Annual Review of Applied Linguistics (2002) **22**, 243–262. Printed in the USA.
Copyright © 2002 Cambridge University Press 0267-1905/01 $9.50

13. DISCOURSE APPROACHES TO ORAL LANGUAGE ASSESSMENT

Richard F. Young

This chapter begins with a careful look at a sample conversation and examines the many layers of interpretation that different academic traditions have constructed in order to interpret it. These layers of interpretation include linguistic forms, nonverbal communication, linguistic context, situational context, and the embodied histories that participants bring to interaction. All are incorporated into a rich definition of discourse. The chapter then reviews recent studies that have compared the discourse of oral interaction in assessment with oral discourse in contexts outside assessment to show how different they are. The next section discusses studies that have related ways of speaking to the cultural values of communities of speakers with a view to understanding the cultural miscommunication that occurs in assessment of speaking in a second language community. The review concludes by stressing the wholeness of face-to-face interaction, listing the layers of interpretation of interaction that have not thus far been considered in oral testing, and setting out a potentially fertile area for future research.

Oral language is a complex phenomenon and understanding how well a second language learner uses oral language is one of the most challenging issues in language assessment. For more than fifty years now, the technology of recording and reproducing speech has allowed researchers to understand it in greater detail; that technology together with theoretical advances in discourse analysis have created a detailed and nuanced picture of speech. In recent years, language assessment researchers and test developers have begun to utilize the new understanding of oral discourse in order to examine existing means of assessing oral language and to put forward new methods of testing. In this review, I will take as a point of departure a conversation that I will then analyze in order to illustrate some of the properties of oral language that discourse analysts have identified. I will then turn to the concerns of language testers to produce a valid assessment of an individual's oral language ability and review comparisons between the discourse of the assessment task and the discourse of conversation. Finally, I will review the manner in which ways of speaking form part of a culture

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

and are valued accordingly and, consequently, how cross-cultural variation on speaking influences its assessment.

The Discourse of Oral Interaction

In order to understand the phenomenon of oral language, I reproduce here a conversation between a child (C) and his mother (M) published some years ago by Ray Birdwhistell (1960, 1970). In order to demonstrate the complexity of oral interaction, I will provide a layered description of the conversation in which each partial layer of interpretation builds on the layers that precede it. Let us first look at the words of the conversation, (Birdwhistell, 1970, pp. 283–285):

C: Mama. I gotta go to the bathroom.
M: [no response]
C: Mama. Donnie's gotta go.
M: Sh–sh.
C: But mama.
M: Later.
C: Ma ma.
M: Wait.
C: Oh mama, mama, mama.
M: Shut up. Will yuh.

By studying only the words of this conversation, there is already much that can be said about the structure of the discourse. The first obvious observation is that there are two speakers and the conversation has a beginning and an end. In this case, oral language is clearly created by more than one person. And although there are words that we attribute to one speaker and not to another, the two speakers are part of one conversation, which we infer from the observation that they take turns and that adjacent turns by C and M appear to be related to one another. The conversation begins with a call from C to which M's response is noticeably absent.

Each of the other adjacent lines forms a pair—a call by C and a response by M. Not only do these adjacent lines form a pair that is topically and functionally related, but the whole conversation from beginning to end is about a single topic: C wants to go to the bathroom. Already, the discourse of this conversation is showing its dialogic nature through orderly turn taking, adjacency pairs, and topical coherence. We can infer from this sample of oral language that doing oral language is about far more than just one person speaking.

Clearly though, the speech patterns of each person in the conversation are important in expressing that individual's meaning; at the same time, interpreting those patterns is important for the other individual in order for both to maintain the

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

orderly dialogic structure that I have shown here. The words that the participants speak are part of what helps them to construct the conversation, but the way in which they say them is also an important ingredient of their meaning. I now reproduce the conversation again, this time indicating the sentence stress, intonation, pausing, and the voice quality of each utterance that Birdwhistell (1970, pp. 293–285) assigns.[1]

1. C: ^{3}Ma^{2}ma ((pause)) ^{3}I ^{2}gotta go to the ^{3}bath^{2}room

2. M: ((pause))

3. C: ^{2}<u>Ma</u>^{3}ma ((pause)) ^{2}<u>Don</u>nie's <u>gotta</u> ^{3}<u>go</u>^{1}

4. M: ^{2}Sh–^{1}sh

5. C: ^{1}But ((pause)) ^{4}ma^{3}ma

6. M: ((softly)) ^{3}La^{1}ter

7. C: ((whining)) ^{3}Ma: ^{1}ma:

8. M: ((rasping voice)) ^{3}Wait^{1}

9. C: ^{1}Oh ^{3}ma^{1}ma ^{4}ma^{2}ma ^{3}<u>ma</u>^{3}<u>ma</u>:

10. M: ((loudly)) ^{3}Shud^{1}dap ((softly)) ^{2}will ^{3}yuh

This second layer of description shows that intonation and voice quality add a further dimension to our understanding of this conversation. Consider how C's intonation on the word "mama" changes as the conversation progresses. In line 1, the tones on the two syllables are "^{3}ma^{2}ma," a slight fall, but when he receives no response, this tone changes to marked stress and a rise in line 3 "^{2}<u>ma</u>^{3}ma." Then, when M responds with a dispreferred second pair part to C's request, the tone changes to a fall in line 5: "^{4}ma^{3}ma." C persists in this request by increasing his pitch range to a high fall in line 7 accompanied by a change in voice quality to a whine, to which M responds with a change of voice quality to a rasp in line 8, but M's response is still not the one C wishes to hear and he repeats his call with increasing volume in line 9. M's closure of the interaction is done with loud volume on the words "shuddap." Such a close analysis of the way the speech patterns of each participant in the conversation change as the conversation progresses illuminates another aspect of oral language: Language provides a context for itself. In other words, the way that a participant pronounces a word is not fixed but varies according to where that word is used in the sequence of

