Duranti A. (1997). Linguistic Anthropology. Cambridge, CUP.

5
Transcription: from writing to digitized images

Boas and Malinowski were both concerned with standards of field research and empirical methods and believed that showing the linguistic sources of their ethnographic descriptions, that is, informants' verbal accounts, was a very important part of an anthropologist's task. Since they did not have the luxury of a machine that could record and then play over and over again what the informant said, transcription meant writing down in a systematic and careful fashion informants' answers to questions regarding traditional knowledge and various aspects of the social organization of their community. The transcription of an actual conversation among native speakers or any other kind of verbal performance at the normal rate of speech was beyond the technological reach of early ethnographers. To capture information about language use, they were forced then to rely on two kinds of techniques. One was to try to catch a word or phrase as it was used in the course of an interaction, make a mental or written note about it, and then wait for an opportunity to ask an informant about it:

> When an exceptionally good phrase [about botany or gardening] occurred I would make a brief note of it, mental or written, and then lead my informant to repeat it, not necessarily as I had first heard it, but so as to reproduce the information it contained and its linguistic character. (Malinowski 1935, vol. 2: 5)

Another technique was to elicit narratives about a given topic and transcribe them. This method relied on native speakers' ability (and patience) to speak clearly and slowly, and their willingness to adapt to the ethnographer's limited understanding of the local language. Ordinary talk, whether formal or casual, was of course a real problem, as reported by Boas himself in a letter to Ruth Benedict in 1930 (Boas was seventy-two):

> I am worrying now about the style of oratory because I do not yet know how to get it down. Anyway I have my troubles with ordinary conversation. Narrative I can understand quite well, if they talk distinctly, but many have the Indian habit of slurring over the ends of their words – whispering – and that makes it difficult.
>
> (quoted in Mead 1959: 43)

Things have changed considerably in the last few decades. With the invention of tape recorders first and video recorders more recently,[1] new research methods have been developed. Linguistic anthropologists in particular have been quick to capitalize on these technological advances. Most linguistic anthropologists have adopted electronic recording of natural speech as a standard practice in their research methods. The introduction of these new media has fostered higher standards of accuracy and an interest in interactional details that would have been overlooked in the past.[2] Linguistic anthropologists have become particularly keen on producing transcripts of stretches of native discourse recorded during spontaneous interactions, ranging from ceremonial events to casual conversations.

In this chapter, I introduce different units of analysis of spoken language and the logic that underlies their use. I dedicate several sections to the "word" as a unit of analysis because it has been so important in both linguistics and anthropology. I then launch into discourse units and the different formats and conventions that have been introduced for their transcription. I also discuss transcription tools other than writing, including drawings and digitized images. Finally, I discuss translation and various formats for its representation.

5.1 Writing

Writing systems have been important for the development of linguistic analysis for at least two reasons: they have been crucial for our understanding of how linguistic sounds change over time (historical linguistics) and for the segmentations

[1] Film, of course, was available much before video technology was perfected to allow for portable cameras and camcorders. With some notable exceptions (e.g. Connor, Asch, and Asch 1986), however, ethnographic filming has run its own parallel course and ethnographers have rarely incorporated film in their analyses. This is partly due to the forbidding cost of film and the technical expertise required, not to mention the common field condition of not having electricity to recharge batteries, high humidity, etc. In addition, however, there has always been in western academia a higher value given to words over images. With the exception of a few visual anthropology programs, anthropology graduate students and junior faculty are usually encouraged to publish printed material rather than spend their time producing films or figuring out how to integrate the two media.

[2] Within different disciplinary traditions such as human ethology and social psychology there is a more established tradition of detailed empirical studies of visual communication (e.g. Argyle 1969, Argyle and Cook 1976; Eibl-Eibesfeldt 1968, 1970, 1974; Ekman 1982; Kendon 1977, 1980, 1990). In sociocultural anthropology, Bateson and Mead (1942) were among the first to encourage fieldworkers to use photography and film but even today most ethnographers do not engage in detailed analysis of audio-visual recordings.

of meaningful strings of sounds into units of analysis such as sentences and these, in turn, into words and their components (morphemes, phonemes) (see below and chapter 6). Written records have allowed linguists to have access to earlier stages of languages (Ancient Egyptian, Hittite, Sanskrit, Old Turkish, Ancient Mayan). By comparing those early records with existing languages – the so-called "daughter languages" of the older, dead ones –, in the eighteenth- and nineteenth-century linguists were able to develop hypotheses about how languages change over time and across space (Bynon 1977; Lehmann 1973; Keiler 1972). The theories developed on the basis of such written records have been then used to reconstruct earlier stages (what linguists call "proto- languages") of currently spoken languages that had no indigenous writing tradition.

But writing systems contain a number of assumptions about language structure. One of the best case studies to make this point is Mark Aronoff's (1985) analysis of the orthography developed by the Masorets for Biblical Hebrew between AD 600 and 800. Aronoff shows that the conventions introduced to mark stress are based on a syntactic analysis of the text that resembles, in some respects, the formalization adopted by modern structuralist and generative syntacticians.

Writing – more specifically alphabetic writing – was essential to the notion and practice of transcription as originally developed by Boas for "salvaging" rapidly disappearing Native American languages and cultures (see section 3.1). Given that writing down the sounds of a language makes us face important decisions on linguistic structures and the organization of a given linguistic system, Boas and his American Indian consultants were not only making a record of the past, they were also presenting an analysis of the language they were transcribing.

Writing is a powerful form of classification because it recognizes certain distinctions while ignoring others. For instance, in English we use the letter "s" for representing the plural of words despite the fact that when we do so we are in fact conflating distinct sounds: the *s* of *cats* is not the same as the *s* of *dogs* (see section 6.3). Native speakers of English "know" this difference, although they might not be aware of it, while literate non-native speakers are often confused by the the fact that same letter is used for what are in fact different sounds.

Writing down a language that has never been written before constitutes a first *description* of that language. By allowing us to see what we hear, that is, by transforming an acoustic phenomenon into a visual one, writing allows for a different type of manipulation of linguistic signals, for different kinds of abstractions, for new types of connections. But like any other powerful analytical tool, writing not only highlights certain properties (Goodwin 1994), it also hides some others (Irvine and Gal in press). First, visual representation in the form of orthographic conventions of any sort (whether alphabetic or syllabic, for instance) reproduces an ideology of interpretation whereby we believe that we know what something

means by a one-to-one match between individual words and individual meanings. This is indeed the theory of interpretation represented by past and present extensions of propositional logic to natural languages. As I will discuss below and in later chapters, this theory presents some problems, especially when faced with speech as used in actual interaction and not under controlled conditions.

Second, since any writing system contains a partial theory of the sounds and units of the language it purports to represent, when we write down the sounds of a language that has never been transcribed before, we bring to it a history of ways of thinking about what linguistic sounds are like and what they are for. Writing is also associated with particular grammatical traditions. Thus, the early missionaries in Africa, Asia, North and South America, and Oceania used the distinctions found in Latin grammars as their guiding principles for grammatical description. This meant that they imposed morphological distinctions such as nominal cases (nominative, dative, ablative, vocative) even on languages in which the noun did not change depending on its place in the sentence (Anderson 1985a: 197–8; Cardona 1976: 37–42).

Writing down a language also establishes a particular dialect or register among the several in use at any particular time as the **standard language**. Such a practice has important consequences not only for the destiny of local dialects that are different from the one chosen as the standard, but also for the type of idealization made by students of language (Finegan 1980; Morgan 1994). Until the birth of urban sociolinguistics in the 1960s, issues of orthographic representations in the West were mostly restricted to the interest of fiction writers who wanted to reproduce (or just give a feeling for) non-standard dialects, usually in dialogue. With the exception of phoneticians and phonologists, western grammarians (syntacticians, semanticists) working on their own language did not seem to have doubts as to how to represent the examples they were creating for their argumentation. Even now, if one opens up a linguistic textbook or a journal of formal linguistics, one discovers that syntacticians working on English assume that there is no problematic relationship between the graphic representation of sentences on a page and their spoken counterpart. In other words, standard orthography is implicitly associated with the idealization of speech that is central to contemporary formal theories of grammar. The uncritical adoption of a particular system of representation is therefore not simply a theoretical stratagem (e.g. we need to assume some basic abstract system to explain language acquisition and shared semantic interpretations), but also an ideological ploy that ends up reinforcing hegemonic assumptions about what any speaker *should* be saying. This means that although writing offers us great opportunities for analysis, it also constrains the range of phenomena we are likely to study and taints them with particular ideological implications. It is therefore crucial that we critically appraise the use of orthographic representation in

linguistic analysis, so that we can exploit writing as an analytical tool while stretching the analytical boundaries established by its past uses.

Finally, recent experiments have suggested that familiarity with a writing system (the practice of reading in particular) might be crucial for developing the ability to segment speech into separate sounds (phonemes) or larger units (morphemes) (see chapter 6). We cannot assume, for instance, that any speaker of English would be able to separate the sounds that linguists see as forming a word like *fly* or *bite*. In a series of so-called *phoneme deletion* experiments, in which subjects are asked to delete a particular sound of an existing word, most non-literate adults could not perform the task. Reviewing the existing literature and their own work on this faculty, Scholes and Willis write:

> Speakers of English are able to manipulate phonemes only if they can read. The acquisition of the alphabetic representation of language enables the language knower to transfer this way of representation (i.e. sequences of discrete sublexical elements) to speech. In short, we know about phonemes because we know about letters. (Scholes and Willis 1991: 220)

The hypothesis that writing has an impact on speakers' ability to perform linguistic analyses on their own or on other people's speech is part of an attempt to link the introduction of literacy to cognitive as well as social changes in individual members of particular speech communities. This is a topic of great controversy given that the role of literacy in linguistic analysis has been underestimated if not altogether ignored by grammarians as well as by philosophers of language, who have been assuming that the type of analysis they engage in is an adequate idealization of cognitive abilities that any speaker of a language (and not only linguists) can make.[3] Although many formal linguists today recognize that what they might be studying is the grammatical competence of an idealized group of speaker (viz. the tacit knowledge of language by an average university professor), they do not readily admit that their culture, including the culture of literacy and the importance that reading and writing have in their daily life, might in fact have an impact on the type of analysis they propose.

