Social scientists in any discipline struggle with the question of how to gather appropriate data to answer their research questions. Because pragmatics is a fairly recent arrival and practiced by scholars educated in different research traditions, data collection in pragmatics draws on the methods and techniques developed in many of the older and better established social sciences, such as anthropology, sociology, psychology, and linguistics. In this article, I will review the types of data collection most commonly used in pragmatics to date (summarized in Table 1).

Table 1
Focus and Procedure in Some Data Collection Formats

<table>
<thead>
<tr>
<th>Focus</th>
<th>Procedure</th>
<th>Interaction</th>
<th>Comprehension</th>
<th>Production</th>
<th>Meta</th>
<th>Online/offline</th>
<th>Interaction with researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic discourse</td>
<td>on</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elicited conversation</td>
<td>on</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td>on</td>
<td>+/-</td>
</tr>
<tr>
<td>Roleplay</td>
<td>on</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production questionnaire</td>
<td>off</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple-choice</td>
<td>off/on</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scaled response</td>
<td>off</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>off</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diary</td>
<td>off</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinkaloud protocol</td>
<td>on</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The columns under *Focus* specify whether the data can inform different aspects of language use—interaction, comprehension, or production—or the participant’s metapragmatic knowledge and subjective theories. In the *Procedure* columns *online/offline* indicates whether data are collected while the participant is engaged in an activity involving language use (i.e., online) or whether the participant is prompted to recall pragmatic information from memory and report rather than use it (i.e., offline). *Interaction with researcher* refers to whether or not researchers-participant interaction is an inherent part of the procedure, as in interviews. Even in the data types marked minus interaction, researcher and participant will have some form of contact prior to and sometimes during the data collection, and this interaction may well influence the data.

*Authentic discourse, elicited conversation,* and *roleplay* are types of spoken interaction; *production questionnaires, multiple-choice,* and *scaled response instruments* are survey methods and thus obtain written responses when self-administered; *interviews* are a specific type of spoken interaction that may or may not be structured by a questionnaire; in its less structured forms, interviews produce narrative self-reports and are thus akin to *diaries* as a story-telling genre. *Thinkaloud protocols* can be related to interviews and diaries in that they, too, produce narrative self-reports; however, in their classic form, they are on-line verbalizations of thought processes rather than stories. *Verbal protocols* have their home in experimental psychology and are thus furthest removed from the conversational interaction in authentic talk activities that opened the list of data collection procedures in pragmatics. I shall now consider each procedure in turn.

**SPOKEN INTERACTION**

The common denominator of authentic discourse, elicited conversation, and open-ended roleplay or simulation is that the data are oral, interactive productions and thus allow examination of a wide range of discourse features, including the overall structuring of talk exchanges, the distribution of turns at talk, sequencing of conversational contributions, speaker-listener coordination, and participants’ joint achievement of transactional and interpersonal goals. They shed light on participants’ production of communicative action and their (mis-)comprehension of the interlocutor’s contributions. The obvious difference between the three types of spoken interaction is that authentic discourse is motivated and structured by participants’ rather than the researcher’s goals whereas elicited conversation and roleplay are brought into being for research purposes. From a sociolinguistic and pragmatic perspective, this fact alone underscores that neither of the two elicited data types can ever be *the same* as authentic conversation since the overall purpose of an interaction is its most powerful structuring force—hence
designations such as *activity type* as a unit of analysis for studying contextualized talk exchanges. But very much unlike the popular perception that *inauthentic* equals *invalid*, interactions arranged for research purposes can be most useful data collection procedures if used judiciously.

**Authentic Discourse**

Two recoding techniques are most commonly used to collect data on authentic spoken discourse: taking fieldnotes and audio- or videorecording. Data on individual communicative acts have been gathered either by notetaking (e.g., on compliments: Herbert, 1989; apologies: Holmes, 1990; invitations: Wolfson, d’Amico-Reisner, & Huber, 1983) or audiorecording (e.g., on compliments: Pomerantz, 1978; apologies: Owen, 1983). In studies of extended speech events, by contrast, notetaking is not an option; they require electronic recording. Research on authentic speech events with nonnative speaker participants has predominantly focused on discourse in institutional settings, such as business negotiations (Ehlich & Wagner, 1995), academic advising sessions (Bardovi-Harlig & Hartford, 1996), and oral proficiency interviews (Young & He, 1998), to name but a few. This choice is motivated by a number of reasons.

Institutional discourse lends itself well to demonstrate the interrelation of text and context, for instance, how institutional structures influence communicative action and are reproduced by it. The relationship of communicative action and social power is thus particularly evident in institutional discourse. Compared to interpersonal conversation, institutional talk has the advantage of being more highly structured, routinized, and recurrent, a direct consequence of the purpose of the institution, role distribution between actors (institutional representative vs. client), and actors’ goals. From a research methodological perspective, this is greatly advantageous because the institutional patterning allows researchers to observe native and nonnative speakers in the same social roles, usually that of a client (one interesting exception being the international teaching assistant, who is an institutional representative). Thus studies of institutional discourse often include an inbuilt control group (e.g., Bardovi-Harlig & Hartford, 1996).

Perhaps the most difficult part in gathering extended authentic data is to gain access to the research site. Institutions are often reluctant to allow any form of observation, and if they do, they may not allow recording. Yet without audio- and preferably videorecordings, the entire research enterprise will be in jeopardy. As we know from microanalytic approaches to discourse, such as conversation analysis (e.g., Heritage, 1997), interactional sociolinguistics (e.g., Gumperz, 1996), and ethnographic microanalysis (e.g., Erickson, 1992), discourse features and strategies exhibit finely
structured co-occurrences and regularities that escape even the well-trained observer and are impossible to fixate in memory. There is thus a real danger that memorization and taking fieldnotes will result in recording salient and expected (or particularly unexpected) facets of the interaction, at the expense of less salient but perhaps crucial (often indexical) material. Fieldnotes are a valuable source for providing contextual information and are indispensable in ethnographic studies, yet they cannot replace electronic recordings (Duranti, 1997, also for practical suggestions for recording interaction). In fact, this is not only true for the study of extended speech events but may also apply to the investigation of individual speech acts. For instance, notetaking has proven to be a productive method for studying compliments in different varieties of English because compliments are most frequently packaged as single-turn utterances with a simple, short, highly formulaic structure (e.g., Wolfson, 1989). But in studying compliment responses, it was only through the microanalysis of audiorecorded exchanges that it could be demonstrated how recipients manage the conflicting pragmatic preferences of agreeing with the interlocutor while avoiding self-praise at the same time. Notetaking from memory cannot capture temporal phenomena such as brief silences, pauses, and draws, yet such material carries crucial indexical information that interacts with recipients’ linguistic choices in creating the pragmatic meaning of compliment responses. Thus the most insightful early study on compliment responses (Pomerantz, 1978) would not have been possible without the microanalysis afforded by the audiorecorded data.

