Interobserver Agreement on First-Stage Conversation Analytic Transcription

FELICIA ROBERTS
Purdue University
JEFFREY D. ROBINSON
Rutgers University

This investigation assesses interobserver agreement on conversation analytic (CA) transcription. Four professional CA transcribers spent a maximum of 3 hours transcribing 2.5 minutes of a previously unknown, naturally occurring, mundane telephone call. Researchers unitized transcripts into words, sounds, silences, inbreaths, outbreaths, and laugh tokens, and then coded each of 1,827 units on as many as 15 transcription dimensions. Agreement was assessed using Cohen's kappa for nominal level data: Speaker designation, unit sequencing, semantics, orthography, cutoff, and plosiveness reached the level of "substantial" agreement (90% or greater accuracy). Pitch, overlap, doubt, and smile voice reached the level of "moderate" agreement (80–89% accuracy), while pace, sound stretch, underline/amplitude, and intonation fell below acceptability except when examined post hoc as presence versus absence of the feature. Silence lengths, examined as ratio-level data, were reliable at the "acceptable" level (alpha ≥ .70) among those using a counting method (as opposed to stopwatch or other mechanical means). We make recommendations for transcription training.

Conversation analysis (CA) is a qualitative approach to the study of the social organization of human interaction (Heritage, 1984; Levinson, 1983). Over the last decade, CA research has grabbed the attention of communication scholars, as evidenced by both the methodological debates it has engendered (Cappella, 1990; Jacobs, 1990; Pomerantz, 1990), as well as its representation in the discipline's core journals (Beach & Metzger, 1997; Clayman, 1992, 1995; Clayman & Heritage, in press; Hopper, 1989; Pomerantz, Fehr, & Ende, 1997; Robinson, 1998; Robinson & Stivers, 2000; Stivers, 2001; Stringer & Hopper, 1998; Thomason & Hopper, 1992). Inductive in nature, the process and product

Felicia Roberts (Ph.D., University of Wisconsin, 1996) is an assistant professor of communication of Purdue University in West Lafayette, IN. *Jeffrey D. Robinson* (Ph.D., University of California, Los Angeles, 1999) is an assistant professor of communication at Rutgers University. Correspondence concerning this article should be addressed to: Felicia Roberts, Purdue University, Beering Hall of Liberal Arts and Education, Room 2114, Department of Communication, 100 North University Street, West Lafayette, IN 47907-2098; email: froberts@purdue.edu.

Human Communication Research, Vol. 30 No. 3, July 2004 376–410

of CA research are principally grounded in communicators' orientations, which are themselves grounded in observable features of naturally occurring interaction. Conversation analysts understand that transcripts of interaction are at least third-generation versions of data (the first being in vivo interaction and the second being audio or videotapes of interaction) (Heritage & Atkinson, 1984; Ochs, 1979), and analyses are principally conduced in conjunction with audio or videotapes (Heritage & Atkinson, 1984; Sacks, 1984); nonetheless, transcripts are the primary source of evidence for and against analytic claims in published research. Even when audio or video data are published (e.g., Jones & LeBaron, 2002), transcripts nonetheless embody claims about what observers witnessed regarding participants' behavior. As such, CA rests fundamentally on transcripts being reliable.

This article reports on the reliability of—or, more accurately, intertranscriber agreement across—first stage CA transcripts. We begin by reviewing briefly the importance of transcription for CA and articulating our definition of first-stage transcripts. We then argue for the importance of establishing intertranscriber agreement and related studies are reviewed.

The Importance of Transcription for CA

CA assumes that "there is order at all points" in interaction (Sacks, 1984) and that "no order of detail . . . can be dismissed a priori as disorderly, accidental or irrelevant" (Heritage, 1989, p. 22). This assumption led, early on, to the formal development of a detailed transcription-notation system by Gail Jefferson (Sacks, Schegloff, & Jefferson, 1974). Using a wide range of orthographic and typographic conventions to represent phonemic, prosodic, and other vocalic (including some phonetic) detail, the system also has conventions for transcribing language's suprasegmental properties (e.g., speech rate), temporal features (e.g., silence), and utterance alignment (e.g., overlapping talk). In transcribing for this level of detail, CA research has demonstrated that communicators rely on (and are held accountable for) a range of otherwise unimaginable interactional practices for making sense in (and of) communication. These sense-making practices continue to refine (and sometimes challenge) our understanding of fundamental topics within the field of communication, such as interpersonal affiliation and distance (i.e., facework), power, identity, and theories about the function of communication. Transcription is thus a core analytic activity, a first step in developing deeper understandings of communication processes. In this study, we assess the reliability of this practice at the point of first-stage transcription.

First-Stage Transcripts

By first stage transcript, we mean a transcript of an audio portion of interaction produced by a trained CA transcriber, for whom the data is completely unfamiliar (i.e., not previously heard, discussed, transcribed, or analyzed) and for the purpose of serious and public analysis, such as a data session among professional colleagues. By first (i.e., "first stage" or "first draft"), we explicitly do not mean rough. A general rule of thumb is that first stage CA transcripts take approximately 1 hour to complete per minute of interaction transcribed.

It is true that some published CA transcripts are products of multiple stages or iterations of transcription (Wrobbel, 1998), and thus may be more reliable (and valid). This, however, must be qualified by two points. First, not all published CA transcripts have been systematically refined beyond the first stage. For example, for a variety of reasons (e.g., the inaccessibility of original audiotapes), researchers sometimes use data from preestablished pools of possible first-stage transcripts and cannot revise them or assess their reliability, in the sense of "correctness." Second, even when first-stage transcripts are refined, it may be done by the same transcriber, leaving open the question of intertranscriber agreement (see Kitsinger, 1998). For the forgoing reasons, establishing reliability of first-stage transcription practices is desirable. Furthermore, having a baseline level of first stage intertranscriber agreement is necessary for the following reasons: (a) It aids transcription training by illuminating, in a manner that is not confounded by subsequent stages of refinement, lacuna in foundational CA training; (b) it allows future research to answer the question of whether or not next stages of refinement improve agreement, such as focused retranscription by the original transcriber or the collaboration of multiple transcribers to produce a consensus transcript (Ferber, 1991); and (c) in a pragmatic vein, when proposing and defending CA transcription budgets to external reviewers, it is necessary to know how much transcription time is required to achieve acceptable levels of reliability.

The Importance of Establishing Intertranscriber Agreement

In Bakeman and Gottman's (1986) terms, transcription is a physically based coding process in that the units of analysis are physical, hearable features of sound production whose sense is largely based in a relatively standardized lexicon and thus whose representation requires relatively little inference. The increased reliability of physically-based coding schemata is relative to ones that involve cultural informants making inferences about social actions (Bakeman & Gottman, 1986), such as coders using transcripts to identify questions, positive affect, interpersonal power,

and so on. As a result of its physical bases, intertranscriber agreement has rarely been questioned. On the other hand, transcription is by no means a purely mechanical–physical process of replication. It is rather one of human re-presentation, with all of its entailing confounds (Ochs, 1979). Talk is also certainly susceptible to mishearings. Perhaps more consequentially, even when different persons hear the same events, there are a variety of reasons why they might listen to, process, and thus transcribe them differently (Ferber, 1991).

In general, intertranscriber agreement is important because it provides an index of the accessibility and reportablity of transcribed phenomena, which lay the groundwork for the reliability and validity of analytic claims based on such phenomena (Clavarino, Najman, & Silverman, 1995; Silverman, 2001). Several scholars have argued (from a radical interpretation of social constructionism) that reliability is not a relevant issue for qualitative research; however, this position has been reviewed and strongly refuted by qualitative scholars (Kirk & Miller, 1986; Silverman, 2001). CA findings need to be based on evidence (e.g., transcripts) that is sufficiently reliable according to widely accepted and expected social scientific standards, which guard against unwarranted empirical and theoretical claims. Despite the fact that these standards are tested quantitatively, they are no less relevant for qualitative research. The reliability of coding schemata decreases with their complexity (Bakeman & Gottman, 1986); therefore, reliability is an especially important issue for conversation analysts, who rely on a highly nuanced transcription system for their understanding of interaction.

Previous Studies of Transcription Reliability

Numerous studies have assessed intertranscriber agreement in terms of whether or not two or more transcribers render the same word (Irwin & Wong, 1983; Pye, Wilcox, & Siren, 1988); the statistic used has been percentage agreement. In the single previous study examining CA transcription, Patterson, Neupauer, Burant, and Koehn (1996) found 94% agreement in terms of verbal content. An immediate objection might be that percentage agreement is inappropriate, because it does not correct for chance agreement, which is promoted by the finite quality of coding schemata (Bakeman & Gottman, 1986). As language users have a virtually infinite capacity to create novel words (and to create them in situ), language contains a virtually infinite number of "codes" (Patterson et al., 1996). Thus, the process of determining whether transcribers reliably rendered the same word is not a matter of determining if transcribers "coded" it on a standardized, finite set of codes (and then testing their reliability

in applying such codes), but rather one of comparing transcribers' renderings in terms of agreement or disagreement.

On the other hand, the CA transcription system does contain finite sets of standardized conventions for rendering how words, sounds, and other types of "units" (e.g., breathing and silence) are produced by speakers (e.g., loudly or quickly). When testing the reliability of these features, Cohen's kappa is appropriate in order to control for chance agreement. Furthermore, if one is interested in whether or not transcribers can reliably sequence units, Cohen's kappa again becomes appropriate (Bakeman & Gottman, 1986). Patterson et al. (1996) tested intertranscriber agreement for intonation and overlap; however, they used only percentage agreement and found moderate results (82% and 86%, respectively). They did not test for agreement on unit sequencing.

