, 2000), the evidence showed that interviewers in which interviewees were licensed to respond to explicit and implicit groundings cues substantively—conversational interviews—improved understanding and response accuracy, particularly when respondents' circumstances mapped onto questions in complicated ways. The evidence from the field study showed that these complicated mappings are frequent enough in a national sample to suggest that when groundings cues are ignored (in strictly standardized interviewing conditions) data quality may be compromised. There was, however, a cost to responding to groundings cues in both the laboratory and field study: providing clarification takes time, and no increased response accuracy is accompanied by increased interview duration.

In another laboratory study (Schorber et al., 2004, Experiment 1), we made a first attempt at disentangling how interviewers' responses to explicit groundings cues (respondents' requests for clarification) and implicit groundings cues might affect data quality. In this study, we compared response accuracy under strictly standardized conditions and four versions of conversational interviewing. In all four versions, interviewers were able to clarify concepts after respondents provided explicit groundings cues (what we called respondents-initiated clarification). The groundings cues differed in whether (1) interviewers could also volunteer clarification in response to implicit cues (mixed-initiative clarification) and (2) they could use their own words to clarify the questions (paraphrased versus verbatim clarification).

As in the Schorber and Cronin (1997) study, respondents answered 12 questions on the basis of fictional scenarios designed to be unambiguous (straightforward) or to include a mapping ambiguity (complicated). For example, one straightforward scenario described a prototypical nuclear family with two parents and two children living in the home. Respondents used this to answer "How many people live in this house?" The complicated counterpart described a similar family in which one child was a college student living away from home. Should this child be counted as living in the house? The definitions of "living in a house" created by the survey sponsors involved the confusion (the child should be counted as living in the house) and so
SPOKEN CUES OF RESPONDENT NEED FOR CLARIFICATION

Just as the explicit requests for clarification occurred more often for complicated than for straightforward mappings, so did reporting and some speech disfluencies. In addition to producing longer pauses, respondents produced fillers, pauses, and repairs reliably more frequently for complicated than for straightforward situations. This suggests that these communication cues—who they indicate processing trouble, uncertainty, or even intentional grounding requests—count as valid markers of need for clarification. Combinations of some cues were even more diagnostic of need for clarification. For example, fillers and repairs, and fillers and pauses, appeared together more often in complicated than straightforward situations, even more than one of these cues alone. Hedges and discourse markers, in contrast, appeared so differently in answers for complicated than straightforward scenarios, which suggests that, at least for these questions, they are not diagnostic of need for clarification.

Obviously, this set of results is based on a relatively small sample (44) of interviews using particular fact-eliciting behavior-bound questions, and so we should be cautious about generalizing their generality. But it is possible that if telephone interviewers can be trained to do and detect the cues that are particularly reliable indicators of need for clarification, particularly in coordination with one another, they might be able to volunteer clarification in particularly judicious ways, explaining the question's meaning when it is actually needed and refraining from offering help when it is not needed.

As an additional set of findings is that the use of various communication cues was affected by what interviewers were licensed to respond to: Respondents used more cues differently in conventional than in standardized interviews. Not surprisingly, explicit requests for clarification were far more likely in conversational (mixed-initiative, paraphrased) than standardized interviews; respondents also seemed more inclined to ask subjective grounding help. And along the same lines, perhaps it is not surprising that respondents produced reports (e.g., “She bought a floor lamp” when asked “Did Dana purchase or have expenses for household furnishings?”) more often in conversations than standardized interviews. As with explicit requests for clarification, reporting is an effective strategy only if interviewers are able to react substantively, perhaps determining and recording the answer on the basis of the report. In a standardized interview, reporting is less likely to help the respondent to ground his understanding; the interviewer is more likely to respond with a nondirective (and nongrounding) probe like “Would that be yes or no?”

