



The Oxford Handbook of Linguistic Fieldwork

Nicholas Thieberger (ed.)

<https://doi.org/10.1093/oxfordhb/9780199571888.001.0001>

Published: 2011

Online ISBN: 9780191744112

Print ISBN: 9780199571888

CHAPTER

6 Reasons for Documenting Gestures and Suggestions for How to Go About It

Mandana Seyfeddinipur

<https://doi.org/10.1093/oxfordhb/9780199571888.013.0007> Pages 147–165

Published: 18 September 2012

Abstract

The reasons for documenting gestures and suggestions for how to go about it, are enumerated in this article. Using gestures, the hand and arm movements speakers make when talking, seems to be a universal feature of human communication. So far there has been no report of a culture that does not gesture. But our knowledge about the cultural diversity of gesture use within linguistic practices is limited, even though gestures are an integral part of those practices. In order to describe linguistic patterns and regularities of language use, understanding gesture is indispensable. This article is intended as an introduction to the *whys* and *hows* of gesture documentation for linguistic fieldworkers. It provides an overview of the multifaceted phenomenon of gesture, and an insight into how gesture interacts with language, cognition, and culture. Different aspects of gestures — like their semiotic properties, their multiple functions in conversation, and the cognitive linkage between gesture and speech — are introduced. The term ‘gesture’ has been used for many different phenomena, from facial expressions to making a verbal compliment. This article uses the term ‘gesture’ to refer only to the hand and arm movements speakers make when they communicate. This excludes other non-verbal behaviours like gaze, head movements, and eyebrow flashes. Note that this is an arbitrary decision, which reduces the breadth of bodily expression to one set of articulators. The study provides an overview of some basic aspects of gesture, demonstrating how crucial it is to consider gestures as an integral part of language documentation.

Keywords: [gesture documentation](#), [human communication](#), [cognitive linkage](#), [facial expression](#), [verbal compliment](#)

Subject: [Linguistic Anthropology](#), [Linguistics](#)

Series: [Oxford Handbooks](#)

Collection: [Oxford Handbooks Online](#)

6.1 Introduction¹

Using gestures, the hand and arm movements speakers make when talking, seems to be a universal feature of human communication. So far there has been no report of a culture that does not gesture. But our knowledge about the cultural diversity of gesture use within linguistic practices is limited, even though gestures are an integral part of those practices. In order to describe linguistic patterns and regularities of language use, understanding gesture is indispensable.

p. 148 Kendon (1986a) shows how gesture is relevant for many areas of study (see also Kendon 2007). This chapter is intended as an introduction to the whys and hows of gesture documentation for linguistic fieldworkers. It will provide an overview of the multifaceted phenomenon of gesture, and an insight into how gesture interacts with language, cognition, and culture. Different aspects of gestures—like their semiotic properties, their multiple functions in conversation, and the cognitive linkage between gesture and speech—will be introduced. This introduction is far from a comprehensive account of the complex phenomenon, but should heighten attention and awareness of a central aspect of human communicative behaviour.

The term ‘gesture’ has been used for many different phenomena, from facial expressions to making a verbal compliment. For the purpose of this chapter we will use the term ‘gesture’ to refer only to the hand and arm movements speakers make when they communicate. This excludes other non-verbal behaviours like gaze, head movements, and eyebrow flashes. Note that this is an arbitrary decision, which reduces the breadth of bodily expression to one set of articulators. The gaze, the face, the head, the body, and the hands form an orchestrated whole together with speech, and manual actions are not necessarily the dominant component of visual bodily action in utterance.

We also exclude behaviours like blushing, self-grooming, straightening clothing, or actions like smoking or giving something to someone. The rationale behind this second class of exclusions follows Kendon (2004a: 15), who defines gestures among other features based on how observers treat gestural actions since they are ‘directly perceived as being under the guidance of the observed person’s voluntary control and being done for the purpose of expression rather than in the service of some practical aim’ (see Wilkins 2006 for a critical review of this analytical position).

The interest in gesture reaches back to the ancient Greeks (for an overview of the history of gesture studies see Bremmer and Roodenburg 1991; Kendon 1982; 2004a). And although in recent years the study of gesture has become more and more prominent, gestures have been mostly studied in western societies (for exceptions see e.g. Barakat 1976; Brookes 2005; Creider 1977; Enfield 2009; Green 2009; Haviland 1993; Kita and Essegbey 2001; Le Guen 2011; Sherzer 1991; 1993; Sparhawk 1981; Streeck 1993; 1994; Wilkins 2003).

This chapter provides an overview of some basic aspects of gesture, demonstrating how crucial it is to consider gestures as an integral part of language documentation. Section 6.4 provides some practical advice on how to go about including gestures in language documentation.

6.2 Gesture basics

p. 149 Following the definition above, gestures, just like words, communicate information. With gestures, speakers may indicate a location, depict the shape or the size of an object, or show how an object moves. A speaker might enact how someone threw a ball and then point to where the ball was thrown. Speakers also do interactive work with their gestures like rejecting, denying, negating, offering, giving, and comparing. Speakers can use gesture to mark discourse structure and to regulate the coordination between speakers in conversation. Communities have repertoires of gestures that are like words and that are used with and without speech (for classifications of gestures see Bavelas et al. 1992; Efron 1941/1972; Ekman and Friesen 1969; McNeill 1992; Müller 1998, Wundt 1973[1921]; and see Kendon 2004a for a comparative discussion).