5.2 The word as a unit of analysis

Alphabetic writing was particularly important for the identification of the word as a basic unit of analysis in linguistics. Although linguists have been searching for writing-independent criteria for establishing the boundaries of words in different languages around the world, there is little doubt that the first impulse for assuming

[3] On Sapir's analytical use of intuitions by speakers of unwritten languages, see section 6.3.1.

the word as a basic unit of analysis in linguistics must have come from alphabetic writing conventions. Among the criteria currently used to isolate words in consistent ways are: pausing, stress, and certain morphological processes or constraints that seem to apply to words but not to larger units (Anderson 1985b).

Languages display a considerable variation in length and shape of words, especially when we use pausing as a criterion for defining word boundaries. Whereas in some languages, one seems to be able to pause after each syllable (Vietnamese is said to be such a language), in other ones, most typically Native North American languages, pauses are allowed only after what appear as full sentences. Another criterion used to distinguish word units is permutability of word order. Words can often be moved in different positions within a sentence (although languages as well as types of words within the same language vary considerably in this respect), but parts of words (morphemes) cannot as easily be moved around. Thus, whereas in Latin sentences like (1) below we can change the order of units such as *lupus*, *vulpem*, and *arguebat* and still produce meaningful sentences (Latin is particularly flexible with respect to word order), the same cannot be said of the meaningful parts of each word.

(1)	Lup-us	vulp-em	arguebat
	wolf-Subject	fox-Object	accuse+Past
	"The wolf was accusing the fox"		
(1)'	vulpem	lupus	arguebat
(1)"	arguebat	lupus	vulpem
(1)"'	lupus	arguebat	vulpem
(1)""	vulpem	arguebat	lupus
(1)""'	arguebat	vulpem	lupus

Thus, we cannot move the ending of lupus (*-us*) or the ending of vulpem (*-em*) to produce **uslup* or **emvulp*. Similarly, we cannot move the part of the verb (*-ebat*) that conveys the information about temporal relations.

Traditional orthographies are not always consistent in the ways in which they recognize words and analysts must develop their own understanding of the status of a particular morpheme or combination of morphemes. This is especially the case with categories like pronouns and tense or aspect markers, which are sometimes treated as separate words and other times as affixes and hence part of larger word units. This is the case, for instance, with so-called "clitic pronouns." They are typically unstressed, unemphatic, short morphemes that do the work of referring to participants in the immediate context (linguistic or otherwise). For these reasons, then, they do not seem to qualify as independent words. Orthographic traditions, however, vary. In written Bantu languages, for instance, clitic pronouns are typically treated as part of the verb. When the full nouns in (2), from

Haya (in Tanzania), are replaced by anaphoric pronouns, as in (3), they are shown to be part of what Bantu specialists call the "verb complex," a string of morphemes that includes subject-verb agreement, tense and aspect markers, causative or instrumental infixes, and other types of syntactic and semantic markers:[4]

(2) *Kat' á-ka-siig-is'* *ómwáán'* *ámajút'* *ékitambâla*
Kato 3sg-Pst-smear-Instr child oil handkerchief
"Kato smeared oil on the child with the handkerchief"

(3) *Kat' á- ka-ki-ga-mú-siig- isa*
Kato 3sg-Pst-pro$_{i}$-pro$_{ii}$-pro$_{iii}$-smear-Instr
"Kato smeared it on him with it" (i=handkerchief; ii=oil; iii=child)

In this case, then, no difference is made between the subject-verb agreement (*á-*) which must always appear on the verb – notice that it is present even though the Subject is expressed by a full noun (*Kato*) – and the pronouns referring to the various other nominal arguments (Object, Goal, and Instrument) in the sentence, which are there when the full nouns are not present.

The orthography of Romance languages, on the other hand, typically treats clitic pronouns as separate words in sentences with finite verbs and as suffixes in sentences with infinitival verbs. This is shown in the Italian examples (4) and (5) below, where the third person singular masculine pronoun *lo* is in one case a separate word and in the other a suffix to the verb *chiamare* "call":

(4) A; Sai dov'è Mario?
know:2ndsing where is Mario
"Do you know where Mario is?"
B; No, ma **lo** vedo domani.
no but him see:1stSing tomorrow
"No, but I see **him** tomorrow."

(5) A; dove posso trovare Mario?
"where can I find Mario?"
B; puoi chiamar**lo** a casa verso le tre.
can:2ndSg call-him at home around the three
"you can call **him** at home around three."

Should *lo* be thought of as a word? It depends. The pronoun *lo*, like other clitic pronouns (*mi*, *ti*, *la*, *li*, etc.) typically participates in the intonational unit of the verb

[4] In these examples the apostrophe indicates that the final vowel of a word is deleted, e.g. the "o" of the name *Kato*, when the next word starts with a vowel; the acute stress indicates a high tone, the circumflex a rising and falling tone, and the absence of stress a low tone (see Duranti and Byarushengo 1977: 63).

with which it co-occurs and does not carry primary stress (*lo-védo* and *chiamár-lo*). Furthermore, clitic pronouns can participate in assimilation processes that indicate a tendency to become part of larger units. Thus, as shown in (5) above, when the pronoun *lo* co-occurs with an infinitive form (*chiamar-lo*), the verb loses its final vowel (becoming *chiamar* instead of *chiamare*). Similarly, when the clitic pronoun precedes a word that starts with a vowel, it tends to lose its final vowel, e.g. *Mario lo imita* (lit. "Mario him imitates") –> *Mario l'imita*. These phenomena show that clitic pronouns can enter the structure of another word and it might make sense to think of them as part of larger word units. At the same time, if we take pausing as a criterion, things do not appear so straightforward. Italian speakers can stop after each of the words in a sentence like *lo vedo domani* (although, again, this ability might be due to writing practices). Furthermore, if there is ambiguity, the clitic pronoun *can* be stressed for emphasis (*la vedi? No, ló vedo* "do you see her? No, I see *him*").

The decision about whether an expression should be granted the status of "word" usually reflects how seriously a researcher has taken the task of analyzing a language and showing the relationships among its different parts. Decisions about word units become particularly important whenever linguists are involved in either revising or establishing orthographies (Romaine 1994; Schieffelin and Doucet 1994). In these cases, a consistent analysis might make a difference for the accessibility of the orthographic conventions to native speakers, children in particular. Furthermore, an understanding of what constitutes an individual word can enter the discussion of the nature of linguistic classifications, especially for anthropologists interested in the evolution of those classifications across time and space.

5.2.1 *The word as a unit of analysis in anthropological research*

The word as a unit of analysis has been particularly important in anthropological research. Key notions in anthropological theory such as the concepts of *potlatch*, *totem*, *mana*, *taboo*, and many others are actual words taken from a particular language and raised to symbols of universal or quasi-universal types of human activities, relationships with the supernatural, and individual or group characteristics. The most important part of traditional social anthropology, namely, the study of kinship systems, is based on the ability that humans have to use individual words to identify social relations among people. But kinship charts are just one well-known example of the interest that ethnographers have always had in native classifications. Lists of names for plants, animals, tools, and places have always formed an important part of fieldworkers' notebooks, reflecting the western view that the first step in knowing something is the ability to write down its name, hence the identification of individual words is crucial. This is demonstrated in an exemplary way by **evolutionary studies** of color terminology (Berlin and Kay 1969) and ethnobotanical nomenclature (Berlin 1992). In these cases, the extent to which the

names for different colors, animals, or plants are derived from the same word has been seen by evolutionists as evidence for how human groups might expand their vocabulary over time.

> In the ethnobiological lexicons of all languages, one is immediately struck by the structural uniformity of expressions that linguistically characterize man's recognition of the basic objective discontinuities of his biological world. These expressions are, for the most part, unique "single words" that can be said to be semantically unitary and linguistically distinct. Examples of such semantically unitary names in English folk biology would be oak, pine, and maple. Primary terms of this sort appear to represent the most commonly referred to concepts of the botanical world and can be referred to as "generic names." (Berlin 1975: 66)

Berlin argues that simple words naming generic classes are the first items in the ethnobotanical lexicon of all languages. The next stage consists of names produced by analogy (by means of expressions meaning "like" or "related to") and after that come processes such as the addition of modifiers (e.g. the adjective "true" or "genuine") with distinctions that are eventually lexicalized and lose the connection with the original generic name. In this kind of evolutionary study the word is the starting point as well as the goal of linguistic classification.

5.2.2 *The word in historical linguistics*

Another area of study that has been largely based on the word as unit of analysis is **historical linguistics**, that is, the study of how languages change over time, including the development of different languages from a common ancestor. The **comparative method**, a technique by which sound similarities and differences across languages are systematically examined and laws are proposed to explain those similarities and differences, started out as a way of matching lists of words. Despite the reluctance many linguists feel about centering their work on words, the comparative method has been very successful in historical reconstruction:

> Linguists have commonly been uneasy about relying on vocabulary. They consider vocabulary to be the least significant part of a language. It may be very unstable and vary widely from speaker to speaker and situation to situation. Phonology and grammar [=morphology, syntax] are more central. Yet there are certain crucial advantages of vocabulary over other sectors of language for comparative work. 1. Vocabulary items are relatively easily found and easily stated. 2. There can readily be obtained a sizeable sample of word-pairs (or glosses that will produce word-pairs) which come close to being independent of each other. [...] 3. Gloss lists can be selected in such a way as to bias our results in certain desirable ways. For example, word resemblances due to language universals are particularly common in a few specific meanings (e.g. child words for parents) and apparently negligible elsewhere. By eliminating such glosses, this source of resemblances can be minimized to the point of insignificance, and hence be safely overlooked in preliminary comparative work. (Gleason 1972: 4–5)

Starting with William Jones's 1784 lecture on the relationship between Sanskrit (an ancient language of India) and European languages and continuing with the work of the European historical linguists of the nineteenth century (Bopp, Rask, Schlegel), the comparison of word lists across languages has been used again and again not only to identify language groups (called "families") but also to reconstruct the origins of certain human groups or races.[5] The comparative method was used, for instance, to posit a southeast Asian origin of the Polynesian people before convincing archaeological evidence was available (Kirch 1984: 42). Examples of the relationship among different Austronesian languages is given in table 5.1, where groups of cognate terms in four modern languages (Tagalog, Malay, Fijian, and Samoan) are derived from the same reconstructed form in an ancient hypothetical language called "Proto-Austronesian." Figure 5.1 illustrates the relationship among some of the main subgroups of Proto-Austronesian.

