Wolfram, W. (1993), 'Ethical considerations in language awareness programs', *Issues in Applied Linguistics* 4, 225–255.

Wolfram, W., Hazen, K., and Ruff Tamburro, J. (1997), 'Isolation within isolation: A solitary century of African American English', *Journal of Sociolinguistics* 1, 7–38.

Yamada, R. (2007), 'Collaborative linguistic fieldwork: Practical application of the empowerment model', *Language Documentation & Conservation* 1(2), 257–282.

Yu, E. S. H. and Lieu, W. T. (1986), 'Methodological problems and policy implications in Vietnamese refugee research', *International Migration Review* 20(2), 483–502.

4

Transcription in Linguistics

Lorenza Mondada

Chapter outline

Transcription is an indispensable practice and tool for linguists studying spoken language: it allows scholars to represent recorded talk in a textual written form and thereby to transform the transient and fleeting nature of spoken language into a visual stabilized object. This chapter presents and discusses principles and problems raised by transcriptions within a diversity of fields in linguistics. The first section 'Introduction: Transcription as research practice' shows that transcripts connect to key issues, touching on practical and theoretical aspects. The section 'Diversity of transcription practices: Different responses to perpetual challenges' discusses several controversies in linguistics, which reveal a variety of possible responses to the challenges of transcribing. The last section 'Practices of transcription in CA' focuses on a particular discipline, conversation analysis. This approach to social interaction has developed the practice of transcription in an exemplary way, first on the basis of talk, then expanding to the transcription of multimodality, integrating language and body conducts. In conclusion, the chapter shows how even small choices concerning the annotation of minute details have big analytical and conceptual consequences.

Introduction: Transcription as research practice

The use of recording technologies in linguistics – and other fields of the social sciences – for documenting spoken language has generated audio (and later video) data requiring adequate forms of representation, annotation and inscription of the recorded sounds (and visual cues). Transcription is one response to this need. The practice of transcription in several fields of linguistics has been extensively developed since the 1970s, thanks to two concomitant factors. On the one hand, the increasing sophistication, miniaturization and accessibility of recording technologies have made it increasingly easy to document communicative events. On the other hand, the development of a diversity of approaches centred on spoken language, discourse and interaction, such as ethnography of communication, discourse analysis (DA), conversation analysis (CA), sociolinguistics and anthropology of language, has prompted a renewed interest in spoken language and favoured the use of these technologies for its study in its actual contexts of use.

Transcription in linguistics has emerged as a necessity for textually and visually representing language as an object of study. Given the labile nature of the original spoken event, which disappears as soon as it is uttered, linguists work on two artefacts that represent it: the first is the recording (often different audio/video sources) and the second is their transcription/annotation/coding. The former constitutes the primary data, the latter a form of selective and interpretive reconstruction that refers to them. As Duranti put it (2006), in reference to Plato's myth of the cave, transcripts are shadows on the wall: they never exhaust either the original data or the original event, which is lost forever.

Therefore, although transcripts are often detached from the original data and circulate autonomously as 'immutable mobiles' (Latour, 1986), they are objects and practices that lie at the core of a long and iterative research process. They crucially depend on recording technologies and what they make available (or not); they result from the work of going back and forth between transcribing and analysing. Thus, transcripts are a reflexive and emergent outcome, making analysis possible and at the same time being generated through some form of protoanalysis. They also which not only the professional hearing and sight of the researcher but also the quality of technical tools like players, headphones and software play a crucial role.

Transcripts and transcription practices are comparable to other practices and inscriptions that characterize scientific research in a broader sense. As demonstrated by the social studies of science, inscriptions (Latour, 1986) are theory laden and have theoretical consequences. In particular, they are crucial in objectifying and rationalizing the phenomena to be studied: they transform them into 'Galilean objects' (Lynch, 1988); that is, into objects that are characterized by observable forms, recognizable patterns and regularities. Scientific practices make use of a variety of these – field notes, minutes, tables, visualizations, maps, transcripts, annotations, coding sheets – which are, in the Western scientific tradition, heavily dependent on textual representations, including not only written language, but also visual and spatialized features (Goody, 1977).

In linguistics, transcription constitutes one such mode of inscription that plays a central role not only within research procedures but also in the constitution, training and manifestation of the researcher's professional identity. These inscriptions are characterized by a paradoxical status: they try to respect the specificities of their object, *orality*, which is a dynamic, labile and evanescent series of sounds (and body movements), within a mode of representation that is a *written, textual, spatialized* and *visualized* fixation. Thus, the challenges of transcription concern the manner in which these two somehow contradictory aspects are managed: transcripts are aimed at preserving the specific features of spoken language as a dynamic temporalized object by using inscriptions that are a static spatialized representation (Bergmann 1985).

In the first part of the chapter, I present and discuss some issues and principles of transcription characterizing a diversity of fields in linguistics, with a special focus on a series of controversies that reveal the variety of possible responses to the challenges of transcribing and their consequences. In the second part of the chapter, I focus on a particular discipline, CA, which has developed the practice of transcription in an exemplary way, and which continuously reflects on its possible expansions, within new challenging dimensions, such as the transcription of multimodality, integrating language and body conducts.

Diversity of transcription practices: Different responses to perpetual challenges

Transcription is used within different fields of linguistics concerned with spoken language – such as phonetics, syntax of spoken language, child studies and psycholinguistics, acquisition studies, sociolinguistics and anthropology linguistics, DA and CA, as well as, more recently, corpus linguistics. Transcription is also practised in other cognate disciplines in the social sciences, such as anthropology and sociology. Each discipline, school or model has its own vision of what a transcript should be.

The practice of linguistic transcription is not new, but can be traced back in the history of linguistics. In the fields of phonetics and dialectology, for instance, transcription represents an important step in the promotion and professionalization of linguistic research. As shown by Bergounioux (1992) in a note about the history of dialectology in France, the use of a specific convention and alphabet for transcribing dialect – introduced by Gilliéron and Rousselot in their new journal *Revue des Patois Gallo-Romains* in 1887, just one year before the official publication of the International Phonetic Alphabet (IPA) – operated the separation not only between speakers and fieldworkers, but also between amateur defenders of dialect and professional academics studying dialect. A technical tool for representing language, the transcription system established a divide between common sense interest in dialect and scientific research about it, achieving a new professionalization of the linguist. This historical example shows that transcription raises issues intertwining scientific and political dimensions, of which this section will give other examples.