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

interaction and according to the attitudinal meaning that the participant wishes to convey.

Thus far, the conversation has been abstracted from the situational context in which it occurred, but of course all conversation occurs in a context that is both linguistic and situational, and a discourse approach to the study of language must take situational or extralinguistic context into consideration. Birdwhistell observed the conversation at about 2:30 p.m., April 14, 1952 on a bus in Arlington, Virginia. The participants were a mother and her son; and Birdwhistell provides the following description of the context in which the conversation occurred:

> Mother and child spoke with a tidewater Virginia accent. The bus route on which the event was recorded leads to a middle-class neighborhood. The way in which the mother and child were dressed was not consistent with the dress of other riders . . . The little boy was seated next to the window . . . The child was about four, and his mother seemed to be about twenty-seven to thirty. (1970, p. 283)

Birdwhistell's description gives contextual information about what Bourdieu (Bourdieu & Thompson, 1991) has called the *habitus* of the participants: their accents and dress, which distinguish them from the middle-class neighborhood through which they are passing and imply that mother and child are from a lower socioeconomic class than the other riders on the bus. The description of the physical and social context tells us that this conversation between mother and child did not take place in private and the context—on a bus in front of other participants from a higher social class—also allows us to interpret the mother's refusal of her son's request. It is not possible to go to the bathroom on a bus and talking loudly about such things can cause embarrassment if the talk is overheard by people from a higher social class. The mother's nonresponse in line 2 to the child's request is an attempt to avoid the conversation, she then attempts to terminate the conversation in line 4, and then to make the conversation inaudible to other passengers on the bus by her low volume in line 6. None of these attempts succeed until she finally manages with loud volume and informal command in line 10 to terminate what seems for her to be an embarrassing conversation.

The final layer of description that Birdwhistell (1970, pp. 283–285) provides is a line-by-line description of the nonverbal aspects of interaction and an interpretation of the participants' intentions. (The descriptions and interpretations precede the speaker's verbal turns, shown in boldface):

1. The little boy . . . seemed tired of looking out of the window, and, after surveying all of the car ads and the passengers, he

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

leaned toward his mother and pulled at her sleeve, pouted and vigorously kicked his legs.

1. Child: ^{3}Ma^{2}ma ((pause)) ^{3}I ^{2}gotta go to the ^{3}bath^{2}room

2. His mother had been sitting erectly in her seat, her packages on her lap, and her hands lightly clasped around the packages. She was apparently "lost in thought."

2. Mother: ((no verbal reply))

3. When the boy's initial appeal failed to gain the mother's attention, he began to jerk at her sleeve again, each jerk apparently stressing his vocalization.

3. Child: ^{2}Ma^{3}ma ((pause)) ^{2}Donnie's gotta ^{3}go^{1}

4. The mother turned and looked at him, "shushed" him, and placed her right hand firmly across his thighs.

4. Mother: ^{2}Sh–^{1}sh

5. The boy protested audibly, clenched both fists, and pulled them with stress against his chest. At the same time he drew his legs up against the restraint of his mother's hand. His mouth was drawn down and his upper face was pulled into a tight frown.

5. Child: ^{1}But ((pause)) ^{4}ma^{3}ma

6. The mother withdrew her hand from his lap and resettled in her former position with her hands clasped around the packages.

6. Mother: ((softly)) ^{3}La^{1}ter

7. The boy grasped her upper right arm tightly, continued to frown. When no immediate response was forthcoming, he turned and thrust both knees into the lateral aspect of her left thigh.

7. Child: ((whining)) ^{3}Ma: ^{1}ma:

8. She looked at him, leaned toward him, and slapped him across the anterior portion of his upper legs.

8. Mother: ((rasping voice)) ^{3}Wait^{1}

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

9. He began to jerk his clenched fists up and down, vigorously nodding between each inferior-superior movement of his fists.

9. Child: ^{1}Oh ^{3}ma^{1}ma ^{4}ma^{2}ma ^{3}ma^{3}ma:

10. She turned round, frowning, and with her mouth pursed, she spoke to him through her teeth. Suddenly she looked around, noted that the other passengers were watching, and forced a square smile. At the same time that she finished speaking, she reached her right hand in under her left arm and squeezed the boy's arm. He sat quietly.