This study examines interobserver agreement on CA transcription along 15 dimensions: *unit type*, *semantics*, *orthography*, *speaker designation*, *doubt*, *overlap*, *sound stretch*, *amplitude*, *intonation*, *cutoff*, *pace*, *pitch*, *plosiveness*, *smile voice*, and *length* (for silences, sound stretches, and breathing).[1]

METHOD

Participants

Transcribers were four faculty members employed at four different research universities; their academic backgrounds represent three different scholarly traditions (communication, linguistics, and sociology). All were formally trained in Jefferson's transcription system, but at different points in time and with different instructors. All have published CA research. Two of the transcribers were the present authors.

All transcribers were aware of (i.e., not blind to) the general purpose of the study (i.e., to study the reliability of CA transcription). There are three reasons why this traditional confound (Bakeman & Gottman, 1986) is less threatening to the current project. First, blinding is less necessary when utilizing more physically-based coding schemata. Second, the authors produced their transcripts prior to discussing and constructing the coding schema. Third, the authors were only two of four transcribers involved, and reliability estimates were averaged.

Data

Transcribers were supplied with an audiotape of a telephone call between two female, college-aged students who are close friends (Tina and Mara), both of whom spoke the North-Midland variety of American English (Wolfram, 1991). The call was taped by Tina in her home. According to an institutionally approved protocol for the Protection of Human Subjects, both women voluntarily consented to the recording of the call and its release to transcribers. One author listened to the call to ensure that it was reasonably acoustically clear; however, none of the other transcribers had previously heard the call. All four transcripts are provided in the Appendix.

Procedure

Transcribers were instructed to (a) transcribe the first 2.5 minutes of the call, beginning with the opening and ending with Tina's "Yeah" (see Appendix); (b) transcribe as they normally would, using their own equipment; (c) not spend more than a total of 3 hours on the project; and (d) use Jefferson's notation system as a baseline. Transcribers were also instructed to develop and use additional symbols if needed, and to provide a key for such symbols.

Construction of the Coding Schema

What follows is a brief summary of our coding schema (Table 1). A more extensive guide to coding decisions can be viewed online (see "Codebook" in the reference list for the URL).

Units of Analysis

Our units, arrived at inductively from the inspection of each transcript, were (a) word/sound; (b) silence; (c) inbreath; (d) outbreath; and (e) laugh token. Each transcript was coded into a sequence of these unit types. Example 1 contains two turns of talk. (Hereafter, representations of talk will be preceded by "T" and "L", which respectively identify the transcriber and the line number in their transcript.)

Example 1
(T3, L09): Uh::m (.) I:: talked ta Justin,
(T1, L66): he sounded pretty hungry: hhh huh huh, .hhh

The first turn was coded as having six units: "Uh::m" (word/sound), "(.)" (silence), "I::" (word/sound), "talked" (word/sound), "ta" (word/sound), and "Justin," (word/sound). The second turn had eight units: "he" (word/sound), "sounded" (word/sound), "pretty" (word/sound), "hungry:" (word/sound), "hhh" (outbreath), "huh" (laughter), "huh," (laughter), and ".hhh" (inbreath).

TABLE 1
Variable List, Definitions, and Symbols

Variable	*Definition/Symbol*
Unit Type	Each unit was coded as a word/sound, silence, inbreath, outbreath, or laugh token.
Word/Sound	Discrete vocalizations, including words (e.g., "hotel"), parts of words (e.g., "Y-"), and sounds ("Uhm").
Silence	Anything transcribed as a period inside parentheses (e.g., "(.)" ; a micropause) or seconds and tenths of seconds of silence within parentheses (e.g. "(1.6)").
Inbreath	Any unit comprising a "period" followed only by "h" (e.g., ".hh").
Outbreath	Any unit comprised solely of "h" letters (e.g., "hh").
Laugh Token	Representations for laughter (e.g., "huh", "hih", and "heh") were too numerous to specify a priori (See code book).
Nothing	In the comprehensive sequence of units, the possibility existed that transcribers did not produce a unit of the above type.
Semantics	Word/sound units were compared for whether they had the same meaning (e.g., "just" and "j'st" were similar; "right" and "I'm" were different).
Orthography	Word/sound units were coded as being transcribed in either standard or nonstandard orthograpy (e.g., "Just" and "for" were standard; "Jus'" and "fer" were nonstandard).
Speaker Designation	All units were coded for whether they were transcribed as belonging to Tina, Mara, or neither (in the case of interturn silence).
Doubt	All units were coded for whether or not their renderings were transcribed as being doubtful (symbolized by placing part or all of the unit in parentheses; e.g., "Gilli(an)").
Overlap	All units were coded for whether they were transcribed as being in overlap (or produced simultaneously) with other units. Units were coded as having no overlap onset (symbolized by the left bracket), onset at unit's beginning, or onset in between unit's beginning and ending.
Sound Stretch	Word/sound units were coded for whether they were transcribed as having pronunciations in which sounds were lengthened (symbolized by placing a colon after the lengthened sound). Units were coded as having stretch on first syllable, second syllable, third syllable, or more than one syllable.

TABLE 1 (continued)
Variable List, Definitions, and Symbols

Variable	*Definition/Symbol*
Underline/Amplitude	All units except silences were coded for whether or not they were transcribed as being stressed in some manner (symbolized by underlining part or all of the unit or using capital letters). Units were coded as having no stress, stress on first syllable, stress on second syllable, or stress on entire unit.
Intonation	Word/sound and laugh units were coded for whether or not they were transcribed as having a unit-final intonation shift. Units were coded as having no shift, falling intonation (symbolized by a period), slightly rising intonation (symbolized by a comma), and strongly rising intonation (symbolized by a question mark).
Cutoff	All units but silences were coded for whether or not they were transcribed as being cut off, or abruptly halted (symbolized by a dash after the cut off sound).
Pace	All units but silences were coded for whether or not they were transcribed as being produced with a different pace relative to surrounding talk. Units were coded as having no pace shift, faster pace (symbolized by surrounded unit with ">" and "<"), or slower pace (symbolized by surrounded unit with "<" and ">").
Pitch	Word/sound and laugh units were coded for whether or not they were transcribed as being produced with a different pitch relative to surrounding talk. Units were coded as having no pitch shift, higher pitch (symbolized by "↑"), and lower pitch (symbolized by surrounding the unit with "↓").
Plosiveness	Word/sound units were coded for whether or not they were transcribed as being produced with plosiveness, or bursts of exhalation, which is frequently associated with talking while laughing, crying, coughing, etc. Plosiveness is symbolized by placing the letter "h" in parentheses within, or at the boundaries of, a word/sound (e.g., "nice(h)").
Smile Voice	Word/sound units were coded for whether or not they were transcribed as being produced while smiling, which is an audible phenomenon, and which is symbolized by surrounding units with the British pound symbol ("£").
Silence Length (Counting)	Silences were timed by the counting method (e.g., "One one thousand . . .").
Silence Length (Machine)	This was mechanically timed, measuring distances in sound wave between endings and onsets of vocal periodicity.
Stretch Length	The length of sound stretches was measured in terms of the number of colons.
Breathing Length	The length of inbreaths and outbreaths was measured in terms of the number of "h" letters.

The Comprehensive Sequence of Units

The four transcripts comprised four separate sequences of units. Some transcribers included (i.e., observed and transcribed) units that others did not, and some transcribers did not include units that others did. When measuring agreement, in order to compare transcribers to each other in terms of the same phenomena, we needed to create a single, comprehensive (or all-inclusive) sequence of units (Stockman, Woods, & Tishman, 1981). For example, compare the same turn of talk as transcribed by T1–T4, where each unit is separated by a slash:

Example 2
(T1, L65): we / might- / go / get / something / t' / ea:t.
(T2, L65): we / might / (.) / go / get / something / to / ea:t
(T3, L60): we / might / (.) / go / get / something / to / ea:t.
(T4, L70): we / might / we / might- / go / and / get / something / to / ea:t,
(C. Seq.): we / might / (.) / we / might- / go / and / get / something / to /ea:t,

T1's transcript was coded into seven units, T2's and T3's were coded into eight units, and T4's was coded into 10 units. The comprehensive sequence (C. Seq.) of units for this stretch of interaction contained 11 units, incorporating all of the possible units as observed by T1–T4 (who only agreed on the presence of seven units). Transcripts 1–4 each contained 475, 437, 453, and 462 units, respectively, totaling 1,827 units. Each unit was then coded for as many as 15 transcription criteria (see Table 1). The comprehensive sequence contained 493 units.

Statistical Procedures

All data were entered into SPSS for cross-tabular analysis and computation of Cohen's kappa. The variables unit type, orthography, speaker designation, doubt, overlap, stretch, stress, intonation, cutoff, pace, pitch, plosiveness, and smile voice consisted of categorical data, and their reliability was tested with the kappa statistic. The variables silence length (counted), silence length (machine), stretch length, and breathing length consisted of ratio data, and their reliability was tested with the intraclass coefficient (two-way mixed model).

We measured reliability as a series of pairwise comparisons between the four transcribers, yielding six comparisons. We derived overall agreement as an average of these scores. For the variable "semantics" we used percentage agreement, because the coding method violated assumptions of Cohen's kappa (i.e., because the number of codes is virtually infinite, transcribers' versions of words and sounds were open-ended).

Prior to the work of Gardner (1995), researchers generally agreed on a single (albeit general) standard of acceptable kappa scores for social scientific research, that being .70 and higher. Even though kappa of .60–.69 was considered to be "good" (Fleiss, 1981) or "substantial" (Landis & Koch, 1977), anything less than .70 was to be treated "with some concern" (Bakeman & Gottman, 1986, p. 82), and .75 and higher was preferred (Bakeman, 2000). The general rule of thumb was that kappa scores of .70–.74 represent an adjusted agreement level of 80–89%, which is on the lower side of "acceptable" (Gardner, 1995), whereas .75 represents an adjusted level of 90% or higher.

