Doing Conversation Analysis: Practical Issues in Recording, Transcribing and Analyzing Data

3.0. INTRODUCTION

I ended the previous chapter by outlining what I believe a CA-oriented methodology for a social interactionist approach to SLA studies should look like. In this short chapter, I backtrack a bit to provide readers who are not trained conversation analysts with a basic, rather practical review of current CA methodological practices. In so doing, I am of course not claiming that anybody who reads this chapter will thereby become instantly qualified to undertake CA research. Learning to become a skilled CA researcher minimally entails completing at least one year of course work in CA (Hopper, 1988), ideally followed up by a continuing apprenticeship with an established CA practitioner. Nonetheless, this chapter provides readers who are not familiar with how CA research is done with a more informed sense of how conversation analysts record, transcribe, and analyze conversational data.

3.1. RECORDING DATA

The analysis of talk-in-interaction minimally requires the use of audio or, preferably, video, recordings of participants' talk to capture the tremendous complexity of conversational behavior. These recordings constitute the primary source of data used by conversation analysts. In the early days of CA, audio recordings were the preferred medium. This choice was in large part dictated by the expense and bulkiness of video equipment in the late 1960s and early 1970s, and it is no accident that a lot of early CA studies used telephone conversations as their principal source of data (see, e.g., Schegloff, 1968). One of the best known examples of telephone talk is the *Two Girls* transcript, excerpts from which may be viewed and heard on the World Wide Web (WWW) at http://www.sscnet.ucla.edu/soc/faculty/schegloff/prosody/.

sis: ing, Data

a CA-oriented studies should ovide readers ather practical oing, I am of r will thereby Learning to deting at least ollowed up by practitioner. familiar with now conversaal data.

e use of audio o capture the ese recordings ation analysts. erred medium. d bulkiness of and it is no nversations as). One of the irls transcript, orld Wide Web off/prosody/.

By focusing on talk that occurs during telephone calls, researchers were able to circumvent two problems rather neatly. First, the expense involved in conducting CA research was reduced to manageable levels because audio was already a mature recording technology. Second, because telephone partners do not have access to each others' facial expressions and gestures, audio recordings are well suited to capturing how participants display their mutual understandings to each other by voice alone. However, with the advent of cheaper and better video equipment, video recordings are now the medium of choice, as they allow researchers to see how phenomena such as the direction of participants' eye gaze, facial expressions, and gestures are coordinated with, and indeed are part of, the structure of talk-in-interaction (see Goodwin, 1979, for a seminal early example of research based on video recordings).

3.1.1. Gathering usable data

When audio or video recordings are made of talk-in-interaction that involves a small number of participants — for example, three or four people talking at the dinner table — the recording problems are relatively simple. Typically, one audio cassette recorder or video camera with an external microphone for better sound quality is used to capture the interaction. "High tech" solutions may use radio microphones or specialized directional microphones supported by booms above or to the side of participants. Obviously, this approach is expensive and may not be within financial reach of many researchers. Fortunately, however, whereas high tech solutions yield clearer data when properly used, there is no reason why lower tech solutions should not also yield acceptable recordings.

Other types of talk-in-interaction are more difficult to record adequately and therefore need to be recorded somewhat differently. For example, in classroom situations where small group work is the principal form of learning activity, recording the interaction with a single video camera (even with an external microphone) does not yield usable data. Although the teacher's announcements and other teacher-to-whole-class interactions (e.g., question and answer routines) may be adequately recorded, the various groups' intragroup conversations yielda babble of untranscribable noise.

In order to record these kinds of interactions, each group must be separately recorded. Depending on the researcher's financial and technical resources, separate recordings of each group can be made by using a stations approach (i.e., having one tape recorder per group) or by recording each student separately. The latter solution is preferable, particularly when students are required to move around the

Conversation Analysis Numa Markee

The final camera copy for this work was prepared by the author, and therefore the publisher takes no responsibility for consistency or correctness of typographical style. However, this arrangement helps to make publication of this kind of scholarship possible.

Copyright © 2000 by Lawrence Erlbaum Associates, Inc.
All rights reserved. No part of the book may be reproduced in any form, by photostat, microfilm, retrieval system, or any other

means, without prior written permission of the publisher.

Lawrence Erlbaum Associates, Inc., Publishers 10 Industrial Avenue Mahwah, NJ 07430

Cover design by Kathryn Houghtaling Lacey

Library of Congress Cataloging-in-Publication Data

Markee, Numa.

