Multiple perspectives

ACCOMMODATION IN ORAL INTERVIEWS BETWEEN FOREIGN LANGUAGE LEARNERS AND TEACHERS WHO ARE *NOT* NATIVE SPEAKERS*

Brian J. Richards & David D. Malvern

Abstract. Previous research shows evidence of discourse accommodation by native speakers in oral interviews. This is seen as essential to their validity. The research described here examines whether *non-native* speakers also adjust their language to the proficiency of interviewees.

Interviews of 34 sixteen-year-old learners of French with two non-native teachers were analysed. Measures of the students' French were related to indices of the teachers' language. Multiple regressions suggest that accommodation to the proficiency of individual students does take place, but that some aspects of the teachers' language are grossly tuned to the general proficiency of students in their class rather than finely tuned to individuals.

1. Introduction

It is a natural feature of authentic conversation that as part of the negotiation of meaning speakers adjust their language to features of their interlocutor. In addition to status, age, social background and dialect, such features include the interlocutor's perceived language proficiency, including comprehension. Fine tuning of input to children learning their L1 (see, for example, Pine 1994) has been extensively investigated leading to parallel research into 'foreigner talk' and 'language teacher talk' (see Wesche 1994 for a review). Here too, speech adjustments to the L2 learner's language level are well documented (e.g. Gass & Varonis 1985).

During the last decade much attention has been paid to the conversational authenticity of the oral proficiency interview. In a much-quoted article, Van Lier (1989) argues that pressure to elicit language and the power differential between interviewer and interviewee afford little opportunity for the negotiation of meaning and genuine conversation. Similarly, Young & Milanovic (1992) show that while interviewers show much greater goal orientation than students, it is the students who show greater 'reactiveness' or conversational contingency to the language of their interlocutor. On the other hand, Ross (1992) and Ross & Berwick

* The project 'Oral Assessment in Modern Languages' was funded by The University of Reading Research Endowment Trust Fund. We should like to thank Li Daguo and Mair Richards for their comments on previous drafts of this paper.

Studia Linguistica 54(2) 2000, pp. 260–271. © The Editorial Board of Studia Linguistica 2000.
Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK, and
350 Main Street, Malden, MA 02148, USA

(1992) demonstrate that not only are the exponents of accommodation prevalent in their interviews, but that the frequency of these is related to the proficiency of students. They conclude that the oral interview exhibits characteristics of both conversations and interviews and that the major threat to validity is either the absence of appropriate accommodation or the use of inappropriate accommodation to candidates at high proficiency levels.

One feature of the research referred to above, and indeed of much of the research into oral assessment, is that the examiners were native speakers. However, in many educational systems foreign language teaching and assessment is conducted by teachers who have themselves learnt the target language as a foreign language, and whose own proficiency in the L2 may vary considerably. Little is known about the relative behaviour of native and non-native speakers in oral interviews. It is not clear, for example, whether non-native speakers are sufficiently proficient in the target language, and sufficiently sensitive to the interaction between the proficiency of the learner and the demands of the test, to fine tune their discourse appropriately. If, as claimed by Ross & Berwick (1992), the validity of the test depends on appropriate levels of accommodation, then interviews conducted by non-native speakers may be particularly problematic.

In the study described below, the extent to which interviewers who are not native speakers adjust their language to the proficiency of candidates was investigated in a study of 34 sixteen-year-old learners of French in an oral examination conducted by two non-native speakers.

