

**Please cite as:**

Enfield, N. J., Levinson, S. C., De Ruiter, J. P., & Stivers, T. (2007). Building a corpus of multimodal interaction in your field site. In A. Majid (Ed.), *Field Manual Volume 10* (pp. 96-99). Nijmegen: Max Planck Institute for Psycholinguistics. doi:[10.17617/2.468728](https://doi.org/10.17617/2.468728).

---

**REGULATIONS ON USE**

**Stephen C. Levinson and Asifa Majid**

This website and the materials herewith supplied have been developed by members of the Language and Cognition Department of the Max Planck Institute for Psycholinguistics (formerly the Cognitive Anthropology Research Group). In a number of cases materials were designed in collaboration with staff from other MPI departments.

**Proper citation and attribution**

Any use of the materials should be acknowledged in publications, presentations and other public materials. Entries have been developed by different individuals. Please cite authors as indicated on the webpage and front page of the pdf entry. Use of associated stimuli should also be cited by acknowledging the field manual entry. Intellectual property rights are hereby asserted.

**Creative Commons license**

This material is provided under a Creative Commons Attribution-NonCommercial-ShareAlike license (CC BY-NC-SA 4.0). This means you are free to share (copy, redistribute) the material in any medium or format, and you are free to adapt (remix, transform, build upon) the material, under the following terms: you must give appropriate credit in the form of a citation to the original material; you may not use the material for commercial purposes; and if you adapt the material, you must distribute your contribution under the same license as the original.

**Background**

The field manuals were originally intended as working documents for internal use only. They were supplemented by verbal instructions and additional guidelines in many cases. If you have questions about using the materials, or comments on the viability in various field situations, feel free to get in touch with the authors.

**Contact**

Email us via [library@mpi.nl](mailto:library@mpi.nl)

Max Planck Institute for Psycholinguistics

P.O. Box 310, 6500 AH, Nijmegen, The Netherlands

## BUILDING A CORPUS OF MULTIMODAL INTERACTION IN YOUR FIELD SITE<sup>13</sup>

N. J. Enfield, S. C. Levinson, J. P. de Ruiter & T. Stivers

- Project:** Multimodal Interaction
- Task:** Collect high quality video recordings of spontaneous, naturally-occurring interaction for transcription
- Goal of task:** To acquire a corpus of video data, for investigating the underlying structure(s) of interaction cross-linguistically and cross-culturally.

### Background

Research on video and audio recordings of spontaneous naturally-occurring conversation in English has shown that conversation is a rule-guided, practice-oriented domain that can be investigated for its underlying mechanics or structure (for overviews see Heritage 1984, Levinson 1983). Systematic study could yield something like a grammar for conversation. In the past 30 or so years, a variety of practices and structures have been identified, including the following:

- 1) A system for TURN ORGANIZATION in conversation (Sacks, Schegloff & Jefferson 1974) has held up to moderate cross-linguistic investigation;
- 2) A description of SEQUENCE ORGANIZATION in conversation has proven critical in identifying alternative organizations of overall structure in conversation (e.g., story telling) and in other speech exchange systems (e.g., news interviews or teacher-student encounters) (Schegloff 2007, Schegloff & Sacks 1973).
- 3) An outline of REPAIR in English has identified practices for managing problems of speaking, hearing, and understanding in conversation (Schegloff 1979, Schegloff 1992, Schegloff, Jefferson & Sacks 1977).
- 4) A description of STRUCTURAL PREFERENCE has shown how the design of a turn can facilitate a particular response such as a “yes” (e.g., “Are you going to the party?”) or a “no” (e.g., “You’re not going?”) (Heritage 1984, Pomerantz 1984, Sacks 1973). This has provided insight into how interaction is fundamentally organised, and also has been utilised in applied situations such as when communication fails to work well in a medical consultation.

Much existing research has relied on telephone calls between English-speaking participants. This project seeks to address two primary gaps in current research:

- 1) Social interaction primarily occurs, and is arguably designed for, face-to-face contexts where people have visual access to each other’s behaviour, and to the common environment. Video recordings of face-to-face interaction provide access to the multimodal aspects of communication which play a role in any ongoing interaction (and hence in any interactional/linguistic practice).
- 2) Principles outlined in social interaction research to date have been claimed to apply universally. For example, the Sacks et al. (1974) model of turn-taking has been assumed, until proven otherwise, to operate in all languages and cultures. When languages other than English have been investigated, they have tended to be other European languages.

---

<sup>13</sup> Previous versions of this entry appeared in the 2003 and 2004 field manuals.

This project aims to investigate structural properties of language use in typologically, areally and genetically diverse languages.

### Research questions

What are the principles that underlie the structure of social interaction? Are these principles the same cross-linguistically/ cross-culturally?

### Task

1) Field researchers are asked to video-tape a range of different “maximally informal speech events” involving a range of different participants. Maximal informality is defined as the situation in which the fewest structural constraints on interaction apply. This generalization is meant to discriminate a maximally informal “genre” from explicitly restrictive environments such as ceremonies, speeches, trials, interviews, requested story tellings, etc. The equivalent of “hanging out”, “gossiping”, “chatting”, or “doing nothing” would qualify. More informal situations might be identifiable in terms of *activities*: casual conversation has no explicit pre-determined goals, is often embedded in other activities (like peeling potatoes), is what you do while waiting to do something more important, and doesn’t need elaborate initiation or termination. One could also use the *participants* as a clue: e.g., the kind of verbal activity characterizing same-sex teenagers of the same hamlet in an idle moment. In particular, the following tend to apply in the kinds of situations we are after:

- a. who the participants are is not pre-determined
- b. the dialect or register is not pre-set to be formal
- c. the order of taking turns at talking is not pre-allocated
- d. what can be said is not pre-determined (e.g., unlike in a marriage ceremony)
- e. where and how the participants are spatially positioned is not pre-determined

Note that the most informal situations are not necessarily the most frequently observed. The researcher is not always party to private activities, and may only see more formal ones. Also note (important!), constraints on interaction are never fully absent – we are looking for the *relatively* least constrained speech situation.