Nowadays, transcription is a central concern for a variety of subdisciplines of linguistics. It is a central tool for studying phonetics and prosody. These fields have developed not only specific phonetic notations, such as the IPA, but also specific visualizations of sound, thanks to transcribing and aligning software, such as Praat. Transcription is also a long-term practice for dialectologists, as well as sociolinguists, where the search for an adequate rendition of dialects and socially stratified varieties has been central to the investigation of the specificities of these varieties as well as to the way in which speakers manifest and negotiate their identity (e.g.

1970s the emergence of a variety of paradigms focused on specific features of spoken language as well as on spoken communication and interaction – such as ethnography of speaking, interactional sociolinguistics, CA, DA and linguistic anthropology – prompted new reflections on transcription. Their focus on the importance of the situated communicative activities in which spoken language is used, the sociocultural context in which they are meaningful and the broader discursive and interactive environments that shape and motivate details of talk had important effects on the practices of transcription (e.g. Sherzer, 1994; Edwards and Lampert, 1993; Du Bois, 1991; Ochs, 1979; Jefferson, 1985, 2004).

The following subsections present some of these key issues and reflections by centring on some controversies that have sparked lively debates in the literature. These controversies reveal how transcribing always relies on choices. These options are both practical and theoretically informed; they are differently framed and justified depending on the linguistic models researchers adopt, and they have very different consequences in terms of how researchers define their empirical object and develop empirical analyses. The issues discussed concern the differences between the focus on content versus form while transcribing (section 'Focusing on content versus form'), the consequences of orthographic conventions (section 'Issues of spelling and respelling: Orthography, eye dialect and phonetics'), the problems of dealing with multilingual data and data that have to be translated (section 'Dealing with multiple languages: Translations and multilingual data'), the way in which transcripts are consequential for the way researchers treat the identity and categorization of speakers (section 'Categorizing speakers') and the effects of how transcripts are formatted and presented on the page for their interpretation (section 'Spatializing talk'). These issues (summarized in section 'Synthesis: Some issues') show that transcribing is a practice involving options and choices at many levels, which lead – sometimes in an explicit, but often in an implicit and tacit way – to different types of categorization regarding forms and units, language(s), speakers and actions.

Focusing on content versus form

Transcripts in the academic literature range from a continuum going from a common sensical textualization of contents to the specific transcription of forms obeying transcription conventions. The two ends of this continuum it delivers (e.g. it is common in content analyses of interviews). The latter characterizes approaches centred instead on ways of speaking (e.g. a particular pronunciation, contrastive prosodic patterns, the choice of a particular syntactical construction or the discontinuities of self-corrected talk). The former generally adopts conventional written norms, including orthography, punctuation and layout used for representing direct reported speech in literary and theatrical texts. The latter manifests a critical stance towards written norms, motivated by a special attention to the difference between written and spoken language. Consequently, in the latter case, scholars search for forms of representation of spoken details that preserve the specific details of orality, without reducing them to written standards – avoiding what Linell (2004) calls the *written language bias*. This is why conventions are used that specify orthographic and phonetic notations, but also specific uses of punctuation (avoiding the implicit adoption of written norms and reusing punctuation in a different way).

A vivid example of the contrast between these two poles is offered by Bucholtz (2007), reflecting on two versions of a transcript she used in two different circumstances. The first was used as an ethnographic testimony with a focus on the information it contained, transcribed in a standard written text, and the second as a piece of data to be studied for its socio-interactional organization, transcribed according to the Santa Barbara Corpus conventions (Du Bois, 1991).

(1a) (Bucholtz, 2007: ex. 1a)
Fred: We're always the nerds. We like it. We're glad to be the nerds and the squares. We don't drink, we don't do any drugs, we just get naturally high, we do insane funny things. And we're smart. We get good grades. (Bucholtz, 1998: 122)

(1b) (Bucholtz, 2007: ex. 1b)
1 Mary: [So]
2 Fred: [We're al]ways the nerds.
3 We like it.
4 Mary: You@'re the nerds?
5 Fred: We're <creaky> {glad } to be the ner:ds,
6 a@nd the squa:res and,
7 Mary: Is that what
8 Fred: [we don't–]
9 Mary: [you say] you are?
10 Fred: <[i?]> Well,

12 We don't always say it,=
13 =I say it. n@
14 Mary: @@[@ !]
15 Fred: [But-]
16 Mary: @ You're [[prou:d.]]
17 Fred: [[you]] know,
18 Mary: [@@]
19 Fred: [we don't–]
20 We just don't (0.5) drink,

Bucholtz contrasts these two extracts, pinpointing that 'Fred's comments are not the product of an autonomous, triumphant voice of nerd pride but are rather the result of considerable co-construction (and obstruction) by me as the researcher. Her stated views, while clearly strongly held, are much more hedged and halting in their expression than my first transcript acknowledged' (2007: 788). If we further focus on the details of these transcripts, we notice how revelatory they are of distinct research practices. In particular, what is missing in version 1a relates precisely to the methodology of the interview: questions are erased and the interviewer is made invisible. In version 1b, the negotiation of not only content but ways of speaking is observable. Line 1, a new sequence, seems to be started both by the interviewer, Mary, and the interviewee, Fred. Fred's claim (2–3) is not said in response to a question, but as a self-initiated turn. This claim, and the use of the category 'nerds,' is repaired by the researcher (4), prompting Fred to reissue her claim in an even more emphatic way (creaky voice, stretched syllables, laughter particles indicated by the symbol @). This is again repaired by the researcher (7, 9) in overlap with Fred's progression into her description – which might already be initiated in lines 19–20, again showing her autonomous organization of the progressivity of her talk, which is not merely responsive to the interviewer. Interestingly, the repair initiation (9) occasions a dispreferred repair (10) not targeting the category of 'nerd,' but the voice to which this category is attributed, introduced by the researcher (using the verb 'you say' [9]). First, the verb is repaired in two negative utterances (11–12) and second, the pronoun is repaired from 'we' into 'I' by Fred now assuming personally the use of the category laughing (13). What Fred does here is reflect and negotiate, and eventually subvert, the way in which the researcher reformulates what she says and attributes claims to her and her group. This lies at the core of what researchers do when using interviews and other reporting methods (see also Edley and Litosseliti's chapter in this volume).

ignored are not only the interventions of the researcher, but the negotiations of meaning, authorship and representativity between her and the informant. Thus, from this example we learn not only about transcription choices, but more radically about field methodologies, the work of researchers behind the scene, and procedures of objectivation (of the informant's talk) and transparentization (of the interviewer's interventions).