Table 5.1 *Some proto-Austronesian terms and their related forms in four modern languages (Pawley 1974: 486)*

	Proto-Austronesian	Tagalog	Malay	Fijian	Samoan
two	*Duwa	dalawa	dua	rua	lua
four	*e(m)pat	apat	empat	vā	fā
five	*lima	lima	lima	lima	lima
six	*enem	anim	enam	ono	ono
bird	*manuk	manok	manu	manumanu	manu
eye	*mata	mata	mata	mata	mata
road	*Zalan	daan	jalan	sala	ala
pandanus	*panDan	pandan	pandan	vadra	fala
coconut	*niuR	niyog	nior	niu	niu

[5] See Irvine (1995) for a critique of the ideology implied in the use of the notion of "family" for genealogical classification, especially as it applies to nineteenth-century studies of African languages.

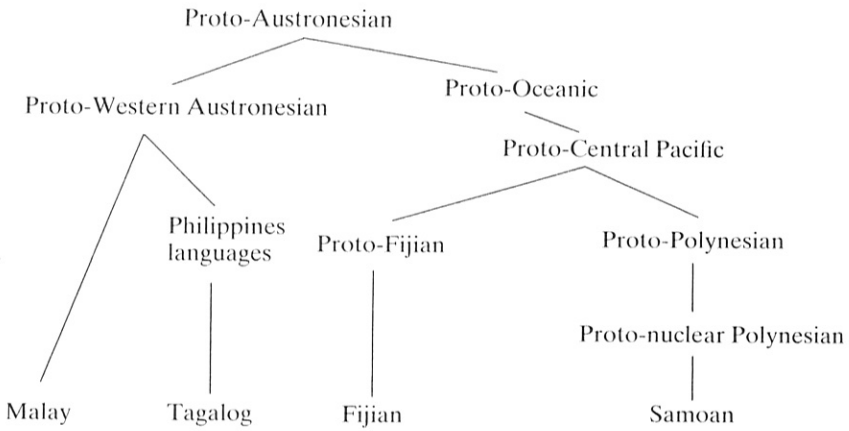


Figure 5.1 A tree-structure (also called "family tree") representing hypothesized relationship among four Austronesian languages (Pawley 1974)

The use of word lists from different languages is a powerful method for reconstructing relationships between different languages. The "family trees" produced from these comparisons, however, do not necessarily represent true historical states or events (Bynon 1977: 67–75). They also ignore variation within the same speech community (Weinreich, Labov, and Herzog 1968) and the possibility of language contact and spread of linguistic forms across linguistic family boundaries (Nichols and Peterson 1996; Trubetzkoy 1939; Weinrech 1953). Unfortunately, the assumption of uniformity and regularity necessary for this kind of reconstruction risks perpetrating the view of the meaning of a word as context-independent. As we shall see, in fact, any word acquires its meaning in the context of larger units such as sentences (chapter 6), speech acts and language games (chapter 7), sequences of turns (chapter 8), speech events and participant frameworks (chapter 9). Finally, what we call a word may in fact be expressing different kinds of "signs." Historical reconstructions tend to be based on one particular type of sign, namely, "symbols" (see section 6.8).

5.3 Beyond words

Despite the great advances in our understanding of linguistic structures and language change on the basis of the word as a unit of analysis, linguists and logicians have long recognized that words only have meaning within the context of a sentence (or a proposition). The "cognitive revolution" of the 1960s included an irreversible shift from the study of sounds and words to the study of full sentences. Especially due to the work of Noam Chomsky and his students, phonology and morphology were replaced by syntax as the most important area of research. At

the same time, researchers in a variety of fields interested in language processing and language use began to explore units larger than sentences. In the 1970s many students of language discovered that there were linguistic phenomena that should be studied in the context of discourse units rather than by looking exclusively at isolated sentences. A particularly active group of typological linguists mostly working in California rediscovered the earlier work on the informational structure of sentences by Prague School linguists and by M. A. K. Halliday and started to apply discourse notions such as "topic" and "theme" to the study of syntax (Givón 1979; Li 1974, 1976, 1978). There was also a renewed interest in language universals based on actual comparison among languages rather than on innate abstract principles (Greenberg 1963; Greenberg et al. 1978; Hawkins 1979; Keenan and Comrie 1977; Edward Keenan 1972, 1976).[6] This research inspired some linguists to look at texts of various sorts to establish the basic word order in a language and its relation to other syntactic and discourse phenomena. More or less at the same time, a group of sociologists soon to be known as "conversation analysts" became interested in the sequential aspects of conversational exchanges as the loci where the constitution of the social order could be studied without what they saw as the pitfalls of classical normative sociology, namely, without the *a priori* acceptance of such concepts as social role, social class, social situation (see chapter 8). Conversation and discourse analysts showed that, contrary to what was argued up to that point by formal grammarians working on isolated sentences, it was possible to engage in a systematic study of the language of conversational interaction.

Within psycholinguistics, language acquisition studies had often been based on discourse and interactional exchanges between children and adults, but this method was seen more as imposed by the circumstances (it was difficult if not impossible to do experiments on very young infants) than as a conscious and happy choice among researchers. In the 1970s, however, child language researchers also became influenced by conversation analysis and began to investigate new types of units, including certain types of interactional routines for grabbing the floor (e.g. *you know what?*, *look!* in English, *¡mira!* in Spanish), maintaining a topic, and building coherence (Ervin-Tripp 1973, Ervin-Tripp and Mitchell-Kernan 1977; Garvey 1984; McTear 1985; Ochs and Schieffelin 1983). It was in this new intellectual climate that discourse analysis gained a momentum and was established as a legitimate field of inquiry, giving birth to several symposia, collections of articles, and journals. Although linguistic anthropologists, given their interest in native texts, narrative, and performance, had been doing discourse analysis all along, they had done so mostly in a theoretical vacuum. This became an occasion

[6] For a more recent collection of articles on language typology, see Shibatani and Bynon (1995).

for some of them to reenter the mainstream of linguistic research without losing their identification with language as part of culture.

In later chapters, I will discuss some of these contributions to our understanding of language as a cultural practice. In what follows I will concentrate on the issue of how stretches of discourse can be collected and represented.

5.4 Standards of acceptability

Given linguistic anthropologists' interest in lengthy, spontaneously produced verbal exchanges, the issue arises as to how to collect samples of lengthy exchanges. A linguist interested in collecting grammatical forms, for example, might elicit **monologic discourse** – that is, a narrative performed by a native speaker for the linguist in front of a tape recorder – or **imagined interactional exchanges**. Although such imagined exchanges may be useful for certain types of linguistic analysis, they cannot be used as hard evidence of actual interactional strategies or, more generally, of patterns of language use. We have sufficient experience with spontaneous verbal interaction now to know that we cannot trust speakers' ability to *imagine* what they would say in a given situation nor can we expect them to *remember* exactly what they said in the past. Social scientists have shown again and again that memory is extremely selective and shaped by the future as much as by the past. Even the most skilled observers may miss or misinterpret what might turn out to be important properties of an exchange. In asking a native speaker to produce an imagined exchange, we are likely to get an **idealized exchange**, which might be accurate in some respects but unreliable in others.

In their analysis of different levels of respect in Nahuatl, Hill and Hill (1978) showed that reported conversations offer poor models of how speakers may show respect to one another. In one of the interviews they collected, a priest reporting what he had said to the president of his community kept alternating among level I ("intimate"), II ("distant"), and III ("honorable") in his use of address forms (pronouns, personal names, titles) and verbal agreement:

> This may in fact be what happens when an ordinary parishioner is transformed into a person of high status; but we have never observed this kind of random variation in actual conversations, which almost always display a very stable level of usage, with any deviation being fairly easy to explain on contextual grounds.
>
> (Hill and Hill 1978: 132)

There are also certain types of features of an interaction that simply cannot be reproduced during elicitation or in a report about an exchange in which the speaker participated. Examples of such phenomena are **pauses** and **overlaps**,

which are absent in idealized exchanges.[7] As has been convincingly shown by conversation analysts, pauses and overlaps are interactionally important phenomena (Sacks, Schegloff, and Jefferson 1974). Their exact place of occurrence and their relative length provide information that is used by participants to interpret what is going on, decide whether to produce further talk, and, if so, of what nature (see chapter 8). Equally important are the re-starts, cut-offs, and other corrections that speakers make of their own talk. A transcript that reflects so-called **false starts** and other apparent "mistakes" may reveal a recurrent organization of such phenomena that is relevant for participants' understanding of each other's past actions and their expectations for future actions (Goodwin 1981). Thus, for instance, Schegloff (1979b) has shown that people regularly correct themselves (he uses the term **self-repair** for this phenomenon) when they introduce a new topic in conversation. The position of the **corrected** (or **repaired**) item is typically at the word that keys the **new topic**. Here are two examples:[8]

(6) B: That's too bad ((very quiet))
A: hhh!
(0.5)
-> B: (I `unno) `hh Hey do **you** see V- (0.3) fat ol' Vivian anymouh?
A: No, hardly, en if we do:, y'know, I jus' say hello quick'n `hh
y'know jus' pass each othuh in the [e hall(way).] still handing
B: [Is she]
aroun' (with) Bo:nny?

(7) B: `hh But it's not too bad, `hh
A: That's goo[d, ((very quiet))
-> B: [Diyuh have any-cl- **Y**ou have a class with Billy this te:rm?
A: Yeh he's in my abnormal class.
B: mn Yeh [how-]
A: [Abnor]mal psy-ch

The recording of actual conversations and the careful transcription of the utterances exchanged during such conversations reveal regularities that deserve analytical attention. How are corrections organized? Why are they recurrent in certain types of discourse environments? How common are they crossculturally? These

[7] For example, overlaps cannot occur when a dialogue is performed by one speaker.

[8] The relevant segments, with the new topic being introduced are the following (in simplified form):
(6)' B: Hey did you see V- (0.3) fat ol' Vivian anymore?
(7)' B: Did you have any cl- You have a class with Billy this term?

same questions can be asked about other interactional phenomena such as overlaps and silences (see chapter 8). Their theoretical importance in understanding human interaction across social contexts cannot be evaluated without a good documentation of their occurrence.

