

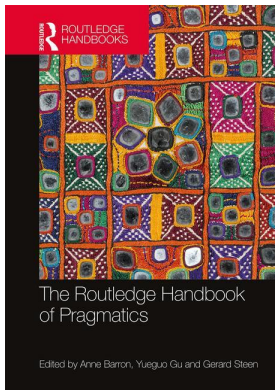
This article was downloaded by: 10.3.98.93

On: 22 Sep 2019

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## **The Routledge Handbook of Pragmatics**

Anne Barron, Yueguo Gu, Gerard Steen

### **Naturally Occurring Data**

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315668925-3>

Andrea Golato

**Published online on: 27 Jan 2017**

**How to cite :-** Andrea Golato. 27 Jan 2017, *Naturally Occurring Data from: The Routledge Handbook of Pragmatics* Routledge

Accessed on: 22 Sep 2019

<https://www.routledgehandbooks.com/doi/10.4324/9781315668925-3>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# 3

## NATURALLY OCCURRING DATA

*Andrea Golato*

### Introduction

All pragmatic research is based on some form of linguistic data, be it invented, elicited, observed or recorded (for a similar argument, see Jucker 2009). Of these, no one method of data collection and analysis is intrinsically better than any other. Rather, the data type, method of data collection and method of analysis adopted in a particular study should depend solely on the research question at hand (for similar arguments, see also Grotjahn and Kasper 1991; Bardovi-Harlig 1999; Turnbull 2001; Barron 2003; Félix-Brasdefer 2007: 159; Golato and Golato 2013). Thus, a researcher interested in language intuition, for instance, goes about data collection and analysis differently than a researcher interested in actual language use. This chapter discusses naturally occurring data and their use in pragmatics research that focuses on spoken language. It starts out by providing a definition of the term and a rationale for the use of such data. It then discusses the two ways in which naturally occurring data are typically collected, namely using field notes and recordings.

### **Naturally occurring data – what are they and why should they be used?**

A useful definition of what counts as naturally occurring data is provided by Potter, who suggests using the '(conceptual) dead social scientist's test'. Potter asks:

Would the data be the same, or be there at all, if the researcher got run over on the way to work? An interview would not take place without the researcher there to ask the questions; a counselling session would take place whether the researcher turns up to collect the recording or not.

(Potter 2002: 541)

As Potter clearly highlights, naturally occurring data are data that are not directly elicited by the researcher; instead, they are data that are observed. Each and every feature of the interaction (e.g., the context, underlying assumptions, linguistic realisations of a given speech event) is talked into being (Heritage 1984) by the participants of the interaction, whereas in elicited data

these features are predetermined by the elicitation instrument designed by the researcher (Félix-Brasdefer 2007; Félix-Brasdefer and Hasler-Barker, Chapter 4, this volume). One important argument for using naturally occurring data is that experimenters may not predict all (or even the most common) situations in which a given speech event (e.g., a compliment or an apology) may be produced (Kasper 2000). Also, different situations may lead to different realisations of a speech event: for instance, a compliment used at the beginning of a meal leads to a different response than a compliment produced at the end of a meal (Golato 2005). Moreover, as Atkinson and Heritage (1984: 3) note: ‘The experimenter is unlikely to anticipate the range, scope, and variety of behavioural variation that might be responsive to experimental manipulation’.

Several studies have compared naturally occurring data with data obtained through other elicitation methods, such as role plays, written or oral discourse completion tasks (DCTs) or questionnaires. Comparisons of naturally occurring data and DCTs have shown that the latter fail to capture how sequences are initiated and then unfold in the interaction (Beebe and Cummings 1985; Wolfson 1989; Beebe and Cummings 1996). Moreover, subjects produce utterances on DCTs that are different in terms of length, complexity, and linguistic realisation than those in naturally occurring data (Bodman and Eisenstein 1988; Hartford and Bardovi-Harlig 1992; Hassall 1997; Turnbull 2001; Golato 2003; Bou-Franch and Lorenzo-Dus 2008; Félix-Brasdefer 2010; Economidou-Kogetsidis 2013). This was also found when DCT data were compared with large-scale corpus data, specifically, the 5 million-word *Cambridge and Nottingham Corpus of Discourse in English* (CANCODE) collected between 1991 and 2001 (Schauer and Adolphs 2006).

However, in a study of requests, Economidou-Kogetsidis (2013) has shown that data from written DCTs and naturally occurring data display similar trends in terms of directness and lexical/phrasal modification, thus indicating that ‘to a certain extent, the WDCT requests represent an approximation to the naturally occurring requests’ (Economidou-Kogetsidis 2013: 33, emphasis in the original).

