**Please cite as:**
Enfield, N.J., Jan Peter de Ruiter, Stephen C. Levinson & Tanya Stivers. 2003. Multimodal interaction in your field site: a preliminary investigation. In N.J. Enfield (ed.), Field Research Manual 2003, part I: multimodal interaction, Space, event representation, 10-16. Nijmegen: Max Planck Institute for Psycholinguistics. doi:10.17617/2.877638.

**You can find this entry on:**
http://fieldmanuals.mpi.nl/volumes/2003-1/multimodal-interaction-in-field-site/

---

## 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 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.

### No redistribution

We urge you not to redistribute these files yourself; instead point people to the appropriate page on the Field Manual archives site. This is important for the continuing presence of the website. We will be updating materials, correcting errors and adding information over time. The most recent versions of materials can always be found on our website.

### Be in touch

The materials are being released in the spirit of intellectual co-operation. In some cases the authors of entries have not had the chance to publish results yet. It is expected that users will share results garnered from use of these materials in free intellectual exchange before publication. You are encouraged to get in touch with us if you are going to use these materials for collecting data. These manuals were originally intended as working documents for internal use only. They were supplemented by verbal instructions and additional guidelines in many cases.

The contents of manuals, entries therein and field-kit materials are modified from time to time, and this provides an additional motivation for keeping close contact with the Language and Cognition Department. We would welcome suggestions for changes and additions, and comments on the viability of different materials and techniques in various field situations.

### Contact

Email us via http://fieldmanuals.mpi.nl/contact/
Language and Cognition Department
Max Planck Institute for Psycholinguistics
Postbox310, 6500AH, Nijmegen, The Netherlands

**Multimodal interaction in your field site: a preliminary investigation**

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

This is a guide for the field worker to collect preliminary data on multimodal interaction, as a first step in comparative work within the Multimodal Interaction (MI) Project.

**Motivation**:

The Multimodal Interaction project is interested in how people coordinate their actions in carrying out joint activities. There are three general guiding questions:

- What are the things that must be achieved for people to successfully organise their interaction, and what are the mechanisms people employ to achieve them?
- Which interactional and coordinative mechanisms are universal and which are culture-specific?
- What could the commonalities and differences tell us about the relationship between language, culture, and human cognition?

We have identified a number of phenomena which we suspect are of special interest, including turn-taking, eye-gaze, semiotic use of space and the physical environment, hand gestures, and facial expressions. In order to begin approaching the questions of interest, we need to gather basic empirical data on interactional organization in our respective field sites. This questionnaire represents a first stage in the project, a stage of basic description, and – in the context of a number of theoretical issues – of hypothesis generation.

There are **three things** this investigation requires you **to do**:

1. Produce a sketch 'ethnography of speaking'; i.e. a basic description of the properties of different *speech event* types conventional in your field site;

2. Identify a type or types of *maximally informal speech event* ('informal conversation, hanging out'), make quality video recordings of this type of speech event (preferably collecting several examples, with different sets of participants), and work with consultants to transcribe selected segments;

3. Make careful observations of a set of further interactional phenomena (turn-taking, greetings, coordinated laughter, back channel signals, eye-gaze, silence, repair markers), possibly also collecting video recordings of these.

## Part 1 — A sketch ethnography of speaking

Members of a given speech community will command a range of different ways of speaking. In making his famous 'case for description and taxonomy', Dell Hymes wrote that 'no normal person, and no normal community, is limited to a single way of speaking, to an unchanging monotony that would preclude indication of respect, insolence, mock seriousness, humor, role distance, and intimacy by switching from one mode of speech to another' (Hymes 1986:38). This intra-community diversity compounds the problem of cross-linguistic comparability. It is therefore important to have a sense of just what is the nature and range of diversity in ways of speaking in one's field community. Crucial background to this issue can be found in Jakobson (1960), Hymes (1967/1972/1986), Slobin et al (1967), and Bauman and Sherzer (1974). Note that these works are concerned with *high-level* sociocultural dimensions of interaction, and tend not to connect these explicitly or directly with *low-level* aspects of performance such as turn-organization, repair, or other phenomena of finer time-

resolution in speech production and comprehension. Such low-level phenomena have been well researched in the Conversational Analysis tradition (Sacks et al 1974, Schegloff et al 1977, Goodwin 1981, etc.), work on 'gesture' (Kendon 1972, 1980, McNeill 1992, 2000, etc.), and work on interaction as 'joint activity' (Clark 1996, etc.), inter alia.