But some of the disfluencies were also differentially present in standardized and conversational interviews—and not always in the direction one might expect. For example, disfluencies (ahs and as) were actually less frequent in conversational interviews than in standardized interviews. Why might this be? Perhaps, when telephone respondents are deferred from requesting clarifications (e.g., as seems to be the case in standardized interviews), their speech is more likely to reflect unprocessed uncertainty. There may be a trade-off between having the ability to ask for clarification and the production of certain communication cues. In any case, these findings sug-
for standardized interviews, where there was no chance they could get clarification. These data suggest that respondents were sensitive to whether the interviewers could provide clarification in response to a visual behavior. Curiously, conversational interviewers did not provide more clarification in response to this behavior, despite glancing at respondents at least once during 80 percent of their looking-away episodes. One explanation is that conversational interviewers simply had not been instructed to treat such cues as indications of respondent uncertainty and that with appropriate training they could provide more and better-timed clarification. Another possibility is that interviewers were so focused on looking at their laptop screens that they were not sufficiently aware of respondents’ gaze aversion to use it as a cue for need for clarification.

In addition to verbally signaling the need for clarification, speakers may supplement these cues visually (e.g., direction of gaze). If so, understanding might suffer on current telephones because, without visual cues, interviewers may miss opportunities to provide needed clarification. Alternatively, respondents may compensate for the limits of auditory-only communication by verbalizing their comprehension problems parsimoniously. Clearly, this warrants further investigation, particularly as video telephony becomes more practical (see Anderson, 2008 and Pache, 2003).

10.4 Interacting with Automated Telephone Interviewing Systems

It is currently unknown whether all interviewers, or only the most socially sensitive interviewers, can use verbal and visual cues of respondent uncertainty as a trigger for providing clarification. The division of attention that may have limited interviewers’ use of gaze aversion in the Cerant et al. (2004) study could be a serious impediment. We propose that technology may be able to help. In particular, diagnostic software could be created that could take some of the attentional burden off interviewers by monitoring for spoken or even visual cues of respondent difficulty. One could even imagine deploying such technology as part of fully automated interviews in the not-so-distant future.

We have begun studying the effectiveness of this kind of diagnosis by simulating such technology. We call it a "Wizard-of-Oz" (WOZ) technique (e.g., Dabbish et al., 1993). In this approach respondents believe they are interacting with an automated system via telephone but are actually interacting with a human (wizard) who presents prompting speech that is created to sound like synthesized speech. Unlike conventional speech recognition technology (as is some of today’s interactive voice response [IVR] systems), the simulated dialogue technology is not limited to utterance recognition but can take into account discourse criteria like whether a concept has already been discussed and whether respondents’ speech contains the kind of markers observed by Schober and Bloom (2004).

Here we describe two experiments using the WOZ technique to simulate automated interviewing technology. Respondents answer questions asked by the simulated interviewing system on the basis of fictional scenarios, just as in our studies of human telephone surveys, so that we have independent evidence about when they have interpreted questions as the survey designer intended. Note one advantage of this sort of study: The behavior of the "interviewer" can be manipulated with algorithmic precision in a way that is far less certain in training human interviewers.

10.5 Diagnosing Respondents’ Need for Clarification from Communication Cues

In the first study (Bloom, 1995; Schober et al., 2000), a Wizard-of-Oz technique was used to simulate a speech interface. Users believed they were interacting with a computer, when actually a hidden experimenter presented the questions and scripted clarification. To enhance believability, we used an artificial-sounding computer voice (Apple’s "Avery" voice); virtually all respondents were convinced they were interacting with a computerized interviewing system, and the data from the few who detected this were removed from the study.

In the first condition, the system could not provide clarification. This was similar to one of our strictly standardized interviews in that a confused respondent could not obtain a definition; if a respondent requested clarification, the system would repeat the question. In the second condition, clarification was based on explicit respondent-initiated grounding cues—the system would provide clarification if the respondent asked for it explicitly. In the third condition, clarification was based on both explicit and implicit respondent-grounding cues (the system was raised)—the system could also "automatically" provide full definitions when users displayed the cues of need for clarification cataloged in Schober and Bloom (2004). These included yeses, arts, pauses, repairs, and talks other than an answer. In the fourth condition, the system always provided clarification; no matter what the user did, the system would present the full official definition for every question.

The results with this simulated system in some respects paralleled those for our studies of human interaction. As in Schober et al. (2004, Study 1), respondents were almost perfectly accurate when they were asked to answer straightforward scenarios. For complicated scenarios, respondents were substantially more accurate when they were always given clarification (80 percent) than when they were never given clarification (33 percent).

