Native Reactions to Non-Native Speech: A Review of Empirical Research*

Miriam Eisenstein
New York University

Recent research considers native reactions to various aspects of non-native speech and associated judgments regarding such speakers. The studies discussed here view listeners, speakers, and language from a variety of perspectives employing both objective and subjective research paradigms. Interlocutor variables which have been found to influence linguistic perceptions include age, social status, degree of bilingualism, and educational level. Even the linguistic sophistication of the listener may be important.

Studies of error gravity, which treat the perceived seriousness of error types in learners' interlanguage, may now be contrasted with data from several related and unrelated languages. The relative intelligibility of language samples has also been investigated as has the role of comprehension in the information of linguistic judgments.

Research reflecting listeners' personal impressions and reactions shows that non-natives tend to be downgraded in contexts ranging from the classroom to the workplace. This area of experimentation would be enhanced by exploration of the issues through studies in natural sociolinguistic contexts.

Introduction

It has been well established that an individual's speech is the source of specific and consistent listener reactions and that linguistic variables influence perceptions of the speaker's personality and status (see Ryan in this issue). This paper will review the extensive empirical research addressing the question of native reactions to non-native speakers and speech forms ranging from interlanguage, the developing system of second language learners (Selinker 1972), to the relatively more stable varieties such as the non-native Englishes described by Kachru (1982) and Smith (1981). The literature considers perceptions of the relative seriousness of particular errors, often referred to as 'error gravity', along with more holistic characteristics including intelligibility, acceptability, and listener irritation. Native feelings about non-native speakers are also the subject of investigation and encompass the evaluation of personality, intelligence, socioeconomic level, and degree of desired interaction with the speaker. Responses may be reactions to qualities inherent in learner language or reflections of stereotypes which are mediated by a linguistic stimulus. A listener's impression of a second language speaker may affect such crucial areas of social interaction as job opportunities.
Native Reactions to Non-Native Speech / Eisenstein

(Kalin, Rayko, and Love 1979), teacher-student relations (Seligman et al 1972), and international business (Inman 1982).

Research Approaches
The relative importance of second language errors has been measured in a variety of ways. Researchers have asked subjects to correct examples with errors, to rank errors relative to each other for perceived seriousness, or to rate the irritability caused by errors. An interesting approach is the operation task suggested by Quirk and Svartvik (1966) in which the subject is asked to perform some grammatical change on a language sample. An analysis of the results is thought to reveal aspects of a hearer or reader's unconscious reactions to the linguistic stimulus.

Reports from actual communicative settings can reveal a good deal about native reactions to non-native speakers. Mauser (1977) for example, has noted that American businessmen have difficulty conducting their affairs in Japan when they try to use humor in their negotiations. While this is quite acceptable in the United States, it is totally inappropriate in a Japanese context. Adelman and Lustig (1981) discuss intercultural communication problems between Saudi Arabian and American businesspeople. Managers find that misunderstandings arise from such diverse areas as negative transfer of intonation patterns, the use of different organizational logic, and differences in social behavior.

Research which is more experimental in nature is based on listener reactions to a language sample presented for the express purpose of examining interlocutor attitudes. There are many alternatives to be considered in creating and presenting a linguistic stimulus. The segment to be judged might consist of words, phrases, sentences, or a stretch of discourse. The mode can be face to face, oral, written, audiotaped or videotaped. Either abstract or contextualized presentation is possible. Remillard et al. (1973) included both taped and written samples, Callary (1974) used a written sample alone, while Williams et al. (1971, in Wolfram and Fasold 1974) used videotapes of speakers. The topic of the speech sample is also relevant. Colquhoun (1978) asked speakers to relate experiences in which they had been in danger, Chachere and King (1976) asked learners to summarize a paragraph they had previously read to themselves, and Granger et al. (1977) had different individuals describe the same picture. Spontaneous monologues were employed by Mulac et al. (1974), masked voice samples from recorded discussions were used by Scherer (1972), and contextualized monologues served as stimuli for Eisenstein and Hopper (forthcoming).

One of the most productive approaches to eliciting linguistic attitudes is the matched guise technique, designed to investigate reactions to languages or dialects by controlling other variables in the experimental situation. This is accomplished by tape recording the same individual reading a given passage in different language varieties. Since the reader is the same, voice quality, actual speaker personality, and passage content are assumed to remain constant in the different stimulus segments. Subjects who are asked to make comments about readers mistakenly assume that the people reading are different individuals. A pioneering study using
this approach was undertaken by Lambert et al. (1960), and Lambert (1967) stated later that the method had good reliability in evaluating language.1

The matched guise technique provides an alternative to the more traditional attitude questionnaires in which subjects are simply asked their conscious opinions regarding members of a certain group of their language. Information about listener's responses has also been revealed through commitment measures asking a subject's willingness to engage in particular types of behavior or relationships with certain speakers [Agheyisi and Fishmen 1970].

An objective means of evaluating attitudes which is widely used in eliciting reactions to language is the semantic differential technique developed by Osgood [Osgood, May, and Miron 1975; Osgood 1964; Osgood, Suci, and Tanenbaum 1957]. On a series of bi-polar adjective scales, the subject rates a concept or speaker for a given set of characteristics. Adjectives used in a study may be directly elicited from a population (Fink 1978) or chosen from the literature [Gardner and Taylor 1968]. Houston (1972) points out that the semantic differential is a good research tool because results are quantifiable and not dependent on the verbal fluency of the subject.