In discussions of how best to collect authentic conversational data, the Observer’s Paradox (Labov, 1972) is often brought up: is the researcher’s presence not going to alter the normal course of the interaction, and will this effect not be exacerbated by a videocamera? As Duranti (1997) argues, in extended ethnographic research, the observer effect is usually temporary. The initial disturbance of routine transactions through the presence of an outsider and her video equipment will subside when the novelty effect has worn off and the routines kick in again. Routinized actions (for instance, teachers’ and students’ classroom practices) are highly overlearnt and difficult to change; they will quickly reemerge once the interaction is under way. However, since initial observer effects are quite possible, researchers should refrain from the get-your-data-and-run type of data collection. In ethnography, prolonged engagement in the field is a fundamental methodological principle (for many reasons besides the observer’s paradox; see Watson-Gegeo, 1988; Davis, 1995 for ethnography in applied linguistics, Davis & Henze, 1998, for ethnography and pragmatics), “blitzkrieg ethnography” (Rist, 1980) amounting to malpractice. In research with a predominantly discourse analytical orientation, it may be helpful if the researcher and her recording device are present in the setting for some time prior to the data collection so that participants can get used to having them around. This
extra time should be calculated as an indispensable part of the data collection and will be rewarded by a better quality of the data.

In addition to the problems of gaining access and observer effects, authentic data may have other drawbacks. Depending on the research purpose, it may take an unreasonable amount of data to obtain sufficient quantities of the pragmatic feature under study, for instance, of a particular speech act. Especially when comparison of nonnative speakers and native speakers of both the learners' L1 and L2 is essential for the research goal, as in studies of pragmatic transfer, authentic data may just not be an option. In this very common situation, elicited conversations or roleplays offer an alternative.

Elicited Conversation

The term *elicited conversation* refers to any conversation staged for the purpose of data collection. Unlike in roleplays, participants do not take on social roles different from their own; however, they assume discourse roles assigned by the researcher. We can distinguish two varieties of elicited conversation.

In *conversation tasks*, participants are requested to converse about a topic or jointly reach a particular goal determined by the researcher. Instructions can be as vague as asking participants to get to know each other (e.g., Scarcella, 1983; White, 1989) or prescribe specific tasks, such as addressing the interlocutor's troubles telling (Kerekes, 1992). Data elicited through conversation tasks has been found useful in studying various aspects of conversational management (Scarcella, 1983) such as backchanneling (White, 1989), the use of indexicals such as the Japanese sentence-final particle *ne* (Yoshimi, in press), and effects of pragmatic transfer in the use of discourse markers and strategies. Kerekes (1992) investigated if and how participants respond to troubles tellings by offering advice or expressions of sympathy. Her study thus focused on specific responding speech acts in troubles telling events. Even though nothing prevents conversation tasks from including differentially symmetric or asymmetric participant configurations, the cited studies featured equal status encounters with fairly balanced participation structures, as far as one can tell from the reports.

In this regard, conversation tasks are systematically different from the other variety of elicited conversation, the *sociolinguistic interview*. As with any kind of interview, the sociolinguistic interview is an asymmetrical speech event in which "one party asks the questions and the other party gives the answers" (Schegloff, 1992, p. 118). Unlike the conversation task, sociolinguistic interviews thus have a genre-specific structure. As part of the standard repertoire in sociolinguistic data collection (Labov, 1984; Schiffrin, 1987), the interviewer asks the informant about her life history, experiences, and attitudes. In
Labov's original design, one important function of the sociolinguistic interview was to ask the informant about highly emotional experiences under the assumption that such topics would trigger vernacular speech. Topic investment has been shown to affect interlanguage performance (Eisenstein & Starbuck, 1989), but its effect on learners' L2 pragmatics has not yet been explored. Sociolinguistic interviews with L2 learners have been analyzed for conversational management and repair (Fæmh & Kasper, 1982) and the acquisition of the Japanese particle ne. Tao and Thompson (1991) examined retroactive transfer in the backchanneling patterns of native speakers of Mandarin Chinese from their interlanguage English, whereas Sawyer (1992) compared learners' production of ne in four interviews, conducted over the period of one year, in order to determine developmental patterns in the use of the particle. Sawyer's (1992) study, one of the first on nonnative speakers' pragmatic development (cf. Kasper & Schmidt, 1996; Kasper & Rose, 1999, for review), raises an interesting design issue. As native speaker baseline data, Sawyer used the frequency of ne in the discourse contributions of the Japanese interviewer. This presents a potential validity problem because the asymmetrical structure of the interview positions interviewer and informant in different discourse roles and this asymmetry is very likely reflected in, and in fact co-constructed by, the use of ne. As a key discourse marker, ne has been shown to index epistemic and affective stance (e.g., Cook, 1992; Yoshimi, in press); as an indexical, its use is, by definition, highly context-sensitive. If comparisons to native speaker use of ne are made, they should be to speakers in the same discourse roles as the nonnative speakers. In other words, baseline data could have been obtained by conducting interviews with native speakers of Japanese that were otherwise comparable to the nonnative informants, allowing examination of the use of ne in the native and nonnative informants' interview responses.

Elicited conversations have the capacity to shed light on such discourse aspects as conversational organization and management, the expression of reference, modality, temporality, and aspect, task-specific communicative acts, and narrative structure. But they are also limited in that they allow investigating only a restricted set of communicative acts and activities, and that participant roles cannot be manipulated.

Roleplay

This limitation of elicited conversations is overcome in roleplays, i.e., simulations of communicative encounters, usually in dyads, based on role descriptions. Roleplays can be defined as "a social or human activity in which participants 'take on' and 'act out' specified 'roles', often within a predefined social framework or situational blueprint (a 'scenario')" (Crookall & Saunders, 1989, p. 15f).

Different types of roleplay can be distinguished according to participant involvement
and extent of interaction. In spontaneous roleplays, players retain their own identities. In mimetic-replicating roleplays, they play the role of a visually presented model, while in mimetic-pretending roleplays, actors assume a different identity (Kipper, 1988). Useful as these categories are as a first rough distinction, they are too broad to capture other potentially important variables that might affect the quality of roleplaying. For instance, a particular type of spontaneous roleplay is the idiographic roleplay, in which actors recall and roleplay specific, recent, and relevant extended interactions (Kern, 1991). The great advantage of this type of roleplay is that actors can rely on recent episodic memory, which will reduce the cognitive load associated with having to invent the action online.

Roleplays also differ in the extent of the interaction. In interlanguage pragmatics, a distinction has been suggested between closed and open roleplays (Kasper & Dahl, 1991).

In closed roleplays, the actor responds to the description of a situation and, depending on the communicative act under study, to an interlocutor's standardized initiation. This procedure has been used to elicit requests (Rintell, 1981; Rintell & Mitchell, 1989), suggestions (Rintell, 1981), and apologies (Cohen & Olshtain, 1981; Rintell & Mitchell, 1989), organized as single turn speech acts. Open roleplays, on the other hand, specify the initial situation and each actors' role and goal(s) on individual role cards, but the course and outcome of the interaction are in no way predetermined. For instance, a fairly complex interaction would be one where actors discover during the interaction that they have conflicting goals and have to negotiate how to manage their goal conflict, as in the example below.