Conventionalized gestures, like ‘thumbs up’, have been termed ‘symbolic gestures’ (Efron 1941/1972), ‘emblems’ (Ekman and Friesen 1969) or ‘quotable gestures’ (Kendon 1988a; 1992). They can be used independently of speech and can have different meanings in different cultures (Morris et al. 1979). Members of a community can quote them and provide verbal glosses for them. Their form–meaning relationship is stable and subject to standards of well-formedness. Changing one feature of the gesture's form changes the meaning. In Britain, the ‘victory’ gesture is made with the index and middle finger extended and spread to form a ‘V’ shape, with the palm turned outwards. If the orientation of the palm is changed by turning it so that the back of the hand faces the recipients, it loses this meaning. It may become an insult when combined with motion component of moving the V up rapidly.

Several dictionary-like lists have been published documenting the meanings expressed by gestures for different cultures (see Kendon 1981; 1984b for a discussion, and for methodological problems see Collett 2004). Context-of-use studies have shown how certain quotable gestures make reference to central cultural concepts of special importance in a given culture (Sherzer 1991; Brookes 2004; 2005; Kendon 2004a).

6.2.1 Semantic interaction

Gestures are coordinated semantically and temporally, and this close coordination has, among other features, led to the view that the two forms of expression are guided under a single aim (Kendon 2004a; McNeill 1992). The following exemplifies how speech and gesture interact at different linguistic levels in the creation of meaning. The relationship between gesture and speech is complex, and speakers have a variety of ways to combine the information expressed in the modalities. Gesture can be the main carrier of information, or it can add and further specify aspects of the referent that is being talked about (Kendon 2004a; Lascaridis and Stone 2009).

p. 150 Speakers combine words and gestures into gesture–speech ensembles (Kendon 2004a) or composite signals (Clark 1996). The components of the ensemble differ in their core semiotic properties. Consider someone describing how she threw a ball and smashed a window. While she says *and then I threw this ball*, she lifts her hand next to her head as if holding a ball the size of tennis ball or baseball in her hand and then moves the hand fast forward, extending her arm. With her hand shape she shows the size of the ball and with the movement she enacts how she threw the ball and the direction in which she threw the ball. The information displayed in gesture specifies aspects of the event that are underspecified in the speech portion of her utterance.

Gestures are also used to depict abstract concepts. Imagine someone talking about an argument he had with a friend. He says *and the argument went on and on and on*. The gesture co-occurring with *went on and on and on* depicts the abstract concept of an ongoing duration visually in multiple circular motions. These gestures are representational in that they depict an abstract entity. In this metaphoric process, an abstract domain

(temporal continuity of the arguing event) is represented in terms of a concrete domain (physically repeated circling motion of hands) (Calbris 1990; Cienki and Müller 2008; McNeill 1992).

While in the above examples speech and gesture provide different types of semantic information about an event, gesture can also become the main carrier of information. Speakers may index their gesture with spoken deictics as in utterances like *she held it like this*, followed by a gesture (Streeck 1993; 1994). These gestures can be, for example, pantomimes of actions someone performed, or they can be elaborate depictions of the shape of an object. This kind of relationship can also be observed in pointing gestures accompanied by demonstratives like *here*, *there*. Without the gesture the recognition of the proposition remains incomplete.

Speakers may use gesture to express something that is socially unspeakable or that should not be overheard, and speakers also substitute words with gestures when they deploy a gesture in the verbal slot of a word not being articulated. Slama-Cazacu (1976) calls these ‘mixed syntax’. She reports a case in which a director is talking about spotlight number 5 on a balustrade. He tells an electrician *Five balcony* and the makes a gesture as if switching on the light. The gesture here is basically a substitute for the verb.

In short, in utterances composed of speech and gesture, both modalities contribute to overall utterance interpretation. The relationship between gesture and speech is complex and speakers have a variety of ways to combine the information expressed in the modalities.

6.2.2 Gesture and pragmatics

Gestures, as semantic entities, are also used for pragmatic effect. Speakers can use gesture to assist the listener in interpreting how the verbal part of the utterance should be interpreted. A speaker who deploys a ‘quotation mark’ gesture during a specific part of the verbal part of an utterance provides the recipient with the information that an utterance is a quote, or should not be taken literally. While many cultures have repertoires of these gestures, their usage for pragmatic effects has rarely been studied (but see Kendon 1995; 2004a for a classification and also Müller 2004; Neumann 2004; Seyfeddinipur 2004; Streeck 2009b; Weinreich 1992).