Unfortunately, even objective tests which try to take the subject's focus off the expression of language attitudes do not always succeed and can result in inaccurate representations of listeners' actual feelings. Schneiderman and Walker (1978) found that raters concealed knowledge when language reactions were measured objectively. Personal interviews served to show that raters modified their responses to language samples because they did not wish to appear bigoted, but the same subjects were willing to reveal their true feelings when given the opportunity to explain their views in a personal and secure setting. The use of open ended questions is also less likely to bias judges' responses than those presenting forced choices. Eisenstein (1979) confirmed objective data by having listeners react to speech samples by choosing a picture they thought represented the speaker and answering open-ended questions about relevant feelings and impressions.

In undertaking a study, it is critical that the researcher give consideration to variables which may influence language perceptions. Listener judgments have been found to be affected by degree of bilingualism [Minardi 1982; Olynyk et al this issue], age, [Magnan 1982], social status [Ryan and Sebastian in press], and sex [Markel and Roblin 1965]. The effect of college education has also been an issue, and while Mulac et al. (1974) did not find significant differences between college and non-college educated judges, Giles (1970, 1971) noted that a college education caused language perceptions to become more liberal and less ethnocentric. Mills and Hemsley (1976) have also found that educational level can influence judgments of grammatical acceptability.

Although many experiments report listener reactions to uncontextualized language samples, the importance of context in affecting responses has been underscored by Greenbaum (1973, 1977) and Ryan and Carranza (1975). Markel and Roblin (1965) found that the content of speech samples influences judgments, as does the congruity of content and voice quality.
The relative contribution of particular linguistic features to listeners' judgments is also of interest. Remillard et al. (1973) reported that syntax, lexicon and phonology are related to acceptability. Callary (1974) showed that judges can assign correct status to a speaker on the basis of syntax alone when phonology and morphology are held constant, and Ellis (1967) found that subjects could correctly identify speakers' social status on the basis of pronunciation alone by simply listening to samples of counting. The prosodic pattern is also an important contributor to linguistic perceptions. Phillipson (1978, in Albrechtsen et al. 1980) observed that the typical intonation of Danish learners elicited negative personality judgments from native English speakers.

Brown et al. (1975) summarized the results of several studies which considered how acoustic variables can influence perceptions of personality from speech. Such acoustic factors include the fundamental frequency of language, the rhythmic pattern, and the power spectrum characteristic of voices. Mehrabian and Wiener (1967) studied the relative contributions of verbal content, vocal characteristics, and visual factors to evaluations and discovered that tone of voice was a major determinant of reactions to utterances. Hart and Brown (1974), using both spontaneous utterances and recitations of standard passages, determined that speech content contained more information about benevolence than did vocal qualities, while vocal qualities were more important for perceived social attractiveness. Brown, Strong, and Rencher (1973) discussed the effects of manipulating acoustic parameters on the perception of personality from speech. Speeding voices made speakers seem less benevolent, and slowing voices caused them to be judged less competent. Finally, Brown, Strong, Rencher, and Smith (1974) showed interactive effects for rate, pitch, and intonation.

**Native Reactions**

**Error gravity**

A considerable body of data has resulted from studies of native speaker reactions to categories of errors produced by second language learners. An assumption of this research on error gravity is that certain kinds of errors are more serious than others which are less extreme in their impact. One of the most thorough research projects in this area was conducted by Johansson (1978, 1975). In his study of native speaker reactions to the English of Swedish learners, Johansson used a variety of research approaches including one in which he added number sequences to interlanguage sentences as a representation of the increased cognitive load that would be caused by natural language activities such as preplanning and attention to nonverbal signals. Listeners, native English speakers, wrote down the sentences in corrected form and included the number sequences. The order of error types found, from most to least irritating was: verb complementation errors, concord errors, and word order errors. School children 12 and 13 produced the same order of error types as university students.

Johansson also used a reading test in which native speakers read correct and incorrect samples and answered questions based on the texts. Incorrect versions took longer to read, and the increase in time corresponded with the ordering of error types in the perception task described above.
In a series of experiments involving different types of phonological errors, Johansson found that mispronounced consonant errors were judged more serious by natives than vowel errors, and mispronounced sounds in isolated words were rated higher than in sentences and texts. Phonological and grammatical deviance were also compared, and Johansson reported that the degree of listener misunderstanding was greater in phonologically deviant samples. Although he stressed the limited nature of these findings, Johansson interpreted the data to indicate the importance of an acceptable pronunciation for non-native speakers. In addition, Johansson evaluated reactions to grammatical versus lexical errors and noted that lexical errors created more problems than grammatical ones.

Johansson also compared the relative contributions of segmental and nonsegmental errors in communication. Bansal (1969) had established the importance of prosodic features in the intelligibility of Indian English, and Dimitrijevic and Djordjevic (1971) had reported native speakers to be more tolerant in assessing individual sounds than continuous speech. In reactions to recorded words, sentences, and a short text, with pronunciation and intonation controlled, Johansson also found individual words were more acceptable than those in context. Native judges ranked individual words higher than non-natives while texts were rated relatively more critically by native judges.

