DISCOURSE COMPLETION TASK: ITS VALIDITY AND RELIABILITY IN RESEARCH PROJECTS ON SPEECH ACTS

Abstract

The article examines the issue of the reliability and validity of the discourse completion task (DCT) in research projects on speech acts. It starts with a review of the main data collection methods in pragmatics. In the paper, the advantages and disadvantages of using a questionnaire for gathering data are discussed. Next, the studies comparing questionnaire data with natural spoken data are presented. The paper argues that relying on one data collection instrument, namely a DCT, does not yield reliable results; therefore the integration of authentic and controlled data is desirable. It is concluded that the study of a particular speech act cannot be carried out without some reliance on naturally occurring data.

Keywords: discourse completion task, questionnaire, speech acts, natural data, pragmatics

1. Introduction

Investigating speech acts requires an analysis of data, which can be obtained by various methods. A serious and long-standing dilemma in pragmatic research concerns the approach to data collection, the validity and reliability of different types of data, and their adequacy to reflect real-life speech acts. The primary focus of this paper is on the discourse completion task, which has been one of the most commonly employed data gathering methods in pragmatics. In this article, we shall discuss how DCTs are designed and administered and examine the strengths and weaknesses of the instrument in question. We shall also try to verify whether DCTs provide reliable results and should be employed for gathering information about speech act production.
2. Data collection methods in pragmatics

Studies conducted in the field of pragmatics have applied a wide array of data collection methods, ranging from data gathered ethnographically to data elicited experimentally. The approaches to studying language can be divided with reference to the locations in which they are used; hence there are armchair, field, and laboratory methods (Clark – Bangerter 2004: 25).

Armchair linguists draw data for their analyses from their intuition and utterances that they can imagine. Within the armchair approach to data collection Jucker (2009: 1615) distinguishes between the philosophical method and the interview method. The former is based on the researcher intuition, whereas the latter makes use of interviews to elicit opinions from the users of a language.

The field method involves observations of naturally occurring conversations which are produced irrespective of the research project. As Jucker (2009: 1616) writes, this particular method comprises one of the following approaches: taking notes of the speech acts encountered in daily life; picking out instances of a speech act under investigation from novels or other fictional material; transcribing actual conversations, or using electronic corpora.

By contrast, laboratory linguists investigate language use by means of experiments carried out in a laboratory. The laboratory method employs the elicitation technique to yield utterances that contain a speech act under investigation. In this procedure, the informants have to imagine communicative situations and say or write how they or other people would respond to these situations. The laboratory method is represented by the technique of discourse completion task (which falls under the type of questionnaires).

Ideally, the data for the studies of speech acts should include a large number of carefully recorded observations of a speech act under investigation by representative participants unaware of the observations, and taking part in authentic interactions (Hinkel 1997: 2). Collecting authentic data, however, has many drawbacks. First, people are reluctant to being recorded and observed, therefore the researchers might have difficulties in gaining access to the research site. Second, the process of obtaining sufficient quantities of a speech act under study can be quite lengthy. Finally, the researcher’s presence might affect the course of the interaction (the observer’s paradox; Labov 1972). Due to constraints pertaining to legal and ethical issues, as well as logistic difficulties in accumulating this kind of spoken data, researchers in pragmatics have had to adopt alternative methods for eliciting speech act data, the most popular of which is the discourse completion task.
3. Discourse completion task

Discourse completion tasks have been widely employed in research projects on speech act production. Brown (2001: 301) offers the following definition of a written DCT: “any pragmatics instrument that requires the students to read a written description of a situation (including such factors as setting, participant roles, and degree of imposition) and asks them to write what they would say in that situation”. The situations in the questionnaire are designed in such a way that a specific communicative act (compliment, apology, invitation, thanking, request, refusal, etc.) is elicited.

There are two types of the questionnaire: an open questionnaire in which a scenario alone is provided, and a dialogue completion task in which a conversational turn is provided. Depending on the research aims, participants have to provide one or two responses to the situation in the case of open-ended questionnaires. Some questionnaires specify that a subject may give a verbal response, a non-verbal response, or no response at all. The DCT can be administered both orally and in writing. The procedure of the oral DCT is as follows: the instructions are either recorded or read out loud by a native speaker and a subject is asked to listen to the scenarios one by one and respond to each orally (Yuan 2001: 274). In the following examples of test items, (1) is constructed to elicit thanks, and (2) to elicit an apology.