**Speech event types**
Hymes defined the *speech event* as an activity that is 'directly governed by rules or norms for the use of speech' (1986:56). We want to consider not just 'speech', but all meaningful aspects of behavior during coordinated co-presence. Note that a speech event may be nested within another activity: A & B chat while peeling potatoes, or whisper in a church service. We can say that chatting is embedded in the activity of potato-peeling or being-in-church, since, for example, a request to pass the potatoes has higher priority than the gossip, or the mode of talking in church (whisper) indicates that talking is 'properly' not done in that setting.

Speech event types come with constraints on the *degrees of freedom* of actions. That's what gives a speech event its characteristic type. There are constraints on participants and the roles they may play, the forms of language (languages, dialects, genres, registers, formality/politeness levels), paralanguage (posture, gesture, prosody), turn-taking (pre-allocation, allocation by chair, self-selection with/without ratification), and action-sequences (pre-specified as in a church service vs. specified at certain points as in greetings, partings, vs. unspecified).

**The task**
This is a task for the researcher, not for your consultants. The general aim is to collect data which will enable project members to find bases for coherent comparison of mechanisms of face-to-face interaction. The task is this: Observe as carefully as possible the full range of different types of situation in which people coordinate their actions – including talk – in some recurrent and/or conventional way. Determine a set of categories of speech event types. For each speech event type, consider a number of variables (explained below), and define each speech event by specifying values for each variable. At the end of this section, below, you will find a form which serves as a checklist for each speech event type you examine.

Begin with the 8 variables which Hymes (1972) listed under the mnemonic acronym SPEAKING:

**S**etting (time, place, physical circumstances)
**P**articipants (who speaks, who listens, whose words and whose ideas are being expressed)
**E**nds (goals and outcomes of the event; e.g. a judgment/decision, coordinated birth of laughter)
**A**ct sequences (form of the message and content of the message)
**K**ey (tone, manner, or spirit: flippant, serious, etc.)
**I**nstrumentalities (choice of *channel* – semaphore, gesture, speech – and *code* – slang, standard, etc.)
**N**orms (of interpretation indicating 'belief systems'; e.g. what is considered good, bad, 'not done')
**G**enres (speech formally marked as distinct in style: poem, myth, tale, proverb, etc.)

Further variables for defining properties of different speech event types are as follows:

- Are there **local terms** for each speech event type?

Cf. English expressions *having a chat, gossiping, having dinner, having a meeting, hanging out, arguing, telling a joke, giving a speech, giving directions*.

- **Newcomer access -** are newcomers to the scene free to join in the speech event?

In some cultural contexts, overhearers or passers-by are not free to simply join in an already ongoing speech event. In other contexts some or any overhearer may be 'licensed' to contribute at will. Can you see any regularities in how people negotiate 'coming in' on an already proceeding interaction?

- **Spatial orientation** of interactants in the situation

Do people sit, stand, lie down? Do they face each other, in a circle, or in some other configuration? How do they physically orient themselves to living space, furniture, objects, other structures?

- **Social make-up** of participants

What is the social make up of interactants present, in terms of age, sex, hierarchical status, or other?

- **Frequency -** how often would each speech event occur?

Note also whether the frequency is seasonal, and if so on what basis. (E.g. *meetings* at MPI Nijmegen don't happen during mid summer.)

- **Duration -** how long does each situation typically last?

There are often differences in how variable this is: at MPI Nijmegen, *meetings* usually last around 90 minutes, and would never be as short as 10 minutes. A *chat in the hallway*, however, can last from 1 minute to an hour or more.

- **Overlapping talk and/or silence**

In each type of speech event, is it possible to say how much people seem to *want to be talking* (i.e. want to have the floor)? Is there much overlap of speaking? How long do overlaps last? Is there someone talking at every moment? Are there noticeable silences? How long are they?

- **Symmetry** of talking-time

Is the time of each individual talking equally distributed between participants? Does one person (or some subset of participants) tend to be talking all the time?

- What **physical activities** are people engaged in while interacting in each type of speech event?

Do physical activities regularly 'compete' with the flow of interaction? If so, how? For example, people might regularly interrupt what they are saying in order to complete some physical task (e.g. while preparing food). Or they might interrupt their practical action in order to talk. How are such task-switches negotiated? Does the content of the talk concern the activity (e.g. as in an office data-session) or not (e.g. as in eating lunch)?

- Are **external representations** other than speech and gesture used in the communication?

Do interactants utilize any kind of external representation such as diagrams, props, maps, sand drawing, or other objects in communicating?