Conversation analysis / Numa Markee

p. cm. — (SLA research)

Includes bibliographical references and index. ISBN 0-8058-1999-1 ©: alk. paper). — ISBN 0-8058-2000-0 (pbk.: alk.

paper).

1. Conversation analysis. 2. Second language acquisition I. Title. II. Series.

P95.45.M35 2000 302.3' 46—dc21

99-39744

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America 10 9 8 7 6 5 4 3 2 1

classroom, as a stations approach is not well suited to capturing the talk that occurs as students move from one station to another. Again, there are both high and low tech solutions available to capture individual speakers' contributions during concurrent intragroup talk Each participant can be equipped with a radio microphone. Alternatively, each individual may be given a Walkman-sized cassette recorder and lapel microphone that records their talk. Because the equipment is highly portable and can be battery-operated, this solution has the added advantage of allowing participants to move around a classroom without having to worry about tripping over cables, etc. Whichever solution is used, the availability of multiple audio recordings means that such technically significant information as the difference between pauses and in-drawn or exhaled breaths, the specific number of laughter tokens, the precise onset and resolution of overlaps, the content of muttered commentaries, and other such fine details can be distinguished and therefore transcribed with a high degree of confidence. A Real Audio file of a recording made using the Walkman recorder solution may be heard on the WWW at http://deil.lang.uiuc.edu/class.pages /EIL367/RealAudio/Coral14.ram. The accompanying transcript may be found in PDF format at http://deil.lang.uiuc.edu/class.pages /EIL367/RealAudio/Coralcollection.pdf.

This multiple recordings technique also has the advantage of providing plenty of insurance. For example, if one group member's recorder jams, is turned off, or yields unintelligible data for any reason,2 the other members' recorders continue to pick up the talk in the group. The disadvantage of this technique is that it yields vast amounts of data, which therefore take longer to transcribe because so much more detail is hearable on the tapes.3 Furthermore, because the resulting transcript for each group is produced from a mosaic of overlapping data, it is possible to introduce errors that do not occur when working from a single tape of the interaction. For example, because there are minor variations in the speed at which each individual recorder records the interaction, the length of a pause on one tape can be slightly different from that on another tape. The reseacher must be consistent in dealing with this type of problem. In my transcription practice, for example, I time pauses by using the current speakers' tape as the definitive data source.

3.2. TRANSCRIBING DATA

In order to produce high quality transcripts, it is essential to use a good transcription machine. Transcription machines have heavy-duty mechanisms, which will not break down under the constant searching for particular passages of talk. Most good machines have foot-

capturing the other. Again, le to capture tragroup talk. one. Alternassette recorder equipment is has the added a classroom Whichever rdings means ence between per of laughter he content of distinguished nce. A Real order solution du/class.pages ranscript may i/class.pages

advantage of sup member's data for any up the talk in it yields vast be because so to because the a mosaic of do not occur. For example, which each of a pause on tape. The problem. In by using the

to use a good heavy-duty ant searching s have footoperated forward and reverse controls, which free up the transcriber's hands for data entry. They also allow voice pitch and tape speed to be manipulated, which can help transcribers understand otherwise unintelligible words or phrases, and may automatically reverse the tape a set distance after the stop button is pressed to facilitate multiple hearings of a difficult passage. In my experience, transcription machines are most effectively used in conjunction with high-quality external speakers rather than with headphones, as external speakers tend to have a better dynamic range than headphones.

Before discussing how CA workers set about preparing recorded data for analysis, I first briefly review how researchers working in the better-known quasi-experimental tradition of SLA studies typically view the job of transcription. Then I comment on the status that transcripts have qua data within a nomothetic epistemology. In this latter approach to doing ACD, it is fair to say that transcription is viewed as a tedious, mechanical task, which may be safely delegated to graduate students or junior collaborators. Furthermore, although producing a transcript is obviously a necessary first step, the resulting transcripts have low status as data. In an approach that privileges the quantification of observed phenomena, transcripts are raw data that must be quantified and statistically manipulated in order to yield valuable information. Indeed, it is the statistical tests and interpretations of why results did or did not reach predetermined levels of statistical significance, rather than the transcripts themselves, that are of primary interest to experimentalists.