The work of conversation analysts with audio and video recordings of ordinary English conversational interactions in the last three decades has introduced new **standards of acceptability** not so much for what a transcript should look like – their conventions are certainly not immune to criticism (see below) – but for the kind of evidence that researchers need to substantiate a claim about patterns of language use. Studies exclusively based on recollections or occasional observations of speech patterns seem no longer acceptable. New standards must apply to old studies as well. Scholars who quote earlier studies should review the research methods used by the authors they cite to establish whether the evidence presented in the past would hold up in the present. Unfortunately, not everyone who writes about conversational exchanges is careful about reviewing the research methods used by the authors they quote. A careful examination of many of the now classic studies of face-to-face communication published in the 1970s will reveal that for some time, linguistic anthropologists, like many of their colleagues in other branches of anthropology and linguistics, did not feel obliged to give information on how they collected the data discussed in their published work (or perhaps editors and publishers found this information trivial and not worth occupying printed pages). Even in those few cases in which authors openly discuss their methods, such discussions were not adequately attended to by readers and colleagues. For instance, despite the fact that in her influential study of greetings among the Wolof, Judith Irvine states at the very beginning of the article that she was unable to record greetings on tape,[9] most of the colleagues I talked to about the article believed that Irvine had audio tapes of the greetings. They had simply *assumed* that she did.

Without implying that one should throw away several decades of observations and speculations about face-to-face encounters that did not have the luxury of magnetic or electronic recording, it is imperative that new generations of researchers learn to read past contributions in the light of current standards of acceptablity. Particular attention should be given to passages in which fieldworkers describe the conditions under which the study was conducted. When such information is not available, one should try to contact the writer. If this is not possible, extreme care should be exercised in generalizing from descriptions

[9] "It proved difficult to persuade informants to act out hypothetical greeting situations for my benefit, and difficult ever to record greetings on tape (hence the 'typical' greeting I have illustrated on p. 171 is a construct of my own experience rather than a recorded text)" (Irvine 1974: 168).

that are not accompanied by a discussion of the methods used in collecting the information presented in the article.

5.5 Transcription formats and conventions

I will hereafter use the term *transcription* for the process of **inscribing** social action and *transcript* for the finished, although by no means definitive, product of such a process. Following Ricoeur's formulation (1971), I will consider inscribing a process whereby some of the characteristics of an action in real time and space (e.g. something someone said) are fixed into a record that will outlast the fleeting moment of real-life performance.

> In living speech, the instance of discourse has the character of a fleeting event. The event appears and disappears. This is why there is a problem of fixation of inscription. What we want to fix is what disappears. (Ricoeur [1971]1981: 198)

Although transcription has been largely used for fixing vocal sounds into graphic representations, in a transcript there are no *a priori* reasons to favor speech over other forms of communication. As I will show later, the more we learn about representing other aspects of communicative behavior, the more we realize how important it is to develop ways of integrating the analysis of speech with other codes and other modes of communication.

Any kind of inscription is, by definition, *an abstraction in which a complex phenomenon is reduced to some of its constitutive features and transformed for the purpose of further analysis.*[10] This applies to alphabets as much as to photographs, X-rays, or any kind of measurement. What changes in each medium is not only the instrument used but also the relationship between the form of representation (writing, black-and-white images, numbers on a scale) and the phenomenon that is being represented through the inscription technology. Thus, when we write on a piece of paper a phrase that someone just said, we are creating a record of his live action of speaking (for a purpose and for an audience) exclusively as a linguistic token, which can be later examined and compared to other similar linguistic tokens either in the same or in different codes. In so doing, we are performing two important analytical operations involving different levels of abstractions:

a) *Selection*. We are concentrating only on a very small subset of the actions the speaker performed. Thus, we are leaving out other aspects of what the speaker was doing, for instance, with his body (eyes, mouth, hands, etc.). We are also ignoring the prior, simultaneous, or subsequent acts that he and the other

[10] The reason for speaking of "*further* analysis" is that transcription is itself a form of analysis.

participants in the scene performed, including further talk that might be relevant to the one segment we decided to make a record of.

b) *Simplification*. We are simplifying the speaker's performance by ignoring certain features of his speech and presenting an abstraction of it that is theoretically informed (some like to call it "biased"). Thus, when we look at an utterance as represented by a spectrogram, we realize that sounds are not as separate as they appear when we write them down. Typically, in casual conversation, there are no spaces (or pauses) between most words that form the same utterance. Linguists have thus been relying on intonation as one of the indicators of discourse units – Chafe (1987), for instance, has introduced the term **intonational units**. Furthermore, the features of what we consider one sound may spread over several sounds, making it difficult to say where one sound starts and where it ends.

The issue here, as always in representation, is the relevance of the information we decide to reproduce on a piece of paper or on a computer disk for a particular purpose. As Ochs (1979) reminds us, the choices we make in preparing a transcript are always influenced by theoretical as well as pragmatic considerations – e.g. readability (see my rendition of Schegloff's example in footnote 8 above). In addition to the goals of the research agenda – a transcript should carefully represent what is of theoretical interest to the author –, there are what we might call aesthetic considerations. A transcript should not have too much information, otherwise it becomes too unpleasant to read and defies one of its purposes, namely, being accessible to others (Ochs 1979: 44–45). A transcript should be inviting, that is, it should make readers feel like they want to read it. Visual display and conventionality have, for this reason, an important part in transcription. Transcripts done according to conventions that are unknown to most people or seem unintuitive do not look appealing and readers are more likely to skip them. This possibility always lingers over the choice between conventional orthography and phonetic symbols. The advantage of using conventional orthography is that it is accessible to a much larger audience. The problem with it is that it comes with a set of prescriptive assumptions about what a language should be like and makes it difficult to represent how it is actually spoken. If one looks at a transcript like the one given in (8) below, it is difficult to imagine what the speaker sounded like, but it is very readable given that there are only a few extra conventions that a reader must learn, mostly about pauses (between square brackets or with two periods) and lengthened sounds (the symbol "–"):

(8) Okay. The movie seemed very [.25] sound oriented [.4] Even though there weren't [.6] there was no dialogue. [3.5] [1.5] A–nd [1.3] the first [.75] thing I noticed ... was ... the sound of the man picking ... pears. (Chafe 1980: 304)

There is, however, a major problem in using standard orthography, namely, that it serves best speakers of the standard dialect – which is after all the variety that the writing system is designed to represent. Speakers of other varieties are implicitly characterized as deviant, proportionally to the number of modifications necessary to represent their speech. Thus, the most used convention in the next transcript – from an interview with a Black teenager – is the apostrophe (') to signal that a sound expected in Standard English is missing.

(9) LARRY: You know, like some people say if you're good an' shit, your spirit goin' t'heaven ... 'n' if you bad, your spirit goin' to hell. Well, bullshit! Your spirit goin' to hell anyway, good or bad.
INTERVIEWER: Why?
LARRY: Why? I'll tell you why. 'Cause, you see, doesn' nobody really know that it's a God, y'know, 'cause I mean I have seen black gods, pink gods, white gods, all color gods, and don't nobody know it's really a God. An' when they be sayin' if you good, you goin' t'heaven, that's bullshit, 'cause you ain't goin' to no heaven, 'cause it ain't no heaven for you to go to. (Labov 1972c: 194)

Given the potential implications of using modified standard orthography, sociolinguists like William Labov who work on non-standard dialects must constantly stress that what they are transcribing is just *another language* and not an impoverished one. After the passage reproduced here in (9), Labov (1972c: 194) goes on to write: "Larry is a paradigmatic speaker of nonstandard Negro English (NNE) as opposed to standard English (SE)." Needless to say, the issue of what a given dialect or language is called can be highly controversial. This can be quite an issue in some cases, for instance, in Native America, where people might insist on a nomenclature that is not attested in previous literature (Jane Hill, personal communication).

Another problem with standard orthography is that it does not do justice to certain paralinguistic phenomena, including **sound play**,[11] and in this way precludes the possibility of generalizations on such phenomena (Ochs 1979: 45).

Alphabets that have been developed by phoneticians have the advantage of building on traditional orthography but favoring actual pronunciation. In

[11] "The use of standard orthography is based on the assumption that utterances are pieces of information, and this, in turn, assumes that language is used to express ideas. In sound play, the shape rather than the content of utterances is foregrounded and the function of language is playful and phatic ... rather than informative" (Ochs 1979: 45).

principle, they do not come with preconceptions about which particular pronunciation is the standard or unmarked one. One such alphabet is the one developed by the International Phonetic Association (IPA), which has enough symbols to systematically cover the total range of linguistic sounds found in natural languages (Pullum and Ladusaw 1986). Anyone who is familiar with the IPA symbols should be able to read them, without having to know anything about the language that is being transcribed.[12] Computer technology with different fonts available on the same screen has made it easier to have access to such alphabets, but their use is still limited to people who have had an extensive training in phonetics or linguistics. As shown in example (10), the knowledge of the Latin alphabet or English orthography is not sufficient (although it helps!) to guess what the symbols represent (Ladefoged 1975: 161).

(10) æplsləmənsəntʃɛriz

Once we are told that (10) represents something that in English orthography would be written *apples, lemons, and cherries*, things start to get a bit clearer. If we try to make it easier by separating the phonetic symbols into "words," we encounter a classic problem of transcription, namely, the need to make decisions that seem arbitrary at first. In this case, for instance, it is difficult to decide on pre-theoretical grounds where to break the sequence əntʃɛriz "and cherries," given that the sound "t" in a sense belongs to both "and" and "cherries" – one could either say that the "d" of *and* has become "t" for assimilation to the next sound (tʃ) or say that it just disappeared. The final choice should be determined by our phonological theory, that is, the kind of phonological processes we consider common in languages in general and in one language in particular.

(10)' æpls ləmənz ən tʃɛriz
apples lemons and cherries

To avoid some of these problems, most people working on spontaneous interaction end up adopting the compromise of adapting traditional orthographies to their descriptive and theoretical needs. There are, however, different ways of doing this, from very conservative to experimental. For instance, in conversation analysis, English orthographic conventions are adapted to reproduce some stylistic and vernacular properties of the participants' speech:

(11) KEN: Hey yuh took my chair by the way an' I don't think that was very nice.

