



Psychotherapy Transcription Standards

Erhard Mergenthaler & Charles Stinson

To cite this article: Erhard Mergenthaler & Charles Stinson (1992) Psychotherapy Transcription Standards, *Psychotherapy Research*, 2:2, 125-142, DOI: [10.1080/10503309212331332904](https://doi.org/10.1080/10503309212331332904)

To link to this article: <https://doi.org/10.1080/10503309212331332904>



Published online: 25 Nov 2010.



Submit your article to this journal [↗](#)



Article views: 3212



View related articles [↗](#)



Citing articles: 102 View citing articles [↗](#)

PSYCHOTHERAPY TRANSCRIPTION STANDARDS

Erhard Mergenthaler
University of Ulm, Germany

Charles H. Stinson
University of California, San Francisco

Although use of psychotherapy transcripts is becoming increasingly important in psychotherapy research, large-scale collaborative work is hindered by lack of suitable transcription standards. Guidelines are presented for the transcription of discourse, such as psychotherapy sessions, for research and educational purposes. Transcripts generated by following these standards will be readable by human judges; they will also be easily submitted for computer-aided text analysis, such as formal concordance.

Research in the fields of psychotherapy and psychoanalysis is increasingly based on primary data (Luborsky & Spence, 1978) such as verbatim protocols of therapy sessions. This can be easily observed at research meetings such as those of the Society for Psychotherapy Research, where the contributors are paying increasing attention to these empirical research materials. For the purposes of this paper, the many different possible research foci can roughly be divided into those using *human judges* and those based on *computational methods*. To the category using human judges belong such analyses as application of rating scales, content analyses, text interpretation, and hermeneutics. Some formal analyses, grammatical analyses, content analyses, and knowledge-based analytic approaches belong to the latter category (Mergenthaler, 1985). Depending on the research question and theoretical perspective, each of these groups has somewhat different needs, and this has been reflected in the development of different specialized transcription formats. Although useful for specific analyses, one transcription format is often of little utility or relevance to other studies. Expert narrow focus, expediency, and limited resources have led to transcripts that are not broadly useful. As extreme examples, a handwritten transcript with dense annotations is as useless to a computer-aided text analytic

This research has been supported by the German Research Foundation DFG within the Collaborative Research Program SFB 129 B2 in 1980–1988.

We wish to acknowledge gratefully the contribution to this work by the Program for the Study of Conscious and Unconscious Mental Processes, which is directed by Mardi J. Horowitz and is supported by the John D. and Catherine T. MacArthur Foundation.

We wish to acknowledge gratefully the influence of discussions with Wilma Bucci of Adelphi University. We also wish to acknowledge the important suggestions and comments provided by the many secretaries, assistants, and scientists who have worked as transcribers, verifiers, and members of transcription committees at both PCUMP and the Ulm Textbank.

Correspondence should be sent to Erhard Mergenthaler, Sektion Informatik in der Psychotherapie, Universität Ulm, Am Hochsträss 8, D-7900 Ulm, Germany

system as a fully indexed data table is to the clinician wishing to read the transcript for content.

One might argue that this is no problem, for specialization is a necessary concomitant of scientific progress. But this does cause problems, especially in comparative, collaborative, or convergent studies. It is difficult to perform comparisons by applying alternative analytic methods to the same sessions, if they are prepared following different transcription standards. Large-scale collaborative approaches with convergent analyses become prohibitively time-consuming, if possible at all. Some standardization is necessary to allow collaborative efforts involving different domains of expertise.

It is surprising that neither authors using transcripts in their research nor reviewers commenting on their articles give much attention, if any, to the use of transcription rules. Stiles and Sultan (1979), for example, made no mention of the process of transcribing, but seem to assume a transcript as a given valid and reliable data set, ready to be used for subtle analyses: "Each transcript was coded independently by two volunteer coders according to a coding manual for the taxonomy" (p. 611). Although the taxonomy these authors are using is sensitive to linguistic phenomena, which also are addressed within transcription standards, they did not consider these issues. To give another example, the only remark concerning transcription in Oxman, Rosenberg, Schnurr, and Tucker (1985) is: "Each patient's speech sample was corrected by the interviewer after transcription" (p. 1151). The reader is left to imagine how stuttered words, speech disturbances, and so forth were handled. Our point, of course, is not to claim that studies like those just mentioned should be disregarded. But we do assert that this is a significant problem: Their findings cannot be adapted to other studies without proper evaluation of the underlying transcription processes. As an illustration, Kächele's (1983) analysis of verbal activity level of therapists in initial interviews and long-term psychoanalysis cannot be compared to work on another sample of transcripts where nonverbal utterances such as "mm-hm" have not been treated as words.

Transcribing verbal exchanges from audio and video recordings presents countless complex and subtle problems to the researcher. Some of the challenges can be approached with common sense; others require knowledge of linguistics or computer science. What makes a transcript well-suited for computer-aided analysis may also make it indecipherable to a human, and vice versa. It is, for example, immediately evident that variations in word form by tense and case offer subtleties important to human understanding, but they cause problems for simple computer word count or concordance programs. A simple program may consider a capitalized word at the beginning of a sentence to be completely different from the same word in all lower case type. Conversely, reducing a transcript to word stems with consistent grammar makes for monotonously mechanical reading and obscures sense and nuance. There are countless other matters to consider: homonyms, homomorphs, plural and possessive forms, capitalization, spelling, anaphoric and ambiguous references, contractions, compound words, punctuation, pronunciation, slang, affected speech, quotations, paraverbal utterances, simultaneous speech, timing, pauses, stuttering, incomplete or incomprehensible words, spoken abbreviations, and others. To the eternal distress of grammar teachers everywhere and to the intrigue of linguists, people rarely if ever use formal English grammar in normal discourse.