Issues of spelling and respelling: Orthography, eye dialect and phonetics

Transcripts involve the representation, in a written form, of what has been said. Orthography represents the most frequently used option (see example 1) – contrasting with the IPA, which is used only in restricted cases, even by phoneticians. The IPA is often considered relevant for a fine-grained phonetic annotation (albeit not in a straightforward way – because it implies other forms of normalization and selection [Local and Kelly, 1989]). However, the IPA is also often considered difficult to use as well as to read, especially for longer transcripts. Furthermore, orthography does not solve all the problems; its use is submitted to very contrasted choices, which have generated considerable controversies. Orthography can be used in a standard way, respecting written norms; it can also be transformed into non-standard and even creative ways, in order to better represent individual, dialectal, ethnic or social particularities of spoken productions. For example, Preston (1982) identifies three categories of what he calls 'respellings' of words in transcripts: *eye dialect* (free adaptation of orthography for capturing phonetic details), *allegro* forms (elision of non-pronounced sounds) and *dialect respellings* (normalization of dialectal features). These adjustments of normative orthography have been diversely discussed and evaluated in the literature (see also Bucholtz, 2000, 2007).

A first issue discussed in the literature concerns the adequacy of written standards for capturing spoken variation. Orthographic adjustments aim to show the difference between spoken and written norms, and give a hint about the specificity, and even uniqueness, of a single production. One example of orthographic adjustment is the use of so-called eye dialect; that is, the spelling of words in non-standard ways in an attempt to represent specific ways of pronouncing them (e.g. 'coz' for 'because' and 'I dunnu' for 'I don't know'). Eye dialect has been used in literary texts, as well as in popular

of the use of eye dialect is the work of Jefferson (1983, 1985) in CA, which captures meaningful distinctions that would have been erased by a standard orthographic rendition. As a criticism of eye dialect, some linguists have argued that it often displays a naïve conception of orthography and its rules and it creates graphemic 'monsters' that are contradictory and inconsistent within its system (Blanche-Benveniste and Jeanjean, 1987). It can also be superfluous, since it is 'unnecessary to indicate phonetic features which are predictable from general rules of the orthography' (Macaulay, 1991: 287). It should be noted that these discussions vary depending on the languages considered, their orthographic rules and the national normative traditions.

A second issue concerns the fact that eye dialect has been considered stigmatizing for certain linguistic communities. Preston (1982) observes that folklorists tend *not* to represent American middle-class 'northern' English speakers with eye dialect, whereas such respellings are abundant for African Americans, Appalachians and non-native speakers. Moreover, he notes that in the latter case, they demote the speakers in terms of social status, intelligence and sophistication. Blanche-Benveniste and Jeanjean (1987) reveal that perception is often systematically biased by sociological variation: some categories of speakers tend to be transcribed in a way that interprets what they said as 'errors' rather than as (correct) grammatical constructions (e.g. in the French double negation system, the first negative particle is often omitted, especially among certain categories of speakers; but the transcriber might contribute to the stigmatization of the speaker by not transcribing the negative particle 'ne' in 'on ø avait rien à manger'/"we had nothing to eat," even when it does not present any audible difference with the standard spelling 'on *n*'avait rien à manger'). This is consequential for the description of the grammar of spoken language and its social stratification. In sociology, Bourdieu et al. (1993) use similar arguments, but leading to different conclusions than the ones generally assumed by linguists. He argues in favour not only of the written standard, but even of normative corrections of the responses given by his interviewees, in order to avoid their caricature and stigmatization. In a response, Lahire (1996) accuses him of producing sociological artefacts by erasing details constituting and revealing the social identity of the speakers. This controversy shows how issues of ethics and politics are at the core of orthographic representations.

A further, very different, issue related to orthography concerns the use of transcripts in corpus linguistic data banks: because search engines find it difficult, and sometimes impossible, to recognize words in non-standard order to enable automatic searches within larger corpora (see, e.g., Leech et al., 1995). In this case, standard orthography constitutes a solution to technological constraints – although algorithmic solutions have also been found to include non-standard spelling in searches.

These controversies show that standard orthography is much more than a representational convention: it is a form of sociocultural technology that characterizes and enables practical uses of texts and contains normative values. Reproducing standards as well as subverting them can produce strong normative effects. Their effects of authority versus stigmatization shape the way in which the original talk and its transcription are categorized and interpreted, and therefore shape the identities of both transcribers and transcribed, by positioning them socially and accentuating their (a) symmetries.

Dealing with multiple languages: Translations and multilingual data

Transcribing embeds implicit norms about written and spoken language, as seen in the previous discussion, as well as multiple tacit assumptions about what (a) language is. These are revealed most explicitly when transcriptions go beyond the monolingual space of a homogeneous community of speakers, researchers and publishers. I will discuss two cases in this respect: the first concerns transcripts in a language that is not the same as the language of the published article, and consequently need to be translated for a recipient who does not know it; the second concerns transcripts of multilingual talk. Both cases question straightforward conceptions of what a language is.

Often data feature in publications that do not use the same language (typically journals in English) and for readers who may not understand the original language of the data. Translating data is a delicate and often frustrating process. A lot of the specificities of the original language are not available in the final translation (Traverso, 2002). Many choices, concerning not only the translated forms, but their visual disposition on the page, have significant implications for how the data will be finally accessed by the reader. The distance from the original can be managed in different ways: if the transcript and its translation are quoted separately, one after the other (or even in a footnote), their distance is significantly greater than in interlinear translation; likewise, an idiomatic translation, trying to give an an autonomous manner, detached from the original, whereas a translation trying to maximally fit with the original invites the reader to return to it and helps them to understand it. To enhance the readability of the original, grammatical glosses are added to the interlinear translation (see excerpt 2a). These glosses are important for highlighting the grammatical structures and sometimes for allowing some elements to be left in their original form rather than translating them. This is the case, for example,with Finnish particles studied by Sorjonen (2001), as illustrated in the following short example:

(2) (Sorjonen, 2001:91)

```
0:   .hhh (0.7)  Joo::. Annap-pa-s   se tarke-mpi   osote
                 PRT    give-CLI-CLI it precise-COM  address
     .hhh (0.7)  Joo::. Give me the more precise address
```

Sorjonen transcribes the particle on the first line in Finnish ('joo'), she glosses it on the second line (categorizing it as 'PRT,' particle), but she refuses to translate it on the third line (actually, she integrates its original form in the English translation), because this would just erase the subtle differentiations that her study is precisely aiming to demonstrate.