But the pattern for grounding cues was somewhat different. Unlike in Schober et al. (2004), requiring explicit grounding cues (requests) in order to provide clarification was entirely ineffective, because respondents almost never asked for clarification despite being instructed to do so if they were "at all uncertain about the meaning of a word in a question." In the respondent-initiated clarification condition, the accuracy of respondents’ answers was no better (29 percent) than when they were never given clarification. Most likely it did not occur to respondents that clarification was necessary; the presumption of interpretability (Clark and Schenker, 1991) probably breaks in computer-administered interviews. What was effective was relying on respondents’ implicit grounding cues: response accuracy was reliably better when
the system provided clarification in response to users’ difficulties and pauses (the mixed-initiative clarification condition) 59 percent, although not as good as when clarification was given always.

When the system provided clarification in response to implicit grounding cues, respondents were actually more likely to ask explicitly for clarification. Respondents asked questions more often in the mixed-initiative condition, presumably because they were more likely to receive that clarification might be useful. These users also spoke less fluently, producing more words and asks—and there is some evidence that this tendency increased over the course of the interview. We speculate that this was because these users at some level recognized that the system was sensitive to their cues of uncertainty.

Why did respondents with the computer speech interface give explicit grounding cues (ask for clarification) so rarely? Perhaps even more than the respondents in the telephone interviews in Schöber et al. (2004) they found it relatively uncomfortable to articulate their confusion or uncertainty to a computer agent. But we cannot conclude this with certainty, as there are other differences that we suspect may have been even more important: Obtaining a definition with this particular speech interface was a more daunting prospect than getting a definition from a human interviewer, because the entire definition—not just the relevant parts—would be spoken, and this was time consuming (up to 108 seconds) and impossible to stop. In contrast, human interviewers can potentially provide just the relevant part of the definition (as in the paraphrased clarification interviews in Schöber et al., 2004) and respondents can interrupt the interviewer if necessary to circumvent the full delivery of the definition. Finally, respondents in the current study could not reject a system-initiated offer to provide a definition because the system did not offer—it simply provided—the definition. In the Schöber et al. (2004) interviews, it was often the case that interviewers asked respondents if they wanted clarification.

As in our studies with human interviewers, clarification took time. The more clarification a respondent received, the more time the interview took. Sessions where clarification was always provided took more than twice as long as sessions with no clarification or when it was (rarely) respondent-initiated (12.8 versus 5.2 and 4.9 seconds per question, respectively); mixed-initiative clarification took an intermediate amount of time (9.6 seconds per question).

Respondents rated the system more positively when it was responsive (respondent or mixed-initiative conditions). When the system was not responsive (no clarification or clarification always), users rated the system less control and felt that interacting with the system was unengaging. Respondents did not report finding system-initiated clarification particularly more annoying than respondent-initiated clarification—which they almost never used.

Overall, these results suggest that enhancing the collaborative responsiveness and diagnostic capability of a speech-interviewing system can improve comprehension accuracy without harming user satisfaction, as long as the system provides help only when it is necessary. But these improvements come at the cost of increased task duration, which raises questions about the practicability of a system with only these characteristics in real-world survey situations.

18.6 MODELING RESPONDENTS’ SPEECH TO PROVIDE MORE TAILORED CLARIFICATION

We propose that systems may be able to provide more precisely tailored clarification to respondents by attending to their grounding cues in a more nuanced way. We demonstrated this in an experiment (Bion, 2005; Bion et al., 2007) in which we modeled different groups of respondents’ relevant paralinguistic behaviors. In particular, the respondent modeling techniques allowed us to distinguish behaviors more likely to signal uncertainty from those that are less likely to do so. For example, someone who regularly says “well” and asks as part of their daily repartee is less likely to be signaling comprehension difficulty with or without someone who rarely uses them. A listener who makes this distinction is modeling the individual speaker. The same logic can apply to groups of speakers. Older speakers have been shown to be less fluent than younger speakers (e.g., Borton, 2001), and so the same fluency rate for a young and older speaker may indicate different states of understanding. In other words, the same level of silencing might indicate problematic understanding for a younger speaker but ordinary speech for an older speaker. We applied this idea to automated interviewing by allowing the system to offer clarification on the basis of a generic model (same criteria for all respondents) and a stereotyped model (different criteria for old and young respondents).