In a recent advancement based on Gardner's research, Bakeman, Quera, McArthur, and Robinson (1997) refined the aforementioned general rules of thumb. Bakeman et al. demonstrated that, for identically fallible observers, acceptable kappa scores vary by the number of codes within a category (k) and the variability of their simple probabilities (π). Specifically, accuracy can be quite high (i.e., 90% or higher) with kappas quite low when k is small (e.g., when a category only has two codes) and when the distribution of codes within a category is disproportionate (e.g., when one code of a two-code category represents 90% of the observations). Bakeman et al. defined three levels of event probability: equiprobable (the rarest case; $\pi_i = 1/k$), moderately variable ($\pi_i = 0.5/k$), and highly variable ($\pi_i = 0.25/k$). They found, for instance, that when $k = 2$ and when the codes' probabilities are highly variable, kappas of .44 or higher represent 90% accuracy; however, when $k = 2$ and when the codes' probabilities are equiprobable, kappas of .61 represent 90% accuracy. When k is 5 or greater, Bakeman et al.'s model suggests that traditional rules of thumb apply. Regarding ratio-level variables, we used traditional standards of acceptability (.70 or greater) for intraclass coefficients.

RESULTS

When evaluating intertranscriber agreement on categorical variables, we used the refined estimates of Bakeman et al. (1997; see above). As seen in Table 2, results for categorical variables are separated into those for which there is "substantial" intertranscriber agreement (i.e., 90% or greater agreement), "moderate" agreement (i.e., 80–89%), and "unacceptable" (i.e., lower than 80% agreement). K and π are supplied for each categorical variable. As seen in Table 4, results for ratio-level variables are separated into those for which agreement is "acceptable" (kappa $\geq$ .70) and "unacceptable" (kappa < .70).

In accordance with principles of CA and its transcription system, coding for each variable was extremely nuanced. Thus, results are presented first in terms of the most conservative measures of reliability (see *Initial Results*), those analyses done at the finest level of coding detail. For instance, "sound stretch" had four codes: stretch on first syllable, second

TABLE 2
Intertranscriber Agreement Scores for Initial Categorical Variables

Transcription Variable	k[a]	π[b]	*T1 & T2*	*T1 & T3*	*T1 & T4*	*T2 & T3*	*T2 & T4*	*T3 & T4*	*Mean*
90+% Accuracy									
Speaker Designation	3	High	.80	.98	.95	.79	.78	.95	.88
Orthography	2	High	.43	.66	.46	.54	.69	.60	.53
Cutoff	2	High	.40	.66	.32	.35	.56	.36	.44
Plosiveness	2	High	.46	.67	.55	.62	.66	.58	.59
80–89% Accuracy									
Semantics	na	na	87%	93%	89%	88%	85%	81%	89%
Unit Type	6	High	.74	.79	.67	.77	.70	.65	.72
Pitch	3	High	.56	.77	.64	.39	.48	.65	.58
Overlap	3	High	.40	.54	.40	.38	.54	.43	.45
Doubt	2	High	.30	.39	.10	.21	.30	.32	.27
Smile Voice	2	High	.22	.77	.09	.23	.04	.06	.24
Below 80% Accuracy									
Pace	3	High	na	.25	na	na	.02	na	.13
Sound Stretch	4	High	.45	.63	.58	.39	.46	.58	.51
Underline/Amplitude	4	High	na	na	.27	na	na	na	.27

NOTE: [a] k = Number of codes within the given category/variable. [b] π = The variability of codes' simple probabilities: Equiprobable (Equi; π_i=1/k), moderately variable (Mod; π_i=0.5/k), and highly variable (High; π_i=0.25/k).

syllable, third syllable, or more than one syllable. After determining agreement levels for all variables, we reexamined those for which accuracy did not exceed 90%. We were interested in determining whether transcribers were attuned to phenomena, and including them in their transcriptions, yet perhaps at less refined levels. For example, although transcribers did not reliably transcribe sound stretch at the syllabic level, did they reliably hear that some sound stretching was occurring within a unit? To address this question we created several derived variables and present them as post hoc results.

Initial Results

Intertranscriber agreement for speaker designation, orthography, cutoff, and plosiveness were substantial. Levels of agreement for semantics, unit type, pitch, overlap, doubt, and smile voice were in the moderate range. Agreement on pace, sound stretch, underline/amplitude, and intonation was unacceptable at the finest level of detail. The level of agreement for silence (counting) was acceptable; levels for silence (machine), stretch length, and breath length were unacceptable.

TABLE 3
Intertranscriber Agreement Scores for Ratio-Level Variables

Transcription Variable	*T1 & T2*	*T1 & T3*	*T1 & T4*	*T2 & T3*	*T2 & T4*	*T3 & T4*	*T4 & C-Speech*	*Mean*
Acceptable								
Silence (Counting)	.69	.90	na	.70	na	na	na	.76
Unacceptable								
Silence (Machine)	na	na	na	na	na	na	.69	.69
Stretch Length	.70	.42	.26	.34	.26	.13	na	.35
Breath Length	.48	.60	.33	.47	.00	.73	na	.44

Post Hoc Results

To address the question of whether transcribers are capturing core phenomena, but in slightly different (or less nuanced) ways, we created a gross derived variable for unit type, semantics, overlap, sound stretch, pitch, pace, underline/amplitude, intonation, and emphasis; the variables doubt and smile voice could not be collapsed further (see codebook).

The variable unit type was originally composed of six categories: word/sound, silence, inbreath, outbreath, laughter, and nothing. Unit type (gross) collapsed the three categories dealing with (vocalized or nonvocalized) breathing (inbreath, outbreath, laughter). These phenomena can be transcribed quite similarly (see Table 1) and are not necessarily mutually exclusive categories. For example, plosive out-breath can be simultaneously heard and transcribed as a laugh token. Indeed, one transcriber marked some "laugh tokens" as simultaneously inbreaths [e.g., ".hheh" (T3, L 64) or ".heh" (T4, L 42)] and these were coded as laugh tokens (see codebook for decision rules).

The variable semantic originally represented all word/sound units, including those transcribed as being in doubt (by placing them in parentheses). Insofar as this convention (i.e., parentheses) specifically indicates doubt about the (semantic) nature of words/sounds, it is arguable that transcribers should not be penalized in cases where their renderings differ, yet the semantics of such renderings were transcribed as being in doubt. "Semantic (gross)" considered transcribers to be in agreement in cases where two renderings disagreed, yet one or both were enclosed within parentheses [e.g., "(Ok-)" and "(So)"].

The original overlap, sound stretch, pitch, pace, and underline/amplitude were transformed into binary gross variables: presence or absence of symbol within a unit. Regarding the original variable intonation, we collapsed the categories of *slight rise/continuing* (symbolized by a comma) and *large rise* (symbolized by a question mark) to produce intonation

TABLE 4
Intertranscriber Agreement Scores for Derived Categorical Variables

Derived Transcription Variable (Gross)	k[a]	π[b]	*T1 & T2*	*T1 & T3*	*T1 & T4*	*T2 & T3*	*T2 & T4*	*T3 & T4*	*Mean*
90+% Accuracy									
Semantics	na	na	94%	98%	96%	93%	93%	97%	95%
Unit Type	4	High	.75	.80	.72	.80	.72	.70	.75
Overlap	2	High	.48	.59	.50	.41	.60	.52	.52
Sound Stretch	2	High	.55	.69	.66	.59	.59	.70	.63
Pitch	2	High	.82	.91	.85	.67	.77	.80	.80
80–89% Accuracy									
Intonation	3	High	.55	.68	.57	.67	.61	.62	.62
Emphasis	2	Equi	.53	.68	.53	.50	.59	.53	.56
Underline/Amplitude	2	Mod	.25	.51	.33	.19	.33	.28	.31
Below 80% Accuracy									
Pace	2	High	.19	.25	.16	.08	.02	.04	.12

NOTE: [a] k = Number of codes within the given category/variable. [b] π = The variability of codes' simple probabilities: Equiprobable (Equi; $\pi_i = 1/k$), moderately variable (Mod; $\pi_i = 0.5/k$), and highly variable (High; $\pi_i = 0.25/k$).

(gross): no shift, downward shift, and upward shift/continuing. (See Discussion for problems with combining period and comma).

In English, at least three acoustic parameters are constitutive of general sound stress: pitch, stretch/duration, and amplitude. None of these articulatory gestures, however, has been found to correlate individually with stress (Wolfram & Johnson, 1982). Thus, it may be that transcribers can agree that a word/sound is generally stressed, yet disagree regarding the articulatory gesture involved. To test this, we created the derived variable "emphasis," which included the presence or absence of a shift in either "pitch," "sound stretch," or "underline/amplitude."

As seen in Table 4, the level of intertranscriber agreement was substantial for unit type (gross), semantics (gross), overlap (gross), sound stretch (gross), and pitch (gross). Agreement was moderate for intonation (gross), underline/amplitude (gross) and emphasis. Agreement was "unacceptable" for pace (gross).

In terms of variability across pairings of transcribers, T1 and T3 (the authors of the study) tend to have higher rates of agreement, but not appreciably so in relation to the next highest pairing on a given variable. To determine the effect of the T1 and T3 pairing on the overall kappa scores, we removed them from the calculation of averages on each variable. In this way we could see if their high level of agreement contributed disproportionately to overall levels of agreement. Only the average scores for cut-off and smile voice were affected. The observed kappa for cut-off dropped to 0.40 overall, putting it at the 89% level of agreement (moderate).