Monolingual and bilingual listeners were included in a study by Minardi (1982) who investigated reactions to the interlanguage of Polish speakers with English as a second language target. She considered the reactions of native English listeners unfamiliar with the learners' group, English-Polish bilinguals, and English monolinguals, some of whom reside in ethnic neighborhoods where Polish people live. Fifteen sentences with grammatical and lexical errors were read by a native English speaker to eliminate phonological and prosodic errors, and subjects were asked to write sentences containing errors in corrected form. Any changes made in addition to the actual error correction were counted by the researcher as a measure of irritation. Although the order of error gravity thus obtained was essentially consistent for all listeners, the number of corrections deemed necessary was greater for those living outside Polish neighborhoods. From most to least serious, errors were ranked as follows: 1) errors caused by use of Polish endings, 2) errors in word building and tense, 3) errors in word order and complementation (mostly of a serious nature). A subjective judgment test listing uncorrected sentences in written form supported this order or errors.

Minardi also considered phonological and prosodic errors by eliciting a series of words, sentences, and short passages from six Polish ESL learners. Listeners judged each stimulus on a scale ranging from very bad to very good English. Highest scores (very good English) were given to the sentences and lowest scores were attributed to the passages. The lower rating of the passages by natives supported Johansson's findings. Minardi suggests that while lexical and grammatical elements are most helpful for interpretation on the sentence level, prosodic errors in discourse may make it difficult for the listener to find a familiar pattern.

Chastain (1980) investigated the reactions of native Spanish speakers in Madrid to 35 written Spanish sentences, each containing one to three errors typical of
English-speaking Spanish learners in an American university. Sentences were rated either comprehensible and acceptable, comprehensible but unacceptable, or not comprehensible. Listeners were asked to underline errors and evaluate them. Comprehension was more seriously affected by word usage than by the choice of a wrong word or the addition/omission of words (with the exception of incorrect past participle forms). Some of the errors rated as unacceptable in descending order of importance include: omitting an infinitive after a preposition, failure to use *estar* in the progressive, not using passive subjunctive with an *if*, incorrect use of *para* and *por*, and failure to use the irregular form of the past participle. Some errors considered comprehensible and acceptable are: omission of the definite article after the verb *gustar* and use of the plural possessive with a singular noun. Results seem to indicate that native speakers are not concerned with correct use of definite articles or agreement of nouns and adjectives in most cases. Surprisingly, few error categories caused serious comprehension difficulty, and ninety percent or more of the native speakers were able to understand material including forty of the forty-eight errors used in the sample.

A study by Guntermann (1978) was conducted in El Salvador to determine the seriousness of interlanguage errors in terms of their relative intelligibility. This investigation began with an analysis of the oral interviews of 30 Peace Corps volunteers whose errors were transcribed in writing and classified according to grammatical categories. Based on this analysis, a recording of 43 sentences was made, some with multiple errors. Sentences were taped by four English-speaking males who had high proficiency in Spanish. Listeners, non-English speaking members of families with whom the volunteers had lived during training, were asked to restate each sentence according to what they thought the volunteers had to say. Most constructions were accurately represented by listeners. Sentences containing multiple errors were most often miscomprehended, followed by those involving substitution errors. Omission and agreement errors were misinterpreted less frequently. In a later phase of the study, article agreement errors were found to be more disruptive than article omission.

As part of a larger experiment, Druist (1977) had fifteen adult Spanish speakers judge the oral Spanish of twelve Puerto Rican-American teenagers who exhibited a wide range of ability in Spanish. The most unacceptable sentences included English syntax and vocabulary errors, verbs incorrectly marked for person or number, the indicative used for the subjunctive, and present or subjunctive substituted for the conditional.

Politzer (1978) compared the perceived seriousness of error types produced by American speakers of German as a second language. He recorded 60 pairs of German sentences in which six error categories were compared with each other. Sentences were tape recorded by a German-English bilingual who assumed a slight American accent in all of the sentences, and judges were 146 German speaking teenagers. While there was considerable variation in the data, the general ranking from most to least serious was: vocabulary, morphology, word order, gender confusion, case endings. Considerable variance was attributed to listeners' schooling, sex, and age.
Native speaker reaction to interlanguage in French has been the subject of several recent studies. Piazza (1980) investigated Frenchmen's tolerance for sentences containing grammatical errors typical of Americans learning French, in terms of comprehension and irritation. Parisian Lycée students rated 20 types of grammatical errors embedded in tape recorded and written language samples which had been produced by an American learner of French. One hundred sentences containing 20 error types (four of each) and 20 error free sentences were used as stimulus materials. Each listener rated the sentences (spoken or written) on a scale of comprehensibility or irritation. Results revealed that more comprehensible errors were generally less irritating, and respondents gave a lower rating for irritation than lack of comprehensibility caused by identical sentences. When error types were combined into six groupings, natives showed greatest tolerance for errors involving verb tense, usage, and agreement, a medium amount of tolerance for errors involving noun markers and word order, and least tolerance for errors of verb form and pronouns. Also, errors were better accepted in written than in spoken language. It is interesting to note Piazza's comment that 'some of the error types investigated are ones made by Frenchmen themselves'. This calls into question whether 'error' is the proper designation for such language forms which may be acceptable in certain native dialects or registers. Non-native use of colloquial native forms was considered by Swacker (1976) whose work is discussed later in the paper and by Magnan (in this issue).