10. Mother: ((loudly)) ^{3}Shud^{1}dap ((softly)) ^{2}will ^{3}yuh

Birdwhistell's line-by-line description shows the coordination of speech and nonverbal communication. In line 3, the child's jerking of his mother's sleeve is in rhythm with his stressed syllables, and in line 7 he digs his knees into his mother's thigh to accentuate his whining call. His mother's verbal response in line 4 is accompanied by a movement to constrain him, and her "wait" in line 8 is accompanied by a slap on his legs, a nonverbal response that follows the child's digging of his knees into her thigh. The participants' facial expressions are also coordinated with their talk; for example, in line 5 the child frowns to complement his protest, and in line 10 his mother frowns and purses her mouth to accompany her "shuddap will yuh." Birdwhistell also confirms the interpretation that there are more than two participants in this interaction by noting that the mother "looked around, noted the other passengers were watching, and forced a square smile." As Bell (1984) has shown, the co-participation of the other participants may result in the mother's designing this interaction (or at least its termination) with this audience in mind.

I have described this conversation in layers in order to give the reader a sense of the procedure that Ryle (1971) and Geertz (1973) have suggested that we use to approach the problem of interpreting interaction in context—what Geertz called "thick description." All these layers contribute to an understanding of the discourse of interaction, which Celce-Murcia and Olshtain define as follows:

> A piece of discourse is as an instance of spoken or written language that has describable internal relationships of form and meaning (e.g., words, structures, cohesion) that relate coherently to an external communicative function or purpose and a given audience/interlocutor. Furthermore, the external function or purpose can only be properly determined if one takes into account the context and participants (i.e., all the relevant situational,

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

> social, and cultural factors) in which the piece of discourse occurs. (2000, p. 4)

A valid and accurate assessment of oral language must somehow index each of the layers of description that we have seen are part of the complex process of interpreting spoken interaction. Such an assessment indexes the interactional structure of the conversation that has been described by Young (He & Young, 1998; Young, 1999) as a complex configuration of interactional features including boundaries such as openings and closings, participation frameworks, sequential organizations of turns and topics, and semiotic structure (Young & Nguyen, in press). It also indexes the context in which the conversation takes places on at least two dimensions: linguistic and situational. Other language in the conversation contextualizes and thus influences the choice of language at a specific point in the conversation. And at the same time the situational context (including the personal histories of the participants, where the conversation takes place, and the invoked presence of other non-focal participants) influences the language of the focal conversation. And finally, the conversation is not constructed through the single modality of speech. As Birdwhistell's description of the participants' hand gestures, body movements, and facial expressions shows, communication through speech is tightly coordinated with action in the nonverbal channel.

The 20th century tradition of assessing oral language as described by Spolsky (1990, 2000) has not taken into consideration the layers of interpretation of speech in any systematic way, and indeed some oral assessment procedures specifically exclude any consideration of oral language as interaction (Bernstein, 1999). Instead, two aspects of the discourse of oral language assessment have been studied at some length: their construct validity and cross-cultural variation in speaking. The question of the construct validity of oral tests is the degree to which discourse in oral tests corresponds to a theoretical model of speaking. Variation in ways of speaking across cultures has been studied because oral second language assessment often involves an assessor from the target linguaculture judging the speech of a candidate from a different linguaculture (Agar, 1994). These are the issues that this review will now address. I will first review recent studies of validity, and I will follow that with a review of studies that have addressed questions of cross-cultural variation in oral testing.[2]

Validity in Oral Testing

Cumming (1996) describes a large number of ways in which the concept of validity has been used in language testing and in psychology, citing 16 definitions given by Angoff (1988). Construct validation, according to Cumming, is the type of validity that has been most important in recent years. Several organizations have said that construct validity is "the most important consideration in test evaluation" (American Educational Research Association, American

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Psychological Association, & National Council on Measurements Used in Education, 1985). This has led to considerable work describing the discrepancies between what a Language Proficiency Interview (LPI) is supposed to measure and what it does measure.

The question of whether the discourse of an LPI reproduces the discourse of natural conversation has been asked by several researchers. Van Lier (1989) and Young and Milanovic (1992) questioned the supposedly conversational nature of OPIs and van Lier proposed an alternative modular approach to the existing OPI format so as to transform an OPI into a conversation. Lazaraton (1992, 1996a, 1997) examined the overall structural organization of the LPI as well as examiners' question design and the interactively co-constructed nature of the assessment of the learners' language ability. She showed that, although LPIs import their fundamental structural and interactional features from conversation, they are identifiably instances of interviews for the participants.

One particular LPI, the Oral Proficiency Interview, used extensively by organizations such as the American Council on the Teaching of Foreign Languages (ACTFL) (American Council on the Teaching of Foreign Languages, 2001; Breiner-Sanders, Lowe, Miles, & Swender, 2000), has been critically examined in some detail by Johnson (Johnson, 2000, 2001; Johnson & Tyler, 1998). Johnson (2001) conducted a discourse analysis of 35 LPIs and compared the discursive architecture of this practice with ordinary conversation. One of the major findings that resulted from her comparison was that the distribution and allocation of turns in LPIs differs markedly from the way that turns are distributed and allocated in ordinary conversation. The kind of turn structure that is typical of LPIs is illustrated by the following transcript from an Oral Proficiency Interview conducted by telephone in English as a second language (Johnson, 2001, p. 94).[3]