In contrast to experimentalists, CA workers regard audio/video tapes and the resulting transcripts as their primary sources of data,4 and thus almost always do their own transcription. This is because transcription is viewed as an essential part of the discipline of doing CA. The fine level of detail demanded by CA transcripts inevitably makes transcription a time-consuming business. For example, a dyadic conversation lasting one hour is relatively easy to transcribe in that there is little extraneous noise. However, even this kind of talk-ininteraction may take up to 20 hours to transcribe (van Lier, 1988). Clearly, the types of talk that occur during small group work are even more difficult to transcribe because analysts have to transcribe each group as a separate conversation. They must therefore exclude overheard talk from other groups, except, of course, when the participants themselves orient to what is going on in another group. Consequently, it may take up to 40 hours to transcribe one hour of recording for each group. Thus, when nine different groups or parties are formed during the course of a single one-hour lesson (see the background information provided in chapter 8), it may take some 360

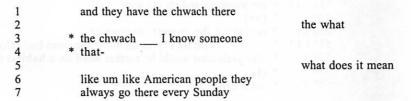
hours to produce all nine transcripts of talk that occurs during this period of time. Even then, analysts may decide to go back to the recordings later to do further transcription work as issues that only emerge after a long acquaintance with the data emerge to the fore. In a real sense, therefore, a transcript is never finished; it is only a working hypothesis about how participants construct talk in real time.

Time-consuming and tedious though transcription may be, it is a necessary part of doing CA. From a practical standpoint, it is the close engagement with the data that enables analysts to know their data in intimate detail. Furthermore, from a more abstract, epistemological perspective, conversation analysts insist that transcripts are, in fact, important preliminary theoretical statements about what talk-in-interaction is (see Ochs, 1979; although Ochs works within a language-socialization perspective that is somewhat different from that of CA in some important respects, language-socialization specialists and CA workers are in complete agreement on this point). Of course, this is not to say that experimentalists therefore deny that transcripts reflect a particular theoretical orientation. After all, all transcripts are designed to represent in a convenient fashion the empirically observed phenomena that are of theoretical interest to the researcher. However, whereas transcripts are merely a means to a statistical end for experimentalists, the notion of transcripts-astheoretical-statements achieves the status of a fundamental tenet of CA.

Given this essential difference in attitudes toward the relative status of transcripts qua data, it is not surprising that SLA transcripts such as the one exemplified in Excerpt 1.2 (reproduced here as Excerpt 3.1) rely on a relatively gross level of detail, which is largely limited to reproducing the words that participants spoke. For example, the only information included in this transcript about the quality of the talk produced by participants consists of the cut-offs at lines 4 and 10 and the pauses of unspecified length (represented by dashes) that occur at lines 3, 9, and 13,).5

Excerpt 3.1

Learner (NNS English) Interlocutor (NS English)



's during this back to the ues that only the fore. In a aly a working ime.

may be, it is oint, it is the o know their ore abstract. insist that al statements h Ochs works what different -socialization in this point). ore deny that After all, all : fashion the nterest to the means to a anscripts-asental tenet of

d the relative A transcripts iced here as ich is largely spoke. For ipt about the he cut-offs at presented by

S English)

mean

8		yes?
- 9	* you knowevery morning	
10	* that there pr- that- the American	
11	people get dressed up to go to um	
12	chwach	
13	*	oh to church I see
	(Pica, 1987, p. 6).	CERTIFICATION TO THE PROPERTY OF THE PROPERTY

This is not to say that experimental researchers who are interested, for example, in quantifying the occurrence of pauses as precursors of repair do not produce much more detailed transcripts that capture this particular phenomenon, it is just that CA transcripts do not predetermine what phenomenon is going to be studied. Instead, CA transcripts routinely provide extremely detailed information about what people say and how they say it. For example, as shown in Excerpt 2.4 (reproduced again here as Excerpt 3.2), CA transcripts provide detailed information about phenomena such as pauses and silences (lines 520–521; 524–525; 531–532; 534–535; 537–539) stress (lines 520, 541), lengthening of vowels (lines 528, 532, and 536) and cut-offs (line 520), overlaps (lines 526/527; 529/530; 532/533; 536/537, and 540/541), laughter tokens (line 537), in-drawn breaths or exhalations (line 526), and the affect that accompanies a particular utterance (line 541).