Another important issue for researchers is transcript segmentation. The single word as unit of analysis is usually not problematic, but consider what problems

might arise if the turn of speech is the unit of analysis. It is possible one analysis might need to ignore content-free utterances and choose not to transcribe them. Another might focus precisely on the pragmatic use of such content-free "continuers" or paraverbal utterances. Nonversatile standards can cause important information to be lost at the initial steps of transcription. Still other issues of clinical importance (such as substitution of pseudonyms for protection of confidentiality) are handled by different groups in dissimilar and inconsistent ways. It is thus hardly surprising that transcript analysis is becoming a highly, specialized research area.

Development of standards is always difficult. While it seems unlikely there will be a complete transcription standard that is universally acceptable, it is also obvious that some standardized approaches will be generally useful to much transcript research. (Those interested in general questions about transcription and coding methods are referred to Edwards and Lampert [1991] for more detail.)

Dahl (1979) did pioneering work in psychotherapy research at Downstate Medical Center, State University New York, when he studied word frequencies of spoken American English. He devised a suitable set of transcription standards, which he used for a corpus of 15 psychoanalytic cases. All words from 15 sessions sampled from each of these cases formed the basis for his dictionary.

Mergenthaler (1985) developed additional transcription standards that would fulfill some basic needs for linguistically oriented analyses of psychotherapy texts. The establishment of the Ulm Textbank (UTB) facilitated further psychotherapy text-analytic approaches; this led to greater standardization of transcription and consequently improved exchange of textual data and empirical results. In the German-speaking countries of Germany, Switzerland, and Austria, these standards are currently in widespread use in psychotherapy transcript research.

There has been little comparable development of standards in the Anglo-American scientific community. Beginning in 1983, however, the Program for the Study of Conscious and Unconscious Mental Processes (PCUMP) at the University of California, San Francisco was given the challenge of encouraging collaborative research among scientists of various disciplines in search of convergent evidence of unconscious influences on conscious mental processing. These studies use transcripts, physiological recordings, and other records from psychotherapy sessions. The authors worked together to adapt the Ulm Textbank standards to the English language and to extend them with more versatile methods for segmentation, timing, and coordination with other data sets. This transcription standard is presented here and is proposed for use by other individuals and groups interested in psychotherapy transcript analysis.¹

It is beyond the scope of this paper to present an exhaustive overview of what is possible when using standards such as those proposed here. In order to give the reader some idea of what can be done, however, we present a brief outline of some selected methodological approaches. Starting from a semiotic view, language is understood as a system of symbols whose structure is determined according to rules based on the relationship between form and content. Accordingly, it is possible to distinguish formal, grammatical, and substantive measurements. Each of these can be further subdivided according to whether it can be applied to a speaker's text or to

¹For those interested in computer-aided text analysis, the authors have available software that uses transcripts formatted according to these standards. With this software one can quickly and easily produce concordances, word lists, various kinds of vocabularies, text listings with word counts, and frequency distributions of specific word categories. There is also software available to help transform text formatted by Dahl's standard to those described here.

the entire speech activity in a dialogue. It is therefore possible to speak of monadic or dyadic values for any of these types of measurements and data, and to use these measures in empirical studies, such as detecting state-related repeating patterns of an individual's discourse or differences between intake interviews of different diagnostic groups of patients (e.g., neurotic vs. borderline patients). The possibilities of formal, grammatical and substantive measures range widely. Of the analyses that might be considered formal, the best known are simple frequencies of occurrence, which form the basis for ratios and distributions. A good example may be a patient's pace (also known as inverted Type-Token Ratio), which is defined as the ability to generate new words in the course of a therapy session or even in the course of a total treatment. Pace is calculated by dividing the number of different words (vocabulary size) by the total number of words (text size) in a given text. As we know from literary research "pace is related to maturity and development" (Baker, 1988, p. 37). In the context of psychotherapy research a patient's increasing power of verbal versatility may be interpreted as a sign of working-through and improvement. Thus pace may be seen as an objective measure for psychotherapeutic process in both macro- and microanalytic perspectives. Transcription standards are essential if such a measure is to be applied reliably to textual materials of different sources (e.g., different speakers, transcribers, and institutions); much depends on what a word is defined as, what spelling is used, what markers may be used, and so on.

Another formal analysis is measurement of speech disturbances, including production of broken words, broken thoughts, repetitions of words or fragments, stuttering, and use of specific interjections and particles. Such dysfluencies may occur in connection with warded-off material and already have proved to be clinically sensitive (Mahl, 1956). A qualitative rather than quantitative approach is obtained with computation of the characteristic vocabulary. When applied to patient's and therapist's speech it yields all words that have been used significantly more or less often by one speaker compared to another. Also interesting is the common vocabulary, that is, all words shared by both patient and therapist. These vocabularies can be examined in a variety of ways, including comparison of their constituent words, sizes, and changes over time. Other interesting formal aspects of discourse might be pause times (silences) and their relation to phenomena of resistance and defense; or emphasized or lengthened word pronunciation, as evidence of affective or emotional tone. Without standardized transcription and notation of such aspects of discourse, quantitative computer-aided text analyses and comparative studies would be impossible.