1.	Inter:	How long does it take you to to get from Salt Lake City to Provo?
2.	Cand:	I took a bus this morning so it took me about an hour and twenty
3.		minutes to get here.
4.	Inter:	Oh you rode the bus?
5.	Cand:	Yeah I did.
6.	Inter:	Did they have a good bus service from between the two cities?
7.	Cand:	Yeah they have UTA Utah Transit Service and it's real good.
8.	Inter:	(clears throat) What kind of buses are they uh do they have? Are
9.		they big ones?
10.	Cand:	It's really big one.
11.	Inter:	Oh I see I see. Interesting! Now, is there any kind of train
12.		connection between the two cities?
13.	Cand:	Uh [c] usually I I think they do but I never take a train. They
14.		have Amtrak from Provo to Salt Lake and: I don't know how
15.		much it costs but they have it a Amtrak [c] from Provo to Salt

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

16. Lake.
17. Inter: Now, (clears throat) you say that you have lived in Provo for four
18. years now?
19. Cand: Yeah.
20. Inter: Is that the only place in Utah that you've lived?
21. Cand: Yeah, I came I came here in nineteen . . . ninety.
22. Inter: Oh nineteen ninety. And from where did you come?

As illustrated in this excerpt, Johnson found that interviewers' turns in LPIs most often consisted of a question, while candidates asked questions relatively infrequently. As Johnson remarked, this one-sided pattern of question and response contradicts the assumption that an LPI represents a conversation because "in conversation, turn unit type (along with turn allocation and turn distribution) is unpredictable" (Johnson, 2001, p. 93). The means by which participants in an LPI allocate the next turn is also illustrated in this excerpt, and in this particular case it is related to the fact that the interviewer's turn consists solely of questions. At transition relevance places (TRPs) in the interviewer's turns, he selects the next speaker (i.e., the candidate). For example, in the interviewer's long turn in line 11 according to Ford, Fox, and Thompson (1996), there are four points at which the syntax indicates a new turn may be taken. A change of speaker is possible after "Oh I see I see," after "Interesting!" after "Now," and at the end of the question about the train. Although after the first three TRPs the candidate may take the floor by selecting herself as the next speaker, she does not in fact take a turn until the interviewer selects her as the next speaker at the end of the question.

Meanwhile at TRPs in the candidate's turns, either the candidate continues her turn or the interviewer self selects. For example, the syntax of the candidate's long turn in lines 13–16 indicates five TRPs: after "uh usually I I think they do," after "but I never take a train," after "they have Amtrak from Provo to Salt Lake," after "and I don't know how much it costs," and after "but they have it a Amtrak from Provo to Salt Lake." At none of these TRPs, however, does the candidate select the interviewer as the next speaker. This happens only when the interviewer selects himself at the end of candidate's turn by introducing a new topic with "Now, you say that you have lived in Provo for four years now?" in lines 17 and 18. There are very few occasions in Johnson's data when the candidate allocates the next turn to the interviewer, and in this way the different and complementary discursive roles of candidate and interviewer contribute to constructing a participation structure of the LPI that differs from ordinary conversation.

Other researchers have criticized the validity of other aspects of the LPI. Using the methods of conversation analysis, Egbert (1998) analyzed 20 LPIs conducted in German as a foreign language with American students and found that the organization of conversational repair was explicitly explained by the interviewer, and that thus the students initiated repair by means of the forms taught

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

to them, which are not found in interaction between native speakers. On the other hand, Moder and Halleck (1998) freely admit that the discourse of the ACTFL oral proficiency interview (OPI) does not measure proficiency in an informal conversation, but that in itself does not establish that it is an inadequate measure of communicative competence. Outside the testing situation, nonnative speakers may engage in other speech events having features that are similar to an interview and thus will have to respond to questions intended to seek information, to check information, and to clarify (Schiffrin, 1994). Moder and Halleck (1998) compared OPI interviews of native and nonnative speakers and concluded that the interview frame is interpreted by both groups in a similar way, suggesting that the examination frame does not override the communicative frame of the event. This suggests that the OPI can be viewed as an authentic instance of talk in interaction. It is not an informal conversation, but it does sample the communicative behavior of interviewees in an authentic speech event.

The discourse of other modes of assessing oral interaction besides the ACTFL OPI has been described, including role play, scripted stimulus-response, picture description, and group discussion with another candidate. A role play is part of the Occupational English Test (OET) described by McNamara (Jacoby & McNamara, 1999; Lynch & McNamara, 1998; McNamara & Lumley, 1997; McNamara, 1996, 1997). In the OET role play, an interlocutor plays the role of a member of the public seeking professional help from the nonnative speaking candidate. Lazaraton (1996b) has described the examiners' questions in the Cambridge Assessment of Spoken English as scripted, allowing none of the spontaneous interaction that is found in conversation and modeling more closely the stimulus-response model of interaction described by Silverman (1976). And Riggenbach (1998) has proposed assessment based on an oral language portfolio that includes a range of speech samples: audio or video recordings of the learner engaged in a variety of orally-communicated exchanges, some of which are monologues, some dialogues, some structured (e.g., read-aloud tasks), some semi-structured (e.g., tasks), some rehearsed (e.g., short lectures), and some spontaneous (e.g., role plays). Riggenbach (1998) believes that "this approach to oral skills testing offers a more holistic and comprehensive assessment that could serve as an alternative to more traditional speaking test formats that rely on a single sample and/or a single genre" (p. 65).