One hundred respondents (50 older than 65 years of age and 50 under 40 years of age, answering questions with the system’s automated clarification, Required Clarification, Responder-Initiated Clarification, Required Clarification, Generic Respondent Model, and Stereotyped Respondent Model. The first two kinds of interviews, similar to their namesake in the Bloom et al. study, generated the respondent speech that was used for the respondent model. In the Required Clarification interviews, respondents first answered each question; after this they were presented the full definition and could change their response if they chose to. These interviews served two functions. First, they provided a test bed for the models. In particular, they allowed us to ask how precisely the model’s predicted comprehension accuracy prior to the definition being presented. Second, they served as a benchmark of comprehension accuracy. Because definitions were pronounced for all questions, response accuracy under these conditions provided an upper bound on the benefit of clarification. In the Generic Respondent Model interviews, the system inititated clarification after certain speech conditions (discussed next) were met, regardless of respondent characteristics. In the Stereotyped Respondent model interviews, the conditions that triggered the system to provide clarification were different for older and younger respondents.

The respondent models were calculated with ordinary least-squares regression techniques in which the predicted behavior was response accuracy and the predictors were filters, benefits, restarts, restarts, repairs, repairs, reports, repairs, clarification picks, and passes. While older respondents produced more spoken cues and longer
A CASE OF COMMUNICATION DIFFICULTY IN TELEPHONE INTERVIEWS

Most of our studies are laboratory-based, relying on small samples of respondents answering questions about nonassertive facts and behaviors, and professional interviewers given brief training in alternate interviewing techniques. How can generalizations from experimental manipulations to actual, large-scale surveys be possible at all? Experiments demonstrate that certain phenomena can happen but that they necessarily do happen under all circumstances. To clarify map experimental results to large-scale surveys, one must know how often the circumstances created in the laboratory actually occur in the wild.

We have focused on the accuracy of respondents' interpretation of questions (the extent to which their answers reflect the same interpretations as the survey designers'), rather than on other important indicators of data quality in surveys, such as response rate, completion, and breakdown rates. Whether the findings will extend beyond the lab to larger samples of respondents, different kinds of questions, different interviewer populations, and additional measures of survey quality remains to be seen.

The effects of attending to grounding cues are apparent particularly for situations where the respondent's circumstances are ambiguous with respect to (well-putt) questions, and to the frequency with which this occurs in real-world settings places a limit on the utility of our findings. Thus far, the evidence suggests that these sorts of "complicated-mapping" circumstances are frequent enough to worry about in broader samples and under more natural conditions—see Conrad and Schöber, 2000; Saarela et al., 2005—and that they apply to attitude and opinion questions as well as questions about facts and behaviors, but again we recommend caution in assuming this about every sample of respondents and questions. We should also note that there are other potential sources of trouble answering questions that we have not been investigating: trouble understanding technical terms, trouble deciding on the response task (e.g., estimate or count?), trouble retrieving answers from memory, and troubles resulting from ambiguous (polysemous) terms in questions. Whether the communication cues surrounding these other kinds of trouble are the same as those investigated here is an open question.

In general, we see the findings described here as raising a set of issues that need to be explored in much greater detail, and we hope that this discussion helps to prompt further research along these lines. In particular, we think it would be important to know the following:

(1) How diagnostic is need for clarification for communication cues across different respondents and circumstances? While we have seen good evidence across our samples that, for example, respondents use am more often in their first turn after a question is asked when the answer is likely to need clarification, it is also clear that there is substantial individual variability in discourse style, dialects, and pronoun use. The use from a respondent who never uses is presumably informative in a way that the same from a respondent who regularly uses is not. Ours aversion from a steady gazing respondent is different in meaning than gaze aversion from a respondent who never looks at the interviewer. To what extent do sensitive interviewers already attend to baseline rates of any potentially communicative cues—delay in responding, reporting, gazing, umming—as they decide whether to probe or clarify?
To what extent should interviewers be trained to attend to the individual variability of such cues, and to what extent should interviewing systems of the future be able to diagnose the individual variability of such cues?