(1) It’s your birthday, 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: 75)

(2) At the professor’s office
   A student has borrowed a book from her teacher, which she promised to return today.
   When meeting her teacher, however, she realizes that she forgot to bring it along.
   Teacher: Miriam, I hope you brought the book I lent you.
   Miriam: ____________________________.
   Teacher: OK, but please remember it next week. (Blum-Kulka – Olshtain 1984: 198)

Studies vary to a great extent with regard to the number of subjects and the number of questionnaire items included in the questionnaire. Kasper – Dahl (1991: 230) notice that there are no hard-and-fast rules about how many items should make up a questionnaire and how many subjects should complete the form. The most ambitious application of a DCT, the Cross Cultural Speech Act Realization Project (CCSARP) investigated requests
and apologies by means of a DCT, consisting of scripted dialogues of sixteen situations, eight items for requests and apologies, and was completed by hundreds of native and non-native speakers of thirteen languages and language varieties. Kasper – Dahl (1991: 230), however, suggest that for most purposes in research on speech act realization, questionnaires with 20 items and 30 subjects should “serve as a rough guide for decisions on these issues.” One might imagine that the number of items affect to some extent the quality and completeness of the responses. Subjects tend to be reluctant to fill in a lengthy questionnaire as they consider it an imposition on their time. Therefore, researchers should take into account how much time it takes to respond to the items in the questionnaire before administering it to the subjects. Wolf (1988: 481), for instance, claims that “a full questionnaire should require certainly less than 30 minutes to complete; and preferably less than 15 or 20.” At the same time, questionnaire designers should bear in mind that lengthy and detailed descriptions of the scenarios increase reading time and may pose a problem for less competent readers (Kasper 2008: 292).

4. Advantages and disadvantages of discourse completion task

The reliability and validity of discourse completion tasks have been discussed at great length and this debate continues. It is undeniable that this method brings with it advantages and disadvantages with regard to data collection and analysis of these data.

The DCT appears to have many administrative advantages. One of the merits attributed to this instrument is that it enables a researcher to gather a large corpus of data on difficult to observe speech events in a variety of situations within a relatively short period of time. As a result, DCT serves to obtain information about what semantic formulas are used to implement a particular speech act (Beebe – Cummings 1996: 80). Furthermore, DCTs provide demographic information about the participants of the experiment. As Houck – Gass (1996: 46) observe, consistency of the context allows for a comparison of responses along a number of variables, such as age, gender or ethnicity. What is more, this data collection method allows the researcher to control the contextual variables that are specified in the scenarios and may influence the subjects’ responses. Nelson et al. (2002: 167) note that DCT can be translated into any language, which enables the researcher to use the same situations for two or even more speech communities, and compare the choice and formulation of speech act strategies by the respondents. Kasper (2000: 329) summarizes that a carefully devised questionnaire constitutes a useful research tool to gain insight into “speakers’ pragmalinguistic knowledge of
the strategies and linguistic forms” which serve to realize communicative acts, and to examine contextual factors “under which particular strategic and linguistic choices are appropriate.” Hence, DCT seems to be an appropriate data collection method when the purpose of the study is data production.

Notwithstanding their appeal, DCTs have been criticized for disturbing discrepancies between data elicited via this method and natural data. In spite of the fact that DCTs offer numerous advantages for the cross-cultural study of speech acts, they yield written data, which cannot be expected to precisely represent naturally occurring talk (Ogiermann 2009: 68). In addition, there have been numerous concerns about the design of the questionnaire. Many researchers question the validity of DCT data on the grounds that the situations described in the questionnaire do not fully reflect the socio-linguistic constraints that operate in actual interactions. Critics argue that the items in the questionnaire are described too briefly, and therefore they fail to give sufficient social and situational information, such as the relationship between the subject and the interlocutor, frequency of interaction between them, context and setting of the speech event. As Jucker (2009: 1618) observes, “some dialogues put the informants into roles with which they are unfamiliar. This may create unnatural utterances.” Undoubtedly, interpersonal and contextual details have an impact on the speakers’ utterances, and DCT is lacking in them. What is more, it has been noted that questionnaires neglect non-verbal features (e.g. gestures, posture, facial expressions) and paralinguistic elements (e.g. pitch, intonation), along with most pragmatic strategies which are inherent in oral interactive discourse, to name but a few: turn-taking, speaker-listener coordination, sequencing of action, or hesitations (Kasper 2000: 326). Bardovi-Harling (1999: 238) argues, though, that DCTs are not designed to examine these areas in cross-cultural pragmatic research.