Comparisons between role plays and naturally occurring data have shown that role plays capture interactional details (e.g., turn-taking features, on-line planning, hesitations, hedges) and thus yield more realistic data than other data elicitation methods (e.g., Bodman and Eisenstein 1988; Kasper and Dahl 1991; Turnbull 2001; Félix-Brasdefer 2007; Huth 2010; Bataller 2013). However, role-play data are still only an approximation of actual conversation as they differ in significant ways from naturally occurring data in terms of length, complexity, and prosodic and other linguistic features (Hassall 1997; Félix-Brasdefer 2007; Bataller and Shively 2011; Baller 2013). For these reasons, researchers generally recommend naturally occurring data for analyses of actual language use (Turnbull 2001; Golato 2003; Félix-Brasdefer 2007; Bou-Franch and Lorenzo-Dus 2008; Bataller and Shively 2011). We turn now to the two methods of collecting naturally occurring data, namely to field notes and audio and/or video recordings.

### Field notes

Researchers take field notes on site either as observers or as participant observers. When they hear or observe a linguistic exchange of interest to their study, they note it down as soon as possible after it has occurred. Typically, information is also recorded about the context of the interaction (e.g., the participants, the location, intonation, any other noteworthy elements). Such contextual information may be indispensable in data analysis, depending on the research question of the investigator (Kasper 2000). A researcher may conduct the field observations on his or her own, but might also send a number of research assistants into the field who are trained with regard to both the phenomenon of interest and data collection methods. Overall, this method allows for collecting naturalistic data quickly and in a wide range of conversational contexts, and

from a variety of sources (Kasper 2000). However, this data collection method is not without limitations, the most obvious one being that researchers have to rely on their memory when noting down the observed exchanges. Research (Labov 1984; Lehrer 1989) has shown that a person's recall of prior utterances is limited in terms of both quantity and quality, and frequently does not include elements which were not tightly integrated into the original sentence structure (e.g., modifiers, intensifiers, hedges). In addition, the lack of recordings means that there is no opportunity for replay and consequently no possibility for other researchers to independently verify the occurrence of the utterance. A second problem associated with this method of data collection is that the researcher may note down only the most common or expected linguistic forms of a given speech act, and may fail to recognise (or recognise only later in the interaction) that a speech event of interest has taken place. As Kasper (2000: 319) states: '[T]here is thus a real danger that memorisation and taking field notes will result in recording salient and expected (or particularly unexpected) facets of the interaction, at the expense of less salient but perhaps decisive (often indexical) material'. Given these limitations, it seems that field notes might be appropriate to use when researchers are interested in the semantic content or situational context of a given speech event, but less appropriate when actual language use is the object of investigation.

### Recordings

Recorded naturally occurring data allows for exact and repeated analysis of the linguistic material, and for data verification by other researchers. Whenever interactants have visual access to one another, as in most face-to-face interactions and through web conferencing services, conversations are typically video recorded so that the researcher has the same access to utterances, gestures, gaze, and other embodiments as the conversational partners themselves. In rare situations, video recordings are not feasible, for instance in certain institutional settings where subjects fear that video recordings may reveal industry secrets (Golato 2003). In these settings, audio recordings are made. Otherwise, audio recordings are used for capturing telephone interactions or non-video web conferencing interactions. Typically, the researcher sets up the recording equipment (at times multiple recording devices are necessary to capture the entire setting) and then leaves the room. Note that in many countries, ethical practices suggest and/or law requires that the researcher seek written consent from subjects prior to any recording.

Working with recordings of naturally occurring data is not without its challenges. As Labov (1972: 209) stated: 'The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation'. With respect to the present discussion, this means that the very presence of recording equipment may alter subjects' speech production. However, if given even a short time to adjust, subjects tend to forget that they are being recorded (Kasper 2000). Depending on the speech event, it can sometimes take longer to amass a corpus of an appropriate size (Kasper 2000), and the time it takes to transcribe the recordings can be considerable (as can be the costs if one pays for transcription services). However, the increasing availability of public corpora for a variety of languages (for English corpora, see Lee 2010) allows for examining data from a diverse set of speakers hailing from a variety of backgrounds (cf. Weisser, Chapter 5, this volume). A further challenge is that certain speech acts may not occur frequently in interaction or may be difficult to obtain (Kasper and Dahl 1991; Billmyer and Varghese 2000). In the latter situation, it may take some extra effort to secure permission to audio-video record in certain settings. However, recordings of interactions in cars at car dealerships in France (Mondada 2009), in families dealing with terminally ill patients (Beach 2009) and of arguments between individuals (Jackson and Jacobs 1980) are just a few of the studies that have investigated data that might at

first glance be considered difficult to obtain. Beebe and Cummings (1996) have stated that it is difficult to systematically compare speech samples from different groups and different points in time. A number of studies since, however, have successfully compared speech events across cultures and groups, even at different points in time (Placencia 1997; Pavlidou 2000; Golato 2002; Taleghani-Nikazm 2002). Similarly, Shively (2011) has demonstrated that naturally occurring data can successfully be used to demonstrate language development over time in study abroad contexts.