A. You are going to move into a new apartment on Saturday. It is Thursday today, and you have just received a call from a friend of yours who was supposed to help you move house, saying that he is unable to help you move after all. You don't have a car or a driver's license, so you depend on the help of somebody who does. You decide to ask B, your next-door neighbor. The two of you are friends, and you have helped each other out before. You go to see B.

B. It is Thursday. You have just made arrangements with some friends to spend the weekend in the country. You and your friends are planning to go in your car, leaving Saturday morning and coming back Sunday night. You are at home, watching TV, when the door rings. You can see through the peephole that it is your friend and neighbor, B.

Unlike closed roleplays, an open roleplay such as one based on these role descriptions will evolve over many turns and different discourse phases. Communicative acts will be organized over multiple turns and their sequencing will be strongly influenced by the interlocutor's uptake. The conversational activity will address interpersonal functions,
such as politeness, and interactional functions, such as coordinating speaker and listener contributions through turntaking and backchanneling. Open roleplays thus allow observation of those aspects of conversation that are fairly independent of particular contexts and goals, but unlike authentic discourse and elicited conversation, they also permit us to design contexts and roles that are likely to elicit specific speech events and communicative acts. Moreover, through the role specifications, they also enable us to observe how context factors, such as power, distance, and imposition in Brown and Levinson’s (1987) politeness theory, influence the selection and realization of communicative acts and how the values of these factors may be changed through conversational negotiation. The rich potential of roleplays for eliciting pragmatic and sociolinguistic features in their full discourse context is evident from interlanguage pragmatics research on communicative acts such as requests (Hassall, 1997), expressions of gratitude (Eisenstein & Bodman, 1993), apologies (Garcia, 1989), complaints (Trosborg, 1995), refusals (Widjaja, 1995), and various face-threatening acts (Piirainen-Marsh, 1995); (mis)understanding in service encounters and institutional discourse (Bremer, Roberts, Vasseur, Simonot, & Broeder, 1996), discourse cohesion (Stemmer, 1981); gambits (Wildner-Bassett, 1984, 1994); conversational organization and maintenance (Edmondson, House, Kasper & Stemmer, 1984); routine formulae (Tateyama, Kasper, Mui, Tay & Thananart, 1997); and pragmatic fluency (House, 1996).

We thus know that roleplay produces all aspects of conversation, but we do not know from the use of roleplays alone whether they provide valid representations of conversational practices in authentic contexts. Whereas validity concerns loom large in behavioral assessment and other social sciences using roleplay as a research tool, only a few studies have examined the validity of roleplay in interlanguage pragmatics. Though not methodological in focus, Eisenstein and Bodman’s (1993) study of expression of gratitude by native and nonnative speakers of English sheds light on the effects of three data collection procedures, production questionnaires, open-ended roleplays, and fieldnotes on expressions of gratitude occurring in authentic interactions. All three data types yielded the same words and expressions, yet they differed in length and complexity. The production questionnaire data were the shortest and least complex, the authentic data the longest and most complex, with the roleplay data coming in between. The oral data included more restatements of thanks and discussions about the received gift or service. Both roleplay and authentic data demonstrated that thanking is collaboratively enacted, involving the giver as much as the receiver. The native speaker roleplays were often longer than the nonnatives’ because the native speakers did not ask for services or favors directly; rather, they would indirectly suggest their needs. In the native speaker dyads, but not in dyads involving nonnative speakers, the length of the speech event increased with
greater indebtedness. The less fluent nonnative speakers were not able to compensate for shorter linguistic expressions by conveying appreciation prosodically. The nonnative speakers performed better in interaction with native speakers than with other nonnatives, who were unable to provide the same kind of support for the less competent partner.

Three studies compared roleplay and written production questionnaires. Margalef-Boada (1993) examined the production of refusals by native speakers of German, Spanish (including bilinguals in Castilian and Catalan), and German learners of Spanish in both data types. In the roleplays, native speakers were paired with native speakers and nonnatives with nonnative speakers. The stimulus contexts were the same for the oral and written condition, each specifying a request, offer, invitation, or suggestion that the participant had to refuse. The analysis showed the same content and range of semantic formulae in both conditions, with most of them occurring with roughly the same frequency. Different distributions were most noticeable in direct refusals, expressions of regret, avoidance, and adjuncts. The main difference between the written responses and the oral interaction was the large number of semantic repetitions in the roleplay (p. 116). Through the interactive nature of the roleplay and the multiple turns over which the refusal event evolved, the roleplays were naturally longer, “richer and more complex” (p. 153) than the written single-turn responses. However, in both conditions, participants produced less polite refusals than would be appropriate in an authentic setting, suggesting that the absence of social consequences may have relaxed their adherence to politeness norms.

Sasaki’s (1998) comparative study of production questionnaires and roleplays differed from Margalef-Boada (1993) in that she investigated two communicative acts, requests and refusals, and administered the two tasks to the same participants, Japanese EFL learners. In the roleplays, confederates were native speakers of American English. Both methods elicited similar Head Acts and supportive moves for requests and types and order of semantic formulae for refusals. But responses varied in length and content, the roleplay contributions featuring longer utterances and a greater variety of strategies.

Obviously, when conversational interaction and the sequencing of communicative action in conjunction with turn-taking is the research focus, an interactive procedure such as roleplay needs to be chosen. On the other hand, if the purpose of an investigation is to inform about the types of strategies by which a communicative act can be implemented, written production questionnaires are an effective means of data collection, as we will see below.

A third study (Edmondson & House, 1991) found that nonnative speakers produced longer and more verbose utterances than native speakers in production questionnaires but not in roleplays. The authors attribute this finding to learners either lacking or not having
sufficient control over routine formulae. But in addition, the fictive world of the roleplay may affect learners (especially those at a lower level of L2 proficiency) differently than native speakers. Whereas in authentic interaction, participants’ planning and execution of communicative action is supported by a rich social context, participants in a roleplay have to imagine their own and their co-actors’ roles. Constructing and interpreting communicative intent can easily overstretch learners’ social imagination and memory and thus reduce their capacity for online input processing and utterance planning. It is therefore possible that most types of roleplay underrepresent learners’ pragmatic ability. It would be worthwhile to examine whether idiographic roleplays improve learners’ performance and validity, as they demonstrably do in native speaker performance.

**QUESTIONNAIRES**

Compared to spoken interaction methods, questionnaires may appear far more restricted in the scope of questions they allow us to study. Excluded from investigation are precisely those pragmatic features that are specific to oral interactive discourse—any aspect related to the dynamics of a conversation, turn-taking and the conversational mechanisms related to it, sequencing of action, speaker listener coordination, features of speech production that may have pragmatic import, such as hesitation, and all paralinguistic and nonverbal elements. Despite these limitations, different forms of questionnaire data are the most commonly used data types in interlanguage pragmatics.

The three types of questionnaire used in pragmatics—production, multiple-choice, and rating scale questionnaires—differ from each other in the type of response they elicit. Production questionnaires are open-ended in the sense that they require a participant-generated textual response that is coherent with the context specified in the stimulus item. Multiple-choice and scaled-response questionnaires, by contrast, provide fixed response alternatives from which the participant has to choose the most appropriate one. We shall examine each of the three types in turn.