- p. 151 In contrast to gestures that show, for example, how something moved, gestures can also mark the ‘speech act’ or interactional move an utterance performs in conversation, like asking a question or offering information. In Italy, the *grappolo* hand shape (see Fig. 6.1) is deployed in a gesture that functions as a marker for a certain kind of question (Kendon 1992; 1995; Poggi 1983). In this function of the ‘finger bunch’, the hand moves upwards and inwards towards the speaker multiple times. This gesture is used ‘when a speaker asks a question about something because he is surprised, annoyed or puzzled by it, or when he is testing another’s knowledge of something’ (Kendon 2004a: 231–2).

Figure 6.1.



Finger bunch hand shape.

Speakers also use gesture to mark the focus of verbal elements. Seyfeddinipur (2004) describes a gesture sequence with discourse structuring usage observed in Persian speakers marking the topic comment structure of the co-occurring verbal utterance (see Kendon 1995; 2004a: 233–6 for Italian). The sequence is a combination of two hand shapes marking the topic–comment structure of the spoken part of the utterance. In the example, first the speaker has the right hand closed to a fist with the tip of the index finger and the tip of the thumb touching each other and forming an oval shape. Then the speaker opens the gesture; the index finger and thumb get extended while the other fingers remain curled, the so-called Pistolhand (as a consultant called the gesture). The speaker performs this gesture sequence—Ring–Pistolhand—in synchrony with the topic–comment structure of his unfolding verbal utterance. The Ring gesture is synchronized with the topic part of the verbal utterance and the Pistolhand with the comment part of the utterance, i.e. the verb complex (see Fig. 6.2).

Figure 6.2.



Example of a gesture combination with discourse structuring or parsing function. Straight lines between gesture names mark duration of the gestural configuration; hold means that the hand is held still.

p. 152 Little is known about the kinds of repertoires and the different uses and pragmatic effects these gestures can have, and how the cultural environment shapes such repertoires.

6.2.3 Interactional organization

Gesture and speech also interact in the organization of the conversation itself. Studies in the tradition of Conversation Analysis have shown how speakers temporally organize their gestures to regulate the interaction.

In a conversation, speakers coordinate their activities in a systematic way. If speakers are interrupted while gesturing, they can freeze their gesture and hold the hand in the air, signalling that they want the floor back (Schegloff 1984; Goodwin 1986). Addressees can yield the floor by lifting the hand, and they can display that they are giving up this intent by dropping the hand again. And listeners can gesture without interrupting the speaker but still indicating how they are taking what the other is saying. It is clear already on this coarse-grained level of observation that gesture is used to organize turn-taking.

Gestures can foreshadow what will be said or done next, which enables addressees to foresee points in time when the turn of the speaker may end and they can take over, so that a gapless switch of speakers is possible (Sacks, Schegloff, and Jefferson 1974; Schegloff 1980; 1984; Streeck 1995; 2009a; 2009b; Mondada 2007). So, for example, when a speaker begins a turn with a counting gesture (e.g. thumb is extended while the other fingers are curled and palm is up), the addressee can infer that the speaker will list something in their upcoming talk. Also, when a speaker encounter problems in finding a word, gesture can play a role in coordinating the conversation during the search (Fornel 1991; Goodwin and Goodwin 1986; Hayashi 2003; Schlegel 1998).

As shown above, speakers use gesture to express semantic and pragmatic information and to coordinate the interaction. Note that a single gesture can simultaneously perform semantic, pragmatic, and interactional functions in an utterance—these functions are not mutually exclusive.

6.2.3.1 Integrated systems for expression

p. 153

Within language communities, certain domains of expression like numerals or demonstratives have a set of gestural conventions for expression. In gestural convention, hand shape, movement, orientation, or location become associated with specific semantic features. These then can become organized into a set of systematic oppositions for a broader semantic domain. As initially described by Foster (1948) for Tzintzuntzan (see Wilkins 2006), then later by Zavala (2000) for speakers of Akatek (a Mayan language spoken in the Cuchumatán Mountains of Guatemala), a gestural classificatory system is parallel to the classifier system in the \hookrightarrow spoken language. When Akatek speakers measure objects *in absentia* with gesture, the orientation and the hand shape of the gesture changes depending on whether the referent is, for example, a plant, a bird, a child, or a serpent. If and how this gestural system interacts with the verbal classificatory system is not known.

For Arrernte, a central Australian (Pama-Nyungan) language, Wilkins (1999; 2003) describes a fully integrated speech–gesture system in the domain of demonstrative expression. Arrernte speakers combine gesture and demonstratives into what Wilkins (1999: 30) calls ‘composite demonstrative signals’ that differ from those known in European languages. Speech and gesture are combined to express different kinds of information that only together provide the basis for a proper recognition of the location of the intended referent (for other conventions on pointing, see Kendon and Versante 2003 for Neapolitan area of southern Italy; for Laos see also Enfield, Kita, and de Ruiter 2007; for pointing in Zinacantán Tzotzil see Haviland 2003). Wilkins (1999) not only shows that pointing, which is often claimed to be universal, is governed by cultural conventions but also that for Arrernte a description of the demonstrative system will remain incomplete and misleading if the gestural component is excluded.