[12] This does not mean that from the symbols alone a reader would be able to sound like a native speaker or like the person whose speech is being represented. There are still considerable limitations to the amount of information that can be encoded on alphabetic representations of sounds.

AL: *I* didn' take yer chair, it's *my* chair.
(from Sacks, Schegloff, and Jefferson 1978: 28)

The problem encountered in (10) with the word *and* is resolved in (11), just like in (9), by writing *an'*, a convention that most speakers (and readers) of English are likely to understand. In some cases, however, the adaptations of English orthography are harder to interpret for those who are not practitioners of this method. Thus, the forms *yuh* and *yer*, two forms that are often used by conversation analysts, represent fast vernacular pronunciations of "you" and "your" respectively that are not so obvious to most readers of English. Things get more obscure with words like *does* and *was*, which are often transcribed by conversation analysts as *dz* and *wz* respectively. In this case, readers must guess that the letters "d" and "z" have syllabic value [d̩z̩] otherwise the sequence would be interpreted as a voiced alveolar affricate (e.g. the first sound of the Italian word *zebra* [dzebra] or in the last sound of the English *lads* [lædz]). Although a native English speaker may at times guess what conversation analysts' choices mean, the lack of internationally available conventions make these renditions much less accessible to readers who are non-native speakers of English. In transcripts like the one in (12) below, the adaptation of English orthography to phonetic transcription is carried to an extreme by marking as special pronunciations those that are quite predictable, such as *iz* for *is* or *he'z* for *he's*. Since the "s" of *is* is typically pronounced voiced ([z]) – see chapter 6 – by native speakers, it is hard to understand the reason for changing its ordinary spelling. The question here is whether a phonetic feature that is predictable from general rules of the orthography should be marked (cf. Edwards and Lampert 1993; Macaulay 1991a; 1991b: 24).

(12) F: 'hh how iz our fri::end
N: Oh: he'z much better I'm, 'fraid –
[hh h h h
F: [Well uh that's *mar*verlous (Pomerantz 1984: 96)

Although for most readers these transcripts are still more accessible than those in IPA format, they require familiarity with their mostly implicit conventions.[13] Usually they turn out to be excellent mnemonic devices for those who have listened to the transcripted interaction a sufficient number of times to imitate them, but they are baffling to everyone else.

One of the issues in this as well as in other transcription systems is the

[13] For a list of conventions used by conversation analysts, see Atkinson and Heritage (1984:ix–xvi), M. H. Goodwin (1990:25–6). These lists do not provide hints on how to read the phonetic conventions.

audience for whom the transcript is produced (Haugen 1980; Macaulay 1991b: 24). Since a transcript is going to be quite different depending on who is seen as its primary audience, we must make conscious and consistent choices. This does not mean that once we opt for one system we cannot change our mind later on. What is important is to follow a criterion that is consistent with our priorities and that can be understood by our readers. Thus, if we are concerned with the ability of native speakers and other people who know the language (especially other social scientists who do not have a linguistic training) to read our transcripts, we might opt for adapting standard orthography to our needs. At the same time, we must be aware that the choice of standard orthography may also cut out some readers or misguide them. This is particularly so for languages that are not likely to be known by most of our readers. My decision, for instance, to follow Samoan orthography in my transcripts and use the letter "g" for the velar nasal ŋ, which is otherwise written "ng" in most orthographies, has meant that almost no one among my colleagues or students remembers that the word written *lāuga* ("ceremonial speech") is pronounced [la:uŋa] – everyone keeps saying [lauga]. The fact that I always have a footnote or paragraph explaining Samoan orthographic conventions in my publications does not seem sufficient, even with linguistically sophisticated readers. Rather than blaming my readers, I should probably rethink the ways I have been trying to communicate with them. I mention this piece of personal history to stress the fact that *the process of transcribing implies a process of socialization of our readers to particular transcribing needs and conventions*. We must decide what is important for us to communicate in our transcripts and devise effective strategies to such ends. For this reason, a transcript that is devised for personal use only will be different from one that we plan to present at a conference or publish. In publishing a transcript, we might need to amplify a certain type of information while simplifying in other areas. The ephemeral character of any version of a transcript is made more apparent in those cases in which researchers move over time to investigate different aspects or different levels of the same exchange. We might then get not only different versions of the same transcripts in different subsequent publications, but also different versions within the same article. This is for instance the case in Goodwin and Goodwin's (1992a) discussion of assessments, where the different layers of interactional complexity of a brief exchange are made evident through slight modifications of the same transcript. I will reproduce here the first four versions (there are a total of eight in the article):

(13) (Version I, Goodwin and Goodwin 1992a: 161)
DIANNE: **Jeff** made en asparagus pie
it wz s::**so: goo**:d.

(13)' (Version II, ibid., p. 163)
DIANNE: **Jeff** made en asparagus pie
it wz s::**so**[: **goo**:d.
–> CLACIA: [I love it.

(13)" (Version III, ibid., p. 166)
DIANNE: **Jeff** made en asparagus pie
it wz s::**so**[: **goo**:d.
CLACIA: [I love it.
((nod nod))

(13)"' (Version IV, ibid., p. 168)
((lowers upper trunk)) *((nod with eyebrow flash))*
DIANNE: it waz s : : **so**[: **goo**:d.
CLACIA: [I love it.
((nod nod))

Although this technique would not be practical for transcripts that cover several minutes or hours of conversational interaction,[14] it does offer a powerful representation of the analytical process the researchers went through while examining different aspects of the information made available to them in the recording (in this case a video tape).

For transcripts of long stretches of interaction, gestures can be incorporated by extending the use of the bracket originally introduced by conversation analysts for overlapping talk. Such a technique has been used by Ochs, Jacoby, and Gonzales (1994) in transcripts such as the following:

(14) STUDENT: [And let me tell you (0.2) there's something (.)
[((*moves toward board; adjusts glasses*))
mo:re I can say: mtsk is [that that (0.2) those gu-
[((*points to j*))

[14] At least not in traditional print. It becomes more feasible with the help of computer technology.

that dynamics starts (0.5) not at the moment you
[reach this point (0.5) [but [at the moment
[((*points to b, looks at PI*)) [((*looks at board*))
[((*points to a*))
(Ochs, Jacoby, and Gonzales 1994: 153)

As Ochs (1979) points out, the visual display of a transcript has important implications and consequences for the way in which readers will process the information and assess the importance of different elements.

The traditional bias in favor of speech and against non-verbal behavior – reflected in the term itself with its negative definition (non-verbal is anything that is *not* verbal) – is something that has become more and more apparent with the increased use of video technology and the richness of the audio-visual display. Researchers are learning to integrate in their representations information available to the interactants but earlier on only grossly recorded in their fieldnotes.

5.6 Visual representations other than writing

Although in face-to-face encounters talk often dominates interaction, a transcript that only shows what people have been saying may leave out some important aspects of what was happening at the time among the participants. However, the kinds of transcripts I have been discussing so far were designed to represent speech and not other forms of communication or social action. Anyone who has tried to represent on a page what people actually do in a stretch of face-to-face interaction knows that traditional orthography is indeed a very poor medium for representing visual communication, not to mention the physical surroundings of the interaction. Verbal descriptions of what people do rarely capture the meaningful subtleties of human action. Furthermore, by transforming non-talk into talk, verbal descriptions reproduce the dominance of speech over other forms of human expression before giving us a chance to assess how non-linguistic elements of the context participate in their own, unique ways, to the constitution of the activity under examination. In many cases, it is still true that a picture is worth a thousand words. Students' reactions to slides and footage of a landscape or social event often reveal how misled they had been by printed words. For instance, there is a big difference between describing what the outside or the inside of a house looks like and seeing an image of it. In some cases, previous ideas about what an event might look like prevent readers from accurately processing what an author might have written. Until they saw a video tape of a Samoan *fono*, some of my students believed that the chiefs would be standing around during such a meeting. To see everyone seated along the periphery of the house was a shock to them.

Several methods have been used by social scientists over the years to visually enhance the printed rendition of fleeting moments of interaction. Each method is grounded in a different tradition and reveals different theoretical interests. I will here briefly concentrate on two traditions: the representation of gestures and the representation of participants' visual access to each other and to their surrounding environment.

5.6.1 *Representations of gestures*

Actio quasi sermo corporis.
Cicero, *De oratore* 3, 222[15]

At least since Darwin's interest in human gestures as a source of insights into human evolution (Darwin 1965), anthropologists, human ethologists, and other social scientists have been fascinated with the issue of the universality vs. cultural relativity of gestures and expressions (Bremmer and Roodenburg 1992; Eibl-Eibesfeldt 1970; Polhemus 1978). Anthropologists have been drawn to this discussion for a number of reasons, including the need to provide an accurate description of communicative events.

Sociocultural and linguistic anthropologists have long been aware of the need to complement traditional ethnographic accounts based on naked-eye observation with more precise and detailed descriptions based on more reliable forms of documentation. Gregory Bateson, for instance, in his "Epilogue 1936" to *Naven* – an ethnography of the Iatmul people of New Guinea that has since become a classic of social anthropology – regretted that he had been forced to use vague and inadequate descriptions of the expressive behavior or "tone," as he called, of social actors: "Until we devise techniques for the proper recording and analysis of human posture, gesture, intonation, laughter, etc. we shall have to be content with journalistic sketches of the 'tone' of behaviour" (Bateson 1958: 276).

Thanks to the work of visual anthropologists, ethnographic filmmakers, ethologists, and visually oriented linguistic anthropologists, the recording and analysis of human gestures have lately become more and more common in anthropological studies.

It is now universally accepted that in face-to-face interaction what humans say to one another must be understood vis-à-vis what they do with their body and where they are located in space (e.g. Birdwhistell 1970; Farnell 1995; Goodwin 1984; Goodwin and Goodwin 1992a, 1992b; Hall 1959, 1966; Kendon 1973, 1977, 1990, 1993; Kendon, Harris and Key 1975; Leach 1972; Schegloff 1984; Streeck 1988, 1993, 1994; Streeck and Hartge 1992). This means that one of the greatest challenges in representing gestures is not just to reproduce a particular posture

[15] "Delivery (is), in a way, the language of the body" (see Graf 1992: 53).

or movement, which can be done with a series of drawings, but how to visually maintain on a page the connection with co-occurring talk. The recurrent interpenetration of verbal and visual communication in everyday interaction has been at the center of some of the work recently done by linguistic anthropologists working with audio-visual records.