Glosses rely on a linguistic model and theory, providing specific grammatical categories that are used for the original language and can be understood by the recipient – therefore often their universal character is presupposed by applying them to very different languages. Moreover, glosses and their level of detail depend on the focus of analysis. Glosses and translations can reveal the theoretical framework of the researcher and how it impinges on the intelligibility of the transcription and the orderly character of the material represented.

Multilingual data raise similar challenges, showing the limits of the categorization of recorded forms as belonging to a single clearly identifiable language (Mondada, 2000). To discuss these issues, I draw on an example from Léglise and Sanchez Moreano (2017), featuring a client talking to an employee at the national electricity company in Cayenne (French Guiana).

(3a) (Léglise & Moreano, 2017: ex. 1: Corpus EDF Clapoty - Nelson/Léglise)

a. **Yèr** **mo** **té** **pasé** **la**
yesterday 1SG PST went here
Yesterday I was here

b. **i** **té** **gen** ***an::*** madame un peu costaud à côté là
3SG TE.PST avoir ART.INDF misses a little.ADV sturdy next to.ADV here.ADV

c. i m' a donné [...] comme **té** ***ni*** **problem**
3SG 1SG have given as if.CONJ TE.PST have problem.N
she game me [...] as if there was a problem

Multilingual data – in which the phenomena of code-switching and code-mixing are observable – raise the issue of how not only to transcribe but also to *categorize* the languages used by the speaker (e.g. being French or English). Often the solution consists in identifying the language in a typographically visible way: in extract 3a, bold refers to Guiana Creole, bold italics to Antillan Creole and roman to French. This kind of notation imposes clear-cut decisions, attributing to each form a unique and full membership of a language. These decisions are implicit in all transcriptions – revealed by their orthographic norms and explicit in the glosses – but are particularly exhibited in transcriptions of code-switched/-mixed talk. These can be problematic, especially for languages that are closely related and in which contact phenomena blur clear-cut boundaries between them. Léglise and Sanchez Moreano (2017) discuss several difficulties raised in this extract by the continuum between French and various Creoles spoken in the same area. For instance, 'yèr' (a.) is written in Creole, but nothing distinguishes it orally from 'hier' in French; likewise, the final adverb 'la'/"là" (end of lines a. and b.) is fundamentally the same, but gets transcribed with two different orthographies, treated somehow differently in the glosses; furthermore, the word 'problem'/"problème" (c.) is pronounced in the same way in the three languages. The third line is particularly tricky, given that both Guiana and Antillan Creoles use 'té,' the pre-verbal marking of the past tense. These possible alternatives show the choices made visible by the first version of the transcript: the first line is treated as being consistently in Guiana Creole; the second line is treated as beginning in Guiana Creole, then using a determinant in Antillan Creole and finally ending in French – where the word 'madame' constitutes a boundary case, since it could be French as well as Creole, although it would have different connotations in both languages. The third line is treated as beginning with a pronoun in Guiana Creole, continuing in French, and ending with a mix of Guiana and Antillan Creole ('ni' is the only form here that is univocally identifiable in the latter language). Thus, the first line supposes the continuity and consistency of the speaker's linguistic choices through spates of talk, but the third line does not.

Given the proliferation of problematic choices in this kind of transcription,

words of the third line: their possible categorizations are indicated both in the orthography and in the typographical convention:

(3b) (Léglise and Moreano, 2017: ex. 2: Corpus EDF Clapoty - Nelson/Léglise)
c. *i* probléme
c. **i** ***té*** ***problem***
c. i(l) m' a donné [...] comme **té** ***ni*** **problem**
3SG 1SG have given as if.CONJ TE.PST have problem.N
she game me [...] as if there was a problem

This example shows the interest of multi-transcription; that is, a transcription integrating various possible variants. More generally, Léglise and Sanchez Moreano (2017) propose computer-supported solutions for their corpus, enabling multi-transcription and alternative labels. The general strategy consists in leaving the choices open for further steps in the research. In this case, the solution is both technologically supported and pragmatically postponed to a later phase in the analytic process.

Categorizing speakers

Transcripts are not only constituted by transcribed talk; they also integrate other information, and most notably the identification of the participants speaking – often in the form of either a name or a category in the left margin. This identification is seldom discussed in the literature (but see Mondada, 2002), although it has important consequences for the interpretation of the transcript. The identification of the speaker is the first element to be read on the left, at the beginning of the line, and this position is consequential for how the text will be read, understood and interpreted. Speakers can be identified by a choice of letters (S, D, R, for example, which has a different impact than A, B, C, the latter imposing a sense of order that the former does not), or of names (Bea and Ahmed vs. Mrs. Baker and Mr. Hakimi, allowing different levels of formality as well as different social and ethnic identities to transpire). This also raises issues of how to anonymize the names of the participants and to choose their pseudonyms, for instance by selecting names that are both distinct and related to the original ones. Speakers can also be identified by categories (Interviewer and Interviewee, Doctor and Patient, etc.), which raises issues of local relevance. As pointed out by Sacks (1972) and Watson (1997), a diversity of categories can potentially be used to describe a person: they might be referentially correct, but the issue is whether moment within the talk, possibly changing as talk unfolds (Mondada, 2002). These considerations show the problematic character of identifications such as Man and Woman (which suppose the a priori relevance of gender and essentialize it), Native and Non-native (much criticized for their linguistic and ideological presuppositions, as well as for excluding other social identities than those manifested by the participants' linguistic competences), Student1, Student2, Student3 (supposing homogeneous characteristics of persons belonging to the same class), etc. These considerations show the importance of the choice of names identifying the speaker for the analysis of their talk and their possible stereotyping consequences, aggravating possible effects of orthographic misspellings.