An example of a grammar-oriented measure is part of speech distribution. It is known from previous studies, that various classes of patients differ significantly in their use of syntactic categories; for example, neurotic patients tend to use more verbs than do psychotic patients (Lorenz & Cobb, 1954). Part of speech distance characterizing the similarity/dissimilarity of therapist and patient use of parts of speech can be calculated using chi-square statistics. In a recent single case study, Mergenthaler (1990) observed a significant interaction effect between part of speech distance and therapeutic alliance. Obviously, the way that nonverbal utterances, non-understandable words and other phenomena are transcribed will affect quantification and subsequent interpretation of any variance. Clearly, then, standards are needed.

Among the substantive measures, content-analytic dictionaries have been applied to study several areas of interest: affective language, anxiety themes, regressive imagery, speech disturbance, referential activity, emotional tone, abstractness,

vagueness, and more. These dictionaries may be used for classification purposes either in longitudinal or in transectional studies by calculating the frequency distribution of these categories or possible subcategories. They can also be used as screening instruments to locate moments of interest in a therapy session or to locate outstanding sessions in a series of transcripts, which might then undergo more detailed analyses. Such dictionaries can be easily distributed among scholars and thus contribute to collaborative research endeavors. However, without underlying standards, findings in various sites may not be comparable. Transcription standards do not only benefit computer-assisted text analysis; such standards can help synchronize convergent multidisciplinary research involving audio and video recordings, physiological parameters, pre- and postsession rating scales, and other data. For example, current multisite research organized by PCUMP entails use of therapy records by several groups analyzing person schemas, defenses and controls, and theme and topic. Although each group has some unique data sets, transcripts following the standards described here form the data set common to all. All of these approaches, with origins in various scientific traditions, are able to use the transcripts prepared from the clinical material according to these standards.

Therefore it is our hope that this paper will offer a significant step in the direction of uniform standards, enable greater sharing of archival records among various sites, and perhaps help in the production and preservation of important large archives for future research.

DEVELOPMENT OF TRANSCRIPTION STANDARDS

GUIDELINES

The standards described here are a part of an effort to guide the creation of transcripts that are useful for both *human readers* and scientific analysis *using computers*. Obviously, development of standards is always difficult: Universal standards are certain to please no one every time or in every case, and standards will inevitably evolve as needs change. Nevertheless, adherence to a carefully designed foundation can make life much easier for researchers working with transcripts, and can provide rational guidelines for consistent future development. Besides the desire to be independent of specific research goals, the development of transcription rules has been guided by seven principles.

1. *Preserve morphologic naturalness of transcription.* The graphemic presentation of word forms, the form of commentaries, and the use of punctuation should be as similar as possible to the presentation and use generally accepted in written text.
2. *Preserve naturalness of the transcript structure.* The printed format should be as similar as possible to what is generally accepted, like the printed versions of radio plays or movie scripts. The text must be clearly structured by speech markers.
3. *The transcript should be an exact reproduction.* The loss of information resulting from the transition from a visual and/or acoustic to a written record of the interview should be as small as possible. A transcript should not be prematurely reduced but should be kept as a raw data form.

4. *The transcription rules should be universal.* The rules governing transcription should, as much as possible, make the transcripts suitable for both human and machine use.
5. *The transcription rules should be complete.* It should be possible for the transcriber to prepare transcripts using only these rules based on his or her everyday language competence. Specific knowledge, such as codings stemming from various linguistic theories, should not be required.
6. *The transcription rules should be independent.* It should be possible to transcribe various kinds of therapeutic discourse with the same set of rules. Transcription standards should be independent of the transcriber, understandable and applicable by secretaries and scientists.
7. *The transcription rules should be intellectually elegant.* The transcription rules must be limited in number, simple, and easy to learn.

These seven principles have guided the development of the transcription rules detailed below. However, these principles are ideals that generally cannot be achieved in practice. Some compromises were necessary; for example, in a preference for psychotherapeutic discourse with two participants. These standards are still applicable for discourse involving more than two persons, but phenomena such as simultaneous speech are difficult to handle clearly. In such cases the reader is directed to ways of parallel transcription as with a musical score and as discussed in (Edelsky, 1981). There is also software available that will assist this kind of transcription task on a personal computer.

The standards presented here might be seen as a common denominator for a variety of scientific and other uses of transcripts. The original German version has been in use for over 10 years in the production of several thousand transcribed sessions. The English version has been applied to several hundred transcripts. An Italian version has recently become available, and a hundred or more transcripts prepared. All these transcriptions have been used successfully for a variety of applications: clinical supervision, clinical training, teaching, and research. In most cases, the standards were completely adequate; in a few research applications involving linguistic questions, some additional rules were needed and were added by the researchers.