**Production Questionnaires**

Items in a production questionnaire include a situational description and a brief dialogue with one turn as an open slot. The specified context is designed to constrain the open turn so that a specific communicative act is elicited. Item formats used in different studies vary in a number of ways. In the classic discourse completion format (a), the exchange is terminated by a provided rejoinder (1 & 2) and can also be prefaced by an interlocutor initiation (1). The rejoinder can be positive as in (1), where Charlie accepts Jim’s (expected) apology, or negative as in (2), where Leslie refuses Walter’s (expected)
request.

(a) 'classic' discourse completion

(1) In the lobby of a university library
Jim and Charlie have agreed to meet at six o'clock to work on a joint project. Charlie arrives on time and Jim is an hour late.
Charlie: I almost gave up on you!
Jim: __________________________________________
Charlie: O.K. Let's start working.

(2) After a meeting
Walter and Leslie live in the same neighborhood, but they only know each other by sight. One day, they both attend a meeting held on the other side of town. Walter does not have a car but he knows that Leslie has come in her car.
Walter: __________________________________________
Leslie: I'm sorry but I'm not going home right away.
(Blum-Kulka, House, & Kasper, 1989)

This basic format has been variously modified. In dialogue construction (b) and the open response formats (c, d), no rejoinder is provided. Dialogue construction can either be initiated by a provided first pair part (3) or the participant has to provide both (or all) contributions (4).

(b) Dialogue construction

(3) Your advisor suggests that you take a course during the summer. You prefer not to take classes during the summer.
Advisor: What about taking Testing in the summer?
You say: __________________________________________
(Bardovi-Harlig & Hartford, 1993)
(4) At a restaurant

Catherine is having dinner at an expensive restaurant. When she is getting up from the table she bumps into a waiter, who spills a tray of food.

Catherine: ____________________________________________
Waiter: ________________________________________________

(Bergman & Kasper, 1993)

The open response formats differ in requiring a verbal response (c) or allowing a verbal, nonverbal, or no response, that is, to opt out (d). The choice to opt out permits to identify sociopragmatic differences in the appropriateness of communicative acts (Bonikowska, 1988).

(c) open item - verbal response only

(5) It’s your birthday, and you’re having a few friends over for dinner. A friend brings you a present. You unwrap it and find a blue sweater.

You say: ____________________________________________

(Eisenstein & Bodman, 1993)

(6) It is not the first time that loud rock music is heard from your neighbor’s apartment quite late at night.

You pick up the phone and say: __________________________

(Olshtain & Weinbach, 1993)

(7) An American classmate offers you a ride to an off-campus meeting which you are both required to attend. You are planning to go, and you need a ride, but you do not want to ride with her because of her unsafe driving.

Classmate: Since we’re both going to that meeting downtown on Thursday, why don’t you ride with me? I’d like some company and besides, you could help me find the place we’re supposed to go to.

You: ____________________________________________

(Robinson, 1992)
(d) open item, free response

(8) You are a corporate executive. Your assistant submits a proposal for reassignment of secretarial duties in your division. Your assistant describes the benefits of the plan, but you believe it will not work.
You: __________________________________________

(Beebe & Takahashi, 1989; Takahashi & Beebe, 1993)

(9) You are going shopping with a friend. She is trying on a pink blouse which she is thinking of buying. In your opinion it does not look good on her at all.
You would: __________________________________________

(Steinberg, 1995)

Does item format, specifically, the presence or absence of a rejoinder and, in case a rejoinder is included, whether it is positive (1) or negative (2), influence responses? A recent study investigated this issue and found that different item formats have an effect on participants’ strategy choices for requests, complaints, and apologies. Results from production questionnaires with different item formats are thus not directly comparable (Johnston, Kasper, & Ross, 1998).

Furthermore, production questionnaire formats differ as to whether they require the participant to imagine how a fictive person would act in the specified situation (1, 2, 4, 8) or as they think they themselves would act (3, 7). In this regard, production questionnaires allow for the same roletaking options as roleplays.

An obvious question to worry about is whether and how the written mode as opposed to spoken production may result in different responses. Rintell and Mitchell (1989) addressed this question by comparing spoken and written requests and apologies from native and nonnative speakers of English, elicited by oral and written versions of the same production questionnaire. They found that nonnative speakers’ but not native speakers’ oral responses were significantly longer than their written responses, and in some situations both groups were more direct in the written than in the spoken mode. But these differences were outweighed by the similarities of the written and oral responses, suggesting that strategy choice and wording of single-turn responses to production tasks may be fairly stable across modalities.

However, not all participants—for instance, young children and low proficiency L2 learners—are capable of reading and providing written responses to production questionnaires. In order to accommodate such learners, Rose (in press) developed the
Cartoon Oral Production Task. This instrument, designed for primary school EFL students in Hong Kong, comprises a series of cartoons, each depicting a familiar scenario. Research assistants worked with each child individually, explaining the scenarios and eliciting the desired communicative acts (requests, apologies, or compliment responses), which were tape-recorded.

A serious concern is how production questionnaires compare to authentic data. Beebe and Cummings (1996, originally presented 1985) compared refusals elicited through a single item questionnaire with refusals performed in telephone conversations in response to the same request. Interlocutors in these interchanges were native speakers of American English. The questionnaire responses did not represent natural speech with respect to the actual wording, range of refusal strategies, and response length, but they modeled the “canonical shape” of refusals, shed light on the social and psychological factors that are likely to affect speech act performance, and helped establish an initial classification of refusal strategies.

Hartford and Bardovi-Harlig (1992) examined the rejections by native and nonnative graduate students of their academic advisers’ suggestions for the students’ course schedules. The production questionnaire elicited a narrower range of semantic formulae and fewer status-preserving strategies than the authentic data, yet it proved an adequate instrument to test hypotheses derived from the authentic interactions. The questionnaire data confirmed Hartford and Bardovi-Harlig’s (1992) hypothesis that the nonnative speakers were more likely to use unacceptable content to reject advice than the native speakers.

When carefully designed, production questionnaires are useful for gathering information about speakers’ pragmalinguistic knowledge of the strategies and linguistic forms by which communicative acts can be implemented, and about their sociopragmatic knowledge of the context factors under which particular strategic and linguistic choices are appropriate. Whether or not speakers use exactly the same strategies and forms in actual discourse is a different matter, but the questionnaire responses indicate what strategic and linguistic options are consonant with pragmatic norms and what context factors influence their choice (although recent studies suggest some qualification, see below). In interlanguage pragmatic research, we may be interested in finding out what L2 learners know as opposed to what they can do under the much more demanding conditions of conversational encounters. For such research purposes, production questionnaires are an effective option.
Multiple-Choice

Multiple-choice is a versatile questionnaire format because it allows eliciting information on production, comprehension, and metapragmatic judgments. Just like production questionnaires, multiple-choice items specify the situational context and include a prompt for a response, but rather than leaving the response selection to the participant, they specify several response alternatives from which one has to be chosen. This is illustrated in an item from a request study.