Spatializing talk

In her seminal article on transcription as theory, Ochs (1979) shows the consequences of the layout of transcripts and the spatial disposition of their text on the page. A transcript is not just a linear continuous text, but a spatialized text, in which the identification of the participants, line numbering, disposition of transcribed talk and representation of time through the positioning of textual strings all contribute to the interpretation of the coherence, consistency and progressivity of the participants' action unfolding line by line. Ochs's argument points out the fact that the successive disposition of the transcript's lines, one after the other, corresponds to the sequential organization of the adults' talk, where a turn responds to the previous one and projects the next. Interactions with or among children might not work in the same way, and their textual disposition should take their specific interactional competences into consideration. For instance, Ochs shows that transcripts beginning with an adult's turn in first position on top of the page and on the left of the line impose a reading of the next (child's) line as dependent on the previous one – and possibly as incoherent, if it does not respond to it. Ochs contrasts the *list format* (in which one turn is followed by the next on the verticality of the page) with a *column format* (in which each turn is disposed beside the other, as distinct columns, and the child's column is placed on the left of the adult's one, favouring the reading of the child's actions as having their own coherence and not only as responding to the adult's ones). Although the column format is now seldom used, this example shows the interpretive effects of the spatialization of talk

Synthesis: Some issues

The variability of transcripts has often been treated by scholars as the result of 'errors,' and as showing the low reliability of the transcribers' work, due to carelessness (Kitzinger, 1998) and inconsistencies (O'Connell and Kowal, 1990). But variation has also been treated as inherent to the linguistic phenomena transcribed, as well as to the practice of transcription, which is a never-ending process (Bucholtz, 2007; Mondada, 2007). The controversies exposed in this section show that transcription is never a mechanical practice, but instead requires constant choices that have analytical consequences.

Transcripts exhibit not only the choices of the transcriber, but also their membership within a theoretical paradigm and a disciplinary community. These choices concern different issues:

- the definition of the phenomena to be transcribed and their preservation, including the accuracy and precision of their annotation;
- the organization of these phenomena and the principles on which their order is based, according to the analytical and theoretical approach adopted;
- ethical and political issues, often associated with problems of stereotyping and erasure of (ir)relevant details;
- technical issues, associated with demands of robustness, consistency, reproducibility of the conventions and their implementation, which are particularly important for the digital treatment of transcripts and the automatization of searches.

Each of these layers implies necessary choices, which can become arenas for controversies and discussions, depending on the principles and objectives governing transcription – which are sometimes discussed, but often remain implicit and tacit. One field in which transcription practices have been explicitly articulated with analytical principles is CA, to which we now turn.

Practices of transcription in CA

CA represents a field within linguistics and the social sciences (see also Baxter's chapter in this volume) in which transcription has been developed in detail over the last decades, in a way that is particularly coherent with focus on this approach in order to show how transcription can be explicitly related to the analytical principles of a paradigm and how new challenges emerging from empirical studies foster solutions that are crucially based on new ways of transcribing.

CA's distinctive way of transcribing is strongly related to its fundamental tenets. In particular, these principles concern a focus on situated action, as it is organized within social interaction and among various co-participants, as it unfolds sequentially, establishing retrospective relations with previous actions and projecting subsequent actions, within a temporality organized in a continuous, emergent and incremental way. These constitutive aspects (see Sidnell and Stivers, 2012, for an extensive presentation) inform a methodology that is crucially based on audio-video recordings of naturally occurring interactional activities and on their fine-grained transcription. The focus on situated actions entails the audio-video recording of interactions in their social context, without being orchestrated by the researcher and in a minimally invasive way. Moreover, the attention to the way participants themselves organize these actions and make them intelligible for others materializes in audio-video recordings that document in the most comprehensive manner the communicative resources used by the participants, including not only language but also body conducts (Heath et al., 2010; Mondada, 2012).

In the following sections, I first detail the principles supporting verbal transcription and show some of their analytical consequences (section 'Verbal transcripts in CA'). Then I show how transcripts have been expanded for multimodal analysis (section 'Multimodal transcripts in CA'); that is, how they have integrated, besides language, details of embodied conducts, such as gesture, gaze and body postures.

Verbal transcripts in CA

CA's transcripts are consistent with CA's conception of social action. The starting assumption is that action is produced in an orderly and accountable way by and for the participants (Sacks, 1984: 22), and thus 'no order of detail in interaction can be dismissed a priori as disorderly, accidental, or irrelevant' (Heritage, 1984: 241). This prompts a textual and visual representation of talk and embodied conduct that carefully considers the orientation of the participants towards the issues of 'why that

incarnate the fundamental principles of temporality and sequentiality on which the organization of social interaction is based. Therefore, transcripts are particularly careful in representing the emergent, incremental, situated and contingent unfolding of action in time. Moreover, they reflect the fact that what makes social interaction intelligible is not a predefined set of forms decided upon by the analyst, but an open-ended indexical arrangement of resources that are mobilized and oriented to moment by moment by the interactants, within an *emic view* (endogenously defined by the participants) (vs. an *etic view*, exogenously defined by the analyst) (Mondada, 2014a).

This endogenous order has been captured by the transcript notation developed by Gail Jefferson, which is currently widely used to represent talk in interaction (although national variants, all inspired by her, exist, such as GAT or ICOR). Gail Jefferson was a founder of CA, a charismatic figure of the movement, who developed some of the most important analytical objects of CA (such as turn-taking, overlap, repair and laughter; see Sacks et al., 1974; Jefferson, 1983, 1985, 2017). Jefferson (2004) also developed a way of transcribing them that would make it possible not only to *represent* them but more fundamentally to *discover* them – to hear and see, notice and capture them (see also Psathas and Anderson, 1990; Hepburn and Bolden, 2017).

An example of analysis of laughter discussed by Jefferson (1985) shows the issues and the payoffs of this kind of transcription. It is a dirty joke, for which she provides a transcript in two versions, the first *describing* laughter and the second *transcribing* it. The first adopts an orthographic transcription; the second an adapted version of it.