The kappa for smile voice dropped precipitously to 0.12, moving it to the unacceptable category (below 80% agreement). The dramatic change in the status of smile voice is likely due to T1 and T3 tending to mark both the onset and offset of the feature; this issue is taken up further in the section on Procedural Recommendations.

DISCUSSION

This investigation assessed interobserver agreement on a wide range of features in four conversation analytic (CA) first stage transcripts. We will discuss results in terms of levels of agreement (substantial, moderate, unacceptable), with brief preliminary recommendations relative to particular features. The discussion of results is followed by an extended consideration of procedural recommendations, the generalizability of the findings, limitations, and future research.

Findings of Acceptable Agreement

Substantial Agreement

Transcribers reliably assigned units to particular speakers (i.e., speaker designation). Along these lines, transcribers reliably assigned intra- and interturn silences to particular speakers and separate (i.e., nonspeaker-designated) lines, respectively, suggesting that transcribers can accurately determine the ownership of silence (Sacks, Schegloff, & Jefferson, 1974).

Initially, agreement on semantics was at the top end of moderate (89%). This, however, included words and sounds that were placed within parentheses, which specifically symbolizes doubt about their (semantic) nature. When transcribers were given credit for agreement when words and sounds were rendered as being in doubt, semantic agreement was substantial (95%). This figure is quite high, and noteworthy considering that semantic differentiation was extremely conservative, including differences between words such as "yeah" and "yes," "It's" and "(th)'ats," and "time" and "times," as well as differences between vocalizations such as "Ah" versus "Uh".

Similar to the variable semantics, agreement on unit type was at the top end of moderate ($\kappa = .72$; adjusted agreement value = 89%). When vocalized and nonvocalized breathing units were collapsed (i.e., inbreaths, outbreaths, and laugh tokens), however, agreement was "substantial" ($\kappa = .75$; adjusted agreement value = 92%). Unit type included all variables in the comprehensive sequence, including those that transcribers "failed" to render (see codebook); therefore, it was also an index of transcribers' agreement on the sequencing of units or their ability to transcribe units in

the same interactional order (e.g., first a word, then a silence, and then a laugh token/breath). Thus, transcribers' agreement on the sequencing of words, sounds, and silences was substantial.

It is not simply that transcribers agreed on the existence, sequencing, and semantics of word and sound units. Additionally, agreement on "orthography" (i.e., word/sound pronunciation) was substantial. For example, regarding the sentence "I: talked ta Dusti:n," it is not simply that all transcribers rendered a "word" for the third unit, but they rendered semantically similar words (i.e., "to") in orthographically similar ways (i.e., "ta"). At least some of the lack of agreement on "orthography" is likely explained by the fact that transcribers' usage of nonstandard orthography is relative to complex relationships between transcribers' and speakers' background, experience, and identities (Preston, 1982).

Transcribers also achieved substantial agreement on the length of silences when using the counting method (i.e., "one one thousand . . ."). Reliability does not equal validity; however, this finding is reassuring given the importance of prior CA research concerning the role of silence in conversation (e.g., preference organization).

Finally, transcribers achieved substantial agreement on rendering the cutting off, or abrupt halting, of units, and on rendering intraword/sound plosiveness, which is often associated with laughing while speaking. It is clear that transcribers hear and note the placement of laughter particles, inbreaths, and outbreaths; nonetheless, transcribing vocalized versus nonvocalized breathing units may require more focused attention.

Moderate Agreement

The variable "overlap" determined whether transcribers reliably rendered units as: (a) beginning in overlap (e.g., "[like"); (b) ending (or partially) in overlap (e.g., "br[eak,]"); or (c) not being in overlap with other units. At this level of detail, transcribers' agreement was moderate (80%). At a more gross level—whether or not any part of a unit was rendered as being in overlap—transcribers' agreement was substantial (93%), which is encouraging given that these were first stage transcripts. Even though there may be some empirical situations in which this gross level of overlap is valid, it is critical to note that overlap is a highly organized social practice whose deployment is often organized (a) more finely than at the level of entire words and (b) by reference to interactional structures other than words (e.g., sequences) (Jefferson, 1983, 1986; Schegloff, 1987, 2000). Thus, it is imperative that transcription training focus on the precise placement of overlap.

At a gross level, transcribers achieved a substantial level of agreement (98%) rendering the presence or absence of pitch shifts; however, their agreement rendering specific types of shifts (i.e., higher or lower) was

"moderate" (86%). This is consequential because type of pitch shift matters for interactants, as for example in the case of pitch peaks, (Schegloff, 1998).

As noted earlier, at least three acoustic parameters constitute general sound stress in English (pitch, amplitude, duration), yet none is found to correlate individually with stress. Given the moderate level of agreement for pitch, and the unacceptable levels for sound stretch and amplitude (see below), we surmised that transcribers might be able to agree that words and sounds were generally stressed, yet disagree regarding the specific articulatory gesture involved. In support of this, there were numerous examples where transcribers differentially rendered stress with symbols for pitch, sound stretch, and amplitude. For example, compare "to͟o" with "too:"; "m͟ore" with "mo:re"; "bi:g" with "b͟ig"; "I͟" with "↑I"; and "th͟at's" with "↑that's". Thus, we created a derived variable *emphasis*, which had two codes: (a) no pitch, sound stretch, or amplitude; and (b) the presence of any of them. Reliability for emphasis was at the top end of the moderate range (89%), and suggests that transcribers are able to discern accurately the presence of at least general (albeit unspecified) stress.

Transcribers' agreement on doubt (registering that a particular rendering or hearing was doubted by a transcriber) and smile voice were on the low end of moderate (84% and 82%, respectively). The lower agreement on "doubt" indicates that different transcribers heard different portions of transcripts more or less accurately. For example, T1 was completely unable to decipher a phrase that T2, T3, and T4 all clearly heard as "watchin' one". Thus, as previous research suggests (Pye, Wilcox, & Siren, 1988; Schriberg, Kwiatkowski, & Hoffman, 1984), having more than one transcriber examine the same transcript can produce more reliable (and presumably valid) results. Smile voice has not been researched in its own right; however, there is evidence that it is affiliated with laughter and joking, both of which are highly organized (Jefferson, 1979, 1984, 1985; Jefferson, Sacks, & Schegloff, 1987; Sacks, 1974, 1978).

Unacceptable Features

Transcribers did not reliably render unit-final intonation shifts at the most refined level (i.e., none, fall, slight rise/continuing, and large rise). One explanation is that the transcription symbol "comma" has been defined in a confounding, double-barrelled fashion as representing either "continuing" or "slightly rising" intonation. These are analytic and objective conceptualizations, respectively; just as the analytic phenomenon of "questioning" can be constituted by myriad features of interaction other than (or in addition to) intonation (e.g., grammar, pragmatics, sequential positioning), so can the analytic phenomenon of turn continuance. Insofar as analysis should be derived from empirical observation, transcription

symbols for intonation should indicate objective (versus analytic) features of interaction. Admittedly, it can be difficult for transcribers to achieve this level of ethnomethodological sensitivity. Extended ear training and awareness of the analytic-objective distinction may help.

Working from the possibility that transcribers oriented to the comma as representing a slight rise, we collapsed it with the question mark (i.e., large rise). Transcribers' agreement rendering this gross level of intonation actually rose to the top end of moderate (88%). This suggests that, at the level of first-stage transcription, transcribers are on their way to reliably rendering at least basic intonational trends. Future transcription training, however, needs to focus on unit-final intonation, which has been, and continues to be, a cornerstone of turn taking (Ford & Thompson, 1996; Sacks, Schegloff, & Jefferson, 1974).

At the finest level, transcribers did not reliably render stress using underlining; even in post hoc analysis at the grosser level (presence versus absence), accuracy was still low (80%). One possible explanation is that the symbol for amplitude (e.g., underlining), like comma, has been defined in a confounding, double-barrelled fashion as representing general stress achieved through either amplitude or pitch. A second explanation is that transcribers may differentially use underlining to indicate objective versus atypical stress. From the atypical position, it is redundant to underline syllables that are normally stressed. Insofar as a goal of CA is to discover "normal" patterns of interaction, transcribers need to strive to transcribe for "objective" stress. Admittedly, this is difficult. The organization of stress is part of transcribers' common sense knowledge, and many occasions of objective stress will go "heard but unnoticed" (Garfinkel, 1967).

Transcribers did not reliably render sound stretching (symbolized by the colon) at the syllabic level (e.g., "Thur::sday" vs. "Thursda::y"). On the other hand, their agreement on whether or not entire words and sounds contained some amount of sound stretching was substantial (94%). This might be impressive for "first stage" transcripts; however, sound stretching is organized at finer levels, perhaps even beyond the syllabic level, as has been demonstrated in some cases of overlap competition and resolution (Schegloff, 2000). Future transcription training needs to focus on the precise placement of all symbols, including those for sound stretch (see Procedural Recommendation for an extended discussion).

At both fine and gross levels, transcribers did not reliably render pace (i.e., speaking faster or slower relative to surrounding talk). In sum, agreement was unacceptable regarding the most detailed operationalizations of some phonological aspects of transcription, including unit-final intonation, amplitude, sound stretch, and pace. This is explained by the fact that most conversation analysts are not formally trained phoneticians. Those who are trained phoneticians, however, have demonstrated

fascinating levels of organization at the phonological level (Kelly & Local, 1989), and future transcription training needs to move in this direction.

At first, it was surprising that the lengths of machine-timed silences were not reliable. Transcribers using software that displays sound waves must still personally examine points of sound offset and onset, which is a difficult task; ascertaining periodicity in a sound wave is a matter of degree. The question of the appropriate way to measure silence—whether to use perceptual (e.g., counting), perceptual-mechanical (e.g., stopwatch), or mechanical (e.g., offset and onset of sound wave) measures—has been an ongoing topic of discussion (O'Connell & Kowal, 1990). As of yet, there is no evidence that mechanical measures of silence are more reliable than other measures, and present results suggest that perceptual measures are acceptably reliable.