Differences between the discourse of interviews and conversation were also noted by Fulcher (1996). Fulcher conducted a comparative study of students taking three different oral assessments: two interviews—one based on picture description, and one discussion on a text—and one non-interview—a group discussion monitored by an examiner. Fulcher discussed the validity of the three tests as perceived by the candidates. He concluded that, "Engaging in a group discussion with a partner gave the students more confidence to speak and say what *they* wanted, rather than having to respond to an examiner" (p. 33). Perhaps for

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

this reason, the group oral "was seen as an enjoyable experience by well over half the students" (p. 34), and students saw the group oral as the most preferable of the three tests.

Apart from the discourse of the conversation itself, the situational context of assessment has also been investigated. Several studies have made comparisons between rating scales used in assessment and intuitive ratings of conversations by nonraters (Milanovic, Saville, Pollitt, & Cook, 1996; Pollitt & Murray, 1996). Pollitt and Murray (1996) found that raters focused on different qualities of talk depending on the overall proficiency level of the candidate. With higher proficiency candidates, raters distinguished among candidates according to stylistic devices the candidates used and according to candidates' content-focused elaboration, creativity, parenthetical statements, and idiomatic expressions. At the lower proficiency level, raters distinguished among candidates according to candidates' grammatical competence; whether candidates were hesitant or staccato; or used form-focused language, set 'textbook' phrases, and rehearsed or stilted speech. In fact, the lack of an empirical basis for rating scales used in the assessment of oral interaction has been criticized by Young (Young, 1995b), and Kenyon (Kenyon, 1998; Stansfield & Kenyon, 1996) has investigated the correspondence between levels of difficulty established *a priori* by rating scales and perceptions of difficulty by students and teachers.

The most fundamental investigation of the wider context of oral assessment is Jacoby's notion of *indigenous assessment* (Jacoby, 1998; Jacoby & McNamara, 1999). Indigenous assessment, according to Jacoby, differs from the traditional activity of examiners or assessors of oral interaction in a second language, whose aim is to evaluate the performance of candidates according to linguistic criteria. Indigenous assessment involves at least one participant who frames another participant's prior communication as good or bad. The response of the co-present participant and any ensuing discussion is also part of the assessment. Indigenous assessment of communication performance can occur in nontesting situations such as an assessment of a child at the family dinner table in response to something a child just said, and it occurs when scientists critique one another's run-throughs of upcoming conference presentations. An indigenous assessment is accomplished among insider participants in some culturally situated activity for their own local purposes and it is not prompted by or designed to serve the purposes of an outside rater. Indigenous assessment has been applied to the assessment of English for specific purposes by Douglas (Douglas, 2000; Douglas & Myers, 2000).

Cross-Cultural Variation in Oral Testing

Most methods of assessing oral ability in a second language involve an assessor evaluating learners' discourse from the perspective of the cultural norms of oral interaction in the target community. Work on the ethnography of speaking

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

has, however, demonstrated that the ways of speaking of a particular group are in effect reflections of the culture of the group and that these cultural patterns transfer into the speech patterns of a second language even among quite advanced learners. According to Gumperz (1982), in different communities participants in speech activities have specific expectations about thematic progression, turn-taking rules, form, and the outcome of the interaction, as well as constraints on what counts as context. Cross-cultural misunderstanding may result from discourse cues that have a certain meaning in one linguaculture being transferred into the second language and being interpreted differently by an interlocutor from another linguaculture (Agar, 1994; Boxer, this volume; Scollon & Scollon, 2001; Young & Halleck, 1998). The effect of such culture-specific discourse organization on the assessment of speaking has been investigated by a number of researchers including Davies (1998) and Kim and Suh (1998) for Americans speaking Korean; Ross (Berwick & Ross, 1996; Ross, 1998) for Japanese speakers of English; and Young, who compared Mexicans and Japanese speaking English (Young, 1995a; Young & Halleck, 1998).

A clear example of the connection between conversational organization and cultural values is provided by Kim and Suh (1998). In Korean OPIs with American candidates, Kim and Suh observed this recurrent question-answer sequence over five turns.[4]

1st Turn (IR)	:	Question
2nd Turn (NNS)	:	Answer
3rd Turn (IR)	:	Confirmation request
4th Turn (NNS)	:	Confirmation
5th Turn (IR)	:	Follow-up

According to Kim and Suh, the interviewer's confirmation request in the 3rd turn is a place where "IR claims his/her right to evaluate prior talk and initiates subsequent talk" (1998, p. 316). The Korean interviewer expects the learner to respond by confirming the interviewer's right to evaluate and to wait for IR to initiate a new topic. The learner's close orientation toward and confirmation of the interviewer's third turn "constitutes a crucial aspect of sociolinguistic competence by indexing that NNS treats IR as a socially higher status person whose assessment he or she respects and values" (p. 317). Kim and Suh mention this as one way in which Korean interviewers assess the proficiency of a conversational partner. Less proficient students respond to the interviewer's confirmation request either by silence or by continuing the topic they began in the second turn, thus ignoring the interviewer's confirmation request. Both of these responses by the candidate challenge the positive face of the interviewer because they fail to recognize the need to recognize IR's higher status. Although such discourse may not of itself determine an assessment of low oral proficiency, Kim and Suh show that such patterns are nonetheless characteristic of low proficiency candidates.