2. Method

We obtained the tape recordings of 34 state secondary school students taking their oral examination in French for the General Certificate of Secondary Education, a national examination taken by students in the United Kingdom at the age of 16. Oral interviews (officially described as 'free conversation'), which averaged just over five minutes in length, had been conducted by two non-native, but experienced, teachers of French, each of whom examined the students in his or her own French class. The interviews were transcribed in CHAT format (MacWhinney 1995) by a native French speaker (Francine Chambers) who was an experienced teacher of French to English-speaking children of this age, and were re-transcribed by the first author of this article (Brian Richards). Discrepancies were resolved with the assistance of a near-native speaker of French who was also an experienced teacher of French as a foreign language in the UK (Mair Richards). Where discrepancies could not be resolved, the utterance or word was coded as 'unintelligible' and omitted from the analysis. CHAT format enabled computer-assisted analyses to

be carried out by the CLAN programs of the CHILDES project (MacWhinney & Snow 1990).[1]

2.1. *The students*

The students had been learning French for five years, receiving four 35 minute lessons a week. In theory, their ability should have been above average because they had all been entered for the 'higher level' examination, designed to extend more able pupils. In practice, it became clear that some students had been inappropriately entered for the higher examination, and proficiency varies in the sample from those who made very little contribution to the conversation to one student who performed at a level comparable to native-speakers. In fact this candidate, whose examination results consisted of the maximum possible scores in listening, speaking, reading *and* writing, had to be omitted from all the analyses reported below because he produced such extreme scores on the student measures that they distorted the regressions. As a result the final sample of students consisted of 33, of whom 12 were interviewed by Teacher A and 21 by Teacher B.

2.2. *Student variables*

Student variables are listed in Table 1. These are of three types. First, there are objective measures taken from the transcripts with the assistance of the CLAN software (MacWhinney 1995). Of these, the vocabulary diversity measure, MSTTR-30, needs some explanation. The mean segmental type-token ratio (MSTTR) has a long history as a measure of range of vocabulary (see Johnson 1944) and, unlike many other measures of vocabulary diversity (Malvern & Richards 1997), is not affected by variations in sample size (i.e. the number of words produced by each speaker). In this case we divided each transcript into 30-word segments of student speech and calculated the average number of *different* words contained in each segment.

The second type of student measure comes from the final results of the GCSE examination. Here, the official examining groups convert the scores on each of the four skills to a mark out of seven, and it is these that are reported here.

Thirdly, six further measures were obtained from the mean ratings of the tape recordings by 24 experienced teachers of French. Range of vocabulary, fluency and complexity of structure were rated on eight-point scales (0–7) and content, accuracy and pronunciation on four-point scales

[1] The 34 transcripts are available from the CHILDES data base on the World-Wide Web. This can be found at http://childes.psy.cmu.edu/database.html where 'bilingual copora', then 'Reading' should be selected.

Table 1: Student (independent) variables (N = 33)

Student variable	Mean	SD
Measures Taken from Transcripts		
Number of Words	164.6	126.6
Number of Utterances	33.2	16.2
Number of Different Words	78.7	42.4
Vocabulary Diversity (MSTTR-30)	22.7	1.7
Mean Length of Utterance (MLU words)	4.4	1.8
Utterances per Turn (MLT)	1.1	0.1
Words per Minute	27.7	11.8
GCSE Examination Results		
Score for Oral Examination (out of 7)	4.0	1.3
Total Score for Listening, Reading & Writing (out of 21)	13.8	3.2
Mean Ratings from 24 Teachers of French		
Range of Vocabulary (0–7 scale)	2.6	1.5
Fluency (0–7 scale)	2.6	1.4
Complexity of Structure (0–7 scale)	2.0	1.3
Content (0–3 scale)	1.5	0.8
Accuracy (0–3 scale)	1.1	0.6
Pronunciation (0–3 scale)	1.2	0.5

(0–3). Full details of the scales and the procedures used can be found in Richards & Chambers (1996).

2.3. *Teacher variables*

Table 2 lists the teacher variables, all of which were obtained directly from the transcripts. These include measures of conversational dominance and control (length of conversation, utterances per turn and mean length of utterance), conversational encouragement (back channels) and conversational contingency (exact, reduced and expanded imitations of the previous student utterance). For the teachers, vocabulary diversity was averaged for segments of 100 words rather than 30. The reason for the lower number for the students was that a small number of the weaker students produced little more than 30 words during their whole conversation.