You are having dinner with your friend’s family. The food that your friend’s mother has prepared is delicious, and you want some more. What would you say or do?

A. I would wait until the mother saw my empty plate and offered more food.
B. ‘Please give me more food.’
C. ‘This food sure is delicious.’
D. ‘Could I have some more please?’ (Rose, 1994)

Comparison of responses to production questionnaires and multiple-choice have indicated differences in the requests provided by native speakers of Japanese in Japanese (Rose, 1994; Rose & Ono, 1995) and advice giving by Chinese ESL learners and native speakers of American English (Hinkel, 1997). In the request studies, the Japanese respondents selected more indirect opting out strategies in the multiple-choice than in the production questionnaire. In the advice study, the Chinese ESL learners chose more direct advice strategies whereas the English native speakers selected less direct and opting out strategies in the multiple-choice but not in the production questionnaire. In all three studies, the multiple-choice results were more consistent with reports on preferences for pragmatic strategies in authentic settings, although no direct comparisons with authentic data were made. The reasons for these differences are far from being well understood, but they call for more research into the validity of both questionnaire types. From a cognitive perspective, the two questionnaire formats impose quite different processing demands, the open-ended production questionnaire presenting a free-recall task, whereas the closed format of the multiple-choice presents a recognition task (Schwarz & Hippler, 1991). Multiple-choice responses require that subjects evaluate a very small number of presented alternatives against their memory structures of compatible events, a much less demanding task than having to conduct a free memory search and make an appropriate selection from a wide array of possible solutions. Moreover, as far as opting out goes, tasks may not be compatible unless the possibility for opting out is expressly specified as a legitimate option in the production questionnaire.

Similar to speech act production, multiple-choice tasks are one means of studying pragmatic comprehension. For instance, in a series of studies, Bouton (e.g., 1988, 1994)
examined how advanced ESL learners understand indirect answers, as in the example below.

Two teachers are talking about a student's paper.
Mr. Ranger: Have you read Mark's paper on modern pirates yet?
Mr. Ryan: Yes. I read it last night.
Mr. Ranger: What did you think of it?
Mr. Ryan: I thought it was well typed.

How did Mr. Ryan like Mark's paper?
  a. He liked the paper; he thought it was good.
  b. He thought it was certainly well typed.
  c. He thought it was a good paper; he did like the form though not the content.
  d. He didn't like it.  

Multiple-choice tasks for examining speech act production and comprehension have in common that in their construction, the designer has to rely on previous research on the communicative act in question in order to make principled selections of the response alternatives. It is not good enough to invent responses intuitively because pragmatic strategies might escape the researcher's attention. Valid sources to select response alternatives from are speech act realization strategies collected through production questionnaires (Rose & Ono, 1995) or from spoken discourse (authentic or elicited) and free responses to comprehension questionnaires (Bouton, 1988).

Finally, multiple-choice can be used to shed light on respondents' metapragmatic knowledge, for instance, how they classify utterances in terms of speech act categories, and what elements in the utterance they use as interpretive resource. The example below is from Koike's (1989) study on learners' recognition of requests, apologies, and commands. Note that in this study, the stimuli are aurally presented so that learners were able to draw on prosodic features as contextualization cues.

Claudia: Oh, Maria, please I would like to go to the concert tonight but I don't have any money. This artist is one of my favorites. Can you lend me five dollar? I can pay you back tomorrow, I promise.

1. This is a(n)
   a. apology
   b. suggestion
   c. request
   d. command
2. What helped you understand?
   a. Certain words
      If so, name one: 
   b. Intonation
   c. Every word was comprehensible to me
   d. Totally incomprehensible (Koike, 1989)

Rating Scales

Often, pragmaticists are not interested in categorical judgments as in Koike’s (1989) example above, but they want to know as how appropriate, polite, deferential, and so forth, people assess strategies of communicative action and their linguistic realizations (usually in specific contexts), a question that corresponds to acceptability judgments in formal linguistics. But we may also want to know how people assess the values and weights of the context variables that influence strategic and linguistic choices, such as participants’ relative power, social distance, and the degree of imposition involved in a linguistic act. The first question raises a pragmalinguistic issue, derived from Hymes theory of communicative competence (see Hinkel, 1996, for a recent study comparing the pragmalinguistic perceptions of ESL learners and native speakers of American English). The second question poses a sociopragmatic problem, addressed in Brown and Levinson’s politeness theory (see Spencer-Oatey, 1996, for a comprehensive discussion of participant variables and Spencer-Oatey, 1993, for a crosscultural study). And just as in formal linguistics, the most common method of obtaining metapragmatic assessments is by eliciting scaled category responses. Such responses represent one form of self-report data, a common data type throughout the social sciences.

Metapragmatic assessments can be obtained for several purposes: as a research issue in its own right; as an additional resource to help interpret performance data; as a preliminary step towards developing the instrument for the main study; or as a combination of the above. In studies using any kind of data elicitation format, such as roleplays, production questionnaires, or multiple-choice, researchers need to know how respondents assess the context variables built into the stimulus situations. Such crucial information cannot be obtained by relying on researcher’s intuition. Sociopragmatic assessments of candidate contexts elicited in a pre-study enable researchers to ground their context constructions empirically and thus to improve control over context variables. Unfortunately, most crosscultural and interlanguage pragmatic studies lack such careful preparation; hence results are difficult to interpret. One exemplary investigation in which the instrument for the main study was developed through a sequence of pre-studies using
sociopragmatic and pragmalinguistic assessments is Takahashi (1995). The example below, from Shimamura’s (1993) study of Japanese EFL learners’ use of supportive moves in requests, illustrates scales for both types of assessment.

Your friend from the mainland is visiting this weekend. You haven’t seen her for a few years and this will be her first visit to Hawai‘i, so you have decided to take her around the island. But your car broke down and you do not want to spend a lot of money renting a car. Then you remember that your classmate who lives in the neighborhood just bought a new car last week. You decide to ask your classmate if you can borrow her car for the weekend. You say...

1) Could I borrow your car this weekend if you’re not using it? My car broke down. I’ll return it with a full tank of gas.

2) Could I borrow your car this weekend if you’re not using it? My car broke down.

3) My car broke down. Could I borrow your car this weekend if you’re not using it?

4) Could I borrow your car this weekend if you’re not using it?

5) Is it your right to make the request to your classmate in this situation?

6) Is your classmate obliged to lend you her car in this situation?
How likely is your classmate to lend you her car in this situation? 