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Another type of culture-specific oral assessment that is not normally considered in discussions of second language assessment is the oral exams that are often part of university assessments, especially in the humanities. Such assessments are prevalent in Italian universities and have been analyzed by Anderson (1999) and Ciliberti (1999), who show how a student's knowledge of subject matter is co-constructed by the student and by the examining professors.

Conclusions

At the beginning of this review, I showed a conversation between a mother and her child in its entirety, and I argued that the wholeness of the interaction could be best understood by considering the layers of interpretation that different academic traditions had formed around the interaction in order to describe and understand it. In the tradition of language assessment, one layer of this description has been thicker than others: the words spoken, their pronunciation, and prosodic contours. In reviewing more recent work in assessing oral language, I have emphasized the layers of interpretation that have not been previously considered in testing—the linguistic and situational context of the interaction, the outcome that the participants intend, the outcome intended by those who make the assessment, and the close relationship between oral interaction and the culture of the participants.

Finally, one important layer of discourse not yet studied in oral language proficiency tests is the role of nonverbal behavior in those tests and the influence of the nonverbal channel on assessment. As Birdwhistell illustrated in his transcription of the nonverbal channel in the conversation between a mother and her child with which this article began, it is a small step from a linguistic discourse analysis to an analysis of interaction at both the nonverbal and verbal levels. Although this has been done in recent work in interactional sociolinguistics (Egbert, 1996; Erickson, 1992; Goodwin, 1981; Key, 1980a, b; Ochs, Jacoby, & Gonzalez, 1994; Streek, 1995; Wells, 2000), and some attempts have been made to understand the role of nonverbal behavior in cross-cultural communication (Adams, 1998; Ekman, 1973; Houck & Gass, 1997; Kellerman, 1992; Stam, 1999; Young, 1994), no studies have so far been published that analyze nonverbal behavior in oral language assessment. This is a potentially fertile area for future research.

Notes

1. Birdwhistell indicated intonation by using the symbols of Trager and Smith (1957) in which the pitch level of the following syllable is indicated by superscript numbers, with [^{1}] indicating a speaker's lowest pitch and [^{4}] representing the highest pitch in their range. Stresses louder than normal are indicated by

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

underlining the stressed syllable. Pauses and voice quality affecting the following words are indicated between double parentheses.

2. The majority of the research on oral assessment in the language testing community has not in fact addressed the two issues of validity and cross-cultural variation. Most research has instead focused on the systematic ways in which variation in scores can be related to the task, the rater, and the rating scale, in other words, to the method of oral testing. Upshur and Turner (1999) provide a very clear overview of this research, which I mention here only in passing because the focus of the present review is on the discourse of oral assessment.

3. The native speaking interviewer is identified as "Inter" and the nonnative speaking candidate as "Cand."

4. IR is the native speaking Korean interviewer; NNS is the American speaker of Korean as a second language.

ANNOTATED BIBLIOGRAPHY

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes, 18*(3), 213–241.

This article goes into detail about the concept of indigenous assessment. Indigenous criteria for assessment are important in performance tests in which some criterion situation is simulated to a much greater degree than is represented in the usual pencil-and-paper test. The criterion for assessment is usually based on a theoretical (usually psychological) and educationally motivated construct assumed to underlie performance, but ethnographic research can reveal what criteria experts in the given field feel are appropriate for assessment. The article describes the assessment criteria used in the Australian Occupational English Test and compares it with the indigenous assessment criteria used by physicists who evaluate rehearsals of upcoming conference presentations by their colleagues.

Johnson, M. (2001). *The art of non-conversation: A re-examination of the validity of the Oral Proficiency Interview*. New Haven, CT: Yale University Press.

The heart of this book is a detailed discourse analysis of 35 oral proficiency interviews conducted over the phone. Based on a close conversation analysis of the discourse, Johnson challenges the construct validity of the OPI. Johnson rejects the uncontextualized communicative competence model of language ability that underlies the OPI. She

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

concludes by proposing a model of interactional competence in which language ability is considered to reflect the contexts in which it is acquired and used.

Young, R., & He, A. W. (Eds.). (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins.

This book is a collection of studies on the assessment of oral proficiency in a second language that combine language assessment and discourse analysis. It is introduced by a chapter in which the editors lay out their framework of interactional competence. The following 13 chapters report empirical studies of oral assessment and are all based on close analyses of audio- and/or videotaped discourse. Questions addressed include: How do participants construct identity and competence through interaction? How do interviewers form their judgments about the candidates' interactional abilities? And how does the meaning of an interview change from one speech community to another?

OTHER REFERENCES

Adams, T. W. (1998). *Gesture in foreigner talk*. Unpublished Ph.D. dissertation, University of Pennsylvania, Philadelphia.