Procedural Recommendations

Utilizing Nonstandard Orthography

There is a longstanding debate over whether transcribers should utilize nonstandard orthography to represent salient features of pronunciation. On one hand, the presence of nonstandard orthography can bias readers' interpretations and evaluations of speakers' social and linguistic identity, and thus of speakers' relative power, status, legitimacy, authority, and authenticity (Jaffe, 2000; Preston, 1982). On the other hand, CA analyzes naturally occurring language, and using nonstandard orthography is one way to embody variation that may be communicatively consequential. Thus, in line with Preston (1985), we suggest that transcribers avoid "eye" dialect, or respellings that do not reflect phonological differences from their standard spelling counterparts (e.g., avoid "sez" for "says" or "wuz" for "was"). We do suggest, however, that transcribers should utilize: (a) speech re-spellings or nonstandard spellings that attempt to capture natural phonetic processes, such as elision, loosening of articulation, and assimilation (e.g., "'cause" as a form of "because"); and (b) dialect respellings, or nonstandard spellings that attempt to capture regional or social differences (e.g., "dis" for "this"). Admittedly, speech and dialect re-spellings are phonologically inaccurate and bound to be inconsistent across transcribers. Pure phonetic transcripts, however, are often inaccessible to non specialists (including many conversation analysts).

Precise Placement of Symbols

All transcription symbols should be precisely placed. For example, consider the use of "colon" (:) to represent sound stretch. On the word "and",

T1–T3 placed the colon(s) after the "a" (i.e., after the pure vowel; "A::nd", "a:nd", and "A:nd", respectively), whereas T4 placed the colon after the "n" (i.e., after the nasal consonant; "An:d"). Similarly, regarding the word "hotel", T1 and T3 placed the colon after the "e" (i.e., "hote:l" and "hote:l", respectively), whereas T4 placed it after the "l" (i.e., "hotel:"). In each case, the differential placement of the colon(s) represented the stretching of different sounds. This is not trivial. Some CA research suggests that stretching different sounds within words can be interactionally consequential. For instance, in the context of a patient responding to a physician's history-taking question, Stivers and Heritage (2001) argued for a difference between "No", "No::", and the patient's actual "N::o".

As another example, consider the use of underlining to symbolize amplitude/stress. Current practice suggests that stressed syllables be underlined incrementally as stress increases, beginning with the first letter of the syllable (Ochs, Schegloff, & Thompson, 1996); however, the latter clause in this suggestion can produce both confusing and invalid results. For example, assuming that there is slight stress on the second syllable of the word "pavilion", the "v" should accordingly be underlined. It is not, however, the "v" per se that is emphasized; it is the vocalic segment of the syllable. (This is because sound pitch, amplitude, and duration are more easily, and thus typically, carried on the vocalic segment). This was represented by T1 ("Pav<u>i</u>lion") and T2 ("Pav<u>i</u>:lion") by underlining the "i", and by T4 by underlining the entire syllable ("Pa<u>vil</u>lion"). Thus, in most cases (but certainly not all), it is more valid to begin by underlining the single vowel (for moderate emphasis), as done by T1 and T2, and progress (for more emphasis) to underlining the entire syllable (including the vowel), as done by T4. A more subtle gradation of stress could be rendered by underlining part, but not all, of the syllable (including the vowel), as in "pa<u>v</u>ilion."

Marking Onset and Offset of Phenomena

It is crucial that transcribers denote both the onset and the offset of transcribed phenomena, such as overlap (with "[" and "]"), pitch shift (e.g., with "↑" and "↓"), pace shift (e.g., with ">" and "<"), and smile voice (with "£"). Without symbols for both onset and offset, it is impossible to know the exact boundaries of transcribed phenomena. In the present study, transcribers did not consistently mark both onset and offset on all features. For example, T3 produced the following:

Mara: He sou:nded pretty hungr(h)[y::
Tina: [heh heh .hh

From this, we know that Tina's initial laugh token ("heh") is in overlap with the stretched "y" of Mara's "hungr(h)y::". However, because the offset of the overlap is not marked, it is impossible to know whether Tina's second laugh token ("heh") and subsequent inbreath (".hh") are also in overlap. Compare this with the following, produced by T1:

Mara: H(h)e sounded pr(h)etty hung(h)ry](h),]
Tina: [hih] hih

It is important to note that transcribers' lack of marking both onset and offset sometimes enforced arbitrary coding decisions (see codebook) that may have decreased levels of agreement. For instance, see the following two renderings of the same stretch of talk:

[T1] Tina: . . . he's ↓like #I <u>a</u>m gunna go £play ba:ll,£#↑
[T4] Tina: . . . he's like ↓I <u>am</u> gonna go play <u>bal:l</u>.

T1 marked both the onset and offset of lowered pitch (e.g., "_" and "_", respectively); therefore, seven word units were coded as being rendered with lowered pitch (i.e., "like I <u>a</u>m gunna go play ba:ll"). Alternatively, because T2 did not mark the offset of lowered pitch, only one unit (i.e., "↓I") was coded as being rendered with lowered pitch, thus producing massive disagreement between T1 and T4. As a byproduct, the merely "moderate" level of overlap and the "unacceptable" level of "pace" were conservative estimates.

Generalizability of Findings

The data used in this study were recorded by placing the microphone of a nonprofessional model analogue tape recorder next to the audio output from a telephone answering machine's speaker phone. Thus, the medium is generalizable to relatively low-quality (albeit audible) audio recordings. Insofar as audio recordings are frequently of higher quality, our findings are conservative. The data are a mundane conversation between two friends; nonetheless, findings should generalize to the transcription of any audio form of talk between a maximum of two persons, including the audio portion of visual media, as well as talk organized by different speech exchange systems (e.g., monologues, interviews, conversation, Sacks, Schegloff, & Jefferson, 1974).

The findings clearly generalize to scholars trained both in CA and Jefferson's (Sacks, Schegloff, & Jefferson, 1974) transcription system

Similarly, findings should also generalize to a range of other scholars, such as a number of different types of discourse analysts (for review, see Tracy, 1995), who adopt (and are trained in) Jefferson's system, yet opt not to transcribe certain features of talk. For example, although action implicative discourse analysts (e.g., Tracy & Ashcraft, 2001) may not transcribe for pitch, volume, pace, or pause length, they are extremely concerned with speaker designation, word and utterance choice, and sequencing. Those features tend to be reliably transcribed; therefore, these findings are also relevant to the transcription practices of those discourse analysts trained in the CA system, but whose research interests do not require detailed transcription of suprasegmentals.

Limitations and Future Research

This was a study of interobserver agreement (or reliability), not validity, which is the ultimate goal of transcription. Nonetheless, reliability provides an index of the extent to which auditory phenomena are perceivable and reportable by trained listeners. In order to test validity, transcripts need to be compared to a "gold standard" transcript, which is extremely difficult to create and defend. If achieved, however, validity is an interesting direction for future research.

Published excerpts are typically, though not always, products of further stages of transcription, such as (a) the same transcriber editing and re-transcribing by repeatedly listening to, and motivatedly focusing on, particular aspects of transcription (e.g., specifically focusing on overlap onset and offset, or on unit-final intonation) or (b) having the transcript evaluated and edited in a data session. The current study, however, allows for future research to determine levels of improvement achieved on intertranscriber agreement from these next stages of transcription effort. The necessity of having transcripts evaluated and edited by multiple transcribers can be both time consuming and costly; therefore, it is of particular interest to know whether agreement can be improved without recourse to other transcribers.

The group of scholars who participated in this study represents the diversity of scholars currently practicing CA. For this reason, while the sample may be small, it is representative of the level of transcription expertise across several fields of study.

CONCLUSION

Overall, these findings indicate that even at the level of first-stage transcription, a representative, though small, group of conversation analysts are reliably capturing auditory phenomena on a number of semantic and acoustic dimensions. Initial results demonstrate that many of these features are being captured and rendered at the finest level of detail. For those features which fell below the level of acceptability, much of the disagreement can be explained by slight falterings in practice (such as not consistently marking offsets) or trying to render complex articulatory gestures (e.g., the production of emphasis) using different symbols to mark the same core phenomenon. Indeed, when allowance was made for slight variations in placement and type of symbols within a syllable or word, reliability scores reached acceptable levels. Given the overall findings, areas for improvement in training and practice have been outlined and avenues for future research are opened. Furthermore, the conventional wisdom that one hour is required to transcribe 1 minute of talk can now be viewed as a reasonable calculation for determining the time needed for producing reliable first-stage CA transcripts.

Even though we advocate that conversation (and other discourse) analysts spend their time conducting primary research and not reflexively studying CA methods, we do believe that this paper provides an important resource for presenting CA findings as based on evidence (e.g., transcripts) that is sufficiently reliable according to widely accepted and expected social scientific standards.

APPENDIX

Transcripts

For transcription conventions, see Table 1. The following transcripts are exact duplicates of transcribers' work, except for six minor modifications. First, actual names were replaced with pseudonyms (all transcription conventions were retained). Second, in accordance with pseudonyms, and for purposes of clarity, speaker identification was altered and standardized to "Tina" and "Mara" (T1 used "TIN" and "MAR"; T2 and T3 spelled names out, but spelled "Mara" differently; T4 used "T" and "M"). Third, the font was standardized to 10-point Courier (T1 and T3 used 10-point Courier; T2 used 12-point Courier-New font; T4 used 11-point Courier New font). Fourth, line length was re-formatted to fit the 1" margins of this page (T1 and T2 used 1" side margins; T2 used 1.25" side margins; T4 used 1.31 side margins). Fifth, for purposes of clarity and consistency, the transcription symbols for "smile" voice and "frog/creaky" voice were standardized to "£" (T1 used "$" and "#", respectively; T2 used "$" and "*", respectively; T3 used a smile face icon and nothing, respectively; T4 used "@" for both, but indicated which in the transcript). Sixth, for clarity of analysis, we numbered each line consecutively (T1 and T3 used consecutive line numbers; T2 began each page with 1; T4 did not include line numbers).