## Part 2 — Collecting video-recordings of a 'maximally informal speech event'

It would be impractical to begin comparative work by taking the full set of different speech event types conventional in Community A and comparing them to the full set of different speech event types conventional in Community B. We therefore want to try to identify a first point of comparison. The sketch 'ethnography of speaking' discussed in the previous section forms a background against which to identify a type or types of *maximally informal speech event* (MISE). In the English-speaking context, this could be denoted by terms such as *casual conversation* or *hanging out*. We want to begin by using this type of speech event as a point of comparison of low-level interactional behavior (turn-taking, eye-gaze, etc.) across cultures. Maximal informality might be defined as the situation in which the fewest constraints on degrees of freedom in interaction apply:

a. participants: do not include persons requiring reserve (in-laws, superiors, strangers)
b. instrumentalities and genre: local dialect, not special registers, levels
c. act sequence: less constraint in paralanguage, more prosodic variation
d. turn-taking: not pre-allocated
e. action-sequences, posture/orientation: not prescribed

Note that the most informal situations are not necessarily the most frequently observed. The researcher is not party to private activities, and may only see more formal ones. More informal situations can possibly be identified in terms of *activities*: casual conversation has no pre-determined goals, is often embedded in other activities (like peeling potatoes), is defined as something 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: the kind of verbal activity characterizing same-sex teenagers of the same hamlet in an idle moment. Other markers of informality might include prevalence of joking and laughter.

**The task**
Once you have identified the speech event types which may qualify as MISEs, the task is to collect video-recordings. These recordings do not have to be very long (5-10 minutes is often enough), but you should certainly collect examples from several different scenes, involving different (sets of) individuals. This will allow some generalization. Once you have made some recordings, you will need to select a few sections 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.)

A note on the 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 the manual. By 'composition of the frame', we mean getting certain things in the actual 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. Also, after you have set the frame composition, you should avoid using the 'zoom' at all costs.

Recordings of this kind, even if they only last 5 minutes (hopefully you can collect 20mins or more of top quality material) will be indispensable in providing baseline information for comparison across cultures (as well as for comparison to material collected using task-based interaction; cf. 'diff-task' entry to this manual). The specific plan for these data is for Multimodal Interaction project members to work on the materials collaboratively, in regular data sessions after the 2003 field season.

**Part 2b:** Recording a 'very formal speech event'
In addition to getting recordings of MISE situations, it would also be valuable to record one or more speech events which are very formal, i.e. which involve significant constraints on the degrees of freedom in interaction. Examples would include temple services, official meetings, interviews, etc. Such data would provide an interesting point of comparison, both within and across cultural settings.

## Part 3 — Focused observation on further issues in interactional organization

This section lists further issues of interest, in order of their priority. All of these issues can only really be studied with reference to primary data, and so you are not expected here to resolve of these questions definitively, nor are you to spend a lot of time on this section. Rather, the point of this section is to alert you to a set of issues which we will want to examine in later data sessions. It is enough simply to start being more observant of these features of interaction, and taking notes which may be useful as clues for more detailed investigation at a later time.

**Please work through the points in the order presented here**. That is, make sure you have collected data on section 3.1 before moving on to 3.2, and so on. This is to ensure that field workers will have greatest possible overlap in data collection upon return from the field.

## 3.1 Turn-taking

In general, to what extent are people observed to 'take turns' in talking?
Do you observe long periods of silence during interaction?
Do you observe extended periods of overlapping speech?
Try to introspect about turn-taking when you are having conversations with consultants. Do you notice or experience anything odd or uncomfortable to do with turn-taking behavior (e.g. do you feel that people in general tend to 'interrupt' you, or wait uncomfortably long before replying to you)?
Are there local terms for concepts related to turn-taking, such as *interrupt, butt in, can't get a word in edgeways, talking over each other, in tune with each other*?

## 3.2 Joint laughter and other multi-party simultaneous actions

Laughter is one type of behaviour which people often perform as a tightly timed simultaneous group action. Collective/simultaneous laughter may also involve accompanying actions such as slapping each others' hands, 'whooping' simultaneously, dramatic posture shift, or applauding together. How is this achieved? What focal actions or signals allow coordination of this kind? If possible, try to collect video-recordings of speech events in which group laughter occurs. This topic is a potential subproject in the MI Project.

## 3.3 'Back channel' signals

So-called 'back channel' signals are signals given by listeners who are not taking a turn but letting the current speaker know that they are understood (and perhaps, signaling that the speaker should continue). English 'back channel' signals include, *m-hm*, *uh-huh*, *right*, *yeah*, and the like. Non-spoken 'back channel' signals include nods and smiles. Such signals, both spoken and otherwise, can be different, in form and in function, across cultures (cf. nodding in Japan vs. India, full-utterance repeats in Tzeltal, etc.). Verbal vs. non-verbal 'back channel' may be correlated with gaze differences. Try to identify conventional signals, spoken and otherwise, which are used as 'back channel' markers in your field site.