QUALITY AND RELIABILITY

Transcriptionists cannot be expected to produce a perfect transcription on the first pass. It is usually important to have a second person verify the transcript by reading it while listening to the audio record (while watching the video recording, if available). At UTB and PCUMP where this is a standard procedure, the number of alterations during verification averaged approximately 3% of the total number of words and punctuation markers in the text. This varies depending on the quality of the recording, the characteristics of the speakers' speech, and the experience, skill, interest, and other personal variables of the transcriptionists and verifiers. But still, this is not a measure or proof of reliability: the choice the first transcriptionist makes in transcribing an unclear statement might bias the verifying person who might have made another choice had he or she been first. To estimate reliability, therefore, a sample text was given to three transcriptionists and verifiers working independently. To calculate coefficients of reliability² these transcripts were pro-

Table 1. Reliability Coefficients (Cohen's *Kappa*) and Percentage of Agreement in a Transcription Task

Verification	Reliability			% Agreement	
	^k AB	^k AC	^k BC	A,B,and C	2 out of 3
before	.71	.70	.76	63	85
after	.74	.74	.77	66	95

Note. Based on one audiotaped psychotherapy session (German origin and German edition of transcription standards), using three transcriptionists A, B, and C. Verifying has been done by three other individuals on different thirds of each original transcript.

cessed in the following way: Each was transformed into a vertical text (each word, punctuation marker, etc., was placed on a separate line), and the transcripts were displayed in parallel and rated for differences (for details, see Mergenthaler, 1991). The major differences found were missing short turns of speech, words, or punctuation markers. Also frequent was the replacement of a word by another one that had a similar meaning in the respective context even though the words were different in pronunciation (e.g., “*natürlich*” vs. “*freilich*”). Table 1 gives the reliability measures calculated before and after verifying.

These data clearly show that verification improves the quality of a transcript. The resulting coefficient of reliability around .75 may be seen as quite acceptable for this kind of task and is clearly above what we would expect without standards or when using different ones. This means that for quantitative approaches like counting words, punctuation markers, and other formal aspects the findings clearly will not be affected by the remaining error.

But from a qualitative point of view, even one transcription error such as omitting the word “not” might have a severe impact on the semantic interpretation of a whole session. Such qualitative concerns raise the question of the need for second or third verification passes in the preparation of transcripts. Evaluation of the types of changes made during transcript verification and their impact on overall meaning was the goal of another study done with transcripts of two different therapy hours.

Each transcript underwent three successive passes of verification editing followed by word processing revision. Two verifiers working independently alternated passes with each transcript, such that each subsequent pass was done by the other verifier (each person performed the first and third verification on one transcript and the second verification on the other). After all verification was complete, a third person serving as a judge, was asked to characterize each alteration mark that had been made and to rate the impact that each change had on the overall meaning conveyed by the transcript (for more detail, see Stinson, 1991). The major findings are given in Table 2.

The number of alterations shows a steep decline over the three passes. In the second pass the total number of additions, deletions, or changes of words, punctuation markers, comments and turn of speech indicators is about 1% of the total number of these elements in the transcript. For usual applications, therefore, one

²The problem of calculating reliability for transcription rules has not yet been dealt with in the literature.

Table 2. Total Number of Corrections Made during a Verification Task with Three Passes

Transcript	Total Units	1st Pass		2nd Pass		3rd Pass	
		Alts.	% ^a	Alts.	%	Alts.	%
T1	5547	270	4.86	69	1.24	27	0.49
T2	6552	175	2.67	53	0.81	13	0.20

Note. English language transcript and English edition of the transcription standards.

^aNumber of alterations expressed as percentages of total number of units (words, punctuation, comments, turns of speech).

pass of verification might well be sufficient. This assertion is strengthened when one looks at the kinds of alterations and their impact on the meaning of the transcript. While additions were the predominant alteration in the first pass (74.8%), corrections were predominant in the third pass (48.1%). The impact of the changes on overall meaning is different in each pass ($p < .01$) but not that important (M 2.21 on a 1 to 5 scale with "1" being "no impact"). Among the categories, alterations of words (additions, deletions, and changes) had the most impact on meaning, and punctuation had the least.

COST OF TRANSCRIPTION

From past experience with these standards, it is a good estimate that a skilled typist will require approximately eight hours to perform the first-pass transcription of one 50-minute therapy session. Another 4 hours is needed for verification and correction (second pass). For speech records with difficult or unclear speech, such as heavily accented speech, idiosyncratic speech, or dysfluent speech with stuttering, transcription and verification may require 50% again as much time. Entering time codes—if necessary—might require another hour. Verifiers need to have good reading and listening skills; the task can be performed by typists, secretaries, research assistants, or scientists.

DESCRIPTION OF RULES

These rules might be applied to transcripts written by hand, on a typewriter, or on a computer using a word processor. More complete versions (English, German, and Italian) including more technical details are available from the authors upon request.

WHAT TO TRANSCRIBE?

Verbal Utterances. All words spoken as whole words or parts of words are reproduced in standard spelling. Dialect forms are transcribed in their corresponding standard spelling forms. For example, if an English speaking person's usual speech sounds like the following:

P: I know she ain't gonna gimme lotsa trouble.

it should be transcribed using standard English spelling as follows:

P: I know she ain't going to give me lots of trouble.

Note that the word “ain’t,” although substandard, is retained in its standard dictionary spelling. For transcribing instances where a speaker deliberately uses dialect forms signaling emphasis or humor, see below.

Paraverbal Utterances. All sounds or sound sequences serving as conversational gap fillers, expressions of feelings of doubt, confirmation, insecurity, thoughtfulness, and so on in English are written in the following standard spellings whenever possible (modified from Dahl, 1979):

- *Affirmative:* mm-hm, uh-huh, yeah, yup
- *Negation:* huh-uh, nah, uh-uh, hm-mm
- *Noncommittal:* hm, mm
- *Hesitations:* ah, eh, em, er, oh, uh, um
- *Questioning:* eh, huh, oh
- *Humor:* ha, haha, ho, hoho
- *Exclamation:* ach, aha, ahh, bang, boom, ech, hey, kerbang, oh, ooh, oops, ow, pooh, pow, uch, ugh, wham, whew, whomp, whoo, whoops, whoosh, whop, wow

Additions to this list might be needed. The authors would be grateful to receive submissions for consideration.

Nonverbal Utterances. All other noise-producing actions of the speaker are recorded where they occur in the text in the form of simple comments within parentheses:

P: (sneeze)(cough) well (sigh), I guess I caught a cold (laugh).

Noises Occurring in the Situational Context. Any other sounds produced by the situational environment are indicated within simple comments:

P: later when I (telephone rings); do you need to answer that?

Pauses. One may use a single dash character surrounded by spaces (-) to indicate a pause of approximately one second. Multiple dashes should be separated by spaces. Pauses of greater than approximately 5 seconds should not be indicated with dashes, but should be timed and indicated using the following coded comment form:³

P: I can think of - . nothing (p:00:03:35) nothing at all.

The example above indicates a pause of approximately 2 seconds and a second pause of 3 minutes and 35 seconds.

SPECIAL TRANSCRIPTION MATTERS

Incomplete Words. Word particles generated by word breaks, including stuttering and stammering, are indicated by the word fragment followed by a hyphen (-) and a space. A broken word is defined as an incomplete word that is not repeated:

³There is hardware available that will allow for automated measurement of pauses and other time intervals.

P: whenever- I can never visit them alone.

Stuttering is defined as: (1) one or more word particles, each sharing the initial letters of the following completed word; or (2) a sequence of more than one word particle, each particle sharing initial letters, but not followed by the completed word:

P: sb- sb- she t- t- t- asked me not to call her again.

Indecipherable Utterances. A single slash (/) is entered in the transcript for every utterance that cannot be clearly comprehended but can be distinguished as a separate word. A slash marking an incomprehensible word may be followed with a coded comment of the form "(?:word)" to indicate possibly correct words. Thus the "?:" indicates that the comment contains a word or words that may have been uttered by the speaker:

P: I was /(?:alone) there all /(?:night) / until he / / /.

If one cannot determine the number of words in an utterance or any of the possible words, this should be simply indicated with the following comment:

P: (incomprehensible)

Quotations. If the speaker directly quotes prior discourse, the text for each speaker is enclosed in single forward quotation marks ('), which is the same character as the apostrophe:

P: I asked 'will you do it?' and he yelled 'stop talking to me like that' and slammed the door.

Changes in Manner of Speaking. If the speaker changes his or her usual manner of speaking and uses a voice differing from the usual way of speaking, the words are enclosed between double quote character (""). In such double quoted text, slang and literal transcription may be used.

P: she tells me not to say "yaw! come back now" and "gimme that". what does she think this is, grammar therapy?

Punctuation. Punctuation markers are used to help the reader reconstruct the original flow of speech. They are not used according to traditional grammatical rules, because normal speech is rarely so well-ordered. The transcriber should use punctuation marks to indicate changes in the way of speaking, emphasis, intonation, and cadence. When in doubt, punctuation marks should not be used. Punctuation markers are always placed at the end of a word and should not split a word. The following situations are differentiated:

1. *Completion of a thought.* The clear period (.) indicates the end of a completed thought and is usually accompanied by a drop in pitch.
2. *Broken thought.* The semicolon (;) indicates a broken thought, followed by another thought, for example:

P: I hate the way you; did I tell you about the wedding?

3. *Hesitation.* The comma (,) indicates a hesitation followed by a continuation of the same thought and is usually accompanied by a slight drop in pitch, for example:

P: you, never seem, to look at me when I am talking.

4. *Question.* The question mark (?) indicates a question, usually accompanied by a rise in pitch, or a clear rise in pitch. It should be used at the end of possible questions indicated by a rise in pitch even if the statement does not contain a clear grammatical question form:

T: Do you dislike it when he does that?

P: I should like! it when he does that?

5. *Emphasis.* The exclamation mark (!) immediately follows words clearly emphasized by the speaker as in the prior and following examples:

P: that may not matter to him! but I do not! like it.

Note that the exclamation mark in transcription is used only for emphasis and does not indicate the end of a grammatical sentence.

6. *Lengthened pronunciation.* The colon (:) is not used in its traditional grammatical way but is used to indicate protracted or extended pronunciation of a word as in the following example:

P: well: I never really: liked that much anyway.

FORMAL AND STRUCTURAL ASPECTS

Transcript Heading. The transcript should contain a header. The following example shows the types of information that one may wish to include. The entire set of information should be enclosed in parentheses as a comment:

(SUBJECT ID: 105, SESSION NO: 32, DATE: 29.SEP.1986, THERAPIST: Dr. Smith, TEXT TYPE: psychoanalytic session, VERSION No: 1.0)

Speaker Codes. Each turn of speech begins on a new line and is preceded by a code indicating the speaker. Speaker codes are of the format *Xn*: wherein *X* is a single letter indicating the speaker's role and *n* is an optional digit (if there is more than one speaker of a certain role). If *n* is omitted, it is assumed to be the digit 1. Thus, in the following example;

T: how did that make you feel?

P1: I felt confused and angry.

P2: you never told me you were angry about that.

the first speaker *T* is a therapist and *P1* and *P2* are two patients. The speaker code *T* has an implicit digit component of 1 and is therefore the same as *T1*. This format can

handle monologue, dialogue, individual therapy, group therapy, and single therapists or cotherapists.

A comment after the transcript header can be used to clarify the role of speakers, for example:

(P = Son, P2 = Mother, P3 = Father, T1 = Therapist, T2 = Cotherapist)

Capitalization. With the exception of proper or personal names or the first person pronoun "I," all words including the first letter of a sentence begin with lower-case letters. This enables the use of even the simplest word-counting programs.

Simultaneity. Simultaneous speech presents special problems, both for comprehension and for representation of text. For two speakers however this can be easily handled by inserting a plus sign (+) at the start of simultaneous speech and continuing transcription of the initial speaker until simultaneity ends. This is followed by the entire simultaneous speech of the second speaker and terminated by another "+". The remainder of the nonsimultaneous speech is transcribed in its natural order. In the following example, the words "refused again" and "yes you" were spoken at the same time:

P: I was going to give John the map but he +refused again

T: yes you+ have told me this once before.

Transcription of simultaneous speech is much easier if the dialogue can be recorded in stereo with separate microphones for the patient and therapist.

Compound Words. Compound words with standard hyphenated spellings are connected by hyphens without spaces:

P: I found the picture taped upside-down on the wall with a band-aid.

Neologisms. Neologisms are spelled as best as possible. Words that are created by stringing other words together should be represented with hyphenation:

P: all this gaming-it-out is confusing me.

Word Division at the End of a Line. If the text is for computer-aided text analysis, words should not be split at margins using hyphens (this creates problems for some computer-aided text analysis tools); the word should be typed in full on the next line.

Contractions. The apostrophe (') should be used to indicate contractions:

P: it's not fair that they'd get to go and I wouldn't.

Text analytic systems can then treat the two parts separated by the apostrophe as separate words (e.g., *wouldn't* becomes *wouldn*, which can be treated as *would*, and *t*, which can be treated as *not*). If a contraction produces ambiguous parts, either the words should be spelled out completely or else the ambiguous parts should be followed with a slash and the clarifying word (or words connected by a hyphen without spaces as described above) as in the following example:

P: he'd/bad not done it and he'd/would never do it.

In the first case *d* stands for *had* and in the second case *d* stands for *would*. Without the additional information following the slashes the two *d*'s would be processed as the same word. If 's is not clarified, it should be assumed to represent the word *is* ; if 'd is not clarified, it should be assumed to represent the word *would*.

Do not use the apostrophe to indicate apheresis (the omission of letters at the beginning or end of a word). The word 'cause, for example, should be spelled out in its standard English form *because*. Do not use the apostrophe to indicate the possessive case. Instead of such forms as *Mary's* and *John's* one should transcribe as follows:

P: that coat is Marys and this one is Johns.

Plurals. The apostrophe should not be used to indicate plurals of letters, numbers, acronyms, or abbreviations. The underscore can be used for clarity, if necessary:

P: he always got As because he was the teachers pet.

P: she only types lower case a_s because her typewriter is broken.

Abbreviations. With the exception of formal titles, abbreviations are not used unless the speaker verbally spells one. Periods are not used in abbreviations; use a space instead:

P: Mrs Smith thinks I made a terrible mistake, for example.

T: mm-hm.

P: and it irritates me that Jane always says "e g".

Numbers, Fractions, and the Like. Numbers and fractions are written out in full where possible. Only typical figures such as dates are transcribed as numbers. The abbreviations for "ante meridiem" and "post meridiem" should be capital letters without spaces (AM and PM):

P: in 1981 I saw the first two-thirds of a James_Bond_007 film at eleven-thirty PM for two dollars and fifty cents.

Mistakes. Slips of the tongue and other mistakes are transcribed in full:

P: I couldn't stand the guilt, uh quilt she gave me for my birthday.

Correct Spelling. Spelling should follow Webster's standards.

Where several marking rules apply, it is necessary to include them all in sequence, with a period or question mark going last:

P: he screamed 'don't shoot until you see the whites of their eyes'!

Some Things To Avoid. Do not use a sequence of periods (...) to indicate ellipsis. Do not use special characters (such as { }) unless needed for special purposes of your own.

ADDITIONAL AND OPTIONAL RULES

The following set of rules can be of help for research settings with special needs.

Names. If confidentiality is an issue, pseudonyms may replace personal names, names of places and other identifiers. To signify that a name has been changed, precede it with an asterisk (*) without an intervening space. It is proposed that a separate list of substituted words be maintained and used consistently throughout all material transcribed for the same speakers:

*P: *Jane told *Fred all about *Elliot and *Mary.*

If more than one word is needed to replace a single word, the multiple substituted words should be joined by underscore characters (_) without intervening spaces. This enables the entire substitution to be counted as a single word in the case of subsequent computer text analysis:

*P: *Albert changed his name and moved to *small_southwest_town.*

If a title is to be used before a name, it should be separated from the name with a space. Apostrophes should be omitted from names containing them; hyphenated names should retain the hyphens. Names (even those not substituted by pseudonyms) should be joined with underscores to form a single entity:

*P: Mr *Arnold_OMalley wants to be on Hollywood_Squares and meet Eva_Gabor.*

Date and Time Coding. The date, time of day, and elapsed times of a transcript may be inserted using special coded comments.

1. *Session date.* The session date is indicated with a coded comment of the following form:

(d:10.JAN.1986)

The *d:* indicates the comment is a session date. The date is entered in the format "DA.MON.YEAR" (a two digit representation of the day of the month, a three-capital-letter abbreviation of the month, and a four-digit representation of the year, separated by period without spaces). Thus "*(d:06.MAR.1986)*" represents "*March 6, 1986.*" The session date should be placed at the top of the session transcript just after the heading (note that the form of this code makes it accessible to computer systems). If the exact date is unavailable, the unknown information should be replaced by zero's.

2. *Time of Day.* The beginning of session time is indicated with a coded comment as in the following example:

(t:10:02:15)

The *t:* indicates the comment is the actual time of day of the session, if available. All time codes are in the format "HH:MM:SS" (two-digit representations of hour, minute,

and second each separated by a colon). (Some facilities may allow the notation of video frames also, in which case the time codes would be in the format "HH:MM:SS:FF"; if this is used, it should be clearly indicated in a simple comment at the beginning of the transcript.) Thus "10:02:15" represents 2 minutes and 15 seconds after the hour of 10 O'clock. It is preferable to use 24-hour clock time. The session time should be placed at the first of the session transcript on the line following the session date. If the exact time is unavailable, the unknown information should be replaced by 0's.

3. *Elapsed Time.* It is often helpful to insert elapsed time codes in a transcript. The relative time within a session is indicated with a coded comment of the following form:

P: we saw the movie (+:00:03:00) after dinner.

The "+:" indicates the comment contains the elapsed time since the beginning of the session. The "00:03:00" indicates this is the start of the third minute following the beginning of the session. If the minute changes in the middle of a word, the time code should be placed before that word. The interval between relative time codes (if they are to be used at all) depends on the nature of the study. For example, these codes can be used to relate the text to other temporally ordered data (e.g., physiological recordings). These might be placed at the beginning and end of specific events or they might be placed at regular intervals, such as every whole minute or every 5 minutes.⁴

Ambiguity. Some statements may be ambiguous in print yet unambiguous when heard in a sound recording. It is to the advantage of both computer-aided analysis and human readers to convert such ambiguous utterances into unambiguous ones. A clarifying alternative word may be placed behind a slash (/). Alternatively, a number placed immediately after the slash can be used to indicate the index number of a word's meaning in a specific content-analytic dictionary. In the case of ambiguous pronouns, it is possible to name the antecedent behind the slash or to include several words connected by hyphens (this rule is primarily for use during the verification and scientific annotation phases of transcript preparation):

P: we/group thought he/James-Joyce had ignored it/rules-of-the-game.

Segment Demarcation. Various segmentations of the transcript may be accomplished by using coded comment structures to indicate the start of a "(s:CODE)" and the end of a segment segment "(e:CODE)." These two are used to bracket a segment of type indicated by "CODE," e.g., "DREAM." Whatever word is substituted for "CODE" must be spelled exactly the same in both start-segment and end-segment coded comments. It is permissible for different segment types to overlap or embed. This approach can be used for many types of segmentation. One might segment by relationship episodes and dreams, as in the following example:

*P: (s:RE) When I told *Jane that last night I dreamed (s:DREAM) I was a butterfly (e:DREAM) she laughed (e:RE).*

⁴There are various technical aids to assist in entering time codes.

The coded comments “(s:RE)” and “(e:RE)” indicate the beginning and end of a relationship episode, respectively. The coded comments “(s:DREAM)” and “(e:DREAM)” indicate the beginning and end of a dream description.

Segment demarcations of the same types must not overlap or be embedded. This will not usually be a problem.

GENERAL ASPECTS

Transcriptionists should be individuals who are attentive to detail, motivated, interested in the work, and aware of its value to research. They should not have rigid time limits for completing transcripts; transcriptionists should be allowed to work without unnecessary interruptions. There will be differences in time to transcribe different sessions, for the transcriptionist will need to become familiar with the speech characteristics of each new speaker. It is also important that transcriptionists and verifiers be aware of the potential to have feelings similar to those expressed in the therapy, especially depression. They must have the opportunity to take breaks and work on other, non-emotional material from time to time.

TRANSCRIPT EXAMPLE

We have included the following mock interview formatted according to the transcription standards described above. Several examples of problems typically encountered in preparation of psychotherapy transcripts for research and education purposes are shown:

(SUBJECT ID: 105, SESSION NO: 1, DATE: 9.JAN.1988)

(d:11.JAN.1988)

(t:11:03:00)

(T = Dr. Jones, P = John Doe)

T: can you, do you recall similar episodes in your /(?adolescence) of / / ?
(microphone drops) you um, mentioned that uh an important person
for you during your adolescence was your school teacher Miss *Green.

P: mm-hm.

T: is there a particular incident that stands out in your mind or
(+:00:01:00) interaction between the two of you that stands out +in
your mind that

P: um I think that I can+ credit her for - - sort of turning me around uh
academically, you know, because I was pretty much of a, I would
not work real hard. and I think that um there was always a recogni-
tion that I had some potential um to do well in school, but never
did well.

T: (+:00:02:00) mm-hm.

P: and uh my brother was absolutely brilliant and everybody loved him
and he was valedictorian, and so I had to sort of uh come in his shadow
through grammar school and high school, you know, because we went

to the same schools. so this Miss *Green, she'd/had been his teacher. I remember her very clearly, I can picture her face very clearly. um and she decided, I guess, that I was not going to slide anymore. I like t- to fool around a lot, you know.

T: mm-hm.

P: uh passing notes, goofing off, uh doing (+:00:03:00) things, I mean, not serious. but Miss *Green, she knew all this, I'm sure. well, it came grade time for the first marking period, and I knew enough never to get a C. and this one marking period she gave me two Ds!, uh (+:00:04:00) one D in math um

T: hm.

P: so she gave me a D, and I was just terrified. um she handed the report card t- to me and looked at me um for a long time, I remember that, and she made the D in red, in uh really thick: red: letters.

T: (incomprehensible)

P: and uh so I was just (+:00:05:00) really flabbergasted. I didn't know what I was going to do, because I knew to get a s- D was just awful.

T: mm-hm. do you remember what you were feeling a- a- at the time?

P: uh oh this really sinking feeling, like (child-like voice) "oh no".

T: mm-hm.

P: I had really, really screwed up and didn't know how I was going to get out of it. (p:00:00:40) (sigh) (+:00:06:00) uh it just seemed like the worst thing in the world that could happen. um and I knew my parents were going to be upset. and I knew that um it would be hard to undo that, you know. so I was feeling really, really um now I was afraid, um I was really dejected by it, I mean, that uh that she had done this. because I, actually m-, I should mention, I didn't feel I deserved a D. - - - I think she gave it to me to motivate me.

T: mm-hm. do you remember; what do you think, was going on in her (+:00:07:00) mind at the ti-

P: yeah, I r- I remember, because I said she looked at me for a long time when she handed my report card to me. her saying um, like around, it was after the report card that um that she said 'you: are not! going to get away with doing no work ' and and that I was really going to have to do well.

T: mm-hm.

P: uh to get grades in her classroom. and she reited, reiterated that again. and at the (+:00:08:00) time I thought she was just the most stern, unreasonable person um, I mean, I do recall um really after that um not liking her and uh. so I do think that she, but uh but see I think what she was doing was something really um very caring and very positive. I mean she had singled me out, I think, or maybe she did other people too, to really just get them on the ball.

DISCUSSION

Verbatim protocols form an increasingly important source of data for psychotherapy research, including both computer assisted and human judgment methods. They may be equally important to research in educational psychology, linguistics, an-

thropology, and other fields. In addition, transcripts provide important information and illustrations for training. It is unfortunately and unnecessarily the case that transcripts prepared for one use are all too often entirely unsuited for other types of uses, even within the domain of research.

The standards proposed above have served well for several years for a number of types of analyses, and are immediately useful to researchers or educators facing the many challenges of converting naturally occurring spoken words to transcripts. Currently available software can be used to perform complex analyses of transcripts prepared in this fashion. Furthermore, the guidelines provided here offer a rational path for refinement and extension to suit future needs, including new advances in computer-aided analysis.

REFERENCES

- Baker, J. C. (1988). Pace: A test of authorship based on the rate at which new words enter an author's text. *Literary and Linguistic Computing*, 3, 36-39.
- Dahl, H. (1979). *Word frequencies of spoken American English*. Essex, CT: Verbatim.
- Edelsky, E. (1981). Who's got the floor? *Language in Society*, 10, 383-421.
- Edwards, J. A., & Lampert, M. D. (Eds.). (1991). *Transcription and coding methods for language research*. Hillsdale, NJ: Erlbaum.
- Kächele, H. (1983). Verbal activity level of therapists in initial interviews and long-term psychoanalysis. In W. Minsel & W. Herff (Eds.), *Methodology in psychotherapy research* (pp. 125-129). Frankfurt, Germany: Lang.
- Lorenz, M., & Cobb, S. (1954). Language patterns in psychotic and psychoneurotic subjects. *A. M. A. Archives of Neurology and Psychiatry*, 72, 665-673.
- Luborsky, L., & Spence, D. P. (1978). Quantitative research on psychoanalytic therapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy research and behavior change: An empirical analysis* (pp. 331-427). New York: Wiley.
- Mahl, G. F. (1956). Disturbances and silence in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology*, 53, 1-15.
- Mergenthaler, E. (1985). *Textbank systems: Computer science applied in the field of psychoanalysis*. New York: Springer-Verlag.
- Mergenthaler, E. (1990, June). *Parts of speech: A measure of therapeutic alliance?* Paper presented to the Society for Psychotherapy Research, Wintergreen, VA.
- Mergenthaler, E. (1991). *Zur Reliabilität von Transkriptionsregeln*. Unpublished manuscript.
- Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., & Tucker, G. J. (1985). Linguistic dimensions of affect and thought in somatization disorder. *American Journal of Psychiatry*, 142, 1150-1155.
- Stiles, W. B., & Sultan, F. E. (1979). Verbal response modes and psychotherapeutic technique. *Journal of Consulting and Clinical Psychology*, 47, 611-613.
- Stinson, C. H. (1991). *Transcript verification analysis*. Unpublished manuscript.

Received January 30, 1990

Revision received March 19, 1991

Accepted June 24, 1991