Agar, M. (1994). *Language shock: Understanding the culture of conversation*. New York: Morrow.

American Council on the Teaching of Foreign Languages. (2001). *The ACTFL Oral Proficiency Interview (OPI)*. Available: http://www.actfl.org/public/articles/details.cfm?id=17. Retrieved October 2, 2001.

American Educational Research Association, American Psychological Association, & National Council on Measurements Used in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Authors.

Anderson, L. (1999). La co-costruzione di competenza negli esami orali e il ruolo della comunicazione metapragmatica [The co-construction of competence in oral exams and the role of metapragmatic communication]. In A. Ciliberti & L. Anderson (Eds.), *Le forme della communicazione accademica: Ricerche linguistiche sulla didattica universitaria in ambito umanistico* (pp. 192–219). Pavia, Italy: FrancoAngeli.

Angoff, W. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 19–32). Hillsdale, NJ: Lawrence Erlbaum.

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Bell, A. (1984). Language style as audience design. *Language in Society, 13*, 145–204.

Bernstein, J. (1999). *PhonePass™ testing: Structure and construct*. Available: http://www.ordinate.com/pdf/StructureAndConstruct990826.pdf. Retrieved October 1, 2001.

Berwick, R., & Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In University of Cambridge Local Examinations Syndicate (Ed.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium (LTRC), Cambridge and Arnhem* (pp. 34–54). Cambridge, England: University of Cambridge Press.

Birdwhistell, R. L. (1960). Kinesics and communication. In E. S. Carpenter & M. McLuhan (Eds.), *Explorations in communication: An anthology* (pp. 54–64). Beacon Hill, NC: Beacon Press.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia, PA: University of Pennsylvania Press.

Bourdieu, P., & Thompson, J. B. (1991). *Language and symbolic power* (G. Raymondson & M. Adamson, Trans.). Cambridge, MA: Harvard University Press.

Boxer, D. (this volume). Discourse issues in cross-cultural pragmatics.

Breiner-Sanders, K. E., Lowe, P., Jr., Miles, J., & Swender, E. (2000). ACTFL Proficiency guidelines—speaking: Revised 1999. *Foreign Language Annals, 33*(1), 13–18.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching: A guide for language teachers*. Cambridge, England: Cambridge University Press.

Ciliberti, A. (1999). Gli esami orali: Tra agentività e dipendenza, tra auto-referenzialità ed etero-referenzialità [Oral exams: Between agency and dependency, between self-referentiality and other-referentiality]. In A. Ciliberti & L. Anderson (Eds.), *Le forme della communicazione accademica: Ricerche linguistiche sulla didattica universitaria in ambito umanistico* (pp. 166–191). Pavia, Italy: FrancoAngeli.

Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. H. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 1–14). Clevedon and Philadelphia: Multilingual Matters.

Davies, C. E. (1998). Maintaining American face in the Korean oral exam: Reflections of the power of cross-cultural context. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 271–296). Amsterdam and Philadelphia: John Benjamins.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, England: Cambridge University Press.

Douglas, D., & Myers, R. (2000). Assessing the communication skills of veterinary students: Whose criteria? In A. J. Kunnan (Ed.), *Fairness and*

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida (pp. 60–81). Cambridge, England: Cambridge University Press.

Egbert, M. M. (1996). Context-sensitivity in conversation: Eye gaze and the German repair initiator *bitte? Language in Society, 25*(4), 587–612.

Egbert, M. M. (1998). Miscommunication in language proficiency interviews of first-year German students: A comparison with natural conversation. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 147–169). Amsterdam: John Benjamins.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 169–222). New York: Academic Press.

Erickson, F. (1992). They know all the lines: Rhythmic organization and contextualization in a conversational listing routine. In P. Auer & A. Di Luzio (Eds.), *The contextualization of language* (pp. 365–397). Amsterdam and Philadelphia: John Benjamins.

Ford, C. E., Fox, B. A., & Thompson, S. (1996). Practices in the construction of turns: The "TCU" revisited. *Pragmatics, 6*, 427–454.

Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing, 13*(1), 23–51.

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York: Basic Books.

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

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, England: Cambridge University Press.

He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1–24). Amsterdam and Philadelphia: John Benjamins.

Houck, N., & Gass, S. M. (1997). Cross-cultural back channels in English refusals: A source of trouble. In A. Jaworski (Ed.), *Silence: Interdisciplinary perspectives* (pp. 285–308). New York: Mouton de Gruyter.

Jacoby, S. W. (1998). *Science as performance: Socializing scientific discourse through the conference talk rehearsal.* Unpublished Ph.D. dissertation, University of California, Los Angeles.

Johnson, M. (2000). Interaction in the oral proficiency interview: Problems of validity. *Pragmatics, 10*(2), 215–231.

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.

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Kellerman, S. (1992). 'I see what you mean': The role of kinesic behaviour in listening and implications for foreign and second language learning. *Applied Linguistics, 13*, 239–258.

Kenyon, D. M. (1998). An investigation of the validity of task demands on performance-based tests of oral proficiency. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium, Long Beach* (pp. 19–40). Mahwah, NJ: Lawrence Erlbaum.