## 3.4 Eye-gaze

Patterns of eye-gaze in interaction vary greatly within and across cultures, and for many reasons. It is difficult to learn patterns of eye-gaze from simple observation. Good quality video recording is desirable. In this questionnaire, we can only ask a few impressionistic questions.

- Carefully observe 5 minutes of relaxed conversation between peers, and try to specify:
  Do interactants look at each other regularly?
  If they only rarely look at each other, can you say when they *do* look at each other?
  Can you detect any differences between the way speakers and listeners look? (For example, one claim is that listeners are looking at speakers for more of the time; another is that speakers look to listeners just when they are going to stop talking.)
  Try to introspect about eye-gaze when you are having conversations with consultants yourself. Do you notice or experience anything odd or uncomfortable to do with eye-gaze behavior?
  Are there any professed beliefs about eye-gaze (e.g. 'evil eye')?
  Is there any prescribed/proscribed behavior with respect to gaze?
  Is gazing considered a threat? Could you get into a fight by 'looking at someone'? (cf. De Niro in Taxi Driver)
  Are there terms for gaze 'avoidance'?
  Is there a notion 'look at me when I'm talking to you'? (cf. Travolta in Get Shorty)

### 3.5 Silence

Under what conditions is long silence tolerated between copresent participants? Can it happen immediately after greetings, or even before (as in Apache)? Is it typical for in-laws, strangers, etc.? Does it require an embedding activity (peeling potatoes, chewing betel)? Do speakers report any 'discomfort' with silence? (Cf. the English term *uncomfortable silence*.)

### 3.6 Openings, closings, (dis)engaging

Any interaction must be *started up* in some way, and there are often explicit conventions for doing this. Different conventions have different implications. For example, you can say *Hi!* while passing each other without starting an interaction, but not, *Hello, how are you?*. In some societies, greetings have different types according to passing, stopping, calling out, etc. In some societies all greetings/partings are dyadic (you have to greet each person one by one), and are repeated even after short absences. In others greetings are only done once in a day on first sighting, etc.

Equally interesting is how people *dis*engage. Try to observe whether there are any consistencies in the way people enter into, and get out of, involvement in speech events.

Note also the many mechanisms for entering new phases within already proceeding speech situations. See the Conversational Analysis literature on 'projecting' or setting up in advance certain courses of interaction. These include pre's such as *Can I ask you a question?* and 'assessments' which set up storytelling and predetermine what will constitute their endpoint (e.g. *I saw something amazing this morning*). (Cf. also Clark's work on 'navigation' across hierarchical levels within joint activity structure.) Try to observe and record the details of recurrent patterns which serve these types of navigational functions in interaction.

### 3.7 Repair markers

How do people request a clarification (other repair)? Do they use question forms (*what?*) or have dedicated forms? How do they do self-repair – do they have *um/er* markers, glottal cut-offs, particular items that get stretched? How do they invite other-repair (e.g. with dedicated forms like *the you know-the whatdoyoucallit, the thingumijig*)?

## References

Bauman and Sherzer 1974, *Explorations in the ethnography of speaking*.
Goodwin 1981, *Conversational organisation*.
Hymes 1972/1986, Models of the interaction of language and social life, in *Directions in Sociolinguistics: the ethnography of communication* (ed. Gumperz and Hymes).
Jakobson 1960, Linguistics and Poetics, in *Style in Language* (ed. Sebeok).
Kendon, Adam. 1972. Some relationships between body motion and speech: an analysis of an example. In *Studies in dyadic communication*, (ed Siegman and Pope).
Kendon, Adam. 1980. Gesticulation and speech: two aspects of the process of utterance. In *The relation between verbal and nonverbal communication* (ed. M. R. Key), The Hague: Mouton.
McNeill, David. 1992. *Hand and mind: what gestures reveal about thought*. Chicago Univ. Press.
McNeill, David. (ed), 2000. *Language and gesture*. Cambridge: Cambridge University Press.
Sacks et al 1974, on turn-taking, *Language* 50, p696ff.
Schegloff et al 1977, on repair, *Language* 53, p361ff.
Slobin et al 1967, *A field manual for cross-cultural study of the acquisition of communicative competence*. UCB.

Checklist for defining speech event types (for explanation of categories, see above):

| *Variable* | **Speech event:** |
|---|---|
| Setting | |
| Participants | |
| Ends | |
| Act sequences | |
| Key | |
| Instrumentalities | |
| Norms | |
| Genres | |
| Local terms | |
| Newcomer access | |
| Spatial orientation | |
| Social make-up | |
| Frequency | |
| Duration | |
| Overlap/silence | |
| Symmetry | |
| Physical activities | |
| External representations | |
| Further comments: | |