(4a) Jefferson (1985: 28-29, ex. 7)

```
Ken:    And he came home and decided he was gonna play with his orchids
        from then on in.
Roger:  With his what?
Louise: heh heh heh heh
Ken:    With his orchids. [He has an orchid.
Roger:                    [Oh he h hehheh
Louise: ((through bubbling laughter)) Playing with his organ yeah
        I thought the same thing!
Ken:    No he's got a great big[glass house-
                               [I can see him playin with his
                               organ hehh hhhh
```

(4b) Jefferson (1985: 28-29, ex. 8)

```
Ken:     An'e came home'n decided'e wz gonna play with
         his o:rchids. from then on i:n.
Roger:   With iz what?
Louise:  mh hih hih[huh
Ken:               [With his orchids.=
Ken:     =Ee[z got an orch[id-
Roger:      [Oh:. [hehh[h a h 'he:h]' heh
Louise:                [heh  huh 'hh] PLAYN(h)W(h)IZO(h)R'N
         ya:h I [thought the [same
Roger           [uh::        ['hunhh'hh'hh
Ken:                         [Cz eez gotta great big[gla:ss house]=
Roger                                               [I c'n s(h)ee]=
Ken:     =[(    )
Roger    =[im pl(h)ay with iz o(h)r(h)g' (h)n 'uh
```

When Ken says that the guy 'was gonna play with his orchids' this generates a repair initiated by Roger ('with his what?'), to which Louise responds with slight laughter – indicated by several aspired syllables – and Ken with a repair ('with his orchids'). Now Roger laughs too, overlapping with Ken's continuation of the turn and before he has completed his explanation. Louise, in the next turn, formulates again the gist of the joke: in the first transcription, which uses in fact a *description* of the laughter, Louise's turn is orthographically rendered in a unique way, making the relation between 'orchid' and 'organ' totally explicit. In the second transcription, Jefferson shows the pay-off of *transcribing* it: Louise's turn overlaps Roger's laughter, and 'O(h)R'N' is a very different form, in which laughter 'invades the talk' (1985: 29). Here both 'uncontainable' laughter and 'difficulty in hearing the punchline' (1985) are constitutive elements of the dirty joke – which get lost in a normalized orthographic transcription and in the description of laughter.

More generally, the careful transcription of turns-at-talk as they unfold in a finely grained way shows the way turn-taking operates, how participants make recognizable and recognize possibly completed units of talk and opportunities to speak and how they exploit them for organizing their participation in the ongoing activity.

Multimodal transcripts in CA

Transcription challenges have become even more important as scholars

With the spread of video technologies, CA has increasingly integrated the study of multimodal resources; that is, language and embodied conduct (Goodwin, 2000; Heath, 1986; Mondada, 2014a, 2016; Streeck et al., 2011). This has raised new challenges not only for transcription, but more radically for analysis. Multimodality (see also Bezemer and Jewitt's chapter in this volume) includes annotations of language, gesture, gaze, head movements, body postures, body movements and object manipulations. This makes multimodal transcription even more complex than linguistic transcription, because it concerns not only the relatively linear unfolding of verbal turns, but also many embodied courses of action emerging and expanding at the same time. If temporality is fundamental in the transcription of talk, it is even more crucial for the transcription of body movements. This has prompted researchers engaging in a detailed representation of multimodality to develop specific conventions for its notation.

Like Jeffersonian transcripts, multimodal transcripts are aimed at showing the ordered details of interactivity, temporality and accountability of action. First, with regard to *interactivity*, multimodal transcripts, even more than verbal ones, allow the researcher to show that all participants are possibly constantly participating in the current action, for example, gazing, nodding, etc., expressing their online embodied responses, or silently displaying their (mis)understanding or (dis)agreements. Second, the *temporality* of multimodal conducts integrates within the transcription not only pauses and overlaps, but more radically a continuous flow of multimodal resources, such as gesture, gaze, body postures and movements. These resources emerge, unfold and are retracted across time, in both simultaneous and successive ways, exhibiting their fine-grained mutual coordination and their responsiveness to previous actions. Third, the *accountability* of action is achieved by resources that have ordered, distinctive, recognizable forms and trajectories in time, annotated as far as their intelligibility and visibility are concerned (the visibility of a gesture, the noticeability of a gaze shift, the transformation of a body posture and the like, as orchestrated by a participant and seen, glanced at, or monitored by one or more co-participants).

Unlike Jeffersonian transcripts, multimodal transcripts confront the analyst and transcriber with new challenges. There are standard conventions for talk, but not yet standardized conventions for multimodality (but see Goodwin, 1981, and Rossano, 2012, for gaze; and Mondada, 2014a, 2016, 2018, for an integrative system concerning all resources). These conventions face different problems, since the linearity and successivity of linguistic production unfolding in a parallel but not synchronous way. Moreover, whereas there are conventional ways of writing and segmenting talk, the same does not hold for the continuous unfolding of embodied conducts, which are often juxtaposed and intertwined: contrary to the graphemic representation of the spoken, the transcription of embodied conducts relies on their description. Finally, questions of relevance and selectivity, present in verbal transcripts, are even more important for multimodal transcripts, in which the issue of relevance is vividly present and locally defined moment by moment. Their relevance varies with the context (resources are *indexical*) and more particularly with the ecological specificity of each embodied resource (in the sense that some resources are particularly fitted to a given material environment, but would not make sense in another: for example, pointing gestures can be made with different prosthetic objects, depending on the material specificities of the ongoing activity – a cook might point with a knife, a surgeon with a pair of scissors, an architect with a drawing pen and so on).

These differences between verbal and multimodal transcripts account for the fact that, while thanks to the standardization of verbal transcripts it is possible to produce a quite homogeneous transcript of an entire conversation; this is practically impossible for multimodal transcripts. In this sense, multimodal transcription is, more than any other, a form of protoanalysis: it is the result of an analytical eye on the data; it allows the researcher to inscribe this protoanalysis and then to further enhance it.

Ultimately, video data remain the primary reference source for any analysis – transcripts being a secondary source. This generates analytical practices in which the researcher constantly moves back and forth between the video and the transcript in progress. This movement is facilitated by some software (called 'alignment software' – such as ELAN), which integrates within the same interface a video player and a writing tool, allowing the researcher to temporally associate details of the video with details of the transcript, as well as to measure and annotate segments of talk as well as of embodied conducts.