1 2 3 4 5  

(Shimamura, 1993)

In constructing scaled response instruments, it is essential to heed the design principles specified in the sociometric and psychometric literature (e.g., Miller, 1991; Bryman & Cramer, 1994). Informativeness, reliability, and validity will be maximized if the following principles are followed:

1. Composite constructs (such a power or imposition) are unfolded into their underlying dimensions (e.g., for imposition in apologizing: severity of offense, obligation to apologize, likelihood for apology acceptance, offender’s face-loss).
2. Each dimension is operationalized by at least two indicators (e.g., for severity of offense: How serious is John’s offense? How upset is Paul by John’s damaging his car? How great is the damage done to Paul by John? How inconvenient is John’s offense to Paul? How costly is John’s offense to Paul?).
3. Rating scales are divided into five to seven steps.
4. The linguistic material used in items is crosslinguistically equivalent. This is achieved by backtranslation, an indispensable process in crosscultural and interlanguage pragmatics research when any kind of linguistic stimulus material is used.

MORE OPEN-ENDED INSTRUMENTS

Scaled response instruments such as rating scales, Likert scales, and semantic differential scales presuppose that the constructs under study are known and well-defined. When this is not the case, more participant-directed, open-ended types of self-report are preferable, such as narrative interviews, diaries, and verbal protocols.

Interviews

Exploratory research goals require open, inclusive, little predetermined modes of inquiry. But also for hypothesis-testing and triangulating purposes, researchers may prefer a dynamic, context- and respondent-sensitive procedure such as narrative self-reports. The most common procedure for obtaining such reports is the research interview.

Briggs (1986) estimated that “90 percent of all social science investigations use interview data” (p. 1). Large-scale sociological surveys are mostly conducted as
interviews. In ethnographic research, interviews are the second most regularly used data gathering procedure, next only to participant observation. But the scope and frequency of their use is not matched by concomitant analysis of just how interviews work and what kinds of data they produce. In order to understand the structure and process of interviews, it is useful to consider them from a discourse-pragmatic and cognitive perspective.

1. Research interviews are a particular kind of speech event. On the one hand, they share certain structural characteristics with other types of interview (such as counseling, job, medical interviews) that are common in many but by no means all speech communities (Fiksdal, 1990). On the other hand, research interviews include different varieties, depending on medium (face-to-face vs. telephone), degree of prestructuring (open-ended vs. based on an interview schedule), length, etc. Since the interplay of context factors shapes interview interaction and outcomes, it has to be taken into consideration in the analysis of interview data.

2. No matter how prestructured by an interview schedule, interviews are ineluctably interactive. The determined efforts to reduce the biasing effects of interaction through detailed interview(er) guidelines are therefore doomed to failure. Rather than attempting to minimize interaction effects, research on interviews should help understand the structure of interview interaction and how it is related to the data produced by it. The only way of gathering such information is through microanalysis of interview discourse (Suchman & Jordan, 1992).

3. The fundamental exchange structure of the interview is the question-answer sequence (Schegloff, 1992; Wolfson, 1976). Consequently, the interviewee’s answers will be “conditionally relevant” (Schegloff & Sacks, 1973) on the interviewer’s questions. The view of interviews as “pipeline(s) for transmitting knowledge” (Holstein & Gubrium, 1997, p. 113) and interview answers as the sole product of the interviewee’s mind is untenable in light of the discourse-structural properties of the interview.

4. Interviews on attitudinal objects and past events are traditionally predicated on a static and passive view of memory. Cognitive psychologists, by contrast, have insisted since Bartlett (1932) that memory is constructive. Previously encoded information interacts with more recently acquired memories, implicit theories, personal interests, and so forth (Pearson, Ross, & Dawes, 1992). A cognitive perspective thus supports the view that interview answers are no immediate revelations of facts but cognitive and interactional constructions.

In pragmatics, offline interviews—i.e., interviews not related to a specific immediately preceding activity—have served as initial exploration of a research issue,
to triangulate the researcher’s interpretation of authentic discourse data, as one among several data types in a multimethod approach, and as the main data source. Different types and applications of interviews in research on communicative acts can be illustrated by Knapp, Hopper, as well as Bell’s (1984) and Miles’ (1994) studies on compliments in American English. Knapp et al. (1984) is a large-scale survey interview with partly closed-ended questions and brief responses, focusing primarily on the forms of compliments and compliment responses. Data were content-analyzed and frequencies reported. Miles (1994) is a qualitative study, based on observation of compliment exchanges occurring in authentic discourse for information about compliment forms and their distribution and on interviews for community members’ emic views on the social meanings and functions of complimenting. Questions were open-ended and respondents engaged in extensive narratives and commentary. Data were analyzed interpretively, with particular attention to respondents’ discourse. The report includes extensive quotes from interviewees’ comments. The methodological differences between Knapp et al.’s and Miles’ studies resulted in major discrepancies in substantive outcomes. One such difference is the preferred response pattern identified in the two studies. According to the observational part of Miles’ study, only 7% of the recipients expressed agreement with the compliment. In contrast, in Knapp et al. (1984), 46% of the compliment responses registered as agreements and only 16% were minimized. Further, Knapp et al.’s interview subjects reported without exception that they felt satisfied with the compliment experience, whereas Miles’ interview participants reported feelings of embarrassment and face threat.

What are we to make of these different findings? As far as the response patterns go, the convergent outcomes of observational studies on complimenting in American English suggest that the problem lies with the self-report data. In a reporting context, subjects are more likely to abide by the prescriptive pragmatic norm of agreeing to compliments or accepting them without mitigation. This explanation is supported by the other discrepancy between Knapp et al.’s and Miles’ study, the absence or presence of reports of negative affect associated with complimenting. Once subjects report substantive agreement to a compliment, it would be inconsistent to express negative feelings about the same compliment event, whereas reports of experiencing such an event as emotionally gratifying agree with the reported behavioral practice. Interview subjects thus seem to engage in reconstructive memory activity in order to tell consistent stories. While these stories are intriguing material in their own right, their validity is compromised if the research goal is to establish actual practices of communicative action.

In order to establish communicative practices (as opposed to what members believe
these practices to be), interviews are the wrong choice, because interview subjects' reports are affected by memory constraints and prescriptive orientations. The right choice is observation, as the converging results of many observation-based studies on compliments demonstrate (e.g., Pomerantz, 1978, referred to above). However, interviews are useful and often indispensable when the research goal is to establish the cultural meanings that communicative practices have for community members, because such emic meanings can only be inferred from observation. For the same reasons that make narrative interviews a procedure with unique potential for obtaining in-depth information about cultural meanings from 'the native’s' point of view, qualitative and ethnographic interviews hold substantial promise for investigating the meanings of communicative practices in the perception of the nonnative. But interlanguage pragmaticists have to resist the temptation of interviewing without thorough preparation. In addition to Brigg's (1986) book, two excellent guidelines for the theory and methodology of narrative interviews are Spradley (1979), another classic on ethnographic interviewing, and Kvale (1996) on qualitative interviewing.

As the comparison of Knapp et al's (1984) and Miles' (1994) studies have demonstrated, different genres of research interview yield different information. However, one important feature shared by such diverse interview genres as the interview prestructured by a detailed schedule (questionnaire) and the open-ended, narrative interview is that they tap respondents’ long-term memories of generalized knowledge states, attitudes, or past events. By contrast, another interview genre, often referred to as retrospective interview, informs about participants’ thoughts while they are engaged in a specific activity. This interview genre is usually categorized as a form of thinkaloud protocol and will be discussed in the section on this topic below.