It is entirely possible that what is a cue of the respondent's comprehension difficulty in one situation reflects a quite different internal state in another situation. Or, worse, the same cue might indicate different states in the same situation on different occasions. Consider the "looking away" cue. Because respondents look away longer in conversational interviews (when interviewers might react to the cue) than in standardized interviews (when they cannot), looking away would seem to be at least a candidate for a cue under respondents' control. But if respondents look away irrespective of interviewers' ability to detect it, this would more likely mean that looking away is an involuntary reflection of comprehension difficulty.

Alternatively, looking away could reflect something other than comprehension difficulty. It could indicate that the respondent is planning what to say next and does not want to be distracted by looking at the interviewer (c.f. Glöckner et al., 1998). Or it could reflect a state almost diametrically opposed to needing help: Looking away could reflect respondents' desire to maintain the floor and not surrender it to the interviewer, a concern absent in conversational than standardized interviews (see the discussion by Clark, 1996, of turn allocation rules, pp. 321-324). Finally, looking away (or any of the cues we have suggested reflect comprehension difficulty) could indicate ambivalence about answering truthfully. If a respondent is concerned that providing a truthful answer might somehow cause harm, for example, by confessing to illegal conduct, he might look away or answer less fluently or pause longer before speaking than if he has no reservations about answering.

(2) How does an interviewer's responsiveness to any communication cue affect the respondents' likelihood of using it? Presumably, when strictly standardized interviewers ignore explicit requests for clarification, they reduce the likelihood that any but the most persistent or conversationally lessactive of respondents will continue asking for clarification. Does the same go for less implicit communication cues? The preliminary evidence reported here hints that rates of disfluency may well be sensitive to interviewers' responsiveness to them. Surely this is not the sort of thing that is under respondents' conscious control, and it suggests a kind of dyadic regulation of individual processes that is not part of many views of communication.

(3) Are all interviewees equally sensitive to grounding cues? Although reliable empirical evidence on this is rare, ordinary intuition and clinical observation of the general population suggest that people can vary substantially in their interpersonal sensitivity: their empathy, perspective-taking ability, and ability to attend to subtle linguistic cues (see, e.g., Davis, 2005, and other chapters in Malhe and Hodges, 2008; Schoder and Schœn, 2005). Presumably, interviewers who are socially more adept do not survive long in the job, overly sensitive interviewers, for whom denying requests for clarification may be interpersonally aversive, may also not survive as telephone survey centers that require the strictest of standardized practice. What is unknown is the extent to which sensitivity to such cues is trainable, or whether adult language users already have an ingrained repertoire of cues to which they attend.

CONCLUSIONS

is resistant to change. Presumably, there are individual differences in interviewers' sensitivity to grounding cues (the approach of Hall and Bernieri, 2001 might allow assessment of this kind of skill), which will contain the effectiveness of training. In particular, for interviewers low in sensitivity to such cues, the additional task of monitoring for them may be unnecessarily burdensome (see Japel, 2005, for a discussion of interviewer burden). Also, external constraints like time pressure to finish interviews may cause even the most interpersonally astute interviewers to ignore potentially useful cues.

(4) How nonverbal are communication cues? Thus far, little is known. In general and in survey interviews—about the extent to which visual cues provide information distinct from that provided by tone of voice, delay, or content of what is said. While one can find clear examples where a particular cue seems to be the only indicator of trouble, we do not know whether a single cue is always sufficiently diagnostic to warrant intervention by an interviewer or an interviewing system. To complicate matters, it is possible that different respondents may have different discourse styles: One respondent's gaze aversion may always be accompanied by a pause and an um, while another's gaze aversion may provide unambiguous nonverbal information. To the extent that cues are redundant, interviewers who already have a lot to attend to might be able to rely on the cues in the most easily available or measurable channel.