Finally, the opponents of DCTs point to the test-like format of the questionnaire and the fact that spoken performance is elicited through the written mode. This, in turn, gives the informants time to think about and plan their answers. Besides, “the space provided on the sheet constrains the length of the utterances, and the follow-up turn which is provided is also unnatural because the informant knows ahead of time how the imaginary interlocutor will react to his or her utterance” (Jucker 2009: 1618). Furthermore, Barron (2003: 85) claims that this data collection method makes the subjects “recall pragmatic information from memory and report rather than use it.” Despite the weaknesses, different types of the questionnaire have been used extensively to elicit speech act data across languages.
5. DCT data and natural data

Having outlined the advantages and limitations of the questionnaire, we will move on to discuss the studies examining how the use of the instrument in question affects the results of the research. The main focus of this section will be on the review of the studies comparing questionnaire data with naturally occurring data.

One of the first studies of this kind was carried out by Beebe – Cummings (1996). In order to examine whether questionnaire data reflect real-life responses, they compared the speech act of refusals elicited via a written discourse completion task with refusals performed in telephone conversations, in English, in response to the same requests. They noticed some similarities and differences between these two types of data. The findings led the researchers to support, with certain caveats, the use of questionnaires to gather data on speech act performance.

Beebe – Cummings (1996: 80) claim that the questionnaire can be a useful research tool as a means of collecting large amounts of data quickly and easily. They observed that the contents of semantic formulas elicited by means of the questionnaire and telephone conversations were similar. Hence, it is argued that a DCT could be a helpful instrument with regard to the initial categorization of semantic formulas and strategies that will be found in natural speech; examining the conditions which regulate how speech acts have to be carried out in a socially adequate way; and investigating social and psychological factors that are likely to influence the performance of speech acts.

However, the findings of the study led Beebe – Cummings (1996: 80) to acknowledge several disadvantages of DCTs. The researchers found that the amount of data received from the oral responses was much larger than in the case of the written data, which indicates that the responses elicited by means of the DCT were shorter than in the case of the natural data. The study revealed that questionnaires “bias the response toward less negotiation, less hedging, less repetition, less elaboration, less variety and ultimately less talk” (Beebe – Cummings 1996: 71). The telephone conversations, contrary to the written questionnaire, allowed the subjects to cooperate and negotiate their responses. Beebe – Cummings observed that discourse completion tasks fail to elicit “the long negotiated sequences” which are found in natural conversations. The researchers (1996: 80) admit that questionnaire data are not an accurate reflection of spoken data with respect to the actual words produced in real-life responses, the variety of formulas and strategies employed by the speakers, the length of the real-life responses, the number of turns necessary for a successful communicative act, as well as the real rate of occurrence of a speech act.
Although the recorded data are more representative of authentic interactions, the questionnaire, according to the investigators, could be a valid data-gathering method. Beebe – Cummings (1996: 81) believe that “native speaker perception of what constitutes an appropriate refusal, apology, or request is valuable information for teachers and researchers in cross-cultural communication”. Nevertheless, they advocate the collection of data through multiple approaches since each data gathering procedure has its own strengths and weaknesses. However, we need to stress that, while the study by Beebe – Cummings (1996) provides some evidence for the usefulness of DCTs, their findings have to be treated with some reservation as the natural data used in their research are context-sensitive. Telephone conversations differ from natural conversations in many ways, for example, they lack face-to-face interactions.

Similar findings were reported by Hartford – Bardovi-Harlig (1992), who also dealt with authentic data and data elicited through the questionnaire. The researchers examined the language of advice rejections used by the native and non-native speakers of English during advising sessions between students and their advisors. The data for the analysis came from thirty-nine audiotaped natural conversations, and twenty-four DCTs consisting of eight scenarios. The items in the DCT were designed to test the hypotheses based on the observations of natural conversations. The comparison of the data gathered through the two instruments proved that, although a DCT has some merits with regard to “availability of large samples and experimental controls” (Hartford – Bardovi-Harlig 1992: 33), the data collected through this method might be biased in certain ways.