**Diaries**

The least prestructured of all types of self-report is the diary, and it is precisely this property that allows diaries to combine most of the features characteristic of the self-report categories discussed in the preceding sections. They share with scaled response instruments and interviews their focus on past experiences and subjective theories while also permitting retrospective reports on specific attended information in the input or in the diarist’s mind during an activity. They distinguish themselves from any other form of self-report in that they are—in the self-study variety at least—entirely participant-directed, since the diarist decides on the substance, form, and timing of entries without being constrained by a particular task, response format, or social interaction.

In second language research, diary studies date back to the second half of the 1970s. The first review of the early studies from a methodological perspective was offered by
Bailey and Ochsner (1983), more recent updates include Bailey (1990) and Peirce (1994). Diary studies are investigations whose primary data are one or several persons' journal entries about their experiences relating to the topic of the study. Because diaries tell about the diarist's experiences and her interpretation of them from an emic perspective in a holistic, open-ended fashion and are kept during extended and often intensive engagement in the field, diary studies potentially meet some of the standard criteria for qualitative interpretive research (Davis, 1995; Lazaraton, 1995). Because of the in-built emic perspective of personal journals, diary studies in second language research have primarily investigated individual differences, learner strategies, teachers' and students' experiences of second language classroom learning and teaching, and sojourners' and immigrants' perceptions of second language learning and communication in particular social and institutional contexts. Two types of diary study can be distinguished: the self-study, in which the diarist and the researcher are the same person, and the commissioned diary study, in which the researcher requests participants (often language learners or teachers) to keep a journal that is then submitted to and analyzed by the researcher (with or without participant collaboration).

At the time of writing, investigations with a focus on L2 pragmatics and learner diaries as a main data source amounted to one published study (Cohen, 1997), one unpublished conference paper (LoCastro, 1998), and one ongoing study (DuFon, in progress). One remarkable feature on that the three studies converge is their focus on target languages other than English, two examining pragmatic development in L2 Japanese and one in L2 Indonesian. Cohen and LoCastro report on self-studies, whereas DuFon's investigation is based on commissioned diaries as one data source.

Cohen (1997) tells about his experience as a student in an accelerated course for beginners in Japanese as a foreign language. His report informs readers about a wide variety of aspects associated with the learning of L2 pragmatics, such as sociopragmatic and pragmalinguistic transfer from other languages, transfer avoidance, the difficulty of achieving control over routine formulae and selecting expressions with the appropriate formality level, and interactions of cultural and learning style factors. The outcomes of Cohen's learning experience dramatically highlight the discrepancy between successful classroom performance according to the syllabus and low ability to use the target language effectively in communication. LoCastro's (1998) main research question was how her social and cultural identity (as senior American faculty at a prestigious university in Tokyo) related to her development of Japanese pragmatics. Her social position within the sociocultural context of Japan proved to be the single most powerful constraint on her opportunities to acquire pragmatic ability in Japanese, both inside and outside of her work.
context and even including the instruction in Japanese that she took on a regular basis. LoCastro’s study provides further evidence for the need to incorporate a sociocultural perspective in theories of the development of L2 communicative competence. DuFon (in progress) is conducting an ethnographic study with commissioned diaries as one of its components. Adopting a language socialization perspective, DuFon examines how L2 learners acquire politeness in L2 Bahasa Indonesian. The six learners who volunteered to participate in DuFon’s study were students in a four-month program on Indonesian language and culture that included homestay in Indonesian families, instruction at the local university, and structured program activities. Participants were required to keep dialog journals (Peirce, 1994) according to specified guidelines. Once a week, group meetings were held to discuss issues brought up in the diaries. Preliminary analysis of the journal data suggests that participants varied individually in the amount of their writing and themes that preoccupied them.

Second and foreign language learners’ social and personal position towards the target language and culture and their opportunities for interaction and input in different societal domains is increasingly recognized as a significant force in L2 learning. As a prime data source for learners’ own perspective on their language learning experience, diary studies have a particularly rich potential for translating this theoretical orientation into research methodology.

**Think Aloud Protocols**

One feature shared by scaled response instruments, interviews, and diaries is that the self-reported data are elicited in isolation from the contexts in which the reported event occurs. Subjects retrieve pertinent information from long-term memory in order to decide on the value of a context variable presented on a rating scale, answer an interview question, or submit a journal entry, but they are not currently engaged in an activity requiring on-line use of the information. In contrast, think aloud protocols (TAP) are verbalizations of thought processes during engagement in a task.

Anyone wishing to learn about TAP is strongly advised to read the expanded edition of Ericsson’s and Simon’s (1984) book, published in 1993 under the same title, *Protocol Analysis: Verbal Report as Data*. The book not only provides the theoretical framework for predicting under which conditions verbal report should be a valid account of thought processes, but also when and why valid accounts can not be expected.

A minimalist version of the theory goes as follows: Information processed in short-term memory while a subject is carrying out a task is reportable and veridical. Information not processed in short-term memory, such as perceptual processes, motor processes, and all automated processes, are not available for report. Veridical report is also possible
immediately after task completion, when the attended information, or traces of it, is still in short-term memory. Once out of short-term memory, information will be lost or encoded in long-term memory, but storage in and retrieval from long-term memory always entails further processing. Therefore, the best reports are concurrent or immediately consecutive verbalizations. Delayed retrospective protocols may only have a tenuous relationship to the original attended information. In addition to type of information and recency of processing, the instruction to subjects for verbalization is crucial. Prompts should only request subjects to say what they’re thinking. Subjects should not be asked to describe, explain, or hypothesize because such requests will prompt different cognitive processes than those required by the task and will interfere with the task-related processes.

Studies using various types of verbal protocols in second language research have been reviewed in various places, for instance in Cohen’s recent book on learner strategies (1998). Cohen (1996) also reviewed the verbal report studies on interlanguage pragmatics published at the time of writing. I will comment on three published studies because they illustrate different types of verbal protocol and design issues.

M. Robinson (1992) asked six intermediate and six advanced Japanese learners of English to think aloud while completing a Production Questionnaire on refusals. In accordance with Ericsson and Simon’s (1993) prescriptions, instructions requested subjects to verbalize whatever they were thinking while focusing on the task, in the language they were thinking in, and subjects were given a practice session. Immediately after they finished the task, the tape-recorded think-aloud protocol was played back to subjects in a retrospective interview. Coding categories were developed inductively from the protocols and an interrater reliability check was run on the coding of one entire protocol by three coders. Methodologically, one of the interesting outcomes of Robinson’s study is the different information provided in the concurrent and consecutive reports. The concurrent reports were entirely task-focused, evincing what information in the stimulus subjects attended to, their planning decisions, considerations of alternatives, the consulted pragmalinguistic and sociopragmatic knowledge, and the difficulties subjects experienced in deciding on their response. In the consecutive reports, despite the stimulated recall, subjects often had difficulties remembering their task-related thoughts, which was predictable since they completed the entire questionnaire before the retrospective interview. But in some cases, subjects provided more complete reports than in the concurrent verbalization and very informative details about the reasoning underlying their planning decisions and the sources of the L1 and L2 pragmatic knowledge they drew on. In the following example, the concurrent report sheds light on the response alternatives that the learner considered, whereas the retrospective report informs about the
learner’s views of social relationships that guided her decision-making.