Magnan (1982) examined which grammatical errors made by Americans speaking French were most irritating to native French speakers in France and to teachers of French in the United States. Examples of 15 error types were embedded in pairs of incorrect sentences which had been recorded by a female with good pronunciation. For each pair, listeners indicated which of the two incorrect sentences they preferred. Judges comprised native French speakers teaching all levels, native English-speaking French teachers at junior high and high school levels, university instructors (M.A.) and university faculty (Ph.D.). The resulting rank of errors from most to least irritating was as follows: verbs, pronouns, definite articles, prepositions, and adjectives. These findings are similar to previous orders noted above for French, Spanish, German and English.

Magnan (see this issue) found a differential reaction to French gender errors depending upon the age of the listeners. Younger students were more sensitive to gender errors than adults, perhaps reflecting a normative attitude associated with the prescriptive orientation of the school environment. Differing sensitivity to gender errors might also relate to social class since younger students in the sample were from a lower socio-economic class than the older ones. With the exception of age, non-native French speakers showed similar reactions to error types across sex, geographical regions and differing amounts of experience with English. Native judges were irritated by two sociolinguistically stigmatized errors, one associated with popular French (e.g. vous disez) and an error associated with a stereotype of non-native speech (moi dois, toi dois). Reactions of American French teachers were much the same as non-teacher native French speakers. Within teacher groups, native speaker teachers were more similar in reactions to the youngest students while university faculty were more like native adults.
Tardif and d'Anglejan (1981) examined the importance of certain characteristics of French as it is spoken by Quebec Anglophones, native English speakers in Quebec, Canada. Typical errors of advanced level students of French as a second language and bilingual speakers were identified and grouped into five categories. In this study, errors produced by native speakers were not included since the authors note that it does not seem reasonable to ask non-native speakers to conform to a norm not present in native speech. Two lists of sentences were generated, one half grammatically incorrect, each with a typical error, and the other with the grammatically correct version. Sentences were randomly mixed and were tape recorded by two non-native French speakers, one with a strong accent, the other with a very slight accent. Both speakers were English Canadian females studying advanced French. Listeners, adult native French speakers who were university graduates and Anglophone adult students attending French as a second language classes, answered a questionnaire evaluating the sentences on intelligibility, grammaticality, acceptability and irritation.

Results indicated that phonology had a significant influence on each of the four variables. Sentences recorded by the more accented speaker were less intelligible, less acceptable, less grammatical, and more irritating than those recorded by the less accented speaker. In fact, one half of the native judges perceived an error in sentences that had none when they were recorded by the more accented speaker. However, as the authors point out, it was not clear whether the accent itself might not have been considered an error. Degree of accent did not have as strong an influence on the judgments of non-native French speakers, and native speaker judgments were much more accurate than those of non-natives. The hierarchical order of error categories was consistent for the four dependent variables, from most to least serious: place of indirect object pronoun and preposition, preposition preceding cities and countries, place of adjective, and noun gender. Lack of seriousness in gender errors has appeared frequently in the studies reported so far, calling into question the extent to which gender should be the focus of formal learning.

Le Picq (1980) examined the linguistic acceptability of French produced by Canadian students in their sixth year of a French total immersion program. Learners spoke English at home and were upper middle class. Learner tapes and control tapes from native French Canadian speakers were judged by French Canadians of varied ages and levels of bilingualism, with an equal number of French monolingual and French-English bilinguals represented. Each judge evaluated one tape on the basis of orally presented open and closed questions. Judgments included the global acceptability of the sample which consisted of an interview done with the experimenter. The tapes were fifteen minutes long, unusual in length for this kind of research. However, the experimenter felt the time involved was necessary because of the considerable variation present in an individual's production of interlanguage. Error gravity was judged for six sentences, each illustrating a particular error type. These judgments included acceptability of the sentences and irritation produced by them. Finally, the judges were asked to specify the criteria on which they had based their decision.
In descending order of gravity, errors were judged as follows: verb substitution, word order, noun substitution, pronoun omission, verb morphology and article gender. Natives and non-natives did not make judgments based on the same criteria, and both age and level of bilingualism influenced listener reactions. These variables interacted with error type, introducing a particular degree of acceptability and/or irritation. Acceptability also depended on considerations including usage of the second language, proficiency of the listeners, communicative situation, and socio-political relationships between linguistic communities. Based on an analysis of each type, Le Picq also concluded that factors significant in native judgments of non-natives speech encompassed the speaker's self confidence, effort put forth to communicate, vocabulary, social attraction, intelligibility (determined in part by hesitations and voice quality), and linguistic content. A comparison of adult and child judges showed the adults to do better at error detection and children to be generally more accepting. Bilingual judges were more tolerant than non-bilinguals on certain psychosocial dimensions, and bilinguals understood better and were more tolerant of hesitations. For all judges, the personality of the speaker as revealed by the content of the interview was the crucial factor in the evaluation.