As Hartford – Bardovi-Harlig point out, the DCT proved to be a useful instrument to test hypotheses derived from the natural data, and to interpret the findings from this type of data. The researchers, however, observed that the DCT generated a narrower range of semantic formulas than in the case of the spoken data. Moreover, some linguistic forms which occurred in one set of data did not appear in the other set. Hartford – Bardovi-Harling (1992: 47) explain that the reason why certain elements were not produced in the questionnaire is that “DCT does not promote the turn taking and negotiation strategies found in natural conversations.” Furthermore, the analysis of data showed that the instrument affects the degree of politeness employed by the respondents, in that fewer status preserving strategies and more bald-on-record responses were observed in the DCT data than in the audiotaped conversations. The researchers assume that the anonymous character of the questionnaire enables the subjects to provide the responses which would normally cause a loss of face in natural interactions.

The findings of the experiments led Hartford – Bardovi-Harlig to the conclusion that the DCT may be used as an important tool to complement
observational data. It has to be stressed, however, that the data collected by them differ from those obtained from everyday conversations. The researchers used the recordings of academic advising sessions between professors and their students; hence the interactions were formal: there was a large power difference between the interlocutors, and the structure of the conversations was highly focused.

The DCT was also contrasted with naturally occurring talk in the study carried out by Golato (2003). More specifically, Golato’s empirical study dealt with compliment responses in German. The data for the analysis came from spontaneous interactions and a questionnaire designed to elicit the same discourse context which was observed in the naturalistic data.

The comparison of the DCT and conversational data demonstrated that the DCT data can be very different from naturalistically collected data. Golato (2003) found that while some responses were identical in both data gathering procedures, the noticeable differences confirmed that questionnaires are not a reliable reflection of actual language use. First, the study revealed that certain strategies and speech patterns elicited in the DCT did not occur in real life conversations at all. Second, the responses in the DCT were longer and they contained more turn-construction units than in the case of naturalistic data. Golato’s findings are at variance with Beebe – Cummings’ observations (1992) about the length of questionnaire-elicited answers. Golato explains that the lengthy responses are connected with the fact that in a written task respondents do not have an interlocutor who might come in and take a turn, therefore they tend to extend their own turns. Golato (2003: 110) observed that subjects spent some time thinking about the situation at hand, therefore she came to the conclusion that “a DCT is not an on-line task in which a person uses language spontaneously and without consciously focusing on linguistic output, but is instead an off-line task in which a person has time for introspection.” The researcher argues that the hypothetical nature of DCTs yields hypothetical language. The respondents are asked not to conversationally interact, but to write what they believe would be an appropriate thing to say in a given situation.

It is believed that one’s experiences in a given context would affect one’s responses. As a result, naturally occurring data are much better suited for finding out what speakers are actually doing when interacting with co-participants, how language is organized, or what turn-taking mechanism are employed. “DCTs do not clearly and reliably inform us as to how talk-in-interaction is organized and realized in natural settings”. They are, however, helpful tools for learning more about speakers’ perception of their language behavior, so they are suited to the study of “what people think they would say” (Golato 2003: 111).
Data collected from a DCT and a corpus were examined in the light of their potential application in a pedagogic context by Schauer – Adolphs (2006). The authors conducted a study with the aim of verifying whether the discourse completion task and corpus data can provide language teachers with expressions of gratitude used by the native speakers of British English. The DCT of eight scenarios yielded a wide range of interactional formulaic sequences conveying gratitude. By contrast, the data derived from the Cambridge and Nottingham corpus of discourse in English (CANCODE) offered a broader picture of how thanking is realized in terms of “repeated patterns of collaborative negotiation of the expression of gratitude”, in the cases when one speaker repeats the utterance of his or her interlocutor, for example during service encounters. What is more, the corpus, contrary to DCTs, captured elements such as hesitations or false starts. Schauer – Adolphs infer from their study that the corpus as a source of data differs considerably from the DCT with regard to language use, which, contrary to a DCT, appears in an uncontrolled context. The DCT, on the other hand, offers insight into the different ways in which a certain speech act is realized in a closely defined context, and controlled for social relations of the interlocutors, their language proficiency, or nationality. It cannot, however, supply such a rich variety of discourse contexts as the natural data. Besides, the corpus captures the language of a wide range of speakers in various contexts and different social relationships. Last but not least, the corpus stands in contrast to the DCT in terms of the way in which data are categorized. “In the case of the DCT the categories inform the instrument while in the case of the corpus the data often tends to inform the categories” (Schauer – Adolphs 2006: 131).