Excerpt 3.2

```
520 L11: * ok (+) excuse me (+) uh: what what does it mean hab- (+) habi-
521
522 T:
            habitats
523 L11:
            habitats
524 T:
          * yeah (+) you had that word as well (+) what do you think it means
            (++)
525
526 L10: *
           <hhh>//hh//
                    //you// all spoke about habitats didn't it
527 T:
528 L10: * uh:m
            the //m//ost important (1) habitat
529 T:
530 L10:
                 //I//
531
532 L10: * I think (+) the habitats is the:[əm] (+) e//nvironment uh// and the
533 L9:
                                                     //environment//
534 L10: * environment and uh (1) uhm
535
            (++)
536 L9:
          * is it is //it the: nearest environment//
537 L10: *
                      //for for (+) for the fish// you (mea be:) (hh)
538 T:
          * <h> yeah what would be another word for a habitat then (+) it's like
539
540 T:
          * //it's hli-//
```

541 L11: * //I ha//ve no idea ((in an exasperated tone)) (NM: Class 1, Group 3)

The transcript in Excerpt 3.2 conveys much more information than what is commonly included in SLA transcripts. However, it is not particularly detailed by the latest transcription standards in CA. As shown by Excerpt 3.3, the advent of ever more powerful computer technology has made it possible to include stills from video clips to illustrate graphically what could otherwise only be conveyed by a verbal gloss (Goodwin, 1999).

Excerpt 3.3

1 Pam: Okay that should be, wet enough.

2 (1.5)

3 Pam: ° Hmph (0.7) ((holding trowel))

4 Jeff: We're lookin at that right there?

5 (0.3)

6 Pam: Mmm,

7 (0.4)

8 Jeff: Much darker than that.

9 Pam: Yeah. I'm not-=

10 Jeff: There

11 Pam: =I'm just tryin ta put it in the:re.=

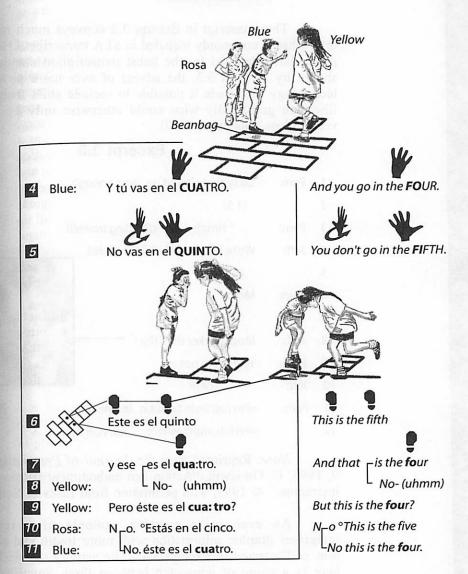
12 =eh hih an(h)ywhere. °hih heh huh

Munsell Book

Note: Reprinted from the Journal of Pragmatics, Vol. 32, No. 9, 1999, C. Goodwin, "Action and embodiment within situated human interaction" © 1999, with permission from Elsevier Science.

An even more complex example of a transcript which integrates graphic information with more traditional textual information is illustrated in Excerpt 3.4. The activity that is being examined here is a game of hopscotch between three Spanish-speaking girls, during which one participant (Blue) accuses another (Yellow) of cheating (Goodwin, 1999). This transcript includes the usual CA information about what is said and how it is said. In addition, it integrates detailed graphic information about how and when the participants use hand gestures to indicate square numbers four and five, eye gaze information, the position of the left hand girl's feet as she

Excerpt 3.4



Note: Reprinted from the Journal of Pragmatics, Vol. 32, No. 9, 1999, C. Goodwin, "Action and embodiment within situated human interaction" © 1999, with permission from Elsevier Science.

e information rever, it is not ls in CA. As ful computer rideo clips to inveyed by a



'unsell Book

, Vol. 32, No. tuated human ice.

script which tual informaing examined peaking girls, (Yellow) of he usual CA a addition, it and when the four and five, 's feet as she

says no vas en el QUINTO, and her performance of what Goodwin calls a deictic stomp (i.e., stamping her foot) to emphasize the square in which her interlocutor should have landed.