Intelligibility studies
Several studies focus specifically on the question of interlanguage intelligibility for native speakers, and some consider non-native listeners as well. Burt (1975) investigated which types of errors might cause the listener or reader to misunderstand the message intended by an EFL learner. Several thousand English sentences containing errors actually made by adult learners were taken from recordings of spontaneous conversations, written compositions, and letters. A series of grammatical sentences containing two or more errors were selected from the corpus. Native English speakers were asked to judge the relative comprehensibility of a sentence as errors were systematically corrected, either one at a time or several at a time. Based on the data, Burt identified two error categories, global and local errors. Global errors are those affecting overall sentence organization and seem to significantly impede communication. These include wrong word order, and missing, incorrect, or misplaced sentence connectors. Local errors affect single elements or constituents in a sentence and do not usually interfere with communication. These encompass problems with noun and verb inflection, articles, auxiliaries, and the formation of quantifiers. Burt also noted learner errors using 'psychological predicate constructions' that require the typical order of the experiencer to be reversed. For example, 'This lesson bores me' exhibits the more common English order. A typical error would be 'Your mother worries you', with the intended meaning that you worry her.

Ervin (1977) reported reactions to Russian interlanguage by three categories of judges: native Russian speakers who knew little or no English and were not teachers, native Russians who taught the language in American schools, and native English speakers who also taught Russian in American schools. American students of Russian enrolled in an intermediate level college course produced oral data elicited through the narration of three picture stories. Comprehensibility
was rated, and results showed generally good agreement for the listener groups on relative learner ranking. The most accepting judges were Russian non-teachers, provided that the interlanguage exceeded a minimum threshold level of comprehensibility. The native English speaking teachers of Russian were most accepting of very low level speakers.

Two experiments on reactions to foreign accent under conditions of masking and filtering were conducted by Lane (1963). Twenty-four American undergraduates listened to recorded articulation lists which had been produced by one American and three foreign born speakers of Punjabi, Serbo-Croatian and Japanese. Non-native examples were much less intelligible under all conditions, and individual differences among the foreign speakers did not appear to affect intelligibility scores. These results with the findings of Smith and Rafiqzad (1979). In a study of the relative intelligibility of nine varieties of educated English [American Standard and eight non-native varieties], they found great consistency in the degree of relative comprehensibility among listeners from eleven countries. unexpectedly, the native speaker was among the least intelligible, and listeners' fellow countrymen were easier to understand than other non-native speakers in only two cases. Intelligibility was assessed on the basis of a close procedure and a listening comprehension test. The results of the study are not conclusive since the speech samples were not balanced for topic or speaker fluency. This problem is acknowledged by the researchers, who note that there was a great deal of variability in the difficulty level of the passage.

Smith and Bisazza (1982) carefully controlled the stimulus material in a study which considered the comprehensibility of three English varieties: American, Indian, and Japanese. A male speaker of each dialect recorded a version of the Michigan Test of Oral Comprehension. Listeners from the United States and six Asian countries were balanced for age, sex, and educational background. Most subjects were freshmen from the University of Hawaii. Intelligibility was measured by having subjects choose one of three pictures corresponding to sentences or paragraphs heard and one of three phrases which best answered the questions after the sentence or paragraph was heard. American English was most comprehensible, followed by Japanese and Indian English. The authors explain the differential comprehension of Indian versus Japanese English on the basis of the listeners' greater exposure to the latter, since Hawaii has a large Japanese population. Results also showed that respondents' subjective judgments of relative intelligibility agreed with the objective test results. This confirms the work of Eisenstein and Berkowitz (1981) who also found that learners' subjective impressions of dialect intelligibility were quite accurate.

The intelligibility of Indian English was also investigated by Bansal (1969). Recordings of 24 Indian English speakers and four R.P. [received pronunciation] speakers were presented to listeners of different nationalities. They heard connected speech, sentences, and isolated words. Respondents were asked to repeat what they heard and understood. Indian and R.P. dialects were not equally intelligible, and there was considerable variation for the Indian dialects depending on word stress sentence stress, rhythm, and intonation.
Nelson (1982) points out that there are both linguistic and non-linguistic factors which may influence intelligibility. The extent of 'tolerable deviation from one's own model may vary according to the attitude towards the user . . . ' Nelson states that deviation in phonology seems to be better accepted than divergence in lexis or grammar. He focuses in particular on the rhythmic patterns of native and non-native English. South Asian languages are syllable-timed as opposed to English which is stress-timed, which results in a distinct rhythm in South Asian English. Stress-timed languages have their stress associated with a breath group, while syllable-timed languages organize their timing with respect to syllable length intervals. In addition to the syllable-timed nature of Indian English, discussed by Nelson, unstressed vowel reduction is not as pervasive as in American English. The question of how prosodic features influence dialect intelligibility is dealt with only to a small extent in the literature and requires further investigation.

Native versus non-native reactions to interlanguage were also investigated by Brodkey (1972) who varied both aspects of language situation and receiver category in distortion tests extracted from lectures and interviews. Samples were based on American college lectures and interviews of Spanish-accented English speakers. Anglo-American, American Indian and foreign university students as well as Anglo-American teachers, some with ESL experience, were asked to write down what they heard on the tapes. Results showed that prior listener experience with the voice of the speaker is a crucial factor in intelligibility, and that a combination of increasing age, education, and experience teaching ESL enhances the listener's ability to understand speakers of interlanguage. Not surprisingly, bilingual Spanish-English listeners had an advantage in decoding Spanish accented English.

**Reactions to speakers**

In addition to judging aspects of interlanguage itself, native speakers often extend their impressions to characteristics of the learners as individuals. Anisfeld, Bogo, and Lambert (1962) evaluated the reactions of both Jews and Gentiles to Jewish-accented English speech in Canada, using matched guise tapes. In general, the Jewish-accented speaker was viewed more negatively than the unaccented speaker by all listeners. These findings differed somewhat from an earlier study by Lambert et al. (1960). French Canadian students in the earlier study downgraded their own group much more than the Jewish group evaluated the Jewish-accented speakers.