3. Results and discussion

3.1. *Descriptive statistics*

Tables 1 and 2 give the means and standard deviations for the student and teacher variables. The standard deviations for the students show that there is considerable variation in their proficiency levels, particularly for the measures obtained from the transcripts and for the teacher ratings of

Table 2: Teacher (dependent) variables (means calculated across 33 students)

Teacher Variable	Mean	SD
Vocabulary Diversity (MSTTR-100)	51.9	4.4
Mean Length of Utterance (MLU words)	5.7	0.5
Utterances per Turn (MLT)	1.9	0.3
Percentage of Utterances which Overlap with the Student	28.0	8.0
Percentage of Utterances which are Imitations of Student	9.7	4.2
Percentage of Utterances which are Exact Imitations of Student	3.3	2.6
Percentage of Utterances which are Expanded Imitations of Student	4.0	3.2
Percentage of Utterances which are Reduced Imitations of Student	2.5	2.4
Percentage of Utterances which are Back Channels	11.4	8.4
Length of Conversation (Seconds)	319.4	113.1

the tapes. Teacher behaviour, on the other hand shows variation in particular on measures of contingency (repetition of the student) and conversational support in the form of back channels. That the most frequent form of repetition is expanded imitation suggests that teachers are, consciously or unconsciously, upgrading the utterances produced by students. When comparing student and teacher measures there is a suggestion of dominance of the conversation by teachers as indicated by their higher MLUs and MLTs.

3.2. *Simple correlational analysis*

Initially, to examine relationships between student proficiency and teacher linguistic behaviour, Pearson product-moment correlations between the student and teacher variables were computed for the pooled data and then for each teacher separately. This preliminary analysis revealed that the pattern of correlations could be very different for Teacher A and Teacher B. Space does not allow us to present the full matrix of correlations of each teacher variable with all the student variables. However, as an example of a case where the teachers differ we take the correlations between teacher vocabulary diversity and the 15 student variables. These are shown in Table 3.

At first sight this set of correlations appears promising. Although no single correlation would explain more than 26% of the variance, all except one are statistically significant, and most reach at least the 0.01 level on a

Table 3: Correlations between teacher vocabulary diversity and student measures (N = 33)

Student measure	*r*
Measures Taken from Transcripts	
Number of Words	.38*
Number of Utterances	.33*
Number of Different Words	.46**
Vocabulary Diversity (MSTTR-30)	.06
Mean Length of Utterance (MLU words)	.44**
Utterances per Turn (MLT)	.34*
Words per Minute	.51***
GCSE Examination Results	
Score for Oral Examination (out of 7)	.49**
Total for Listening, Reading & Writing	.48**
Mean Ratings from 24 Teachers of French	
Range of Vocabulary	.44**
Fluency	.46**
Complexity of Structure	.39*
Content	.48**
Accuracy	.46**
Pronunciation	.37*

* $p < .05$, ** $p < .01$, *** $p < .001$

one-tailed test. As expected with this teacher variable, all are positive. This would be consistent with the teachers using more complex language with more proficient students. However, a separate analysis for each teacher shows a very different pattern. As can be seen from Tables 4 and 5, only eight correlations are now statistically significant, and seven of these are negative, suggesting more complex language with weaker students. The only significant positive correlation is with student lexical diversity for Teacher A.

One explanation for the discrepancy between the results in Table 3 and Tables 4 and 5 would be that overall proficiency levels differed significantly between the two groups of students and that this covaried with differences between the teachers in their use of French. A comparison of the mean GCSE oral scores for each teacher's students supports this interpretation. Teacher A's students scored 2.8 points out of 7 on average, while Teacher B's students obtained a mean score of 4.7. This difference is statistically significant on a *t*-test ($t = -5.72$; $d.f. = 31$; $p < .001$). Figure 1 shows how, for the strongest correlation for the pooled data, that between Teacher MSTTR and Student Words per Minute (see Table 3), a significantly lower vocabulary diversity for Teacher A corresponds with a significantly lower fluency rate for her students. In this case we see

Table 4: Correlations between vocabulary diversity for Teacher A and student measures (N = 12)