Transcriber #1 (T1)
1 ((Ring))
2 Mara: Hello,
3 (.)

Tina: Mara?
(.)
Mara: Yeah,
Tina: Hey it's Tina.
(.)
Mara: He:y=(r).=
Tina: =.hh Uh:m: (.) I: talked ta Dusti:n,
Mara: >Uh huh,<
Tina: A::nd he's j'st gunna: (1.2) ca:ll you: an'- after they get done with thuh mo:vie.
Mara: Oh: okay.
(0.5)
Tina: Nkay.
(0.5)
Mara: Are they ( ) or='re they=h
Tina: They're gunna go ta: £l(h)ebanin:(h) cineple:x. or whatever.£ [hh]
Mara: [( ) (thuh)] pavilion. [hh]
Tina: [hh] hu Ah::.
Mara: ( ) (Thet's) all h
Mara: Have you been there y(h)e:t,
Tina: Yea:h it's really ni:ce. [like I] go[:-
Mara: [(ge)-] [Like a hote:l,
Tina: (h)I (h)kno(hh)w .hh we could go there instead of indy now.
(0.4)
Mara: *Yeah:* No=hhh .h
Tina: Mm: eh- (0.4) its (a nice inn),
Mara: ((h)Hello:,) ((to someone else/pet))
(0.4)
Tina: I think like (.) he's ↓like £I am gunna go £play ba:ll,£_
(.)
Tina: hh [(.) (h).h]h £I'm like okay you wan'ed da hang out with
Mara: [-hhhh]
Tina: me: what's that abo:u(h)t,£
Mara: hh hh
(0.6)
Tina: Oh: well.
(0.6)
Tina: He's a bo:y,
Mara: Yea:h
(0.4)
Mara: .hh He is. a boy.
(0.2)
Mara: It's true.(=hh)
(1.0)
Mara: (So:/Uh:) Did John go home fer(s)- (0.2) br[eak,]
Tina: [Y-] Yea::h
Tina: (h)=He went back ta Michigan frida:y, >er actually (th)-< (0.3) <thursday:::>=ni:ght.
Mara: Oh.
(.)
Tina: Afternoo:n.
(0.3)
Tina: 'Cause he didn't have any classes (.) Friday.

(.)
Mara: Ah. ((like "Oh"))
Mara: Well that's good,
(.)
Tina: ↑Yea:h.↓
Mara: be nice(h), hh
(1.2)
Tina: Uh:m I told Dustin like we might (.) go get something to ea:t
Tina: An' I just talked ta Cliff too, he sounded pretty hungry: hhh
huh huh, [.hhh]
Mara: [heh heh .h]h [H(h)e sounded pr(h)etty
Tina: [h
Mara: hung(h)ry[(h),]
Tina: [hih] hih
Tina: .hhhh Yea:h,
Mara: That's funny,
Mara: Okay,
(1.0)
Tina: ↑A'right.↓
(.)
Mara: S[o:
Tina: [S-
Mara: I think we're gunna have cake over here too so:,
(0.2)
Mara: If anybody wants cake- (.) it'll- (0.2) we'll have some of that left. I'm su:re.
Tina: Are Clark an' Kelly 'unna be there,
Mara: Yea:h,
(0.4)
Mara: I'm not sure how late they'll be here though,
(.)
Mara: (Uh-/Eh-) they're over here right now.
(0.7)
Tina: ↑Okay.↓
Mara: Oh- ((Not "Oh"))
(0.3)
Tina: How big £is Joey.,
(0.2)
Mara: uhh hh He's bi:g.
[(0.4)/(.) .hhh] ((Inbreath is Mara's))
Mara: He c'n: like hold hands and walk.
Mara: An' 'e has, a tooth.
(0.4)
Tina: ↑Eh:::?↓= ((a vocal but not verbal sound; cooing?))
Mara: =(So), (0.2) Yeah, he's getting' pretty big,
(0.2)
Tina: Coo:l,
(.)
Mara: (So) >We got a new couch today,=hh
Tina: Sw↑e::t.↓
Mara: Yea:h.=hh
(.)
Tina: Wha[t co]lor.
Mara: [( )]

Mara: (Im) my gre(f)- my gran'parents got a new couch so they gave
us their old one.
Tina: Ah::[£:£]
Mara: [I j's] went an' got that a couple days ago an' Clark's
Helpin' 'em move it in right no:w.
(0.4)
Mara: £So:£
(0.3)
Tina: Coo:l.
(0.2)
Mara: £Yeah£ (I) j's figured I warn ya so you don't walk inta(h)
th(h)uh l(h)iv(h)ing roo:m an'
Tina: hehh FREAK O:UT,
Mara: £(Pa[:ck), Yeah,£]
Tina: [.h .h h]
Tina: Well like I'm excited ta: you kno:w try out thuh new co:uch.
Mara: £Yeh- ah.£ (("Yeah"))
(.)
Tina: Uhh huh huh [h
Mara: [(Uhm)=hhh
Mara: And like (.) two: m:ore bright orange chair(hh)s.
(0.2)
Tina: ↑Swee:t.↓=
Mara: =So now we have three:, (.) bright (.) orange (.) chair(h)s.
{(0.8)/(0.2) _hhhhh (.)} ((Inbreath is Tina's?))
Tina: Good times.=
Mara: =£Y:eah£ ( ).
(1.2)
Mara: So: {(0.4)/(.) .hhh} Okay. So:: is Cliff gunna come over here
like six thirty then too,
Tina: Y:eah.

Transcriber #2 (T2)

((ring))
Mara: Hello¿
Tina: £Mara¿£
Mara: Yes.
Tina: Hey it's Ti:na.
(0.3)
Mara: Hey::a
(0.4)
Tina: u::m. I::v talked t' Dustin¿
Mara: Uh huh
Tina: a:nd he's jus gonna:: (1.3) ca:ll you:: and- after they get done with the mo:vie.
Mara: Oh: okay.
(0.7)
Tina: ↑°Okay.°
(0.8)
Mara: Are they watchin one ur are they::hh
Tina: They're gonna go: t:: "Le£bano(h)n Cineple:x.='r whatever.=
Mara: =Oh in the Pavi:lion(hh)
Tina: A:h
Mara: (uh) tha:t's all hh=

Tina: Have you bin there ye(h):t.
Mara: Y£e:h. It's really n↑i:ce.=
Like [I go:]
Tina: [.hh Like] a hotel
Mara: I(hh) kno(h)ow Like I go the:re instead of Indy now.
(0.5)
Tina: Yeah:: no:(hhh).
(0.7)
Tina: mm
Mara: but- It's nice there
Tina: £I (d)know:.
Tina?: I think like- he's like "↓right I=m gonna go play baw::.
Mara: (uh[h)
Tina: [(.hh) like- ↑Okay¿=you wanted to hang out with me::. What's
tha:t abou:t.=
Mara: =(hhh)
(0.7)
Tina: O:h well.
(0.8)
Tina: He's a boy:.
Mara: Ye:s=
((woo [woo sound in background))]
Mara: [He i : s] a boy:.
(0.3)
Mara: 'ts true.hh
(1.0)
Mara: >uh uh< Did John: go ho:me fer- (0.5) break=
Tina: =Ye::s. He went back t' Mi:chigan.
=F'riday.<'r actually
(0.5)
Tina: Thurs:day:: ni:ght.
Mara: Oh- Oh. ((Sound produced far back in her throat))
(0.3)
Tina: afternoo:n.
(0.5)
Tina: cause he di(d)n't have any cla:sses.
(0.3)
Tina: Friday.
Mara: O:h. Well 'at's good.
(0.3)
Tina: iYah.
Mara: Pretty ni:ce. (hh)
(1.0)
Tina: U::m I told Dustin like we might- go get something t' ea:t.=an I jus' talked t' Cli:ff
too:. He sounded pretty hu:ngry::¿
Mara: (hhhh) hu[h
Tina: [eh heh
Mara: He sou:nded pretty hungr(h)[y::
Tina: [heh heh .hh
£iYeah::£
Mara: That's funny.
Tina: Okay.
(1.0)
Mara: ↑°All right.° S[o:] I think we're gonna=

Tina: [So]
Mara: =have ca:ke over here too:. So:.
(0.4)
Mara: if anybody wants ca:ke. Then- w- we'll have- some a that le:ft, I'm sureh.
Tina: Are Clark and Gilli(an) gonna be there¿
Mara: Yeah:¿
(0.5)
Mara: I'm not sure how la:te they'll be here thou:gh. But they're over here right now¿
Tina: ↑Okay.=
Mara: =(Ok-)
Tina: How big is Joey.
(0.3)
Mara: mm(hh)(hh) He's bi:g.
(0.5)
Mara: He ca:n- like hold hands and wa:lk.=n' he has a too:th.
Tina: ↑mm↑Ih[h
Mara: [£So-£Yuh- He's getting pretty big.£
(0.4)
Tina: Cool.
Mara: *>Yuh< We got a new cou:ch t'day hh
Tina: ↑Swee:t.
Mara: Yea:h(h)
Tina: W[hat]color
Mara: [en- [en my gran- my gra:nparents got a new cou:ch.=so they give us their old one.
Tina: Uh::.
Mara: I (worked that) out a couple days ago:. 'n Clark's helpin 'em move it in right now:.
(0.5)
Mara: so:
Tina: Coo:l.
Mara: Yeah. Just figured I'd warn you. So y'don't walk in to our l(h)ivi(h)ng roo(h)oom
Tina: ↑mm=
Mara: =.hh↑freak out £'n pa:ck. Iyuh
Tina: Well like I'm exci:ted t' y'know: try ou:t the new cou:ch.
Mara: £Yeah.
Tina: mmhmh[mh
Mara: [u:n.(hh) a:nd like two:- mo:re- bri:ght orange chai:r(h)s(hh)
(0.4)
Tina: £Swee[t
Mara: [So now we have three:. bri:ght orange £chairs.
(1.0)
Tina: Good ti:me.
Mara: £Ye:s.
(.)
Tina: °Right.°
(0.5)
Mara: .hh So:.
(0.3)
Mara: .hh O:kay. So:: is Cli:ff gonna come over he:re.<like six thirty then too.=
Tina: =Yeah.