A study which used both matched guise and attitudinal ratings was attempted by Giles in 1970. In this experiment, listeners from two age groups (12 and 17 years of age) from working class and middle class backgrounds responded to thirteen different foreign and regional accents. (Most but not all of the accents were done by a single speaker.) Listeners reacted to characteristics relating to the pleasantness of the sample, how comfortable they would feel interacting with the speakers, and the prestige associated with the accent. Respondents first judged the speech samples and later rated lists of the same accents. Results showed the older subjects were better able to recognize the accents represented, and accents were regarded slightly more favorably when judgments were elicited directly from the listeners. The working-class group was more prone to 'accent loyalty'. The
ranging of accents on different dimensions showed R.P., French, and North American in positions of higher prestige as compared with British regional accents. German was middle or low, and Italian was intermediate. Accents considered included R.P., North American, Welsh, Italian, Indian, German, Cockney, and Birmingham among others.

Palmer (1973) had linguistically unsophisticated volunteers judge speakers of four unrelated languages on tasks of reading, retelling, and narration. Thirty-six speakers represented Arabic, Lingala, Spanish and Vietnamese native speakers and ranged in English proficiency from poor to excellent. Each respondent evaluated twelve taped segments on a five point scale and was asked to comment on why specific ratings were given. The judges were not successful in guessing language background except for Spanish, but an ANOVA showed a significant difference in how the groups were rated. While it is not possible to be certain to what extent variance related to particular group stereotypes as opposed to second language ability, this study shows that linguistically unsophisticated judges were reliable in their assessments.

Ensz (1976) presented native French listeners with five guises representing Americans speaking French. Errors in grammar, vocabulary and pronunciation (phonology plus intonation) were manipulated so that different interactions could be considered, and reactions were measured on semantic differential scales. The outcome revealed grammatical deviance to be most downgraded with errors in vocabulary and pronunciation causing roughly equal but less extreme negative responses. The guise with slight pronunciation errors only was rated the highest. Results led to the conclusion that sex, age, and occupation did not cause different reactions to deivance in this case. A questionnaire to examine native attitudes towards Americans was also administered, but this factor was not found to influence responses.

The impact made by phonemic and prosodic features of European-accented English upon the attitudinal judgments of American listeners was determined by Mulac, Hanley, and Prigge (1974). Stimulus speakers were from Norway, Italy, Eastern Europe, and the United States; listeners were middle-aged, middle-class townspeople and university students. Spontaneous monologues were elicited by means of photographs, and each was rated on the extent of dialect demonstrated and on a series of bipolar adjective scales. Measures of syntax and semantics were used as covariates in the data analysis. Results showed significant differences as a function of the speaker's country of origin, and all listeners downgraded accented speakers. The native speaker was rated highest in socio-intellectual level, followed by Eastern Europeans and Norwegians. on aesthetic quality, natives were again rated highest, but there was no difference among foreigners. On dynamism, East Europeans were rated high than Italians. There was a significant interaction between sex of speaker and country of origin influencing judgments made.

Albrechtsen et al. (1980) gathered data from Danish pupils who had been studying English for five to six years. During interviews, they talked about a limited set of topics, and eight sets of tapes were selected on the basis of 'error density' which
is the number of errors per fixed number of words. (Two tapes had low error density, four average and two high.) A ninth tape of a native speaker was included for control and one interlanguage tape was repeated to check for consistency of response. Listeners were 300 native English speakers from three regions of Great Britain consisting of adults from nonacademic spheres and teenagers aged 16 to 17. Respondents listened to extracts, answered two questions and replied to bipolar scales regarding linguistic aspects of the tapes, content, intelligibility, and learner personality.

Results showed no difference between geographic groups of listeners but a significant difference between teenagers and adults. A factor analysis resulted in four factors relating to language, personality, content and comprehension. An error analysis was performed on the learner data which considered grammatical categories, intonation, hesitation phenomena, communication strategies, and rate of speech. It was discovered that learners who made fewer lexical and syntactic errors also made fewer segmental errors, had good intonation, limited hesitation, and fewer communication strategies (e.g. message abandonment, language switch, appeal for assistance). Learners' rate of speech correlated significantly with communication strategies. In general, texts with fewer errors were evaluated positively; however, there were positive evaluations of texts which contained a fair number of syntactic and lexical errors provided they displayed few communicative strategies. One limitation of this analysis noted by the authors is that the various communicative strategies were grouped irrespective of their quality. It is also the case that communication strategies are normally used in an interchange with listeners, an experience different from passively hearing tapes. Errors interfering with intelligibility included wrong choice of conjunctions and anaphoric pronouns with ambiguous references. Low comprehensibility was also related to errors at the discourse level and extensive hesitation. An additional finding was that once a certain level of comprehensibility is achieved, increasing correctness does not necessarily improve the attitude evoked by the learner from the native speaker.

The interactive effect of non-native speech and regional dialect is the subject of a study by Swacker (1976). She had native speakers evaluate standard English, East Texas native English, Arabic-accented English devoid of regional markers, and Arabic-accented English containing Texas regionalisms. While the two native English speakers were rated similarly, and the accented speaker was judged lower, the most negative reactions were elicited by the accented speaker whose English contained markers of Texas pronunciation and grammar. This shows that 'certain dialectal markers may be ... acceptable ... when coming from a native speaker, but quite offensive when spoken by a foreigner' [Swaker 1976:17]. This implies caution in teaching the productive use of regionalisms to second language learners.