Student measure	*r*
Measures Taken from Transcripts	
Number of Words	.24
Number of Utterances	.25
Number of Different Words	.36
Vocabulary Diversity (MSTTR-30)	.62*
Mean Length of Utterance (MLU words)	.14
Utterances per Turn (MLT)	.39
Words per Minute	.24
GCSE Examination Results	
Score for Oral Examination (out of 7)	.09
Total for Listening, Reading & Writing	−.52*
Mean Ratings from 24 Teachers of French	
Range of Vocabulary	.25
Fluency	.26
Complexity of Structure	.21
Content	.29
Accuracy	.23
Pronunciation	.14

* $p < .05$

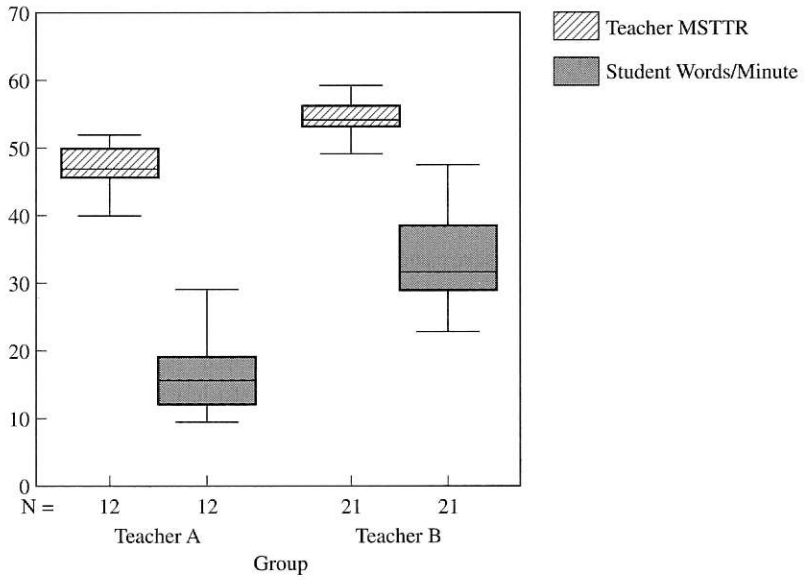

Figure 1: Boxplots of teacher vocabulary diversity (MSTTR-100) and student words per minute for Teacher A and Teacher B

Table 5: Correlations between vocabulary diversity for Teacher B and student measures (N = 21)

Student measure	*r*
Measures Taken from Transcripts	
Number of Words	−.30
Number of Utterances	−.32
Number of Different Words	−.32
Vocabulary Diversity (MSTTR-30)	−.18
Mean Length of Utterance (MLU words)	−.26
Utterances per Turn (MLT)	−.13
Words per Minute	−.28
GCSE Examination Results	
Score for Oral Examination (out of 7)	−.31
Total for Listening, Reading & Writing	−.19
Mean Ratings from 24 Teachers of French	
Range of Vocabulary	−.43*
Fluency	−.46*
Complexity of Structure	−.45*
Content	−.46*
Accuracy	−.47*
Pronunciation	−.44*

* $p < .05$

gross levels of accommodation between the groups rather than appropriate finely-tuned accommodation within each group. Because differences in proficiency between groups could account for significant relationships between teacher and student variables, the group the students belonged to was entered as a student (independent) variable in the multiple regression analyses reported below.

3.3. *Multiple regression analyses*

In order to obtain a fuller picture of possible influences on teacher language, we treated each teacher variable as the dependent variable in a series of stepwise multiple regression analyses in which the 15 student variables, plus the variable Group, were treated as independent variables. The only exception to this was for the dependent variable Length of Conversation where those student variables which were a *function* of length (i.e. number of utterances, number of words, and number of different words) were excluded from the analysis. A total of 10 multiple regressions were carried out, one for each of the teacher measures.