The following example, taken from my own data, shows how video data can be annotated and the analytical issues raised by their transcription. The fragment is extracted from a guided visit of a famous architectural site in France, comprising an ecological garden. The visit is guided by the chief gardener, Luc, who leads three visitors, Jean, Yan and Elise, through the garden. We join the visit as Luc is explaining general problems related to the use of pesticides, with

(5) (Mondada, ARCHIVIS/argus)

```
1  LUC    y a des limites, quoi:, là là on est# on est un
          there are limits,right, there there we are we are
 fig                                             #fig 4.1
```

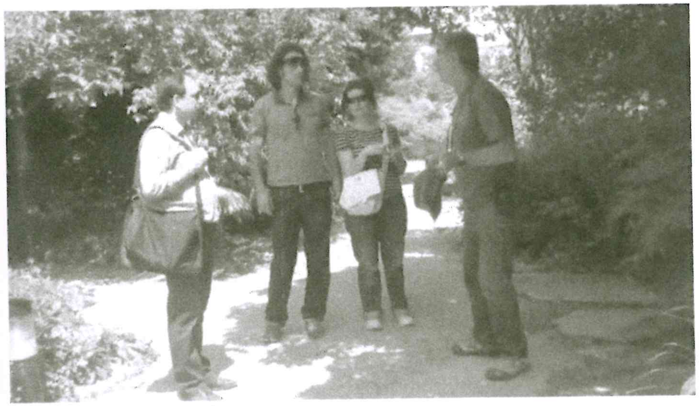

Figure 4.1 Transcript: Luc explaining how he manages the garden.

```
2      p§eu- *r'gardez +%le: #*§ .hβh+h #le papi%β©llon% +bleu là: *#+
       bit-  look at    the:    .hhh    the blue butterfly there
 luc   §.............................§points twd insect------------------>>
 luc       *one step forward*another step forward----------------*
 eli                   %looks--------------------%pivots%
 yan                  +looks------------------------------+pivots----+
 jea                          βturns H back---βpivots-------------->
 cam                                              ©pans to right--->>
 fig                        #fig 4.2  #fig 4.3              #fig 4.4

3      c'est unβ argus.  >voyez<?
        it's  an argus. >see<?
 jea          ->β
```

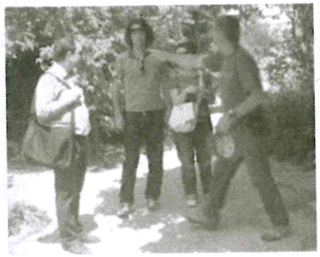

Figure 4.2 Luc begins

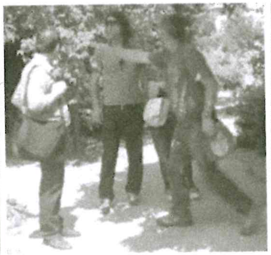

Figure 4.3 Luc

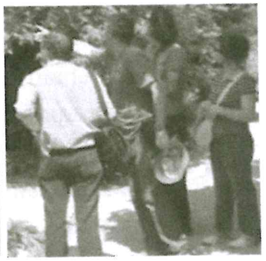

Figure 4.4 All look

The principles of the convention used (Mondada, 2014a, 2016, 2018) consist in paying special attention to the *temporality* of the multimodal conducts, as well as to their *form*. The temporality is indicated by precisely bracketing the beginning and the end of a movement (with the same symbol: for example, Luc's gesture ξpointsξ is bracketed by two ξ, which are reproduced within the talk in order to show how pointing relates both to the turn – here it is synchronous with the in-breath between two occurrences of the article ξ.hhhξ – and to the other conducts of the speaker or the co-participants). The form is described between these two symbols delimiting the movement and is augmented by images (screenshots extracted from the video) – which are also precisely temporally positioned within the transcripts (with the symbol #). This allows for a fine-grained analysis of the ordered contributions of the multimodal resources used.

At the beginning of the extract, the participants' bodies are arranged around Luc, looking at him (Figure 4.1) while he is engaged in his explanation (1–2). When he spots the butterfly, he interrupts his ongoing talk ('peu-'/'bit' [2]), and instead utters an imperative ('r'gardez'/'look' 2) while beginning to point at the object referred to. We notice that Luc does not just use a gesture that co-occurs with its verbal reference. Rather, he moves his entire body, stepping forward (Figures 4.2–4.4) in such a way that he extends not only his finger but also his body – his entire body contributing to the visibility of his pointing gesture. Moreover, the way he utters the name of the insect (2) is characterized by a stretched article, an in-breath, the self-repair of the article, followed by the name and a locative deictic ('le:.hh le papillon bleu là'/'the:.hh the blue butterfly there' [2]). His turn is formatted in such a way that it is expanded during the time it takes the co-participants to change their bodily and spatial positions. They progressively look at him and pivot their bodies, so that by the end of Luc's utterance (end of line 2) they are all reoriented towards the referent. The last to move is Jean (Figure 4.4). The cameraperson moves too, panning towards the left. This annotation integrates the action of the cameraperson in the transcription, treating her as another participant adjusting and responding to the ongoing action, interpreting it in real time (Mondada, 2016).

Thus, the progressivity of Luc's turn is *reflexively* organized with respect to the co-participants' responses: he adjusts to them as they respond to it. As soon as they all look at the butterfly, he names it and ends his utterance with an accelerated '>voyez?<'/'>see?<' (3) – both forms expecting that the organization of Luc's turn and the responses of the co-participants shows that temporality and sequentiality are the fundamental principles governing social interaction. Sequentiality might not be organized turn by turn, strictly successively, but rather in parallel flows of action, as emergent embodied conducts respond to a previous action and unfold simultaneously with it. Contrary to interactions analysed within an exclusive focus on talk, this implies a plurality of temporalities and sequentialities progressing at the same time. Temporality is crucial for the understanding of language and the body in interaction: syntax and body movements are finely organized as emergent *multimodal Gestalts* (Mondada, 2014b), temporally coordinated within and between speakers. On the one hand, these Gestalts are organized in systematic ways: it is possible to identify their recurrence and to describe their multimodal praxeological grammar. On the other hand, they are deeply and indexically embedded in the specific ecology of the activity, since they are adjusted to its material and spatial environment (including the way the socio-institutional context is materialized).

This has consequences for the study of language and grammar. Multimodal transcripts contribute to a better understanding of topics for which linguists have classically recognized the importance of the interplay between language and the body (such as deixis, which has always been understood as articulating speech and gesture, but for which detailed multimodal interactional accounts are still scarce). But they also contribute to a wider range of linguistic topics (such as spoken syntax, which has been recently understood as temporally emergent and incremental, but for which a multimodal account remains to be provided).