An American classmate sometimes sleeps late and misses a class that you share with her. This happened again today and she asks if she can borrow your lecture notes. You have the notes but you don’t want to lend them to her.

Classmate: I missed class again today. Do you think you could lend me your lecture notes? I’m really getting behind in that class and I’d sure appreciate your help!

You: I don’t have it with me now. Well, you may want to ask someone else because my notes are terrible and I would feel embarrassed to show it to you. Please ask someone else.

Concurrent report:
well um - for now I should s tell her that - I don’t have it with me - and uh - um - um and so that she she will ask somebody else - hm - mm [R: what are you thinking?] but um - she may ask again - what should I say - um - I cannot tell her that I - I I was absent from the class - I did attend - um - what I’m gonna say - it’s really hard - um - well - how can I refuse - mm - I can I can just - well I’ll just tell her that I don’t have it with me - and I have to - tell her that - my my notes are - are not good - um - it’s not sufficient for her - oh okay I I should tell her that we- she should ask someone else - = um - well - I don’t I don’t have to ex explain that my notes are good - are not good =

Retrospective report:
R: what was important about this situation?
S: mm - well - I I don’t want to lend a lazy person uh my notes [laughs] [R: hmm] but like I said before it’s also important to make other people happy - as long as I can do things for them - so - it’s it’s easy for me to um to to let other people use my notes - but since I have to refuse um - I just have to say that my notes are not good [R: hmm] yeah - or uh because in as long as uh - notes go um I - have a hard time no uh writing down what my professor says and stuff so I may need some help from someone so - as long as I can help I’d like to um - let the people use my no see my notes so that I can expect someone else to help me some other time. (Robinson, 1992)

Cohen and Olshain (1993) had 15 advanced EFL learners interact in six brief roleplays with a native speaker of English. Each of the roleplays was designed to elicit an apology, a complaint, or a request. After two roleplays involving the same speech act, a retrospective interview was conducted, based on playback of the videotaped roleplays. The retrospective interview was based on a schedule, including questions about the
sources of the chosen linguistic material, utterance planning, and language of thinking. Coding categories reflected these questions.

Widjaja (1997) examined how American and Taiwanese women refused American men's invitations (referred to as date refusal) in three different contexts. Her study shared the main methodological features with Cohen and Olshtain (1993): a combination of videotaped open roleplay with a retrospective interview based on stimulated recall. Both the American and Taiwanese participants' retrospective comments demonstrated that social distance to the interlocutor was the decisive variable in choosing refusal strategies. The appropriacy of particular refusal strategies, on the other hand, was assessed very differently by the Taiwanese and American participants. A strong point in Widjaja's study is the inclusion of native speaker respondents. While presence or absence of native speakers is obviously mandated by the research questions addressed in a study, including native speakers is highly recommendable for verbal protocol research. As long as this line of investigation is as much in its infancy as it is in pragmatics, it is vital for researchers to understand what kinds of information native and nonnative participants attend to, what and how they report, and to what extent they experience concurrent verbalization as intrusive and retrospective reporting as overtaxing their memories. In the nonnative speaker group, close attention must be paid to respondents' L2 proficiency.

In verbal report studies involving nonnative speakers, the question arises in what language the verbal protocol should be delivered. In Cohen and Olshtain's and Widjaja's retrospective interviews, the language of reporting was participants' native language or language of daily communication. Robinson's retrospective reports, on the other hand, were elicited in participants' L2. Even though Robinson's respondents were explicitly instructed to use either Japanese or English during the concurrent think-aloud (p. 81), they reported only in English (p. 65), presumably out of a courtesy to the researcher, who spoke no Japanese. Future studies must ensure that participants actually use whatever language comes to their minds during concurrent verbalization, to minimize the additional processing involved in recoding. For the same reason, immediate retrospection should be initiated in the language used during the think-aloud. Unless participants are advanced enough to think and speak effortlessly in the target language, the experimenter should be bilingual in the L2 and participants' primary language of communication (which can be, but doesn't have to be, their native language). A schematic decision to ask for reports in participants' native language would be psycholinguistically unsound because the native language may not be the language with the lowest activation threshold and thus may not be the language of thought.

The combination of authentic or simulated interaction with retrospective interviews is
a common procedure in interactional sociolinguistics. For studies of miscommunication in interethnic encounters, Gumperz and Cook-Gumperz (1982) recommended commentary elicited through playback of a preceding recorded conversation as a technique for evaluating "how participants reflexively address the social activity that is being constituted by their ongoing talk" (p. 19). In the European Science Foundation Project on Second Language Acquisition by Adult Immigrants, different types of authentic and simulated spoken discourse were supplemented by feedback sessions, which informed about participants' understanding of the recorded interaction, their attitudes, intentions, and experience (Bremer et al., 1992).

In her study of crosscultural gatekeeping interviews, Fiksdal (1990) examined the temporal dimension and uncomfortable moments through microanalysis and focused playback. Participants first watched the videotaped interaction they participated in and provided any commentary they wished to make. In a second viewing, the researcher then stopped the tape and asked the participants for comments "at all moments that seemed uncomfortable because of the topic or because of specific comments of the subjects while viewing it; and (...) at all moments of postural change" (p. 66f). The comments during the playback session provided a crucial source of information about participants' understanding and intent at those particular points in the discourse. In several respects, the use of retrospective interviews in interactional sociolinguistics and ethnographic microanalysis is more akin to analytic induction than to protocol analysis in the information processing approach (cf. Smagorinsky, 1998, and Ericsson & Simon, 1998, for a recent discussion).

**CONCLUSION**

For reasons of exposition, this article has focused on the design features of individual data collection procedures and their applications in pragmatic research. But as I mentioned several times in passing, studies often combine two or more methods. Retrospective interviews will always be recorded in conjunction with data on the participant's completion of the primary task. In ethnographic studies, a multimethod approach is standard, including participant observation, interviews, audio- and videorecordings of interactions, and collection of documents. In fact, researchers in different disciplinary traditions advocate the use of multiple data collection procedures as a means to offset the instrument or observer bias that is necessarily involved in each technique. Material collected by means of complementary techniques and from different sources allows triangulation, which may be necessary or desirable in order to increase the validity/credibility of a study.
All of the data collection methods discussed in this article have usefully illuminated different aspects of pragmatics and will continue to do so. Very likely, we will soon see new techniques, especially those utilizing innovations in computer technology. In a field as complex as crosscultural and intercultural pragmatics, researchers have to borrow from neighbor disciplines as well as design their own methodologies suitable to study different research objects and questions. Given the decisive impact of data collection on substantive findings and theory construction, research into adequate data gathering methodology remains a lasting concern in pragmatics research.
REFERENCES


Gabriele Kasper
Department of ESL
1890 East-West Road
Honolulu, HI 96822

gkasper@hawaii.edu