Taking everything into consideration, the authors conclude that “the aim of corpus research is to provide evidence of patterns in the language use of a population, the aim of discourse completion task (DCT) research is to investigate a linguistic act within highly predefined parameters” (Schauer – Adolphs 2006: 120). They argue that a combination of both data sets might be beneficial in a pedagogic context and facilitate language teaching and learning by providing the learners with patterns of language use.

6. Conclusions

In conclusion, it needs to be emphasized that the questionnaire has both strengths and weaknesses with regard to gathering data. Although opinions diverge greatly on the issue of the validity and reliability of this instrument, researchers agree that the usefulness of this data collection method should be evaluated in terms of the research questions under study. As Jucker (2009:
rightly writes, all methods “can be used to increase our knowledge of language and language use as long as they are used judiciously with a clear understanding of their respective strengths and limitations and without any undue expectations that any of them might be able to solve all the problems in linguistics.”

It has been demonstrated in the article that DCTs fail to show the interactional aspects of a speech event, and that they elicit hypothetical language – that is, what people think they would say in a particular situation. However, a DCT appears to be a suitable data collection instrument for finding what semantic formulas are used to realize a particular speech act and what linguistic behaviour is considered appropriate in a given language.

The studies discussed in the article have reported on the weaknesses of relying on one single data collection method and demonstrated that different data gathering procedures may generate different results. Data gathered by means of controlled elicitation procedures do not allow the researcher to find out how a particular speech act is employed by members of a given speech community in spontaneous interactions. It would appear then that in order to achieve greater reliability of the study, more than one data-gathering method should be used, and one of the data sets should consist of natural speech. “It is only through comparing the artificial data with authentic material that the researcher can assess whether any regularities found in the former are significantly different from those encountered in the latter” (Geluykens – Kraft 2008: 94). One fact seems to be undeniable: the analysis of authentic talk in interactional settings is a prerequisite for any statements about the pragmatic choices speakers make in natural interactions.

REFERENCES

Bardovi-Harlig, Kathleen

Barron, Anne

Beebe, Leslie M. – Martha C. Cummings

Blum-Kulka, Shoshana – Elite Olshtain
Bouton, Lawrence F. (ed.)
1999 Pragmatics and language learning 9. Urbana, IL: University of Illinois at Urbana-Champaign.

Brown, James D.
2001 “Pragmatics tests: different purposes, different tests”. In: Kenneth R. Rose – Gabriele Kasper (eds.), 301–325.

Clark, Herbert H. – Adrian Bangerter

Eisenstein, Miriam – Jean Bodman

Gass, Susan M. – Joyce Neu (eds.)

Geluykens, Ronald – Bettina Kraft
2008 “The use(fullness) of corpus research in cross-cultural pragmatics”. In: Jesus Romero-Trillo (ed.), 93–117.

Golato, Andrea

Hartford, Beverly S. – Kathleen Bardovi-Harlig

Hinkel, Eli

Houck, Noel – Susan M. Gass

Jucker, Andreas. H.

Kasper, Gabriele


Kasper, Gabriele – Merete Dahl

Kasper, Gabriele–Shoshana Blum-Kulka (eds.)
Keeves, John P. (ed.)
1988  
_Educational research, methodology, and measurement._ Oxford: Pergamon.

Labov, William
1972  
_Sociolinguistic patterns._ Oxford: Blackwell.

2002  

Noveck, Ira A. – Dan Sperber (eds.)
2004  
_Experimental pragmatics._ Houndmills: Palgrave Macmillan.

Nurani, Lusia M.
2009  
“Methodological issue in pragmatic research: is Discourse Completion Test a reliable data collection instrument?”. _Jurnal Sosioktologi Edisi 17 Tahun_ 8: 667–678.

Ogiermann, Eva
2009  
_On apologising in negative and positive politeness cultures._ Amsterdam, Philadelphia: John Benjamins Publishing Company.

Romero-Trillo, Jesus (ed.)
2008  

Rose, Kenneth R. – Gabriele Kasper (eds.)
2001  
_Pragnatics in language teaching._ Cambridge: Cambridge University Press.

Schauer, Gila A. – Svenja Adolphs
2006  

Spencer-Oatey, Helen (ed.)
2000  

Spencer-Oatey, Helen (ed.)
2008  

Wolf, R. M.
1988  
“Questionnaires”. In: John P. Keeves (ed.), 478–482.

Yuan, Yi
2001  