Several studies have investigated American reactions to speakers of Spanish-accented English. In general, these speakers have been negatively viewed by white, black, and even Mexican American listeners. Arthur, Farrar and Bradford (1974) reported that speech approaching Chicano English was negatively stereotyped by Anglo university students on scales relating to success, ability, and social awareness. In this study, all accented speakers were identified to listeners as Mexican Americans.
Ryan and Carranza (1975) elicited the reactions of Mexican American, black, and Anglo adolescents to standard and Mexican accented English on scales of status and solidarity in home and school contexts. Although standard English speakers were more favorably rated in every case, differences were significantly greater in school as compared to home context, and a foreign accent was better tolerated in the home environment.

In a study by Ryan, Carranza, and Moffie (1977), a hundred college students reacted to varying degrees of Spanish accentedness in the English of bilinguals reading a formal passage. Small increments in accentedness were found to be associated with gradually less favorable ratings of status, solidarity, and speech characteristics.

Social judgments of speakers with differing degrees of accentedness was also the subject of an experiment by Sebastian, Ryan, and Corso (1978). Middle class Anglo American undergraduates listened to readings in standard English and three degrees of Spanish accentedness. Reactions were elicited with respect to accentedness, intelligibility, desirability of five social relationships with speakers, and the extent to which listeners thought speakers would agree with them on five social issues. An F scale measure of authoritarianism was also administered. Results showed that Spanish-accented speakers were thought to be lower in social class, less similar in beliefs, and less desirable in a range of relationships. Furthermore, evaluations became more negative with increasing accentedness. Social class assumptions were found to be important mediators of other judgments, and ethnicity became more important for listeners as possible relationships became more intimate. The more favorable judgments accorded to Spanish accented speakers perceived to be higher in social class was referred to by the authors as 'the Ricardo Montalban effect.' Interestingly, when Mexican Americans were misidentified as Anglos, they tended to be rated lower than when they were accurately identified.

Ryan and Sebastian (in press) investigated further the effects of speech style and social class background on social judgments of non-native speech. Middle class Anglo-American undergraduates listened to tape recorded readings of the same formal passage in standard American-and Spanish-accented English under two conditions: one in which no background was given and one in which a speaker introduced the tape by setting the social class background of the reader. One middle class and one lower class introduction was provided for each speech style, standard versus Spanish-accented English. Eighty subjects heard tapes with an introduction, and forty listened to voices alone. Each speaker was rated on characteristics including status, solidarity, social class, ease of understanding, and comfort in listening. Social distance judgments were also included. Ryan and Sebastian found that evaluative reactions and social distance judgments were significantly influenced by speech style and announced social background. Speakers of standard English and announced middle class speakers were rated higher than Spanish-accented speakers and announced lower class speakers. Also, announced social class had a greater influence for accented speakers than for non-accented ones. Furthermore, for most intimate interactions (date my sister, have as a close personal friend) standard speakers were preferred.
Sebastian, Ryan, Keogh, and Schmidt (1980) had subjects listen to tape recorded descriptions of six difficult-to-describe colors in accented and non-accented versions of English and in noisy and non-noisy conditions. Respondents selected the color described and then completed a questionnaire about the speaker and the experiment. Noisy speech was rated harder to understand although accentedness ratings did not differ in the noisy versus non-noisy condition. However, speaker characteristics in the noisy condition were rated more negatively than in the normal condition. The authors explain this in terms of a 'negative affect mechanism' in which the listener associates feelings of discomfort with the speaker as well as with the recording.

The studies above essentially referred to attitudes or impressions evoked by non-native speakers in various categories of interlocutors. But the extent to which experimental measures translate into actual behavior has often been called into question. An interesting but disturbing experiment was reported by Buttino and Sebastian (Sebastian and Ryan in press) which bears on native reactions to Spanish accented speakers. While outside the realm of most linguistic research, it illustrates the potential strength of behavior mediated by accented speech. Subjects believed they were administering shocks of different intensity and duration to a partner, and speech style was one of several cues to the partner's Spanish or Anglo ethnicity. Although subjects' reactions were not statistically significant, findings suggested that more aggression was directed toward individuals of Spanish ethnic background by angered subjects.

Accent and employment
A crucial area of native reactions to non-native speech is that of prospective employer judgments of job applicants. In general, research to date shows that non-native speech causes problems for individuals seeking employment due to natives' negative views of the speaker's ability to perform a job.

In 1973, Hopper and Williams concluded a study of speech characteristics and employability which included taped samples of simulated job interviews representative of standard English, Black English, Spanish-influenced English, and Southern-white-dialect English. Prospective employers rated the samples on a series of semantic differential scales and stated the likelihood that they would hire the speakers for each of seven job categories ranging from executive to manual laborer. Hopper and Williams found that, in this case, the employers' judgments were concerned with whether an applicant appeared confident, agreeable and self-assured. Employer ratings of these perceived speech characteristics were most predictable for the employment of executives and supervisory categories and not for the manual laborer position. A surprising conclusion of this research was that the ethnicity of the speaker was not important in employment decisions.