## Conclusion

A major challenge facing researchers who work with naturally occurring data is a lack of control of speaker and context variables, such as age and/or social backgrounds of the speakers (Yuan 2001). Researchers for whom such controls are important may follow Turnbull's (2001) approach to data collection: here, subjects are selected according to speaker variables and receive certain tasks to perform that likely involve the speech act of interest. Crucially, however, subjects are not made aware that their talk is the object of the investigation. For instance, when studying negotiation strategies, teams of subjects engaged in a computer game in competition with other teams, subjects were under the impression that their game moves were the focus of the experiment, whereas in reality the researcher was interested in their language use.

While it may seem that obtaining data in the form of field notes and recordings is more involved than collecting data via elicitation techniques, these two methods of data collection yield the most naturalistic forms of data as defined earlier. Such data are vital when researchers are interested in actual language use, whereas other methods of data collection may be useful for different research purposes (cf. Golato 2003; Jucker 2009).

## Suggestions for further reading

- Sidnell, J. and Stivers, T. (eds.) (2014) *The Handbook of Conversation Analysis*. Malden, MA: Blackwell. Chapter 3 presents factors to consider during the actual recording of naturally occurring data, while chapter 4 discusses transcription techniques.
- Tracy, S. J. (2013) *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact*. Malden, MA: Blackwell.
- This book provides a very detailed overview of the factors to consider when conducting field work.

The following books provide overviews of various research methods in (applied) linguistics:

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Litosseliti, L. (2010) *Research Methods in Linguistics*. London/New York: Continuum International.

## References

- Atkinson, J. M. and Heritage, J. (eds.) (1984) *Structures of Social Action. Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Bardovi-Harlig, K. (1999) 'Researching method', in L. Bouton (ed.) *Pragmatics and Language Learning* (Monograph Series, vol. 9). Urbana: University of Illinois at Urbana-Champaign. 237–264.
- Barron, A. (2003) *Acquisition in Interlanguage Pragmatics: Learning How to Do Things with Words in a Study Abroad Context*. Amsterdam/Philadelphia: John Benjamins.
- Bataller, R. (2013) 'Role-plays vs. natural data: Asking for a drink at a cafeteria in peninsular Spanish', *Íkala, Revista de Lenguaje y Cultura*, 18(2): 111–126.
- Bataller, R. and Shively, R. L. (2011) 'Role plays and naturalistic data in pragmatics research: Service encounters during study abroad', *Journal of Linguistics and Language Teaching*, 2: 15–50.
- Beach, W. (2009) *A Natural History of Family Cancer: Interactional Resources for Managing Illness*. New York: Hampton Press.