6.2.4 When gesture is organized into a linguistic system itself

Speakers also use gesture to communicate when speech is not possible because of environmental circumstances like distance or loudness or sociocultural circumstances like taboos or cultural practices that prohibit speech. Given the circumstances, ‘alternate kinesic codes’ (Kendon 2004a) are developed, which can be observed in stock exchanges, sports (baseball, diving), or in guiding the actions of a crane driver (Brun 1969). While the domains of expression and the level and complexity of codification in such kinesic codes are restricted in the above-mentioned examples, there are also more complex systems. Sawmill workers in British Columbia, due to the specific technicalities of the work and the loud environment, developed a gesture system not only to coordinate work flow issues but also to communicate about private matters (Meissner and Philpott 1975).

Members of some religious orders that observe a rule of silence (e.g. Cistercians, Cluniacs, Trappists) are able to communicate for long stretches of time through the use of a limited number of gestures (Barakat 1975; Kendon 1990; Stokoe 1987; Umiker-Sebeok and Sebeok 1987). Kendon suggests that contextual circumstances influence strongly how such systems develop and become elaborated.

p. 154

The restricted gesture systems described above can become elaborated to what Kendon (2004a) calls ‘alternate sign languages’. They are used as an alternative to speech. The Plains Indians of North America used a sign language as a means of \hookrightarrow communication between tribes that did not share a mutually intelligible spoken language (Farnell 1995; Mallery 1987[1880]; Taylor 1978). Such alternate sign languages have also been described for Indigenous Australians of central Australia (Kendon 1988b). For example,

among the Warlpiri of Central Australia, bereaved women traditionally observe a speaking taboo (Kendon 1984a; 1986b; 1988b), and communicate via an alternate sign language.

The systems described above are developed by hearing people who have access to spoken language. Deaf people with no access to spoken language create full-fledged sign languages (for an overview see of the circumstances and the languages created, see Meir et al. 2010). Sign languages demand their own documentary tools, but gesture can also be found in sign language (Liddell and Metzger 1998; Emmorey 2002; Johnston et al. 2007; Kendon 2008).

6.3 Cultural Differences in Gesture

Many aspects of gesture are shaped by cultural conventions, but systematic comparative studies are sparse (for an overview see Kita 2009). Kita suggests four factors influencing gestural behaviour at different levels: conventions on form–meaning associations; pragmatics of communication; language; and cognition.

Cultural conventions on the semiotic structure of gestures determine the size and the content of the repertoire of quotable gestures like the thumbs-up gestures in a given culture (Efron 1941/1972; Kendon 2004b). Pointing, often assumed to be universal and quite simple, is shaped by conventions, and these conventions have to be acquired. Such form–meaning conventions apply to the type of pointing: whether to use lip pointing or the hands, which fingers to point with, and what hand shape and orientation to use (Kita 2003).

Pragmatic conventions on gesture use can be based on considerations of politeness. In the Ewe-speaking region of southeastern Ghana, speakers restrict the way they point because of a left hand use taboo (Kita and Essegbey 2001). When giving route directions, speakers take a conventional respect position with the left hand held behind the back, while the right hand is used for pointing even when pointing to the left. Speakers only use the left for pointing when it is accompanied by the use of the right hand, which is not considered rude.

Other dimensions of gesture like the size of gestural repertoires, the types of gestures, or the extension of gestures are influenced by cultural patterns of interaction, as Kendon suggests (2004a). The first systematic cross-cultural study on gesture was conducted by Efron (a student of Boas) in the late 1930s (Efron 1941/1972). He refuted prevalent Nazi theories stating that gestural behaviour is determined by genetic characteristics of different races by investigating the how gestural practices changed across generations of Italian and Eastern European Jewish immigrants to New York. Efron showed that Eastern European and Sicilian Jewish immigrants to New York differed in their gestural practices from each other, while the second-generation immigrants had assimilated their gestural behaviour in terms of size, number, and types used.

Cross-cultural differences have also been shown for the number and type of hand shapes used in gesticulation and the amount of gestures accompanying speech phrases (e.g. Kendon 2004b on British vs. Neapolitan speakers).

Differences in gesture rate have rarely been demonstrated, although one of the most prevalent beliefs about gesture is that some cultures gesture more than others. Taiwanese mothers when playing with their children produced almost three times more gestures than American mothers. Goldin-Meadow and Saltzman (2000) suggest this is because they have a stronger interest in instructing their children.

A comparative study of Spanish and German speakers found differences not in the amount of gesturing but in the use of gesture space (Müller 1998). While the Germans gestured mostly from the wrist, the Spanish speakers gestured mostly from the elbow and the shoulder (see also Efron 1941/1972 and Kendon 2004b). It

has been suggested that more expansive gestures may be perceived as more prominent, and that this may lead to the idea that certain cultures gesture more than others (Müller 1998; Kita 2009; and see Kendon 2004b for methodological consideration of assessing gesture rates).

Many factors may be responsible for the described differences. Kendon (2004a; 2004b) suggests that the ‘communicative ecology’ of a given culture has to be taken into account to understand the cross-cultural differences.