Table 6 shows the model obtained for each dependent variable, that is to say the combination of student variables which best predicted each aspect of teacher behaviour. The cumulative proportion of variance

Table 6: Stepwise multiple regression models obtained for each dependent variable

Dependent variable	Entry Order	Independent variables retained in the model	Total *r*-square	Adjusted *r*-square	Direction of influence
Teacher MSTTR	1	Group	.601	.588	positive
Teacher MLU		*None*			
Teacher MLT		*None*			
Teacher overlaps	1	Student MLU	.189	.163	positive
Total teacher imitation	1	Student MLT	.201	.175	negative
Teacher exact imitation	1	Student MLU	.200	.174	negative
	2	Listening reading writing	.310	.264	positive
Teacher expanded imitation	1	Student MLU	.126	.098	negative
	2	Group	.262	.213	positive
	3	Listening reading writing	.356	.290	negative
Reduced imitation	1	Student MSTTR	.243	.218	negative
Teacher back channels	1	Content	.520	.504	positive
Length of conversation	1	Range of vocabulary	.672	.662	positive
	2	Fluency	.795	.781	negative
	3	Content	.850	.834	positive

explained after entering each successive variable is shown by the total *r*-square. Since the latter tends to overestimate the population value of *r*-square, the more conservative adjusted *r*-square is given in the following column. It can be seen from these results that some degree of tuning to the students' language was taking place for all the teacher variables except Teacher MLU and Teacher MLT, which showed no evidence of either within- or between-group accommodation.

The result for teachers' vocabulary diversity (MSTTR-100) confirms the exploratory analyses described above. The group the students belonged to (Teacher A or Teacher B) is the only variable retained in the model, and this explains a substantial proportion of the variance (59%). One other dependent variable, teacher expanded imitation retains Group in the model. This will be discussed below together with the other imitation measures.

As might be expected, teacher utterances which overlap with those of the student are predicted by the length of the students' utterances (MLU). The longer the student utterances the more likely the teacher is to speak over them. This is the only variable retained in the model and explains 16% of the variance.

The teachers' imitation of their students' language presents a more complex picture. The combined category of all kinds of imitation is predicted negatively by students' length of turn (MLT), explaining 17% of the variance. That is to say, when students' speaking turns consist of a smaller number of utterances, some part of the language contained within them is more likely to be repeated by the interviewer. This would be consistent with repetition being used as a means of scaffolding the contribution of weaker candidates or, as suggested above, upgrading their contributions. Disaggregating the imitation variable and analysing the three sub-categories separately supports this interpretation, although it is utterance length (for exact and expanded imitation), the GCSE score for listening, reading and writing (for expanded imitation) and student vocabulary (for reduced imitation) rather than turn length that predict the frequency of imitation. For expanded imitation, the retention in the model of student group as a positive predictor appears to reflect a greater overall tendency for the teacher in the more able group (Teacher B) to use expansion as a conversational strategy regardless of individual student proficiency. One initially puzzling result is that the combined GCSE score for listening, reading and writing is a *positive* predictor of teacher exact imitation. On the other hand, this may result from exact imitations with a questioning intonation being used as a prompt for more elaboration from students whose overall language proficiency is thought by the teacher to justify longer and move informative conversational contributions.

Most of the previous analyses have explained between 10% and 30% of the variation in teacher behaviour. As with teacher MSTTR, however, the final two analyses explain over half the variance and over 80% for the

length of conversation. Firstly, the frequency of teacher back channels is positively predicted by the quality of the content of the students' contribution to the conversation. This explains 50% of the variance. Our definition of 'back channel' is in the sense of contributions which consist of nothing more than 'oui', 'mm', 'mm hmm', etc. and function to signal continuing interest, attention or understanding (see Schegloff 1982:78). In our analyses back channels were not counted as conversational turns because to do so would have distorted the students' MLT calculations by artificially reducing turn length. Rather, they were interpreted as a response and encouragement to successful contributions by students. Their substantial relationship with the rating of quality of content is therefore to be expected.