- Beebe, L. M. and Cummings, M. C. (1985) 'Speech act performance: A function of the data collection procedure?' Paper presented at the TESOL Convention, New York.
- Beebe, L. M. and Cummings, M. C. (1996) 'Natural speech act data versus written questionnaire data: How data collection method affects speech act performance', in S. M. Gass and J. Neu (eds.) *Speech Act Across Cultures: Challenges to Communication in a Second Language*. Berlin: Mouton de Gruyter. 65–88.
- Billmyer, K. and Varghese, M. (2000) 'Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tasks', *Applied Linguistics*, 21: 517–552.
- Bodman, J. and Eisenstein, M. (1988) 'May God increase your bounty: The expression of gratitude in English by native and non-native speakers', *Cross Currents*, 15: 1–21.
- Bou-Franch, P. and Lorenzo-Dus, N. (2008) 'Natural versus elicited data in cross-cultural speech act realization: The case of requests in Peninsular Spanish and British English', *Spanish in Context*, 5: 246–277.
- Economidou-Kogetsidis, M. (2013) 'Strategies, modification and perspective in native speakers' requests: A comparison of WDCT and naturally occurring requests', *Journal of Pragmatics*, 53: 21–38.
- Félix-Brasdefer, J. C. (2007) 'Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish', *Spanish in Context*, 4: 159–185.
- Félix-Brasdefer, J. C. (2010) 'Data collection methods in speech act performance. DCTs, role plays, and verbal reports', in A. Martínez-Flor and E. Usó-Juan (eds.) *Speech Act Performance: Theoretical, Empirical and Methodological Issues*. Amsterdam/Philadelphia: John Benjamins. 41–56.
- Golato, A. (2002) 'German compliment responses', *Journal of Pragmatics*, 34: 547–571.
- Golato, A. (2003) 'Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk', *Applied Linguistics*, 24: 90–121.
- Golato, A. (2005) *Compliments and Compliment Responses: Grammatical Structure and Sequential Organization*. Amsterdam/Philadelphia: John Benjamins.
- Golato, A. and Golato, P. (2013) 'Pragmatics research methods', in C. A. Chapelle (ed.) *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell. Published Online: 5 November 2012.
- Grotjahn, R. and Kasper, G. (1991) 'Methods in second language research: Introduction', *Studies in Second Language Acquisition*, 12: 109–112.
- Hartford, B. S. and Bardovi-Harlig, K. (1992) 'Experimental and observational data in the study of interlanguage pragmatics', in L. Bouton and Y. Kachru (eds.) *Pragmatics and Language Learning* (Monograph Series, vol. 3). Urbana: University of Illinois at Urbana-Champaign. 33–52.
- Hassall, T. (1997) 'Requests by Australian learners of Indonesian', unpublished doctoral dissertation, Australia National University, Canberra.
- Heritage, J. (1984) *Garfinkel and Ethnomethodology*. Cambridge: Polity Press in association with Blackwell Publishers.
- Huth, T. (2010) 'Can talk be inconsequential? Social and interactional aspects of elicited second language interaction', *Modern Language Journal*, 94: 537–553.
- Jackson, S. and Jacobs, S. (1980) 'Structure of conversational argument: Pragmatic bases for the enthymeme', *The Quarterly Journal of Speech*, 66: 251–265.
- Jucker, A. H. (2009) 'Speech act research between armchair, field and laboratory. The case of compliments', *Journal of Pragmatics*, 41: 1611–1635.
- Kasper, G. (2000) 'Data collection in pragmatics research', in H. Spencer-Oatey (ed.) *Culturally Speaking: Managing Rapport through Talk across Cultures*. London/New York: Continuum. 316–341.
- Kasper, G. and Dahl, M. (1991) 'Research methods in interlanguage pragmatics', *Studies in Second Language Acquisition*, 13: 215–247.
- Labov, W. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1984) 'Field methods of the project on linguistic change and variation', in J. Baugh and J. Sherzer (eds.) *Language in Use: Readings in Sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall. 28–66.
- Lee, D. Y. W. (2010) 'What corpora are available?', in M. McCarthy and A. O'Keeffe (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge. 107–121.
- Lehrer, A. (1989) 'Remembering and representing prose: Quoted speech as a data source', *Discourse Processes*, 12: 105–125.
- Mondada, L. (2009) 'The embodied and negotiated production of assessments in instructed actions', *Research on Language and Social Interaction*, 42: 329–361.
- Pavlidou, T.-S. (2000) 'Telephone conversations in Greek and German: Attending to the relationship aspect of communication', in H. Spencer-Oatey (ed.) *Culturally Speaking: Managing Rapport through Talk across Cultures*. London/New York: Continuum. 121–142.
- Placencia, M. E. (1997) 'Opening up closings: The Ecuadorian way', *Text*, 17: 53–81.

- Potter, J. (2002) 'Two kinds of natural', *Discourse Studies*, 4: 543–548.
- Schauer, G. A. and Adolphs, S. (2006) 'Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences, and pedagogy', *System*, 34: 119–134.
- Shively, R. L. (2011) 'L2 pragmatic development in study abroad: A longitudinal study of Spanish service encounters', *Journal of Pragmatics*, 43: 1818–1835.
- Taleghani-Nikazm, C. (2002) 'A conversation analytical study of telephone conversation openings between native and nonnative speakers', *Journal of Pragmatics*, 34: 1807–1832.
- Turnbull, W. (2001) 'An appraisal of pragmatic elicitation techniques for the social psychological study of talk: The case of request refusals', *Pragmatics*, 11: 31–61.
- Wolfson, N. (1989) *Perspectives. Sociolinguistics and TESOL*. Boston, MA: Heinle & Heinle Publishers.
- Yuan, Y. (2001) 'An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral DCTs, field notes, and natural conversations', *Journal of Pragmatics*, 33: 271–292.