Transcriber #3 (T3)

((Ring))
Mara: Hello?
(.)
Tina: Mara?
(.)
Mara: Yeah?
Tina: Hey its Tina.
Mara: He:y. hh
Tina: Uh::m (.) I:: talked ta Dustin,
Mara: Uh huh?
Tina: A:nd he's jus gonna:: (1.2) ca:ll you an- >after they get done with th< mo:vie.
Mara: ↑Oh:↑ okay.
???: (0.6) (XX here?) ((voice in background during "silence"))
Tina: Nkay.
Tina: .h[hh
Mara: [Are they watchin one er are the(h)y: hh
Tina: They're gonna go ta: £I(h)eb(h)anen Cineplex er whatever hh£=
Mara: =Oh in the Pavilion. [heh
Tina: [hh huh Ah::
Mara: That's all. hh
Mara: Have you been there £y(h)et,£
Tina: Yeah, it's ↑really nice.=Like I [go:
Mara: [(It's) like a hote:l.
Tina: (h)I kn(h)ow(h) .hh Like I go there ,instead of Indy now.£
(.)
Mara: Yeah. No. hhh .hhh
Tina: Mm:
Mara: Eh-
(0.4)
Tina: ↑It=s nicer.
Mara: (hal o:,)
(0.2)
Tina: I think like (.) he's ↓like ↓↓I am gonna go ,play ba:ll.↓↓£ hhh=
Mara: =hhh=
Tina: =.hh £I'm like okay you wanted to hang out with me:: what's that abou:(h).£
Mara: hhh
(0.2)
Tina: Oh well.
(0.2)
He's a boy.
Mara: Yea:h.
Child: eh: [eh:. (("whooping" in background))
Mara: [.pth ((lips opening)) He is. a ↑boy.↓'=ats ↑true.↓= nhh ((out breath through nose?))
(1.0)
Mara: °So:°. Did John go home fer (0.6) brea[k,]
Tina: [y-] Yea::h.
He went back ta Michigan Friday.>Er actually< (0.2) <Thur:sday::> n:i:ght.
Mara: Uhn uhn.((sounds like a "no" token would))
Tina: >After< noo:n.

(0.2)
Tina: Cuz he didn't have any classes (.) Friday.
Mara: Ah. Well ↑that's↑ good.
Tina: ↑Yeah.↑
Mara: Be nice, hhh
(1.0)
Tina: Uhm I ↑told↑ D(u/e)stin like we might (.) go get something to ea:t.=And ↑I just talked to Cliff too he sounded pretty hungry. hhh heh heh .hhuh=
Mara: =heh heh £He sounded pretty hungr(h[h)y?£
Tina: [heh heh .hhh
£yea::h.£
Mara: 'ats ((glottalized "that's")) funny, okay.
(1.0)
Tina: ↑Awright.↓
Mara: S[o
Tina: [So-
Mara: I think we're gonna have cake over here too: so:
(0.2)
Mara: if anybody wants cake it'll- we'll have some of that left I'm sure.
Tina: >Are Clark and Kelly gonna be there?<
Mara: Yea:h,
(0.2)
Mara: I'm not sure how late they'll be here though. But they're over here right now.
Tina: ↑Okay↓=
Mara: =O .hhh
Tina: How big is Joey.
Mara: .nhhh ((outbreath through nose)) He's big.
(0.6)
Mara: He can (.) like hold hands and walk and he has a tooth.
Tina: ↑hnnn↑= ((falling/rising pitch movement-cutesy baby relevant sound.))
Mara: =So yeah. He's gettin pretty big.
(.)
Tina: Cool.
(.)
Mara: °So.° We got a new couch today. .nhnhnh ((nasalized outbreath))
Tina: ↑Swee::t.↑
Mara: Yeah. .hhh
Tina: What c[olor.]
Mara: [eh] In my grem- my granparents got a new couch so they gave us their old one.
Tina: Ah::[::
Mara: [(Denton) went an got that one a couple days ago an Clark's helpin im move it in right now.
(0.2)
Mara: So.
Tina: ↑Cool.↑
Mara: Yeah. Jus figured I'd warn ya so you don't walk inta the h(h)iv(h)ing room an:
Tina: huh FREAK OUT.
Mara: Attack. Yeah.
Tina: Well like I'm excited ta: ya kno:w try out the new couch.
Mara: Yea:h.

Tina: he he heh
Mara: Fun hhh
Tina: .h heh
Mara: And like two: mo:re bright orange chairs hhh
Tina: ↑Sweet.↑
Mara: So now we have three:, bright (.) orange chairs.
(0.2)
Tina: .hhh Good times.
Mara: Yeah.°It's great.°
Tina: .hhh=
Mara: =So:.
Tina: hhh=
Mara: =Okay. So: is Cliff gonna come over here like six thirty then too.,
Tina: Yeah.

Transcriber #4 (T4)

((ring))
(1.1)
Mara: hello?
Tina: Mara?
Mara: Yeah,
Tina: Hey it's Tina.
(.)
Mara: Hey:.h
(.)
Tina: Uh:m, I: talked ta Dustin?
Mara: >uh huh,<
Tina: An:d he's just gonna::(0.9)call: you when-=after they get done with the movie.h
Mara: Oh:,=okay.
(0.6)/((talk in background))
Tina: Mkay.
(0.7)
Mara: Are they watchin'one or are they:£heh£-(0.2)
Tina: They're gonna go ta: L(h)ebanodn: ↑Cineplex£, or whatever, [£eh heh£
Mara: [Oh the pavillion, [eh
Tina: [£eh heh
Tina: Ah::
(0.4)
Mara: °£that's all,£ Have you been there?hh
Tina: Yeah, it's ↑really nice, Like I go:=
Mara: =like a hotel:.hh
(.)
Tina: ↑I kn(h)ow(h),huh .h
Tina: Like I go there instead of indie now.
Mara: [Yeah.£ know(h)£huh£]
[((baby crying?))
(0.7)
Tina: Mmm:=
=((burp?))
(0.6)

Tina: It's nicer,
(0.3)
Mara: (Hello:,)
?: yah ((baby?))
Tina: I think like->eh-< he's like ↓I am gonna go play bal:l.
Tina: £he[hh£ .heh
Mara: [ehh.hh
Tina: .h 'mlike ↑okay you wanteduh hangout with me:.what's that abou:t.
Mara: ehhhh
(0.8)
Tina: O:h well.h
(0.7)
Tina: He's a bo:y.
Mara: Yeah£m£.
Baby: eñhhe [hñ
Mara: [He is. A boy.
Mara: hh/(0.3) ((pause or breathing sound))
Mara: (°t's°)true. hhh
(0.8)
Mara: £So:. Did John go home for-(0.5) break?=
Tina: =Yeah:,hh he went back to Michigan. Friday=er↑Actually-th-
(0.5):Thurs:day::n:nigh:t.
Mara: £O-oh,£
Tina: Afternoo:n.
(0.6)
Tina: Cause he didn't have any classes.
(0.3)
Tina: Friday.
(0.3)
Mara: Ah:, Well ↑that's good.
Tina: Yeah.
Mara: Be nice. *hheh*
(1.2)
Tina: Um I told Darse and (Mike)/(Link), we might we might- go and get something to
ea:t, And I just talked to Cliff, too:, he sounded pretty (.) hungry:.
Tina: (hhh)heh he[h
Mara: [>heh heh< hh
Mara: .hh He s(h)ounded pr(h)etty hungr(h)y?=
Tina: =huh huh .hh eyea:h.
Mara: That's funny.
Mara: Okay:.
(0.4)
Tina: .hh .↑Alright.
Tina: [So,£
Mara: [°£So°. I think we're gonna have cake over here, too.So,hh(0.5) If anybody wants
cake, (it'll)- (0.2) we'll have some of that left, I'm sure. .h (.)
Tina: Are Clark and Kelly gonna be there?
(0.3)
Mara: Yea:h,
(0.6)
Mara: >I'm not °sure° how late they'll be here< though:.
Mara: *but* they're over here right no:w,
(0.8)
Tina: ↑Ok↓ay:=.