In an attempt to extend the boundaries of conversation analysis beyond verbal communication, Goodwin (1979, 1981) introduced a series of conventions that were explicitly designed to integrate information on eye-gaze patterns with sequences of turns at talk. In the following segment, for instance, Goodwin (1979, 1981: 131–3) tries to visually capture the relationship between the reshaping of an utterance as the speaker's eye gaze moves from one participant to another.

(15) (Goodwin 1979, 1981)

JOHN: . . . . . Don , ... Don ____
[[
I gave, I gave u p smo king ci garettes::,
[
DON: . . . . . X _____

DON: =Yeah
(0.4)
JOHN: . . . . . Beth _____ , ... Ann
[[
I-uh: one-*one* week ag o to*da:* y. actua lly,
BETH:
ANN: . . . . . Beth ___ , ... John

In this system, the gaze of the speaker is marked above the utterance and the recipient(s) below it. Dots mark movement of one party's gaze from one participant to another. A solid line indicates gaze by one party toward the other. Commas indicate the withdrawing of gaze. By means of these conventions, Goodwin is able to show how the utterance produced by John (*I gave up smoking cigarettes one week ago today actually*) is shaped by (a) whether the selected recipient makes eye contact with the speaker (John changes the utterance in moving from one recipient to the next and finally adds an adverb *actually* which allows enough time for Ann to gaze back at him) and (b) the extent to which and the manner in which the recipient knows the event reported by the speaker (Beth is John's wife and already knows about John's attempt to give up smoking, hence his attempt to make the announcement into an anniversary by saying *one week ago*).

In his comparative study of the symbolic structuration of space and movement by members of two different speech communities, Haviland (1996) uses a combination of transcription, verbal description of gestures, and figures to illustrate how, in telling a story, Guugu-Yimithirr speakers keep track of cardinal points – this ability and practice make their orientation system more "absolute" than "relative" (Haviland 1996: 285).

. . . . . ! . . . . . .
mathi *past-manaathi*
rain + ABS past-become-Past
"The rain had passed over."
right hand: palm out, pulled towards E then
push out W, slight drop.

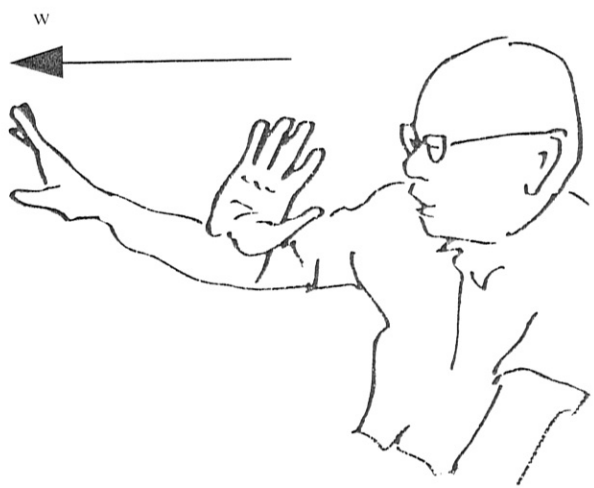


Figure 5.2 Text and picture of storytelling episode (I)
(Haviland 1996: 310)

. . . ! . . . . . .

and yuwalin *nguumbaarr* *guthiirra* *nhaathi*
beach-LOC shadow + ABS two + ABS see-PAST
gadaariga
come + RED-PAST-SUB

"and (he) could see two shadows coming along the beach."
right-hand: pointing with straight arm W,
moving S to rapid drop to lap.

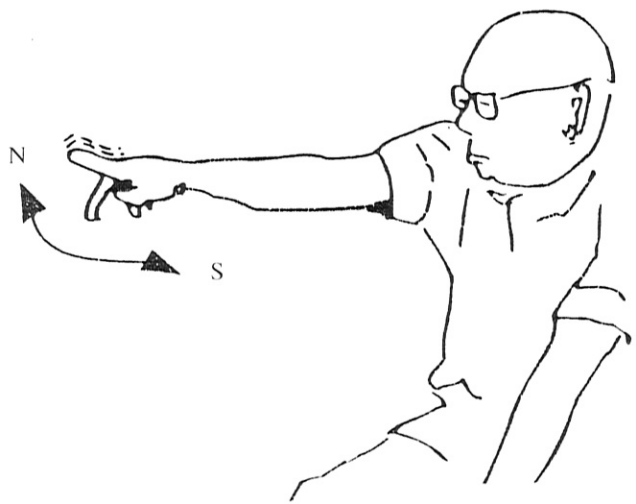
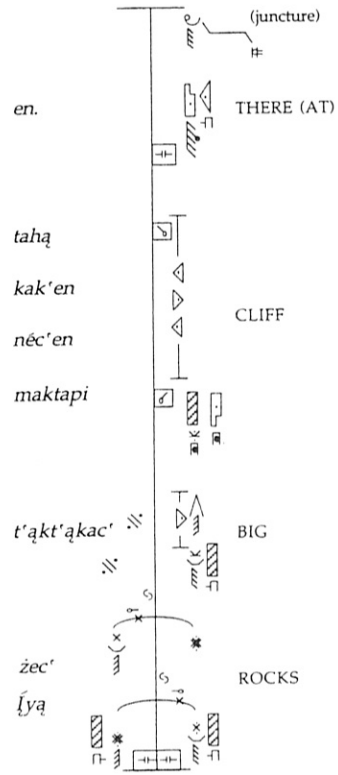


Figure 5.3 Text and picture of storytelling (II) (Haviland 1996: 311)

In these and other cases, linguistic anthropologists have been particularly interested in the unique ways in which gestures that accompany or replace talk contribute to the flow of interaction and rely on the participants' shared knowledge. In her study of Plains Indian Sign Talk and other gestures that are an integral part of Nakota (or Assiniboine) narratives, for example, Brenda Farnell characterized the use of the lips in place of the pointing index – a gesture that is common among many Native American communities (e.g. Sherzer 1973) – as a gesture that provides participants in an exchange with a sense of intimacy and shared history:

> The performative value of this gesture lies in its potential for discretion as a smaller and less obvious gesture, often serving to preserve a degree of intimacy between speaker and addressee that would be lost if a finger-pointing gesture or speech were used instead. (Farnell 1995: 158)

To capture the complex and yet systematic relation between speech, gestures, and space, Farnell uses the Laban script (or Labanotation), a complex system of symbols invented by Rudolph Laban (1956) to describe dance movements. This system of transcription allows Farnell to match words (on the left column) with actions on the right column.



Įyą *żec'*, *t'ąkt'ąkac' maktapi* *néc'en* *kak'en tahą*, *en.*
Rocks there big cliffs [cut edge] this over there at
"There are large rocks that form a cliff, over there."

Figure 5.4 Transcription in the Laban script of Plains Indian Sign talk (Farnell 1995: 94)

Another transcription system for body motion and for prosodic and paralinguistic aspects of talk was devised by Birdwhistell (1970), a pioneer in kinesics, the study of how humans use their body for communicating. These graphic conventions are particularly valuable to the analysts for seeing patterns in their data but remain difficult to decode for the reader without intense training and practice.

As frequently lamented by those who work on gestures, the relatively little attention that gestures have had compared to speech in the study of human communica-

tion continues a tradition of research that identifies the basic structure of communicative acts with grammatical units. This is only partly due to technological limitations or to the recognized centrality of speech in human societies. It is also a consequence of an ideology of communicative events that takes writing (and hence texts) as the highest form of human communication and iconic representations as less sophisticated (Farnell 1995: ch. 2). Writing (especially alphabetic writing), however, is more adequate for the structural analysis of segmentable sound sequences (see section 5.1) than for other forms of communication, especially gestures.

5.6.2 Representations of spatial organization and participants' visual access

Video and computer technologies are rapidly making the task of analyzing and documenting the interplay of speech and gesture much easier. For example, it is possible now to represent the spatial organization of an interaction and the participants' visual access to one another by transfering to the page (or the computer screen) a video image. This can be done by digitizing a frame taken from a video tape. Figures 5.6 and 5.7, for instance, show very different forms of participation in the same narrative event. In 5.6, the man on the left (M) is a **peripheral participant** – a term I am borrowing from Lave and Wenger (1991) –, who listens to the story being told by the woman at the table (R) but is not directly involved in the narration. In 5.7, on the other hand, we can see R the narrator (on the left, smiling) directly address and get sympathetic response from the woman on the far right, D, whom she identifies as her primary recipient by her gaze and body position (she is facing and addressing her among the various participants).

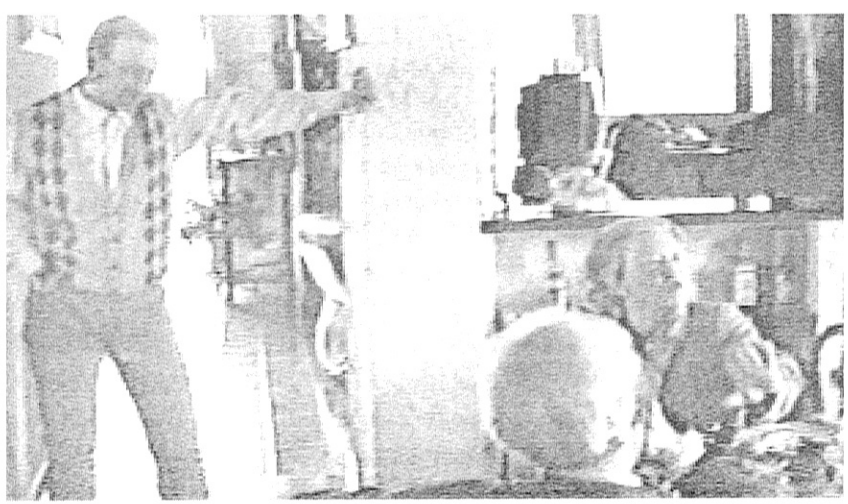


Figure 5.5. The man standing on the left (M) listens to R's story as a peripheral participant

Figure 5.6. A funny line gets a laugh from the story's primary recipient (D)

5.6.3 Integrating text, drawings, and images

Despite their power to communicate the feeling of a fleeting moment in ways that an audio recording or linguistic transcript could never do, images like the two above still do not have in them much of the information that is available to the participants and that researchers might find relevant to their analysis. For instance, camcorders do not record people's names or social relations unless the participants themselves refer to them in their talk. They also do not show a 360° view of the setting and where everyone in the scene is located with respect to one another. In addition to frames like the two above, then, it can be useful at times to display for the reader a diagram with some of the information that is not available in the video or on the sound track. Figure 5.7 shows an example of how a computer graphic program can be used to represent the seating arrangement of the participants around the table and the kinship relations among them (for a similar technique, see Goodwin and Goodwin 1992b).