6.3.1 Linguistic diversity

p. 156

There are cross-cultural differences in what kind of information speakers of different languages express in speech and gesture when talking about the same event. In Japanese, for example, the action of swinging cannot easily be expressed in one word as in English. Speakers often use simple verbs like ‘go’ that do not encode the arc trajectory of the motion. When Japanese speakers describe a swing event they could, in theory, compensate in gesture for this lack of information in their vocabulary by displaying the arc trajectory visually. While Japanese speakers combined speech and gesture in this way, Kita and Özyürek (2003) also found a tendency for them to actually perform gestures that move in a straight line. English speakers, in contrast, who can readily express the shape of the trajectory with the word *swing*, also depict the arc trajectory in their gestures. So gesture does not always compensate for, but often parallels, linguistic packaging. This suggests that ↪ the form of the gesture can be influenced by the way information is expressed in a specific language.

The same seems to be true for syntactic structuring (Kita et al. 2007). In Turkish, the way an object moves (manner) and the path of motion are usually expressed in two separate clauses. When Turkish speakers express how something rolled down a hill, they make a rotating movement when they say ‘it rolled’, followed by a straight, slightly downward movement as they say and ‘it went down’. In contrast, English speakers mostly conflate manner and path information in one gesture by making a rotational movement downwards while they say *it rolled down*. The information expressed in gesture is adapted to the way information is expressed in a specific language.

Kita (2009) suggests that these differences are a reflection of diversity in ‘thinking-for-speaking’ (Slobin 1996). It appears that certain aspects of linguistic typology in sentence packaging and lexicalization may have consequences for the structure of co-speech gesture.

6.3.2 Cognitive diversity

There are cross-cultural differences in how spatial information is conceptualized and expressed, and this kind of cognitive diversity is also reflected in gesture. When speakers express location and directions, they differ in how they anchor spatial relationships (Levinson and Wilkins 2006). Guugu Yimithirr speakers of Hopevale in Queensland anchor spatial relationships at all levels of scale with regard to the cardinal directions. Haviland (1993) reports two instances in which a Guugu Yimithirr speaker is describing how his boat flipped in a storm. The first time he is facing west. While he says, ‘The boat was lifted up and starting to turn’, he brings up his arms with the palms facing him and rotates them in a flipping motion forward from east to west. Two years later the speaker tells the same story in a different location facing north. He brings up his hands with the palms facing each other and rotates them again in a flipping motion from east to west (clockwise). The directionality of the flipping motion matched the absolute orientation of the incident, and the speaker modified his gesture to preserve the absolute orientation of the turning of the boat.

There are also cross-cultural differences in how an abstract concept like time is represented in gesture. Time tends to be represented in concrete spatial terms. So concepts like past and future are expressed in English

p. 157

by metaphorical expressions like *the event is ahead of us* or *the event is behind us*, mapping the temporal relations onto the sagittal axis, which passes from the front to rear of the body (Boroditsky 2000). There are cultures that reverse the directionality. Nunez and Sweetser (2006) report that speakers of Aymara (Chilean Andes), when talking about the future, often use expressions containing the word ‘back’ and they often use the word ‘front’ when talking about the past. Accordingly, especially older speakers with limited Spanish would gesture forward to represent an event in the past while they ↴ gestured backward to represent that something will happen in the future. In contrast, French speakers, for whom the future is ahead and the past is behind, gesture forward for the future and backward for the past (Calbris 1990; see also Kita, Danziger, and Stolz 2001 for a comparison between Mopan and Yucatec, Mexico; Boroditsky and Gaby 2010 for the use of cardinal directions in the expression of time in the Australian Aboriginal community of Pormpuraaw).

The way speakers think about space or the way they map temporal relations onto space is reflected in gestural representation. Comparative systematic studies are needed to gain a better understanding of how metaphorical mappings in gesture and speech take place.

6.3.3 Summary

Gesture is an integral part of the linguistic practices of a given speech community. The practices are manifold, and extend from communicative events like ritual speech and monologs to conversations. In everyday interaction, speakers use gesture–speech ensembles to express diverse types of meaning, including referential meaning. Gestures can be used to mark focal elements of the discourse or to provide an interpretation framework for the verbal part of the utterance, and they are used to organize the conversation. Speech and gesture together form composite signals (Clark 1996) that have to be taken into account in order to fully understand and describe certain domains of linguistic expression. When speech is not possible, gestures are used to form kinesic systems for communication.

Cultural differences shape many aspects of gesture, like the repertoire, the size, the content, orientation, the types, the use in context, and so on. Linguistic, cognitive, and social factors underlie these variations. To gain a broader understanding of how cultural differences in gestural behaviour emerge, it is necessary to study gesture in its various contexts of use.

6.4 How to Start Including Gesture in Documentation

p. 158

It may at first seem daunting to include another modality in the documentation of language. But the documentation of the role of gesture in language use and of how the gestural system and the verbal system interact in communication can easily draw upon methods and techniques already established for language documentation. The work of a researcher studying gesture does not drastically differ from the ↴ work of a descriptive linguist documenting and describing a hitherto undescribed language. Gestures, like any sign, have different aspects—form, meaning, and context—that can be analysed and described. To enable such descriptions, a minimal approach can be taken by ensuring that the recordings of spoken language is of a quality which allows for (at least some) analysis of gestures at a later stage. This involves using video as the basic recording tool and following certain standards for the filming—such as framing the picture so that it minimally includes the speaker's head and upper body rather than only the face or the hands (see below for more detail). A second more elaborate step is to expand the corpus compilation by including recordings of topics and genres which are gesture-prominent. This minimal approach may be all that can be managed by a researcher whose focus is on other matters.