Mara: =£>(so)<,
(0.2)
Tina: How big is Joey?
(0.4)
Mara: eh(h)hh ((breath through nose))He's bi:g.
(0.7)
Mara: He can:-like hold hands and wa:lk, and he has a too:th.h
(0.4)
Tina: mhm:_::_:=
Mara: =£So, yeah, He's gettin pretty bi:g.
(0.5)
Tina: Cool.
(.)
Mara: £So,=We got a new couch toda:y,hh
(0.3)
Tina: ↑Swee:t.
Mara: Yeah,hh
(.)
Mara: [(A:f-)
Tina: [What color,
(.)
Mara: An' my gram- my grandparents,got a new couch, and they gave us
their old one.=
Tina: Uhah::.
Mara: =(Ed) went and got that a couple of days ago, and Clark's
helpin' him move it in, right now,
(0.5)
Mara: So:,
(0.3)
Tina: Coo:l,
Mara: Yeah, I figured I'd >(maurice),< yaknow (alter)/(all through)
th(h)e l(h)iving roo:m, an'
Tina: °ehheh°
?: ↑FREAK OUT,
Tina: (un)pa:k, [yeah,
Mara: [ehheh
Mara: Well it got me excited ta: yaknow try out the new couch,
Mara: Yeah.
Tina: eh(h)hm[hm
Mara: [(ts)fu:n,hh
Tina: [ehm
Mara: [A:n like- two m:ore bright orange chairs(h).
(0.4)
Tina: ↑Sweet,
Mara: (So) Now we have three,(.) bright orange chairs.
(0.5)
Tina: .hh Good ti:me°s°.
Mara: £Yeah.
Mara: (Sgreat).
(1.0)
( ): hh
Mara: *So ↑Ok↓ay:,So:,is Cliff gonna come over her:e, like
six- thirty then too:,=
Tina: =Yeah,

NOTES

1. This investigation is not designed to test agreement on the application of a transcription system per se. Rather, it is designed to test transcribers' abilities to reliably hear and render the same phenomena as indexed by particular transcription conventions. For example, the CA transcription symbol for sound duration is the colon (e.g., ":"). While breathing is also audible and varies in duration, breath duration is not indexed by the colon, but by the addition of the letter "h" (e.g., "hhh" represents a longer outbreath than "hh"). We therefore did not test intertranscriber agreement on the application of a symbol by giving transcribers "credit" when they did not (mis)apply it. Rather, we tested transcribers' abilities to reliably hear/render symbols where they were relevant. Thus, in the example of the use of the "colon" to symbolize duration, we only examined units (i.e., words and sounds) where the application of the symbol was possibly relevant.

REFERENCES

Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138–159). New York: Cambridge University Press.

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Bakeman, R., Quera, V., McArthur, D., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods, 2*, 357–370.

Beach, W., & Metzger, T. (1997). Claiming insufficient knowledge. *Human Communication Research, 23*, 562–588.

Cappella, J.(1990). The method of proof by example in interaction analysis. *Communication Monographs, 53*, 236–242.

Clavarino, A., Najman, J., & Silverman, D. (1995). Assessing the quality of qualitative data. *Qualitative Inquiry, 1*, 223–242.

Clayman, S. (1992). Caveat orator: Audience disaffiliation in the 1988 presidential debates. *Quarterly Journal of Speech, 78*, 33–60.

Clayman, S. (1995). Defining moments, presidential debates and the dynamics of quotability. *Journal of Communication, 45*, 118–146.

Clayman, S., & Heritage, J. (in press). Questioning presidents: Journalistic deference and adversarialness in the press conferences of Eisenhower and Reagan. *Journal of Communication*.

Codebook. Guide to coding decisions for inter-observer agreement on "first-stage" conversation analytic transcription. Retrieved from http://web.ics.purdue.edu/~froberts/research.htm

Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research, 20*, 105–121.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). Cambridge, UK: Cambridge University Press.

Gardner, W. (1995). On the reliability of sequential data: Measurement, meaning, and correction. In J. M. Gottman (Ed.), *The analysis of change* (pp. 339–359). Mahwah, NJ: Erlbaum.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge, UK: Polity Press.

Heritage, J. (1989). Current developments in conversation analysis. In D. Roger and P. Bull (Eds.), *Conversation: An interdisciplinary approach* (pp. 21–47). Clevedon, UK: Multilingual Matters.

Heritage, J., & Atkinson, J. M. (1984). Introduction. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 1–16). Cambridge, UK: Cambridge University Press.

Hopper, R. (1989). Sequential ambiguity in telephone openings: What are you doing? *Communication Monographs, 56*, 240–252.

Irwin, J., & Wong, S. (1983). *Phonological development in children 18–72 months*. Carbondale: Southern Illinois University Press.

Jacobs, S. (1990). On the especially nice fit between qualitative analysis and the known properties of conversation. *Communication Monographs, 57*, 243–249.

Jaffe, A. (2000). Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics, 4*, 497–513.

Jefferson, G.(1979). A technique for inviting laughter and its subsequent acceptance/declination. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 79–96). New York: Irvington Publishers.

Jefferson, G. (1983). Notes on some orderlinesses of overlap onset. In V. D'Urso & P. Leonardi (Eds.), *Discourse analysis and natural rhetoric* (pp. 11–38). Padua, Italy: Cleup Editore.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 346–369). Cambridge, UK: Cambridge University Press.

Jefferson, G. (1985). An exercise in the transcription and analysis of laughter. In T. A. van Dijk (Ed.), *Handbook of discourse analysis, Volume 3* (pp. 25–34). New York: Academic Press.

Jefferson, G. (1986). Notes on "latency" in overlap onset. *Human Studies, 9*, 153–183.

Jefferson, G., Sacks, H., & Schegloff, E. A. (1987). Notes on laughter in the pursuit of intimacy. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 152–205). Clevedon, UK: Multilingual Matters.

Jones, S. E., & LeBaron, C. D. Research on the relationship between verbal and nonverbal communication: Emerging integrations [Special issue]. *Journal of Communication, 52*(3).

Kelly, J., & Local, J. K. (1989). On the use of general phonetic techniques in handling conversational material. In P. Bull & D. Roger (Eds.), *Conversation: An interdisciplinary perspective* (pp. 197–212). Clevedon, UK: Multilingual Matters.

Kitzinger, C. (1998). Inaccuracies in quoting from data transcripts. *Disourse & Society, 9*, 148–156.

Kirk, J., & Miller, M. (1986). *Reliability and validity in qualitative research*. London: Sage.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Levinson, S. C. (1983). Conversational structure. In S.C. Levinson (Ed.), *Pragmatics* (pp. 284–370). Cambridge, UK: Cambridge University Press.

Ochs, E.(1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43–72). New York: Academic Press.

Ochs, E., Schegloff, E.A., & Thompson, S.A. (Eds.). (1996). *Interaction and grammar*. Cambridge, UK: Cambridge University Press.

O'Connell, D.C., & Kowal, S. (1990). Some sources of error in the trancription of real time in spoken discourse. *The Georgetown Journal of Languages & Linguistics, 14*, 453–466.

Patterson, B. R., Neupauer, N. C., Burant, P. A., & Koehn, S. C. (1996). A preliminary examination of conversation analytic techniques: Rates of inter-transcriber reliability. *Western Journal of Communication, 60*, 76–91.

Pomerantz, A. M. (1990). Conversation analytic claims. *Communication Monographs, 57* 231–235.

Pomerantz, A., Fehr, B. J., & Ende, J. (1997). When supervising physicians see patients: Strategies used in difficult situations. *Human Communication Research, 23*, 589–615.

Preston, D. R. (1982). Ritin' Fowklower daun'Rong: Folklorists' failures in phonology. *Journal of American Folklore, 95*, 304–326.

Pye, C., Wilcox, K. A., & Siren, K. A. (1988). Refining transcription: The significance of transcriber "errors." *Journal of Child Language, 15*, 17–37.

Robinson, J. (1998). Getting down to business: Talk, gaze and body orientation during openings of doctor-patient consultations. *Human Communication Research, 25*, 97–123.

Robinson, J., & Stivers, T. (2000). Achieving activity transitions in primary-care consultations: From history taking to physical examination. *Human Communication Research, 27*, 253–298.

Sacks, H. (1974). An Analysis of the course of a joke's telling in conversation. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 337–353). Cambridge, UK: Cambridge University Press.

Sacks, H. (1978). Some technical considerations of a dirty joke. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 249–269). New York: Academic Press.

Sacks, H. (1984). Notes on methodology. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 21–27). Cambridge, UK: Cambridge University Press.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking oragnisation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 70–85). Clevedon, UK: Multilingual Matters.

Schegloff, E. A. (1998). Reflections on studying prosody in talk-in-interaction. *Language and Speech, 41*, 235–263.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society, 29*, 1–63.

Schriberg, L. D., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by concensus. *Journal of Speech and Hearing Research, 27*, 456–465.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction*. Thousand Oaks, CA: Sage.

Stivers, T. (2001). Negotiating who presents the problem: Next speaker selection in pediatric encounters. *Journal of Communication, 51*, 252–282.

Stivers, T., & Heritage, J. (2001). Breaking the sequential mold: Answering "more than the question" during comprehensive history taking. *Text, 21*, 151–185.

Stockman, I. J., Woods, D. R., & Tishman, A. (1981). Listener agreement on phonetic segments in early infant vocalizations. *Journal of Psycholinguistic Research, 10*, 593–617.

Stringer, J. L., & Hopper, R. (1998). Generic he in conversation? *Quarterly Journal of Speech, 84*, 209–221.

Thomason, W. R., & Hopper, R. (1992). Pauses, transition relevance, and speaker change. *Human Communication Research, 18*, 429–444.

Tracy, K. (1995). Action-implicative discourse analysis. *Journal of Language and Social Psychology, 14*, 195–215.

Tracy, K., & Ashcraft, C. (2001). Crafting polices about controversial values: How wording disputes manage a group dilemma. *Journal of Applied Communication Research, 29*, 297–316.

Wolfram, W. (1991). *Dialect and American English*. Englewood Cliffs, NJ: Prentice Hall.

Wolfram, W., & Johnson, R. (1982). *Phonological analysis: Focus on American English*. Washington, DC: Center for Applied Linguistics.

Wrobbel, E. D. (1998). A conversation analyst's response to Patterson, Neupauer, Burant, Koehn, and Reed. *Western Journal of Communication, 62*, 209–216.