Obviously, these kinds of transcripts are pushing the limits of what information is technically possible to communicate to readers on However, there is no doubt that WWW-based electronic transcripts are able to convey even more information, even more more elegantly than conventional paper transcripts can today. As seen Excerpts 3.3 and 3.4, electronic transcripts are still in their infancy. However, as electronic journals gain more acceptance in academia as a serious means of disseminating scholarship, the kind of information contained in Excerpts 3.3 and 3.4 will become both common-place and interactive. Not only will researchers be able to listen to the original data, and thus be able to evaluate the accuracy of the original researcher's transcription, but hot links from verbal glosses will call up streaming video files of how participants executed a particular hand gesture or deictic stomp. In classroom contexts, hot links will allow researchers to see what materials teachers and students are using, and read, hear, and see how learners interpret and perform the tasks set by the teacher or materials.

Of course, SLA researchers will need to critically evaluate whether displaying data in this way adds anything to a substantive understanding of fundamental issues in SLA studies or whether researchers are merely being seduced by the considerable attractions of a technological Brave New World. I believe that these very detailed descriptions of both conversational and gestural context are an important key to a better understanding of SLA processes. For example, if the talk that occurs in Excerpt 3.4 were between a NS (Blue) and a low level NNS of Spanish (Yellow), we would be able to document how the participants' use of gestures and language combined to provide Yellow with highly contextualized, and thereby presumably comprehensible, input about the number system in Spanish and the language of spatial relationships.

Also in need of discussion and development are clear ethical standards concerning what information is suitable for dissemination via the WWW and how this information is to be displayed. Whereas it is relatively easy to camouflage participants' true identities with paper transcripts, information that can be included in electronic transcripts makes the problem of protecting participants' identity much more complicated. In the end, however, whatever consensus emerges on these issues, it is safe to say that the trend toward greater complexity in CA transcripts will also affect how data are displayed in SLA

odwin calls e square in

ne limits of readers on electronic even more ly. As seen eir infancy. idemia as a nformation n-place and he original ie original will call up cular hand will allow using, and asks set by

y evaluate substantive or whether ractions of ry detailed ext are an sses. For 'een a NS be able to combined bresumably th and the

ear ethical ination via nereas it is with paper transcripts nuch more merges on complexity transcripts, even if CA is not ultimately accepted as a viable methodology for SLA studies.

Finally, I address one aspect of CA transcripts that is somewhat controversial. In most of the excerpts that are reproduced in this book, standard English spelling, supplemented by the occasional use of phonetic script to capture non-standard pronunciations of words, is used to represent participants' talk (see, e.g., line 532 of Excerpt 3.2). However, the usual practice in most CA transcripts is to avoid using phonetic script, on the grounds that this is a tool of etic research. Consequently, as shown in Excerpt 3.5 from the *Two Girls* transcript, the characteristics of the participants' New York accents and various sandhi phenomena are represented through phonetic spellings (see lines 1, 2, and 5) invented by the transcriber.

Excerpt 3.5

1 Ava:

* I'm so:: ti:yid.I j's played ba:ske'ball t'day since the

* firs' time since I wz a freshm'n in hi:ghsch[ool.]

3 Bee:

5 Ava:

* I'm so:: ti:yid.I j's played ba:ske'ball t'day since the

[Ba::]sk(h)et=

[Ba::]sk(h)et=

[Yeah fuh like an hour enna ha:lf.

This has led a number of writers to criticize the use of "funny English" in CA transcripts on the grounds that it is demeaning to the participants (Preston, 1982, 1985) or an inconsistent means of representing participants' talk (Edwards, 1992). The issues surrounding how to transcribe participants' talk are in fact quite complex. (For a general overview of the technical issues involved in transcribing oral data, including critiques of CA transcription conventions, see Du Bois, 1991; Edwards & Lampert, 1993; Green, Franquiz, & Dixon, 1997; Ochs, 1979; Roberts, 1997). It is important for anyone working with L2 data to make a principled decision about how to transcribe nonstandard pronunciations of English words. I find Preston's arguments in this regard to be compelling. My own preference when transcribing the talk of second language speakers of English, therefore, is to try to avoid any suggestion of ridiculing how participants talk by using standard English spelling, supplemented by phonetic script as appropriate.

3.3. ANALYZING DATA

The best way of understanding how CA researchers construct a conversation analysis is to follow an analysis as it is being constructed. I offer readers this opportunity in chapters 7 and 8. A few brief

pointers that highlight the most characteristic ways in which arguments are constructed may be useful. As already noted in chapter 2, there are four defining characteristics of CA:

 It adopts a radically emic approach to research which, unlike ethnographic approaches to ACD, avoids the use of secondary data

It generally avoids all but the most basic forms of

quantification.

trelies on analyzing prototypical examples of a particular phenomenon, using different kinds of text-internal, convergent evidence to establish the credibility of an analysis (this, to use ethographic terminology, amounts to what might be called text-internal triangulation of the data).