When we match the information in figure 5.7 with the visual record and the transcript of what is being said – a narrative about R's first encounter with her mother-in-law some thirty something years earlier –, we are in a better position to make sense of the organization of this event. I, for instance, is often integrating R's story with comments and clarifications, which he addresses to P, right across from him. He also anticipates some of what R is about to say. Once we put together the information that I and R are husband and wife with the theme of R's narrative, for instance, we can better understand the ways in which I participates in the event. He is the only one in the scene who had independent access to the events and characters R is telling about in her story. In some moments, in fact, he is one of the characters of R's narrative. These features warrant his participation as an ideal *co-narrator* but not as a *primary recipient* – he already

knows the story (see chapter 9).[16] We can also see some differences between D and P, on the one hand, and D and M on the other that might help us make sense of R's choice of D as her primary recipient. D is R's only female affine in the scene. This means first of all that she is less likely than P to have already heard R's story. Furthermore, her structural position is similar to R's position in her own story. D is a young woman who married into R's family. It might thus be easier for her to identify with R's position or perhaps better appreciate R's reactions to the treatment she received from her mother-in-law.

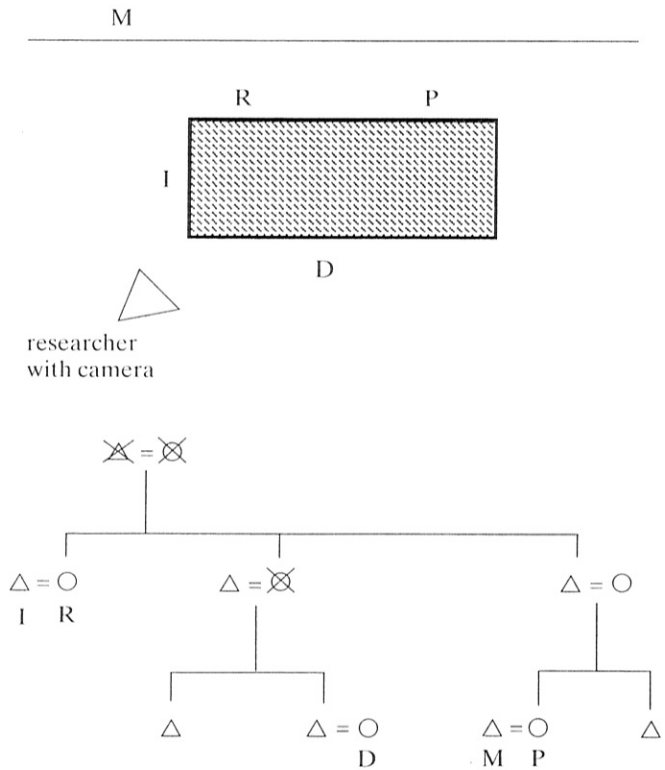


Figure 5.7 Spatial arrangement and kinship relations among participants in event shown in figures 5.5 and 5.6 above.

In performing this kind of analysis it is important to remember that whether or not certain facts about the participants will be relevant to whatever we have to say about what they say is not something that can be decided *a priori*. We cannot say once and for all that kinship is *always* important in social interaction or that

[16] This does not mean that people do not tell each other stories they already know, because they do at times, but that when this happens the story is framed differently.

gender is (see section 8.3.2). There are moments when kinship (or some other social attribute such as gender, social class, ethnicity, profession) might not be relevant to what is going on. The relevance of different attributes or background knowledge participants have of one another is an empirical question that must be addressed on each occasion. At the same time, it should be obvious that having access to such background knowledge about the social actors in a scene opens up for the analyst a wealth of interesting questions which allow for more complex (or, in Geertz's terms, "thicker") accounts (see section 2.3.2). For one thing, new hypotheses are possible that could not be thought of before. This is indeed the strength of linguistic anthropology with respect to other approaches to the analysis of discourse: its commitment to finding appropriate ways of integrating the information on a transcript with other kinds of knowledge that is being shared or accessed by the participants.

Of course, when more ethnographic information is revealed about a particular context, we grow hungrier, that is, we want more. We are thus faced with the fact that unveiling cultural knowledge about the participants becomes a potentially endless process. Geertz was hinting at this aspect of doing ethnography when he told the by now famous "turtles" story:

> There is an Indian story – at least I heard it as an Indian story – about an Englishman who, having been told that the world rested on a platform which rested on the back of an elephant which rested in turn on the back of a turtle, asked (perhaps he was an ethnographer; it is the way they behave), what did the turtle rest on? Another turtle. And that turtle? "Ah, Sahib, after that it is turtles all the way down."
>
> Such, indeed, is the condition of things. ... Cultural analysis is intrinsically incomplete. (Geertz 1973: 28–9)

This property of cultural analysis is seen by some as discouraging. If we can never get near to the bottom of things, what kind of science are we engaged in? But this is precisely the distinctive feature of human life, namely, that there is a potentially infinite number of layers of meaning in what we do. In fact, if our science is to look into such layers, our science is included in what we study – it is intrinsically reflexive (Luhmann 1981) – and therefore our science is just as infinite as the object of our study. The issue is not how to avoid getting into the potentially infinite layers, but how to find order in them, sometimes an order that is similar to the one proposed by the participants themselves, some other times a different order, that would be alien or even appalling to them. Once more, the difference among different approaches to social interaction lies in the ways in which each discipline moves along the different interpretive paths. Rather than

restricting their analytic boundaries to specific forms (as grammarians and conversation analysts often do) or to specific contents (as psychologists often do), linguistic anthropologists are interested in exploring ways of integrating information made available through a *variety of interpretive procedures*, including traditional participant-observation and fieldnotes, drawings, digitized images, transcripts with translations, and kinship charts.

5.7 Translation

> Let me start with the apparently paradoxical and yet perfectly plain and absolutely true proposition that the words of one language are never translatable into another. (Malinowski 1935, vol. 2: 11)

Most linguistic anthropologists work on languages other than their native language and must present what is recorded on a tape to an audience who is not likely to know the language spoken by the participants in the interaction. This means that for many linguistic anthropologists an important part of preparing a transcript consists of **translating**. This activity involves more than going from one language to another. It implies a long series of interpretations and decisions that are rarely made apparent in the final product, which might just look like another line of text. In fact, as Malinowski theorized a long time ago (1923), translation assumes an ability to match words with the context in which they were uttered. It is an activity that for anthropologists is intimately linked to ethnography. It implies an understanding not only of the immediate context but also of more general assumptions, such as a people's worldview, including their ways of relating the use of language with social action. If we conceive of translation as the mere exercise of matching words or phrases in one language with those of another, we are likely to miss one of the main contributions of the anthropological study of language, namely, the idea that for anthropologists the activity of translating is intimately related to ethnography, to the contextualization of words within the activity and the larger sociopolitical and cultural systems in which their speakers participate.

Translation starts in the field, when the linguistic anthropologist works at producing an **annotated transcript** (Schieffelin 1979, 1990). The annotated transcript not only contains contextual notes written during the recording sessions (see chapter 4), but all kinds of interpretative statements made during the transcription process. In preparing the transcripts of 83 hours of spontaneous speech between Kaluli children and their mothers, siblings, relatives, and other villagers, Schieffelin soon discovered that the mothers' comments on the tapes, including their laughter at situations they found humorous, constituted an important source of information on how they saw the situation. These comments, accompanied by the interpretation produced by a male assistant who had not been present during the recording, were integrated in a transcript that had a lot more than the words that had been exchanged among participants. It is transcripts of this sort that provide the basis for future translations. Several techniques are now available for keeping track of these on-going commentaries. The interaction with the knowledgeable participant/informant/research assistant can be recorded on tape, hand notes can be written on the side of the transcription pages, or (when a computer is being used) footnotes can be added to the text.

There are several formats for presenting transcripts with translation. All of the formats I will present here are currently adopted by linguistic anthropologists and each of them has different implications and consequences. The reason for discussing each one of them is to give readers an opportunity to judge which method is best suited to their needs. There is no such a thing as a *perfect* transcript, but there are transcripts that are *better* than others for certain specific needs!

Format I: Translation only.

The first format is to give the translation only. This is usually done when the researcher wants to concentrate on the content of what was said or feels that the original text might be unimportant or distracting. Here is an example from a transcript of a segment of a Kuna ritual greeting between a chanting "chief" (CC) and a responding "chief" (RC) inside the "gathering house":

(16) CC: Yes you appear as always.
RC: Indeed.
CC: In truth.
You still appear.
In good health.
RC: Indeed.
CC: In truth evil spirits.
In truth I do not want.
I utter.
RC: Indeed.
CC: Powerful evil spirits, see.
Then I do not want them to enter.
RC: Indeed.
CC: Now I am still in good health, say.
In truth still this way.
RC: Indeed. (Sherzer 1983: 75)

What is made apparent in this example is that even when only a translation is provided the visual display of the text is still important in conveying a number of

important assumptions about how the material should be interpreted. In this case, the format assumes a notion of **line** and a notion of **verse**. As discussed by Sherzer in a separate chapter, "lines are marked grammatically by means of an elaborate set of line initial and line final affixes, words and phrases" as well as by syntactic and semantic parallelism and intonational patterns (Sherzer 1983: 41). Verses are identified – in Kuna as in American Indian oral narratives in general – "not by counting parts, but by recognizing repetition within a frame, the relation of putative units to each other within a whole" (Hymes 1981: 318). In other words, texts like the one in (16) above presuppose a fairly complex theory of local poetics, which the researcher must find ways to make explicit, whether in the same text or elsewhere. In some cases, linguistic anthropologists have been experimenting with printing conventions to convey in the translation some of the prosodic features of the original oral performance. Thus, Tedlock (1983) used capitals to mark what was said in a loud voice (a convention also used in other traditions), long dashes or repeated vowels to indicate lengthening, and different heights to convey tonal structure, with "spilling letters" indicating a glissando:

(17) The girl went inside and PUT MORE WOOD ON, the fire was really blazing, then it came CLOSER.
It came closer
calling
hooooooooooooooooooooooooohaaaaaaaaaaaaaaaaaa
a
a
a
a
y it said.
The girl heard it very clearly now. (Tedlock 1983: 84)

Although there is no question that having only the translation makes reading easier, one of the disadvantages of not having the text in the original language is that other researchers are not given an opportunity to validate or question the author's decisions in the translation process. This is the reason for most linguistic anthropologists to go against the wishes of journal editors and printing presses and argue for the need to present both the text in the original language and the translation. There are several ways of doing this.