This section provides an introduction to such a minimal approach—namely, how to go about including gestures in language documentation, in terms both of what to record and how to record, and of how to

minimally annotate to make the data accessible.

6.4.1 What to record

A first guiding question for the documentation of gesture is: how do speakers coordinate their verbal and gestural resources in everyday interactions? To answer this question, a starting point is to add to the corpus a variety of topics, and a range of different communicative events (i.e. different text types), which can approach the known range of gesture types (see Seifart 2008 for methods of corpus compilation). Different speakers should be recorded producing the same types of text to control for speaker variability in gesturing.

The fundamental basis is to record gestures in everyday interaction. The tradition of Hymes' (1974) ethnography of speaking can provide a framework for the dimensions and types that should be taken into account (Himmelmann 1998; Hill 2006; Seifart 2008). A guiding view is that using language is a form of joint action (Clark 1996) in which speakers are solving coordination problems by using their expressive resources in different ways. For example, how do speakers draw the attention of their interlocutor to a certain entity in a given situation? Depending on the context, a speaker may use a pointing gesture to direct the attention of the interlocutors to a certain location in conjunction with a demonstrative or they may use a description (like *X is in the kitchen*). Observing these practices and focusing on bodily deployment provides a basis for noticing the role gestural expression may play in a community.

p. 159 Note that, as an outsider, one is not equipped with the cultural knowledge of the conventions regarding the form–function relation of gestures, as Wilkins (1999; ↵ 2006) points out. Community members should be involved in data elicitation and analysis (see Wilkins 2003 for a detailed methodological description).

A starting point is to record situations in which speakers talk about topics that are rich in spatial content, like route directions (Where is x? How do you get to y?) and spatial descriptions (How did the village/area look 50 years ago? What changed and how?). In such situations, speakers are likely to use pointing gestures to indicate locations and they lay out spatial relations with their gestures.

Topics that are rich in spatial content and are very likely to yield a variety of gestures can also be elicited in storytelling. Also, procedural texts about (for example) how artefacts or foods are prepared and used, or how certain rituals take place, typically contain gestures depicting shape, size, use, and concrete aspects of an object or action, and may also contain gestural representations of abstract features like the order of events. This type of discourse can also be recorded in instructional context when a teacher instructs a novice.

Topics dealing with abstract relationships like time (past and future), or kinship (Enfield 2009) are also promising for evoking gesture behaviours. Sometimes sessions set up for some other purpose, such as consultants engaged in discussion within elicitation or translation sessions, can provide interesting data on the use of metaphorical gestures when they (for example) discuss syntactic structures, making them visible through their gestures (Mittelberg 2002).

Gestures with pragmatic function, like structuring the discourse or marking the speech act of an utterance, are most likely to be observed in animated discussions about issues the community cares about and where members may be in disagreement. Political discussions and negotiations often are situations where speakers use these gestures, which may also depict abstract relations (Cienki and Müller 2008).

Gestures can be also elicited in more directed elicitations. Tasks that have proven useful in gesture elicitation are picture book descriptions (e.g. the Frog Story, cf. Berman and Slobin 1994) and cartoon narrations (the Tweety and Sylvester cartoon, cf. McNeill 1992; Road Runner, cf. Bavelas et al. 1992). Using picture books or cartoons has the advantage that the data collected can potentially be compared systematically cross-culturally. Cartoons are rich in motion events of different types (rolling down a hill,

entering, exiting). Because of the relatively static nature of picture books, retelling may or may not be successful in eliciting gesture.

More focused gesture elicitation tools have been developed for motion events (e.g. the Tomato man movie: Özyürek et al. 2008) and for spatial reference (see Majid, Chapter 2 above). However, as mentioned above, the applicability of the stimuli depends on the familiarity of a given community with such materials and tasks, and so some stimuli and tasks may not be appropriate for the cultural context, or for use with certain members of the community.

For those who are more interested in undertaking gesture studies, some examples in the literature of gesture research by field linguists are Enfield (2009), Green (2009), Haviland (1993), Kita et al. (2001), Kita and Essegbey (2001), Le Guen (2011), Sherzer (1991), Wilkins (1999). See also Goodwin (1986) and Kendon (2004a).

6.4.2 How to record

Gesture, just like speech, is a fleeting phenomenon. Because of its ephemeral nature, certain aspects are impossible to catch by casual observation alone. The temporal unfolding over time of the gesture form, its trajectory of motion, and its synchrony with specific elements in speech can only be captured systematically by video. Video recording, however, can be problematic for ethical reasons (Rice, Chapter 18 below), is technically challenging (Margetts and Margetts, Chapter 1 above), and speakers may alter their behaviour in the presence of a recording device.

Video need not only be shot by the fieldworker, but may in fact be better collected by trained community members. Having community members video record interactions can also provide insights into what they consider important in the interactions (for uses of video in documentation, see Ashmore 2008).