These recordings do not have to be very long (5-10 minutes is often enough). Ideally 10 recordings involving 3-5 participants of 10-30 minutes in length would suffice for an initial corpus from which to draw data for sub-projects.

*A note on visual quality:* Please read the instructions for video-recording at the beginning of this manual, and please pay special attention to exposure and to composition of the frame. By ‘exposure’ we mean getting the settings right for the level of lighting available. Try to avoid situations in which speakers are in dark areas where the background is bright. If you must film in such a situation, make sure you set the ‘backlight’ option on the camera. Read that manual. By ‘composition of the frame’, we mean getting certain things in the shot. Do not film close-up shots, as you will miss a lot of important information. People’s whole bodies are important in interaction, especially their hands and arms. You will therefore have to leave enough space in the frame for large/wide gestures not to be cut off. Also, you should try to keep all participants in the shot, even when they are not talking. It is best if you can have the camera set up on a tripod, but if you need to film hand-held, that’s okay too. Just be very careful to keep the camera as steady as humanly possible (e.g. by propping yourself against a post or wall if one is handy). Also, after you

have set the frame composition, you should avoid using the ‘zoom’ at all costs. If you are interested in looking at eye gaze, we recommend you use more than one camera – contact Federico.Rossano@mpi.nl

2) Once you have made some recordings, you will need to select a few sections which are of good quality (i.e. both in sound and visual quality), and work with consultants in transcribing the linguistic material in detail. (For convenience, you may want to do the transcription using the audio signal only – i.e. by first copying the sound from the videotape onto a mini-disc or cassette.) Note that if you are interested in one particular phenomenon (e.g. repair), you might search video data for instances and transcribe just those sequences of talk (be sure to include any possibly relevant talk before and after the target reference term or question, for instance). Most importantly with the transcription is to capture as many details in what is said as possible, including glottal cut offs, stretching of sounds, or other types of hesitations and perturbations, and changes in pronunciation or syntax. When working with consultants be sure that they do not “fix” ungrammatical or odd sounding turns but assist with capturing all details of the talk as produced.

### **Analysis**

The data (the media and transcriptions of interactions) will be used to investigate the use of systematic interactional practices. In the short term, we are interested in data for the sub-project on Repair (Enfield this volume).

Step 1: Having chosen the phenomenon of interest (based on an independent ‘noticing’, or following a sub-project rationale such as ‘repair’; see below), you need to prepare a collection of instances of these phenomena in your corpus. Any number of examples will be useful, but each researcher preferably should have over a hundred examples (ideally several hundred) of each behaviour. The examples should not be hand-picked, but should include ALL examples from a pre-decided stretch of data (e.g. ten minutes from the middle of each of your interactions).

Step 2: For each collection, categorizations must be made regarding the prevalence of one type of behaviour over others, the organization and ordering of behaviours, etc. This will be specified for each collection/subproject.

Step 3: Once collections have been categorised and generalizations can be made, a series of MMI meeting slots will be reserved for presentation of these analyses to colleagues. The categories should be presented, with multiple examples of each (also as media files), and with full support of the analysis using appropriate evidence. As these meetings continue, members of the subproject may sharpen their analysis, and together some generalizations may emerge from comparison across languages.

### **Outcomes**

1) Researchers should aim for an analysis of how their language/culture “does” the task at hand (e.g. the repair of trouble in speaking, hearing, or understanding). Publication plans for each sub-project should be determined in advance.

2) In the short term, the Multimodal Interaction project aims to collect results for the sub-project on repair for contribution to a workshop and subsequent publication project (see next entry), examining and comparing the practices for accomplishing repair cross-

linguistically. In the longer term, conversational corpora form a basis for all subsequent sub-projects (see, for example, *Person Reference in Interaction*, CUP 2007).

## References

- Heritage, J. (1984). *Garfinkel and Ethnomethodology* Cambridge: Polity Press
- Levinson, S.C. (1983). *Pragmatics* Cambridge: Cambridge University Press
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J.M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 57-101). Cambridge: Cambridge University Press.
- Saks, Harvey. 1987[1973]. On the preferences for agreement and contiguity in sequences in conversation. In G. Button and J. R. E. Lee (Eds.) *Talk and Social Organisation* (pp. 54-69). Clevedon, England: Multilingual Matters.
- Sacks, H., Schegloff, E.A., & Jefferson, G. (1974) . A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696-735.
- Schegloff, E.A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289-327.
- Schegloff, E.A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361-382.
- Schegloff, E.A. (1979). The relevance of repair for syntax-for-conversation. In T. Givon (Ed.), *Syntax and Semantics 12: Discourse and Syntax* (pp. 261-288). New York: Academic Press.
- Schegloff, E.A. (1992). Repair after next turn: the last structurally provided for place for the defense of intersubjectivity in conversation. *American Journal of Sociology*, 95, 1295-1345.
- Schegloff, E.A. (2007). *Sequence Organization in interaction: A Primer for Conversation Analysis*. Cambridge: Cambridge University Press.