Secondly, the length of the conversation, something entirely within the control of the teacher, was positively predicted by student vocabulary and ratings of content. However, the 24 teachers' rating of students' fluency was a *negative* predictor. While it is to be expected that longer conversations would be a feature of the more able linguist and a more interesting conversation, this may be tempered to a lesser degree if teachers terminate the conversation once fluency has been adequately demonstrated.

4. Conclusion

The analyses described above suggest that for eight out of ten indices of interviewer language, discourse accommodation does take place in response to student proficiency. In three cases we were able to explain a very substantial proportion of the variance in teacher behaviour. These results offer a degree of support for the validity of interviews carried out by, at least, these particular non-native speakers even though in most cases less than 30% of the variance could be explained. The result is also reassuring in the light of the fact that the tests were conducted by the students' own class teachers. It has long been a source of concern to teachers and examiners that oral interviews which contain highly predictable topics (family, school, holidays, friends, hobbies, etc.) lose their communicative validity because the interlocutors already know each other too well, and that the content of the 'conversation' may have been extensively pre-rehearsed in the classroom. Despite this, and in spite of the concern of language teachers to treat students equally, and therefore fairly, during oral tests (Chambers & Richards 1993), it seems that their linguistic behaviour *is* adjusted to a certain extent according to students' proficiency. It is worth noting, however, that the concern of teachers to treat students equally and the concern of testers to ensure that the oral proficiency interview is an authentic conversational experience marks the classic tension between reliability and validity. It may well be the case, therefore, that teachers' support for valid conversation is overridden when faced with the need to standardize examination conditions. It is

interesting that one variable – teacher vocabulary diversity – showed evidence of gross levels of accommodation between groups rather than to individual students within the group. Presumably, here teachers are pitching their language to the average level of proficiency of the group and varying it as little as possible within the group.

References

CHAMBERS, F. & RICHARDS, B. 1993. Oral assessment: the views of language teachers. *Language Learning Journal* 7, 22–26.

GASS, S. & VARONIS, E. 1985. Variation in native speaker speech modification to non-native speakers. *Studies in Second Language Acquisition* 7, 37–57.

JOHNSON, W. 1944. Studies in language behavior: I. A program of research. *Psychological Monographs* 56, 1–15.

MACWHINNEY, B. 1995. *The CHILDES Project: Tools for analyzing talk.* Hillsdale, NJ: Erlbaum.

MACWHINNEY, B. & SNOW, C. 1990. The Child Language Data Exchange System: an update. *Journal of Child Language* 17, 457–472.

MALVERN, D. & RICHARDS, B. 1997. A new measure of lexical diversity. *Evolving models of language*, ed. A. Ryan & A. Wray, 58–71. Clevedon: Multilingual Matters.

PINE, J. 1994. The language of primary caregivers. *Input and interaction in language acquisition*, ed. C. Gallaway & B. Richards, 15–37. Cambridge: Cambridge U. Press.

RICHARDS, B. & CHAMBERS, F. 1996. Reliability and validity in the GCSE oral examination. *Language Learning Journal 14, 28–34.*

ROSS, S. 1992. Accommodative questions in oral proficiency interviews. *Language Testing* 9, 173–186.

ROSS, S. & BERWICK, R. 1992. The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition* 14, 159–176.

SCHEGLOFF, E. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. *Georgetown University Round Table on Language and Linguistics. Analyzing discourse: Text and talk*, ed. D. Tannen, 71–93. Washington DC: Georgetown U. Press.

VAN LIER, L. 1989. Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly* 23, 489–508.

WESCHE, M. 1994. Input and interaction in second language acquisition. *Input and interaction in language acquisition*, ed. C. Gallaway & B. Richards, 219–249. Cambridge: Cambridge U. Press.

YOUNG, R. & MILANOVIC, M. 1992. Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition* 14, 403–424.

Received September 28, 1999
Accepted January 20, 2000

Brian Richards
The University of Reading
Bulmershe Court
Earley
Reading, RG6 1HY
United Kingdom
email: B.J.Richards@reading.ac.uk