Key, M. R. (1980a). Language and nonverbal behavior as organizers of social systems. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 3–33). The Hague: Mouton.

Key, M. R. (Ed.). (1980b). *The relationship of verbal and nonverbal communication*. The Hague: Mouton.

Kim, K., & Suh, K. (1998). Confirmation sequences as interactional resources in Korean language proficiency interviews. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 297–332). Amsterdam: John Benjamins.

Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic perspective. *System, 20*, 373–386.

Lazaraton, A. (1996a). Interlocutor support in oral proficiency interviews. *Language Testing, 13*(2), 173–190.

Lazaraton, A. (1996b). A qualitative approach to monitoring examiner conduct in the Cambridge assessment of spoken English (CASE). In University of Cambridge Local Examinations Syndicate (Ed.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium (LTRC), Cambridge and Arnhem* (pp. 18–33). Cambridge, England: University of Cambridge Press.

Lazaraton, A. (1997). Preference organization in oral proficiency interviews: The case of language ability assessments. *Research on Language and Social Interaction, 30*(1), 53–72.

Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing, 15*(2), 158–180.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18*(4), 446–466.

McNamara, T. F., & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing, 14*(2), 140–156.

Milanovic, M., Saville, N., Pollitt, A., & Cook, A. (1996). Developing rating scales for CASE: Theoretical concerns and analyses. In A. H. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 15–38). Clevedon: Multilingual Matters.

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Moder, C. L., & Halleck, G. B. (1998). Framing the language proficiency interview as speech event: Native and nonnative speakers' questions. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 117–146). Amsterdam: John Benjamins.

Ochs, E., Jacoby, S., & Gonzalez, P. (1994). Interpretive journeys: How physicists talk and travel through graphic space. *Configurations, 1*, 151–171.

Pollitt, A., & Murray, N. L. (1996). What raters *really* pay attention to. In University of Cambridge Local Examinations Syndicate (Ed.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium (LTRC), Cambridge and Arnhem* (pp. 74–91). Cambridge, England: University of Cambridge Press.

Riggenbach, H. (1998). Evaluating learner interactional skills: Conversation at the micro level. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 53–67). Amsterdam: John Benjamins.

Ross, S. (1998). Divergent frame interpretations in language proficiency interview interaction. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 333–353). Amsterdam: John Benjamins.

Ryle, G. (1971). The thinking of thoughts: What is 'Le Penseur' doing? In G. Ryle (Ed.), *Collected papers*, Vol. 2, Collected essays 1929–1968 (pp. 480–496). London: Hutchinson.

Schiffrin, D. (1994). *Approaches to discourse*. Cambridge, MA: Blackwell.

Scollon, R., & Scollon, S. W. (2001). *Intercultural communication: A discourse approach* (2nd ed.). Malden, MA: Blackwell.

Silverman, D. (1976). Interview talk: Bringing off a research instrument. In D. Silverman & J. Jones (Eds.), *Organizational work: The language of grading, the grading of language* (pp. 133–150). London: Collier Macmillan.

Spolsky, B. (1990). Oral examinations: An historical note. *Language Testing, 7*(2), 158–173.

Spolsky, B. (2000). Language testing in *The Modern Language Journal*. *The Modern Language Journal, 84*(4), 536–552.

Stam, G. (1999, September). *Speech and gesture: What changes first in L2 acquisition?* Paper presented at the 19th annual Second Language Research Forum, University of Minnesota, Twin Cities.

Stansfield, C. W., & Kenyon, D. M. (1996). Comparing the scaling of speaking tasks by language teachers and by the ACTFL Guidelines. In A. H. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 124–153). Clevedon: Multilingual Matters.

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press

Streek, J. (1995). On projection. In Wissenschaftskolleg zu Berlin & E. N. Goody (Eds.), *Social intelligence and interaction: Expressions and implications of the social bias in human intelligence* (pp. 87–110). Cambridge, England: Cambridge University Press.

Trager, G. L., & Smith, H. L. (1957). *An outline of English structure*. Washington, DC: American Council of Learned Societies.

Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing, 16*(1), 82–111.

van Lier, L. (1989). Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly, 23*(3), 489–508.

Wells, G. (2000). Modes of meaning in a science activity. *Linguistics and Education, 10*(3), 307–334.

Young, L. W. L. (1994). *Crosstalk and culture in Sino-American communication*. Cambridge, England: Cambridge University Press.

Young, R. (1995a). Conversational styles in language proficiency interviews. *Language Learning, 45*, 3–42.

Young, R. (1995b). Discontinuous interlanguage development and its implications for oral proficiency rating scales. *Applied Language Learning, 6*, 13–26.

Young, R. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics, 19*, 105–132.

Young, R., & Halleck, G. B. (1998). "Let them eat cake!" or how to avoid losing your head in cross-cultural conversations. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 355–382). Amsterdam: John Benjamins.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*(4), 403–424.

Young, R. F., & Nguyen, H. T. (in press). Modes of meaning in high school science. *Applied Linguistics, 23*(3).

https://doi.org/10.1017/S0267190502000132 Published online by Cambridge University Press