Hopper and Williams' conclusion has been contradicted by several subsequent studies. Rey (1977) found that a Spanish accent had a negative effect on South Florida employers considering prospective employees for a variety of jobs. Kalin and Rayko (1978) reported on discrimination and evaluative judgments against
foreign accented job candidates. In this experiment, English Canadian students acted as personnel consultants and imagined they were hiring employees at four levels of job status, ranging from industrial plant cleaners to foremen. Respondents were asked to predict how well each of the applicants would perform on these jobs. A brief biography and thirty second recording of each candidate was provided. Five speakers had a standard English Canadian accent, and another five were accented but fluent English speakers representing Italian, Greek, Portuguese, West African and Slovak influenced speech. To control for content, two versions of each stimulus tape were made, one accented and one unaccented. Biographies were created in order to make foreign-accented and non-accented speakers appear equal in personal and social backgrounds, education, and intelligence. Listeners' ethnic attitudes and authoritarianism were also measured. Results showed that accent and job status strongly influenced evaluative ratings, and there was a significant interaction between these two factors. Foreign-accented speakers, in general, were rated lower than English-Canadian speakers, and the difference in these relative ratings was greatest at highest and lowest levels of employment but in opposite directions. For the two higher status jobs, foreign accented speakers were rated higher. There were also positive correlations between measures of prejudice and extent of discrimination against foreign speakers.

An extension of the study described above is reported in Kalin, Rayko, and Love, (1979). In this case, accents which had been previously evaluated in terms of the relative status of the ethnic groups represented were chosen. In order of rating from high to low these were English, German, West Indian, and South Asian. English Canadians were excluded to remove the contrast effect between Canadian and foreign speakers present in the previous study. In the first part of the experiment, fifty Canadian undergraduates listened to recordings of fluent but accented English speakers, conversing informally on a current topic. Results showed that students correctly identified groups separated by accent with varying accuracy, but results were better than chance. Comprehensibility ratings corresponded to the evaluative hierarchy for these ethnic groups listed above. In the second stage of the experiment, students played the role of personnel consultants as in the previous study. At each job status level considered discrimination was based upon ethnicity. For the highest status jobs, the order was English accent preferred, then German, South Asian, and West Indian. For the lowest status job, this order was reversed. Relative comprehensibility ratings cannot explain this result since English-accented speakers were rated lower for low levels jobs. Theoretically, however, being understood should be helpful at any level of employment. Familiarity with English is also not an appropriate explanation since highest and lowest ratings were assigned to two native English speaking groups, British and West Indian. The experimenters conclude, therefore, that "the discrimination was triggered by the ethnicity of speakers" (Kalin et al, 1979:200).

Conclusion
Many questions regarding native judgments of non-natives and their interlanguage remain to be answered. It is still unclear whether relations to language per se and to people who speak a particular variety represent two separable categories or whether such responses are inextricably intertwined. Virtually all elements of
language have been found to be influential in affecting judgments; non-native aspects of phonology, syntax, lexicon and intonation have each been shown to evoke negative reactions. Responses reportedly stem from intelligibility, negative affect, or inferences about speakers' membership in particular social or ethnic groups. It is evident from the research reported here that all of these factors can potentially be involved. In general, it seems reasonable to state that non-native speakers tend to be downgraded by natives not only in academic experiments but also in real world contexts ranging from the classroom to the workplace.

Future research must focus on how these negative feelings come about and why variables such as age, sex or social status mediate judgments in some cases but not in others. Researchers must begin to verify their laboratory finds in more naturalistic settings. In most experiments judges are cast in passive roles, reading or listening to word lists, individual sentences or various types of discourse. But most natural communication involves interactions between interlocutors and combines both verbal and nonverbal aspects of communication. There is a need for studies conducted in real contexts allowing for give and take among participants.

Many researchers stress the importance of developing second language skills of learners so that their speech will more closely approximate that of natives. However, the development of effective teaching methods to achieve this goal is far from complete. Indeed, stressing a particular target language form in the curriculum does not insure its assimilation by the learner. Language proficiency testing is another area which might draw on the judgments of native speakers, both in setting priorities for proficiency testing and in test construction.

Listeners' negative feelings regarding non-native speech and speakers have serious consequences for our contemporary society in which there are many opportunities for native/non-native contact. It is to be hoped that future research will address the possibility of modifying adverse listener reactions of the kind reported here. As more is learned about responses to language varieties and, more importantly, to the people who speak them, this knowledge may help us to communicate better and to reserve our judgement of others for the facts.

Notes

*I wish to express my thanks to Alison d'Anglejan, Braj Kachru, S. Sridhar, and Jane Zuengler for their invaluable help in locating material for this paper. Appreciation also to Gail Verdi and Michael Robinson for their insightful comments.

1. Critics of matched guise such as Kramer (1964) have commented that it is unnatural to repeat the same message over and over, and that tapes lack social context. Agheyisi and Fishman [1970] point out that judges in a real setting react to congruity between topic, speaker and variety, interactions not considered in most matched guise experiments. Nevertheless, several studies have corroborated the results of matched guise research using other approaches [See Taylor and Gardner, 1969, d'Anglejan and Tucker, 1973 and Bourhis and Giles, 1976]

2. This approach has the advantage of control but might affect listeners' judgments by creating the expectation that a speaker with such a good accent should not make basic errors.