In any case, there are some guidelines for recording which best facilitate later analyses of gesture speech interaction. An ideal recording frames the conversational situation with the speakers' upper bodies, with at least an arm-length space around their head. If possible, in framing the span of the interlocutors' space, it is best to overcompensate so that the maximum breadth of possible movement (assuming fixed position) is taken into account. The camera should be positioned at 9 o'clock or frontal, and angled slightly down from eye level of the speakers. With such a framing of the speakers, we can observe the hand shapes and movements along with facial expressions, gaze, and eyebrow and head movements. This is important because these all tend to form elements in an orchestrated whole.

An obvious problem for this setup is that the way speakers are oriented towards each other depends on cultural as well as situational restrictions. In many places it is not very natural to have two or more people side by side facing the camera rather than each other, and any artificial composition placing the speakers in culturally unnatural positions should be avoided. So we always have to find a compromise, such as framing only the main speaker in this way while the other speakers' faces may not be fully visible.

Ideally, the social space created by the interlocutors (including all interlocutors and not just a single speaker) should be recorded. This is because the verbal and gestural behaviour of speakers is contingent on the behaviour of their interlocutors, where they are located, whether they look or not, whether they backchannel their understanding or not, etc. (Schegloff 1984; Streeck 1994; Goodwin 1986). Speakers in some cultures are known to orient their gestures and the direction of movement depending on the location of the interlocutors (Özyürek 2002). In addition, for Canadian English-speaking college students, it was found that speakers' gestures are more pronounced and better articulated when the speaker thinks the recipient does not have prior knowledge of what the speaker is talking about: when the information to be

conveyed is not yet in common ground, gestures are more elaborated and complex (Gerwing and Bavelas 2005).

A framing including all interlocutors may not be possible for communities who employ a communal broadcast model of talk (see Walsh 1991; 1997 for Aboriginal communities) rather than a dyadic or face-to-face model of talk. In such a situation, the social space of interlocutors should be framed such that the main actors are captured by the recording.

The video should be coupled with high-quality audio recording. Hence the choice of the right microphones and their placement is crucial (Margetts and Margetts, Chapter 1 above). Only this combination of video and audio data allows us to analyse the way speakers coordinate gesture and speech and other modalities like gaze or head movement. Field notes should include metadata details like identity of the speaker(s) and their relationship to each other, time and location of the recording, and type of speech event.

To obtain good data, the recording setup should be thought through and if possible prepared in advance. This includes choosing the position of the camera and finding a compromise between obtaining good recording quality and minimizing intrusiveness. Setting up in advance has the advantage that sound and picture are adjusted to the location and that one can start the recording beforehand without having to adjust the camera later on. The camera should be fixed on a tripod or placed on a steady straight surface to avoid jittery images; moving the camera in the middle of the recording should be avoided. The less one has to handle the camera during the recording session and in the presence of the consultants, the better. Fiddling with the camera attracts the attention of the people being recorded to the recording device and may lead to very controlled behaviour (Labov 1972a; for a discussion see Himmelmann 1998). Adjusting the camera, for example for panning and zooming, during a recording should be avoided also because there is a high risk of missing crucial parts of the interaction. Static continuous shots are recommended.

If one chooses to prepare a recording as suggested here, this involves two steps. The first step is to film the recording location itself, i.e. where the speakers will be located and what surrounds them. This constitutes important metadata that may be required for the analysis. In addition, a map of the location can be drawn, including again the position of the speakers and the objects or landmarks surrounding them. This may make it possible to analyse the location to which someone is pointing and/or how gestures are oriented spatially (Haviland 1993; Levinson and Wilkins 2006; Kita et al. 2001; LeGuen 2011).

p. 162 In the second step, the camera is zoomed in to frame the interactional space. In staged communicative events, typically one or two speakers are present and the scene is relatively easy to frame. But natural situations pose a challenge for the recording, ↴ since the number of participants and their spatial arrangement change over time. When there are multiple interlocutors, ideally all of them should be recorded. This may be difficult when the participants are located too far apart for the camera to capture them all. Depending on the camera's lens (a wide-angle lens can capture a wider space), the camera needs to be moved further away, but increased distance from the speakers reduces the detail visible in the recordings (such as facial expressions) and potentially also the sound quality, depending on the recording setup and equipment. One way of handling such a situation is to focus the camera on the most active interlocutors and disregard the more distant overhearers who are not actively participating. Even when the focus is on gesture, the intelligibility of the speech should never be compromised. As mentioned above, speech and gesture need to be analysed together to assess meaning and function, and unintelligible speech minimizes the amount of specific information that can be taken from the recording.

One way of being relatively unobtrusive is to set up the camera and leave it running without attending to it (it can help to switch off the little red light that indicates that the camera is running). Speakers, who must be informed about the recording in advance, are more likely to forget or at least not focus on the fact that they

are being recorded (for ethical issues that need to be addressed before recording speakers, see the chapters by Rice and Newman, Chapters 18 and 19 below).