(5) How multivariate are communication cues? Our approach thus far has been focused on the cognitive aspects of comprehension questions and how interviewers' difficulties and other cues provide relevant evidence. But every cue we have discussed—explicit requests for clarification, reports, delays, and so on—is also a potential indicator of the respondent's emotional state, level of attention, and likelihood of continuing the interview. Respondents could delay or rush responses not only because they have trouble understanding the question or have not thought it through enough, but also because they feel a question intrusive, because they feel the interview has gone on too long, or because the interviewers' nonresponsiveness to a request for clarification is becoming trying. To what extent do grounding cues also provide evidence about the rapport and emotional alliance between interviewers and respondents? We suspect that although grounding cues and rapport cues are conceptually distinct, in practice they can be quite intertwined. For example, an interviewer's apology for the stiffness of an interview ("I'm sorry, I can only repeat the question") can be a response to cues that the interview is going offtrack both on affective dimensions (the respondent's frustrated tone of voice) as well as on grounding dimensions (explicit and implicit indicators of need for clarification).

As the space of new technologies available for telephonic interviewing, telephone interviews are beginning to share more features with face-to-face interviews (see Schoder and Coull, 2005 as well as other chapters in Coull and Schœn, 2005). Will additional visual information help improve comprehension and thus data quality? Whitaker (2003) has observed that across various (nonverbal) domains there is little evidence in support of the bandwagon hypothesis: the idea that adding visual information to speech will improve the efficiency of communication. It may
be that the total amount of usable information about a communicative partner’s need for clarification is the same with or without video. It remains to be seen what the facts are for surveys with different populations of respondents, with individually and culturally variable communication styles, with different domains of questioning (sensitive and nonsensitive questions), and with different interviewing agents (human versus computer) with different capabilities (diagnosing and responding to requests for clarification versus having interpretation up to respondents). How these questions are answered will help shape future debates about how telephone interviews should be conducted.

CHAPTER II

Oral Translation in Telephone Surveys

Janet Harkness
University of Nebraska at Lincoln, USA and ZUMA, Mannheim, Germany

Nicole Schoebi and Dominique Joye
SIDOS, Neuchatel, Switzerland

Peter Mohler, Timo Faass, and Dorothee Behr
ZUMA, Mannheim, Germany

11.1 INTRODUCTION

In the context of growing cultural diversity, telephone studies are regularly faced with the challenge of collecting data from respondents who do not speak the majority language of a sampled population. In order to avoid problems related to coverage, nonresponse, and measurement, projects need to approach these respondents in a language they can use comfortably. Three basic options are available to interview such respondents: projects can have written translations prepared, can use interpreters, or can hire bilingual interviewers to translate and interview at the same time. In this chapter, we focus on the last option, when interviewers translate orally while they are interviewing. We present and discuss transcripts of interviews where translators were translating orally and compare these to transcripts of the same interviews when they were translated using a written translation. Interviews for which there was a written version of the questionnaire in the language of the interview are called here scripted interviews and stand in contrast to the orally translated interviews for which interviewers did not have a written

---

We would like to thank the Schweizerische Doku für Dauer in der Sozialforschungsform (SIDOS), the Swiss Social Science Data Archive, in Neuchatel, Switzerland, and the Zentrum für Umfragen, Methoden und Analysen (ZUMA), in Mannheim, Germany for financing the study reported here.

Advances in Telephone Survey Methodology. Edited by James M. Lindsay, Clyde Tucker, R. Michael Bickel, Robert de Leeuw, Lilli Joppa, Paul J. Lovaas, Michael W. Last, and Robert L. Bangert
Copyright © 2005 John Wiley & Sons, Inc.
THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation’s journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it.

WILLIAM J. PIERCE  PETER BOOTH WILEY
PRESIDENT AND CHIEF EXECUTIVE OFFICER  CHAIRMAN OF THE BOARD

Advances in Telephone Survey Methodology

JAMES M. LEPKOWSKI  LILLI JAFEC
Institute for Social Research  Department of Research and Development
University of Michigan  University of Stockholm
Ann Arbor, MI  Stockholm, Sweden

CLYDE TUCKER  PAUL J. LAVRAKAS
Bureau of Labor Statistics  Nielsen Media Research
U.S. Department of Labor  New York, NY
Washington, DC

J. MICHAEL BRICK  MICHAEL W. LINK
Wassat  Centers for Disease Control and Prevention
Rockville, MD  Atlanta, GA

EDITH DE LEEUW  ROBERTA L. SANGSTER
Department of Methodology and Statistics  Bureau of Labor Statistics
Utrecht University  U.S. Department of Labor
The Netherlands  Washington, DC

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION