The Entextualization of Talk

Richard F. Young

University of Wisconsin-Madison

www.wisc.edu/english/rfyoung

Speaking in context

The view of speaking that I want to discuss in this paper differs from the views presented by my colleagues. Many scholars, including those who have presented here today, focus their attention on speaking practices in such a way that what speakers say is examined in isolation from the context in which the practices occur. That is to say, many scholars have accepted a view of speaking as a self-contained system of structures that form patterns and meanings within the system itself. I wish to problematize this view of speaking and, instead, to examine the ways in which a speaking practice relates to the context in which it occurs and which the practice itself helps to create. The key objective of treating an instance of speaking as a practice in context is to understand how cultural meaning is produced and understood in the practice.

A part of the context in which every practice of speaking occurs and constructs is a reaction from a listener. In most instances of a common speaking practice such as a conversation, a speaker produces speech that in turn produces a sequential response in terms of more speech from another speaker (Example from Hanh). Other reactions are of course possible and include nonverbal physical actions (Example from Hanh), or nonverbal assessments and evaluations of the speaker (Example) and of the speech itself (Example).

There are, however, several speaking practices in which the reaction of a listener to speaking is not more speaking. In a semidirect test of speaking such as the Test of Spoken English (Educational Testing Service, 2001) for example, a candidate records her speech and the speech provokes no immediate sequential response from a listener. The candidate's speech is recorded and an assessor responds to it at a much later time. And

the assessor's response to a TSE recording is not more speech, but instead a written evaluation.

In a direct test of speaking on the other hand, such as the ACTFL oral proficiency interview (American Council on the Teaching of Foreign Languages, 2001), an examinee responds to speech from an ACTFL tester. The examinee's response takes the form of more speech and the tester in turn responds to the examinee's speech. The nature of the tester's response, however, is determined only in part by the examinee's speech. The topics and nature of the tester's speech are also a response to the tester's previous reading of the established ACTFL protocol of warm-up, repeated level checks and probes, and wind down. Whether the tester is responding to the examinee's speech or instead following the protocol is often apparent from the tester's talk. And there are occasions when the tester's response to the interview protocol appears to be very different from what we would expect from a response to the examinee's speech (Example from (Johnson & Tyler, 1998)). A further response to speaking occurs, according to the ACTFL protocol, because the interview is recorded and a second tester responds to the recording at a later time. As with the TSE recording, the second ACTFL tester responds with a written evaluation.

These three examples of how assessors respond to speech are worth considering because they show that in assessing speaking assessors respond to speaking in a way that is similar to how we often respond to writing. Although in many speaking practices participants respond in real time with more speech to a person speaking, assessors who evaluate a recorded speech sample respond in ways that are similar to the ways in which readers respond to writing. And even the live examiner in an oral proficiency interview

responds according to a script. In comparisons with speaking, responses to writing may not be so overt or so immediate. When speaking is recorded it becomes possible for us to respond to the recording of speaking in the same way that we respond to writing. That is, we can ignore the speaking, we can go to sleep during it, we can play it over and over again, and we can examine aspects of the recording in great detail. We can also respond to it, but because we do not respond immediately we can choose to be more considered in our response.

Another way in which an assessor's assessment of speaking becomes a response to a text is in the different role that context plays in responding to speaking and in responding to text. Speech among individuals takes place in real time, and when it takes place face to face it happens in a shared social, historical, and physical context. Because speaking is located in context the participants may index the context and thus co-participants may pragmatically infer meanings (Example). An assessor of a spoken or written text, on the other hand, is located in a different social, historical, and physical world from the world in which the speaking takes place.   Skilled writers are well aware of the different worlds that they and their readers inhabit and thus a skilled writer will index co-text but not context. A reader cannot easily pragmatically infer meanings from the world in which the writer lives and writes. It is for this reason that scholars of literature spend their careers researching the social, historical, and physical context of a writer in order to uncover meaning in text (Example: Memorial resolution for Mert Seals).

Instead of responding to speaking as we do in conversation or real time discourse, a recording is a first step toward entextualization of speech, that is, of transforming

speaking into a cultural object that can be evaluated. I thus wish to argue that the act of assessing speaking transforms it from a practice that takes place in real time and in a context with other participants into what is effectively a *text*. Speaking has become a cultural object that we can evaluate and criticize and attempt to reproduce. And I wish to argue that an assessor's response to a text differs significantly from a participant's reaction to talk.

The process by which a speaking practice becomes a text has been studied by anthropological linguists as transcription. Elinor Ochs was perhaps the first anthropologist to study entextualization of speech in her groundbreaking 1979 article "Transcription as Theory" (Ochs, 1979). Ochs showed how the theoretical interests of a transcriber affect the transcriber's attention to dialogic talk involving young children. More recently, John Haviland described how the multi-speaker conversations of Tzotzil speakers from the highlands of Chiapas in southeastern Mexico, were first detached from their indexical surround—their natural home—and how members of the same community repackaged the speech in written words in what they believed to be an appropriate textual form (Haviland, 1996). And Greg Urban described how members of the same indigenous community as speakers in southern Brazil make transcriptions of monologic talk (Urban, 1996).

I will take the process of entextualization described by these anthropologists as a partial model for the process of assessment by testers. Although the processes of transcribing and assessing appear different on the surface, there are parallels between the two that help us to illuminate the process of assessment of speaking. In the first place, both entextualization and assessment require the transcriber and the assessor to attend

selectively to speaking. An assessor and a transcriber react to speech with writing, and writing that is not a response to speaking but rather a transformation. Secondly, the indexical ground of speaking—that is, the physical, social, and historical context in which speaking takes places in real time—is not directly available to the copier nor to the assessor. And finally in both entextualization and assessment, speaking is no longer a fluid and volatile activity but instead an activity that has been captured as a text, a text that has become a cultural object in its own right that can be evaluated and criticized and reproduced again.

The normalization of interactional talk

Ochs (1979) described the entextualization of young children's talk by adult researchers and described the ways that the way the transcriber's biases influence the way that a transcription is made. Underlying Ochs's discussion was her observation that adult researchers attempt to make children's talk look like conversations among adults. Of the many biases that Ochs described when spoken interaction among children is transformed into text by adults, I will focus here on three: the transcriber's expectation that a turn-at-talk is contingent on prior turns, the expectation that conversational roles include those of an initiator and a responder; and the strong tendency to foreground verbal over nonverbal behavior.

Ochs pointed out that a transcript that is written and reads from top to bottom is based on the adult transcriber's belief that a child's utterance is contingent on the utterances that precede it. In reality, Ochs remarked, "young children frequently 'tune out' the utterances of their partner, because they are otherwise absorbed or because their attention span has been exhausted, or because they are bored, confused, or

uncooperative" (p. 46). In a similar way, an adult transcriber who writes a transcription of dialogue in two or more columns must assign one speaker to the leftmost column and, as Ochs remarked, "whichever speaker is assigned to the leftmost column has a better than average probability of being an initiator of a sequence of talk" (p. 50).

In addition to a transcriber's selective attention to contingency between sequential utterances and a bias in identifying conversation as a pattern of initiation and response, Ochs pointed out the overwhelming bias in transcription to foreground verbal over nonverbal behavior. Writing in 1979 Ochs remarked, "In nearly all … treatments of adult-adult speech behavior, nonverbal considerations in the immediate situation are minimized or ignored" (p. 52). Today, two decades later, there is much greater interest in the nonverbal aspects of interaction, as has been demonstrated by new strands of research in psychology and linguistics (Duranti & Goodwin, 1992; Goodwin, 1981; McNeill, 1992; Tannen, 1994; Young, 1994). However the textual bias in favor of words over gestures still persists, and this bias leads us to selectively delete an important part of the discursive practice. As Ochs remarked, "we frequently observe entire acts that are carried out nonverbally, for example, a tug used as an attention-getting device, a grimace as a rejection" (p. 55). Similar important functions are born by eye gaze although again these are rarely entextualized: "In some locations eye contact may function as a summons to attend to the speaker …; in other locations, it may function as a confirmation check, as a signal that the speaker is relinquishing the floor, as an invitation to assume the floor, and so on" (p. 58).

Many of Ochs's observations of how dialogic discourse becomes normalized in text are confirmed by the work of Haviland's (1996) study of entextualization of a corpus

of multiparty conversation in which groups of men gathered together with explicit

instructions to gossip about their neighbors. Haviland's focus differed from that of Ochs

because Haviland was interested in how Indians who had never seen a written text in

their local language of Tzotzil approached the process of creating a text.   Nonetheless

Haviland observed that transcribers attempted to eliminate such processing difficulties

apparent in the original as hesitations, false starts, filled pauses, and other dysfluencies.

Many of Haviland's other observations echo those that Ochs noticed in transcriptions of

children's talk. As Haviland put it,

> There is … a breathless, almost competitive creativity about the
> conversational moment: speakers vie with each other for the floor, the
> word, and the moral, pushing topics in edgewise and interlocutors aside.
> In the written text, all of these features are peculiarly bleached, or …
> *normalized*. (p. 50)

A second issue observed by Haviland is that a text is talk that "has been ripped

from its physical setting" (p. 58) and the social, historical, and physical environment in

which the original talk happened is absent in the text. Thus the pragmatic features of the

original speech context, especially the organization of its participants and the social

relations between speaker and audience are normalized in the process of entextualization.

And finally, Haviland described the same process noted by Ochs whereby the turn

structure and other interactional features of the multiparty original are "smoothed" or

normalized in the newly fabricated text.

Entextualization and power

While Ochs focused on the expectations that adults predicate of children's speaking,

Urban (1996) and both Haviland and Ochs described the normalization of multiparty talk,

Urban stressed the importance of the power relationship between the originator of monologic talk and the copier of the original. Urban's data was drawn from observations he made among an indigenous group living in southern Brazil. The copies of tape-recorded monologic discourse were made by two members of the group: a young man whom Urban had trained to write phonemically and an elder man who repeated for Urban syllable-by-syllable what was on tape for Urban to transcribe. Because of the different status that the two informants held in the community, Urban was able to examine the effect that the social relationship between the originator and the copier has on the copy. Urban summarized his conclusions in four propositions relating to what a copy adds to the original, what a copy omits from the original, whether a copier responds to the original, and what the copier chooses to share of the original.

Urban's first observation was that his younger transcriber often inserted syllables and words at the end of utterances in order to make the transcription conform to a grammatical ideal of completion. The general effect of these additions was to make semantically explicit in the text what was only pragmatically inferable from the spoken form. Urban summarized this observation in Proposition 1 as follows.

> When replication occurs in relatively deliberate contexts (such as that of transcription), the copy may differ from the original by including segmentable forms not found in the original that explicitly encode meanings that are only pragmatically inferable from the original. (p. 30)

Evidence of the converse of Proposition 1 was also found in Urban's data. That is, a copier may omit portions of the original in constructing a text. Urban found that both younger and elder transcribers omitted from the text items that the speaker indicated

as errors or as diversions from what the speaker intended. Urban summarized these phenomena in Proposition 2.

> A copy may differ from its original by lacking portions of the original that are metadiscursive instructions, especially indications of mistakes or deviations from an intended "text," and by "correcting" the mistakes so indicated. (p. 33)

The most original of Urban's observations, however, relate to the effect that the social relations between the originator and the copier had on the text that the copier created. The elder and younger transcribers responded quite differently to a folk myth spoken by another elder in the village. The younger transcriber appeared to acknowledge the authority of the speaker and to copy as much as he could of the myth into the transcribed text. The elder transcriber, however, "had heard and himself told this myth on numerous occasions, he had a sense that the discourse … was a copy of other originals to which he had had equal access" (p. 36). He thus saw the original speaking as yet another copy and as such it was not necessarily more definitive than other copies. He responded to the original by correcting, adding material, and emphasizing certain aspects over others while the younger transcriber hewed more closely to the speaking that he heard. Urban summed up this difference between replication and response in Proposition 3.

> The more symmetrical and egalitarian the relationship between originator and copier (or the more authoritative the copier with respect to the person for whom the copy is being made), the greater will be the divergence between copy and original, and the more likely it will be for the copier to respond to the originator. (p. 37)

In his final observation, Urban focused on the nature of the original discourse itself and the relationship between the nature of the discourse and how shareable it is, that is, how appropriate the transcriber believed the process of transcription to be. The main source of Urban's texts were myths and historical narratives that were common knowledge in the community, but the transcribers were reluctant to transcribe expressions of personal emotions, in particular when a ritual lament came close to an expression of actual grief. Urban summarized these observations in Propositions 4a and 4b.

> The more the discourse is overtly coded as a unique instance, produced by its originator, and linked to a present context and circumstances, the less likely will the copier be to (want to) replicate it or metadiscursively to acknowledge the copy as a replication, and the more likely will the copier be to respond to it.

> The more the discourse is overtly coded as nonpersonal, that is, not as something generated by the originator but as transmitted by him or her, and the less it is linked to a present context and circumstances, the more likely will the copier be to replicate it; hence the more shareable it is. (p. 40)

Urban's four propositions underline the importance of the social relations between copier and originator in the process of entextualization. These social relations affect what is included in the text, what is excluded from the text, how the text is altered to reflect the copier's view of what is being said, and what the copier regards as appropriate and inappropriate to share.

Assessing talk or assessing text?

These three studies of how talk becomes text have indicated several ways in which the colors of the conversational moment are bleached. The interactional structure of multiparty conversations is normalized so that a turn-at-talk is shown as contingent on preceding turns, and what is copied is a sequence of initiations and responses. Speakers' dysfluencies are ignored and text focuses overwhelmingly on the verbal channel at the expense of the nonverbal. In text, the indexical ground of the talk is reproduced by means of adding words to the text, which co-participants in talk must infer pragmatically. And the relationship between the copier and the talker influences whether and how the text is produced. If the copier perceives the talk to be publicly shareable then he is more likely to produce a text than if he perceives the talk to be personal. And if the copier's status is high with respect to the talker, then he may respond to the talk by changing and correcting it.

How do these insights into the relations between talk and text relate to the tasks of defining and assessing speaking in the context of assessment? Although the cultural processes of entextualization and assessment share characteristics that give an impression of similar cognitive processes at work, the assessment of speaking has not been analyzed in the detailed ethnographic fashion that Ochs, Haviland, and Urban have brought to bear on the processes by which texts have been constituted from discourse. Such ethnographies of assessment are needed and may perhaps be a focus for future research by language testers. And one aspect of the process of assessment that has not been given sufficient attention by ethnographers is the degree to which the individual making the transcription is a member of the community of speakers. In many instances of speaking

assessment (but not all), the assessor is a native speaker and the speaker is a nonnative speaker, and thus the assessor is not a member of the speaker's cultural group. In the case of assessors who assess recorded speech, their lack of membership in the community of the speaker compounds the difficulty that an assessor has in inferring pragmatic meaning from context, because not only does a distant assessor not share the physical context with the speaker, but the assessor and speaker do not share a social context nor do they share a similar historical trajectory that has lead to the moment of speaking.

Nonetheless an understanding of entextualization is key to understanding the processes by which speaking is assessed. The insights into entextualization derive from studies of performance but they are nonetheless systematic for that. The cultural value of texts for linguists is very great, so great in fact that Chomsky (1965) in an oft-quoted remark rejected the value of the talk on which texts are based.

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (p. 3)

But for assessors, the cultural value of the original talk is greater. In defining and assessing speaking it is important that the colors of the conversational moment are seen and valued for what they are, not devalued as irrelevant when assessors respond to text.

References

American Council on the Teaching of Foreign Languages. (2001). *The ACTFL Oral Proficiency Interview (OPI)*, [World-wide web page]. http://www.actfl.org/public/articles/details.cfm?id=17.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Duranti, A., & Goodwin, C. (1992). *Rethinking context: Language as an interactive phenomenon*. New York: Cambridge University Press.

Educational Testing Service. (2001). *Test of Spoken English (TSE)*, [World-wide web page]. http://www.toefl.org/abttse.html.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Haviland, J. B. (1996). Text from talk in Tzotzil. In M. Silverstein & G. Urban (Eds.), *Natural histories of discourse* (pp. 45-78). Chicago: University of Chicago Press.

Johnson, M., & Tyler, A. (1998). Re-analyzing the context of the OPI: How much does it look like natural conversation? In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 27-51). Amsterdam and Philadelphia: John Benjamins.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

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

Tannen, D. (1994). *Gender and discourse*. New York: Oxford University Press.

Urban, G. (1996). Entextualization, replication, and power. In M. Silverstein & G. Urban

(Eds.), *Natural histories of discourse* (pp. 21-44). Chicago: University of Chicago

Press.

Young, L. W. L. (1994). *Crosstalk and culture in Sino-American communication*. New

York: Cambridge University Press.