Transcription also plays a central role in defining current challenges for the study of multimodality in interaction. These challenges enlarge the array of resources that have been included in transcripts until now, as well as the variety of activities that can be analysed in this way – such as mobility, multi-activity, writing and the use of technologies, as well as silent activities. As can be seen in the extract above, mobility constitutes an important dimension complementing the rather sedentary vision of social life favoured in many audio and video recordings (Haddington et al., 2013), opening up questions around the organization of sequentiality, multimodality and language on the move. Multi-activity, as a particular praxeological configuration in which participants engage in more than one activity at the same time, constitutes another challenge (Haddington et al., 2014), since simultaneous lines of action make temporality more complex, multiplying the relevant practices involving manual writing, typing at the computer or using other technologies in interaction, constitutes a further challenge. They embed embodied conducts that are privately designed and public activities, as well as different forms of public visibility and accessibility (such as writing private notes and reading them aloud during a collective working session) (see Mondada and Svinhufvud, 2016). For instance, the study of writing as an embodied conduct in interaction allows scholars to integrate analyses of textuality within analyses of social interaction by considering how co-participants actually manage and coordinate the production, transformation and reading of texts in embodied ways. The same can be said of writing and communicating by means of digital technologies (e.g. Luff and Heath, 2015). Finally, video analyses and multimodal transcripts allow scholars to go beyond the limits of language by considering moments in which participants silently engage in collective action. Silence – not only in the form of pauses or lapses within talk in interaction, but more radically in the form of silent collective activities – represents a form of embodiment without language, the analysis of which multimodal transcripts make possible (Mondada, in 2018).

Conclusion

This chapter has discussed several aspects of transcription practice and conventions, showing that they articulate both empirical and theoretical aspects. Ochs's (1979) claim that transcription is theory is evident in the consequentiality of many details of transcripts, such as the choices of orthographic or non-orthographic notations to use, the identification of participants, the translation and glossing of the language transcribed, as well as the annotation of interactional and multimodal features of talk and action. These theory-laden dimensions of transcriptions account for the fact that transcription practices are so different and specialized across academic groups – and are even a distinctive sign of membership in specific scientific communities. On the other hand, interdisciplinary dialogue and cross-fertilization, as well as standardization constraints coming from big data banks of spoken corpora, motivate the negotiation of common standards. Computer-supported interfaces can use multiple standards within multilayered transcripts integrating possible alternatives in different lines

versions, intonation curbs, and other prosodic features, glosses, translations and multimodality. But although technology can open up transcription choices, in the end analysts have to make the ultimate decisions when selecting details to be analysed and shown in published extracts.

As has been seen in this chapter, the choices motivating transcripts – at very different levels – have practical/technological, conceptual/analytical and ethical/political consequences. They emanate from conceptualizations (which might be very explicitly stated or remain largely implicit) of what language, orality and interaction are and allow scholars to elaborate empirical analyses that are in line with these conceptions. They also integrate a certain view of who a speaker or a participant is, encapsulating issues not only of identity but also of voicing – by recognizing and respecting the speakers' voices or by rewriting and revoicing them within complex relations of authority and subordination, embedded within choices of norms and standards. This shows how serious the practice of transcribing is.

Acknowledgements

I warmly thank Mary Bucholtz and Isabelle Léglise for allowing me to use their transcripts and for their inspiring thoughts about transcription.

Further reading

Bucholtz (2007)
A comprehensive presentation of the challenges of transcription in sociolinguistics.

Hepburn and Bolden (2017)
A very clear and pedagogical introduction to the Jeffersonian way of transcribing in CA.

Mondada (2018)
Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality.

Online resources

Transcription conventions mentioned

Santa Barbara Corpus – see Du Bois (1991) and http://www.linguistics.ucsb.edu/projects/transcription/representing Conversation Analysis – see Jefferson (2004) and http://www.liso.ucsb.edu/liso_archives/Jefferson/Transcript.pdf

Multimodal transcription – see Mondada (2018) and https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf

GAT (Gesprächsanalytisches Transkriptionssystem) – see Selting et al. (1998, 2009)

ICOR (Conventions de Transcription, ICOR group) – see http://icar.univ-lyon2.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf

Other websites

IPA/International Phonetic Alphabet:
http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/

Leipzig Glossing Rules:
https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

Transcription Module/Conversation analysis (Schegloff):
http://www.sscnet.ucla.edu/soc/faculty/schegloff/TranscriptionProject/

Tools for alignment of transcriptions

Praat: http://www.fon.hum.uva.nl/praat/

ELAN: https://tla.mpi.nl/tools/tla-tools/elan/

CLAN: http://talkbank.org/software/

Transana: https://www.transana.com

Discussion questions

1. How are transcription conventions and linguistic theories related?
2. Search for examples of transcripts in the linguistic literature: Which options are visible in the way extracts are spatially disposed on the page, speakers are identified and their voices are textually represented?
3. In what sense can transcribing be seen as an *analytical* act? In what sense does it also have a *political* dimension?
4. Record a short moment of talk and experiment with different conventions for transcribing it. Also try different audio and video players and software for listening to and inspecting the data. Reflect on the differences you experience in using them.

References

Bailey, G., Tillery, J., and Andres, C. (2005), 'Some effects of transcribers on data in dialectology', *American Speech* 80(1), 3–21.

Bergmann, J. R. (1985), Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie, in W. Bonss and H. Hartmann (eds.), *Entzauberte Wissenschaft*. Göttingen: Schwarz, 299–320.

Bergounioux, G. (1992), 'Les Enquêtes de Terrain en France', *Langue Française* 93, 3–22.

Blanche-Benveniste, C. and Jeanjean, C. (1987), *Le Français Parlé*. Edition et transcription. Paris: INALF.

Bourdieu, P. et al. (1993), *La Misère du Monde*. Paris: Seuil.

Bucholtz, M. (1998), 'Geek the girl: Language, femininity, and female nerds', in N. Warner, J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, and M. Chen (eds.), *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group, pp. 119–131.

Bucholtz, M. (2000), 'The politics of transcription', *Journal of Pragmatics* 32, 1439–1465.