6.5 Basic Annotation of Gesture

As discussed in §6.2, researchers studying gesture investigate the coordination of speech and gesture to answer questions concerning the forms and functions of gestural deployment, to determine the role of gesture in communication, and to explore what gesture reveals about cognition. A basic and minimal annotation of gesture consists of marking the periods of time in which the hands are gesturing. This type of basic annotation enables the analyst to locate the stretches in the video which are relevant for gesture analysis, and can be used to assess the time speakers spend gesturing. To identify gesture in the movement stream, the basic structure of gesture has to be taken into account.

p. 163 Gestural movement can be segmented into four distinct phases (Kita, van Gijn, and van der Hulst 1998; McNeill 1992; Kendon 2004a): preparation, stroke, hold, and retraction. Consider a speaker describing how a ball rolled down a hill by making a spiral movement downwards with the extended index finger. First the hand is lifted from the lap to chest height. This phase is the preparation: the hand is brought into the starting position for the stroke. The stroke then consists of the hand moving downward in a spiralling movement. The stroke is the only obligatory phase of a ↴ gesture, and it can be characterized as the most articulated phase during which the most force is exerted. A stroke can be preceded or followed by a hold (pre-stroke or post-stroke), a static phase in which the hand is held still in a position. The phase during which the hand is moved back into rest position is the retraction. A stroke can be static or dynamic, with meaning being expressed through a static configuration (e.g. when size or extension is indicated) or through motion (rotating movement downward). Note that this established segmentation scheme mixes formal (hold) and functional categories (preparation, stroke, retraction). To circumvent this problem one, can distinguish in the annotation between static and dynamic strokes.

The phases can be distinguished based on changes in the direction of movement and changes in the speed of motion (acceleration or deceleration) and changes in the articulation of the motion (e.g. tension of the hand, directedness of motion). The segmentation of the movement into discrete phases can be accomplished by looking at the video frame by frame to see when the points of change occur. When the starting and the end points are determined, the stretch can be tagged and categorized (for a discussion of how to determining moments of transition systematically, see Seyfeddinipur 2006).

The time speakers spend gesturing is the basic gesture unit (Kendon 2004a). This is when the speaker moves his hands away from a rest position (e.g. lap) or an activity (e.g. cutting, smoking) and begins to gesticulate until the hands return to a different activity again.

For multimedia annotation, software like ELAN² or ANVIL³ should be used. These annotation tools allow setting up multiple tiers for different levels of annotation as in a musical score. In each tier a different type of annotation can be provided which is time aligned with the video (see Fig. 6.3). Relevant phases in which the speaker gestures can be marked based on a frame-by-frame inspection of the video. Once a period is identified, it can be tagged and replayed.

For each hand, one tier should be set up, and the times during which gesturing occurs should be marked for each hand. During such a gesture unit, the speaker may use multiple gestures, and the next level of annotation for a more detailed analysis entails a breakdown of the stream of gestural movements into gesture strokes, the semantic nucleus of gestures. This level of annotation can be pursued when the researcher has a clearer idea of which gestures are of particular interest.

For those interested in how gesture can be described in terms of form features, it can be briefly mentioned that four parameters are taken into account: hand shape, orientation, movement, and location (Stokoe 1960). It should also be pointed out that there is as yet no conventional transcription system for the form of a gesture and its temporal unfolding. For some examples of how different researchers ↴ transpose the temporal unfolding of the movement in relation to the speech into a graphical representation, see Duncan's annotation and coding procedure described in McNeill 2005; see also Kendon 2004a: 362; Calbris 1990; Mittelberg 2007.

Figure 6.3.



ELAN screen. Example for a simple tier set up for basic gesture annotation. Note that the stroke tier for the left and right hand is inserted for those who are interested in more detailed annotation.

6.6 Summary

The first aim of this chapter was to show why gesture documentation is an important part of language documentation. The second aim was to describe the methodological issues, the procedures, and the results of gesture documentation within language documentation. A minimal approach to the documentation and later analyses of gestural practices has been suggested. This approach requires the researcher to follow certain recording standards, and to broaden the corpus ↴ compilation by including recordings of a range of topics and genres which are likely to evoke different types of gestures. A further step is making the video recordings accessible for later analyses by annotating the parts of the video in which speakers are found to be gesturing. This minimal approach should be possible as part of a fieldworker's normal recording and transcribing activities. The result, a record of gestural practices of a language community, is invaluable.

Much more can be said about the relevance of gesture in the study and documentation of communicative practices, the methods of data collection and analyses. This chapter provides a first overview of the approaches and the methods used in the study of gesture in communication, culture, and cognition. Gesture is an understudied phenomenon, despite its tight linkage with speech and its complex role in human communication. Accounts of gestural deployment in different cultures are still sparse, especially compared to the descriptions and documentations available for the verbal component of language. Gesture is integral

to a better understanding of the use and the interaction of the expressive resources speakers have at their disposal, and of how these shape communicative practices.

Notes

- 1 I would like to thank David Wilkins, Anna Margetts, Nick Thieberger, Birgit Hellwig, Adam Kendon, and two anonymous reviewers for valuable comments and suggestions.
- 2 www.mpi.nl/tools
- 3 <http://www.anvil-software.de/users.html>