Format II. Original and subsequent (or parallel) free translation.

This format is designed to maintain the unity of the text in each language. In (18), for example, the two versions are placed next to each other, with an attempt to maintain an horizontal parallelism.

(18) (Dispute at board meeting)

Disputant:	
1. Lo que se NECESITA ...	What is NEEDED ...
2. Yo soy de ese opinión.	I am of this opinion.
3. A mí no me importa	It doesn't matter to me
4. quien es usted,	who you are,
5. de comisión o como	a board member or whatever
6. quiera que SEA.	you want TO BE. (Briggs 1986: 78)

The use of line numbers helps the reader to compare the original text with the translation. An additional convention for long poetic lines is that of utilizing indentation, as proposed by Joel Kuipers in his transcription of Weyewa ritual speech:

(19)

oruta koki	gather the monkeys	1
ta kalunga	in the field	2
ka ta mandi'i teppe,	so that we can sit on the mat,	3
wandora-na wawi	summon the pigs	4
ta maredda	in the meadow	5
kai terrena pa-mama;	so that you get the quid;	6

(Kuipers 1990: xvi)

In this variation on this format, "lines one through three and four through six are each part of a single poetic line" (ibid.).

These formats continue to assume the notion of a "line" (see above) and are more apt for poetic and ritual speech, but awkward or more arbitrary for ordinary speech. Things are also made more complex by those cases in which the original is in a language with polysyllabic words and complex morphology. In these cases, translators are forced to split words arbitrarily and cannot maintain the parallelism between left and right side of the page:

(20)

S76: Neh, solamente nimo-yōlcocoa para cē, para cē demanda, para cē crreclamarōz cē cosa īhuāxca, quihtōz, 'Pos xiquitta ōnēchcuilihqueh in noāxnoh.' 'Pos ¿quēn ōmitzcuilihqueh?' Pos ihquīn huān ihquīn, [etc.]	As for me, I am sad only for such a, such a demand, for one who would claim something of his own, who will say, "Well look, they took away my donkey." "Well how did they take it away from you?" Well like this and like this, one is sad in his soul because they took it [etc.]

(Hill and Hill 1986: 86)

The logic here is that the authors want the readers to see the original text but are not expecting them to pick out which word is which just from the transcript. When they want to achieve such a goal, they must shift to a different format.

Format III. Parallel free translation and morpheme-by-morpheme[17] *gloss under the original.*

Hill and Hill use this format when they discuss specific grammatical processes. In the following example, for instance, it is important to see that the word *tlaxcal* "tortilla" has become part of the verb, that is, it has undergone the grammatical process called *noun-incorporation* (Mithun 1986; Sadock 1980):

(21) ni-tlaxcal- chīhua "I am making tortillas"
I TORTILLA MAKE (Hill and Hill 1986: 251)

In this case, the left side of the page gives the original text and, on a different line, a literal, in this case morpheme-by-morpheme translation, and the right side gives the free translation. The distinction between the two is important not only because the morpheme-by-morpheme translation may use different words from the free translation, but also because the words in the original language might have a different order from, in this case, English and might make a decoding based on word-by-word translation difficult. When the text is more than one line long, the parallel format becomes awkward and yet another format is advisable.

Format IV. Original, interlinear morpheme-by-morpheme gloss, and free translation.

This format utilizes three lines, one on top of the other, as shown in the following Samoan example:

(22) *523 Mother:* *`ua uma na `ē `ai?*
Pst finish Comp you eat
"Have you finished eating?"
524 Son: *((nods))*
525 Mother: *alu ese lā`ia ma igā.*
go away then from there
"Then get away from there."
526 Son: *`o lea e sau e avaku le mea lea.*
Pred this TA come Comp take-Dx Art thing this
"I've come here to take this thing."
(Duranti 1994: 156, slightly modified)

[17] For a discussion of morphemes, see section 6.4.

The words on the first line (with the original text) can be spaced in such a way to allow for a one-to-one match with the interlinear glosses on the second line. This format is particularly appropriate when the author wants readers to follow the translation process more closely. It is the standard format for most linguistics journals. Its only drawback is that it crowds the page with lots of written material and requires some time to get used to reading it.

The last two examples also show that word-by-word glosses imply a minimal level of grammatical description; they force the linguist to assign particular grammatical, functional, or denotative meanings to each morpheme in the text. Abbreviations such as "Pst" for "past tense," "Comp" for "complementizer," "Art" for "article," and "Dx" for deictic particle assume a theory of Samoan grammar that may not be the focus of the discussion but needs to be attended to before providing the glosses.

The exposure to these different formats is a necessary part of any linguistic anthropologist's training not only because students should get used to the different conventions, but also because in their work they need to be aware of the need for a format that, while meeting the current standards of the research community, can also fullfil their expository needs. In some cases, a *range of transcription formats* might be necessary within the same article or book, depending each time on the specific point made by the author(s). In some cases, if the researcher only wants to identify a morpheme or a word on a line of transcript, there might be no reason to gloss every word and attention to the linguistic form can be achieved by underlining or boldfacing. An example of this method is provided in (23) from a transcript of Tzotzil conversation in which the author, John Haviland, is examining the use of the particle *a`a*:

(23) p: xlok' ono nan **a`a** yu`van
Indeed there will be enough, of course. (Haviland 1989: 45)

In this case the use of boldface points to the only linguistic feature the author wants the reader to focus on.

Other times, researchers might be faced with a situation that requires new conventions. In his study of language socialization in a multilingual village in Papua New Guinea, Don Kulick (1992) devised conventions that would make clear which language was being spoken at any given time. He used italics for words in Tok Pisin, italics and single underlining for the local vernacular, Taiap, and roman for the English translation. Underlining of the roman helps the reader keep track of which variety is being spoken by only following the English translation.

(24) Sopak: *Sia. ŋa ruru sɛnɛ ia kirwmbri wakarɛ. ɛnd-ɛ karɛ, ɛndɛkarɛ* [turns to Mas] *mm. Masito. Kisim spun i go givim papa* [hands Mas a spoon] *Spun.*

Sia [exclamation]. These two poor kids I just don't know. Hungry, hungry. [turns to Mas] Mm. Masito. Take the spoon and go give it to Papa. [hands Mas a spoon] Spoon.

(Kulick 1992: 203)

5.8 Non-native speakers as researchers

The question is at times raised outside of anthropology, especially among formal linguists working on their own language and conversation analysts working in their own society, about the feasibility of working on a language of which the researcher is not a native speaker and hence of the validity of generalizations made about meaning by non-native speakers. Although these doubts seem at first quite legitimate, they often start from the wrong assumptions.

One of the reasons to reject work that is not done by native speakers on their own language stems from the methodological preferences of the different researchers. Thus, for linguists working on native speakers' intuitions, it would seem very suspicious that a non-native speaker would make hypotheses on meaning. To this objection, there are two answers: (i) much of the work of linguistic anthropologists is *not* based on intuitions and introspection but more likely on correlations (tendency, for instance, for certain forms to appear in certain contexts); (ii) linguistic anthropologists rely heavily on native speaker's intuitions and judgment in preparing their transcript, that's what the concept of annotated transcript (see above) is about. Finally, it should be said that the assumption that a researcher–native speaker is the ideal condition is itself suspicious. It assumes that a native speaker has privileged access to theory building, hypotheses, and thick description. Although this might sometimes be the case, it goes against one of the tenets of anthropology, namely, the idea that one of the ways to describe culture is to look at it from both the inside and the outside. Whereas it is hard (and often impossible) for non-members to see things from the inside of the culture, it is equally hard for members to see things from the outside. The problem with many sociologists' view that one needs ethnography only or *especially* when working in another culture is based on the fact that when working on one's culture and within one's society one can leave much knowledge implicit (see chapter 8).

5.9 Summary

Here are some of the main points made in this chapter:

(i) transcription is a *selective* process, aimed at highlighting certain aspects of the interaction for specific research goals;

(ii) there is no *perfect* transcript in the sense of a transcript that can fully recapture the total experience of being in the original situation, but there are *better* transcripts, that is, transcripts that represent information in ways that are (more) consistent with our descriptive and theoretical goals;

(iii) there is no *final* transcription, only *different, revised* versions of a transcript for a particular purpose, for a particular audience;

(iv) transcripts are *analytical products*, that must be continuously updated and compared with the material out of which they were produced (one should never grow tired of going back to an audio tape or a video tape and checking whether the existing transcript of the tape conforms to our present standards and theoretical goals);

(v) we should be as *explicit* as possible about the choices we make in representing information on a page (or on a screen);

(vi) transcription formats vary and must be evaluated vis-à-vis the goals they must fulfill;

(vii) we must be critically *aware* of the theoretical, political, and ethical implications of our transcription process and the final products resulting from it;

(viii) as we gain access to tools that allow us to integrate visual and verbal information, we must compare the result of these new transcription formats with former ones and evaluate their features;

(ix) transcriptions change over time because our goals change and our understanding changes (hopefully becomes "thicker," that is, with more layers of signification).

We must keep in mind that a transcript of a conversation is not the same thing as the conversation; just as an audio or video recording of an interaction is not the same as that interaction. But the systematic inscription of verbal, gestural, and spatio-temporal dimensions of interactions can open new windows on our understanding of how human beings use talk and other tools in their daily interactions.