It seeks to demonstrate that potential counterexamples have been anticipated and encourages other researchers to replicate

initial findings with different sources of data.

In this section, I wish briefly to explain how CA researchers usually set about unpacking the structure of conversation.

3.3.1. Unpacking the Structure of Conversation

Generally speaking, CA unpacks the structure of conversation by analyzing either single cases or collections of talk-in-interaction. With single cases, the objective is to provide an in-depth analysis of a particular phenomenon that facilitates a deep understanding of how the phenomenon under study works. Analyses based on collections of similar data (e.g., particular types of repair) enable the analyst to see whether the practices to which participants are thought to orient are robust enough to account for a broad range of data gathered in different conversational contexts. Occasionally, as Schegloff (1968) demonstrated, these accounts have to be changed in order to account for one recalcitrant example that does not fit the pattern established for the overwhelming majority of other cases.

In a variation on this basic distinction between single-case and collection-based analyses of talk-in-interaction, I use in chapters 7 and 8 what may, at first glance, look like a conventional collection-based methodology. However, the collections I use as the databases for the analyses that are constructed in these two chapters are rather different from their traditional brethren and are also used to fulfill different analytical purposes. More specifically, these collections consist of thematically related sets of talk that occur during the course of two

ys in which ted in chapter

which, unlike of secondary

c forms of

a particular rnal, converlysis (this, to ight be called

amples have s to replicate

ers usually set

n

versation by raction. With analysis of a ding of how collections of malyst to see to orient are gathered in egloff (1968) er to account n established

ngle-case and napters 7 and lection-based bases for the ther different fill different is consist of ourse of two different lessons. The point of constructing these collections is not to collect similar conversational objects for the purpose of comparing how individual examples of a particular phenomenon demonstrate participants' orientations to common behavioral practices; rather, it is to document the longitudinal elaboration of members' understanding and learning behaviors over specific periods of time and to demonstrate how the same methodology is powerful enough to identify both successful and unsuccessful attempts to understand and learn new language.

3.4. CONCLUSION

In this short chapter, I have briefly outlined the techniques conversation analysts use to record, transcribe, and analyze CA data. I have also briefly touched on some ethical issues that need to be discussed as on-line publishing of conversational data on the WWW becomes more feasible and more common. Part II discusses how the construct of interactional competence is constituted in and through participants' talk.

NOTES

- 1. I am assuming here that the talk that is being recorded consists of face-to-face interaction in a conventional classroom. In electronic classrooms, however, students may be recorded by using each computer station's microphone. Furthermore, if the focus of the research is to analyze participants' turn-taking practices as they use synchronous conferencing software, the program's archiving facility provides a built-in way of generating transcripts of the interaction.
- 2. For example, in one class that I recorded, one of the participants, a Muslim woman, was wearing a veil, which frequently brushed against her lapel microphone when she moved her head. This produced a lot of white noise, which, in places, severely affected the intelligibility of the talk recorded on her tape. However, I was eventually able to transcribe most of her talk by listening to her partners' tapes, which also picked up her contributions to the conversation.
- 3. Of course, there is no guarantee that recordings that are made using the low tech Walkman solution in particular will not include talk from other groups, which can make transcribing such talk a nightmare. However, this low tech recording technique does have the advantage of giving the researcher a sense of where one group is in relation to

another because it picks up parts of other groups' conversations.

4. Occasionally, constructed examples or, alternatively, examples based on talk that the researcher overheard but was unable to record (sometimes identified as field notes; see, e.g., the discussion of the "What is death"? sequence in chapter 7, which is an example of how field notes are used by conversation analysts) are also used. There is no real substitute for attested talk, however, and the use of these other kinds of data is best avoided whenever possible.

5. If Ochs' (1979) argument is valid (i.e., that the left hand position in a transcript iconically represents the most important speaker in a conversation), then this transcript would identify the NNS as the most important member of this NNS–NS dyad. Compare this representation with how the speakers are represented in Excerpt 1.1. Here, the speech of T, the adult NS, is in the left-hand column, whereas that of H, the child NNS, is located in the right-hand column.