The purpose of this study was to: (1) describe differences in performance by non-native learners of English, when writing in different genres; (2) determine communicative value of grammatical errors as judged by a panel of native speakers; and (3) demonstrate inconsistencies in native speaker judgment of error gravity. Subjects were 20 English-as-a-Second-Language (ESL) college students. Data were drawn from 40 expository and creative writing assignments. Three native speakers evaluated the effects of student errors on comprehension. Results indicate that the subjects committed more errors in expository than creative writing, and that variation in essay length did not correlate with error quantity. Frequency of error type differed significantly in the two modes. Mistakes with unlear antecedents impeded comprehension most, while pronouns impeded comprehension least; ten other grammatical error categories fell in between in gravity. Although the most recurrent errors and the most serious errors did not present the same hierarchical order, they did reflect the same error categories; there was no correlation between the degree of error gravity and error frequency. The native speakers often could not reach consensus on error gravity. Implications are seen in these results for communicative second language teaching, instructional material development, error and discourse analysis, and classroom treatment of errors. Contains 23 references. (Author/MSE)
Inconsistencies in Error Production by Non-Native English Speakers and in Error Gravity Judgment by Native Speakers

Mhhood T. Arani, Ph.D.

Saint Michael's College
School of International Studies
Abstract

Inconsistencies in Error Production by L2 Learners and in Error Judgment by Native Speakers

The purpose of this study was a) to describe the differences in performance by non-native speakers when writing in different genres; b) to determine the communicative value of grammatical errors as judged by a panel of native speakers of English, and c) to demonstrate inconsistencies in native speakers judgment of error gravity.

The subjects for the study comprised of twenty ESL college students in the U.S. The source of data was forty essays written in the expository and imaginative modes. A panel of three native speakers evaluated the effects of the students' errors on their comprehension on a five point scale.

The study finds that a) the subjects commit more errors in the expository mode than in the imaginative mode and that variation in the length of writings is not responsible for variation in the number of errors committed; b) the frequency of error-type occurrence differs drastically in the expository and imaginative writings; c) mistakes with unclear antecedents impede comprehension the most while pronouns impede comprehension the least. In between these two categories come ten other grammatical categories; d) although the most recurrent errors and the most serious errors do not present the same hierarchical order, they do present the same error categories; e) there is no correlation between the degree of error gravity and error frequency, and f) native speakers often cannot reach a consensus on error gravity.

Given that developing communicative competence is the most essential goal of language learning and teaching, the presenter believes that the findings of the present study offer great insights for communicative language teaching, materials development, conducting error and discourse analysis, and the treatment of errors in second language classroom.
Introduction

In the last two decades there has been a drastic shift in the attitudes of researchers and language educators toward L2 learners' errors and error correction. Notions of error-free L2 production and error-preventive techniques in second language teaching advocated by the proponents of the audiolingual method and contrastive analysis have been discredited and have given way to a cognitive approach recognizing the development of communicative competence as the most inclusive goal of language teaching and learning.

With the increased interest in communicative competence and communicative language instruction (Savignon 1972, 1983, Canale & Swain 1980, Candlin 1975, Munby 1978), error interpretation, differing error perceptions, the communicative value of the learners' errors and the native speakers' degree of tolerance of L2 errors have taken on tremendous significance. Chastain (1980 a) maintains that the teacher perception of learner errors not only affects the instructional approach, but it also influences the selection of content, class activities, grading and error correction procedures.

While the focus of attention has shifted away from mere grammatical accuracy to error perception, error interpretation, and the degree of comprehensibility and acceptability, very often there is much inconsistency in the native speakers' judgement of the gravity of L2 errors. Even with members of a relatively homogenous speech community, considerable variation is found as to what constitutes acceptability or an error (Ludwig 1982). In
this regard, Piazza (1980) suggests that the likelihood of an error being made by a native speaker could also have an effect on the degree of acceptability. Furthermore, the problem of error evaluation is compounded when the number and the types of the L2 learners' errors and the resulting problems in comprehension differ from task to task.

Previous Studies

With the increased interest in teaching language for communication, the number of studies dealing with native speakers' reaction to the L2 learners' errors has also increased in recent years. Most of such studies have used comprehensibility, acceptability and irritation as major criteria for error evaluation. Ludwig (1982) mentions several non-linguistic variables which also influence native speakers' reaction to the L2 learners' production. To her, these variables include the personality of the speaker or writer, the use of communication strategies, and the possibility of cultural stereotypes or cultural clashes. Among the studies that have used comprehensibility as a criterion are works conducted by Burt and Kiparsky (1972, 1974), Piazza (1979), Olsson (1972), Guntermann (1978), Chastain (1980), and Tomiyama (1980). Studies done by James (1977), Politzer (1978), Piazza (1979), Chastain (1980), and Hughes and Lascaratou (1982) employed the criterion of acceptability, the degree to which a given error deviates from the language norm. Irritation—i.e., the result of the form of the
message intruding upon the interlocutor's perception (Ludwig 1982)—was a criterion in studies by Piazza (1979), Chastain (1980 a), Galloway (1980), Magnan (1981), and Ensz (1982). The general conclusion of almost all of these studies is that native speakers do not judge L2 deviant utterances in absolute terms; rather, they rank them in a hierarchy.

With regard to the errors that most impede communication, there are a number of studies cited in the literature. Many of these studies corroborate the suggestion made by Burt and Kiparsky (1974) that errors violating rules involving the overall structure of a sentence, the relation among constituent clauses, or, in a simple sentence, the relations among major constituents (global errors) cause more problems in communication than errors causing trouble in a particular constituent, or in a clause of a complex sentence (local errors). In their study of the communicative value of errors, Burt and Kiparsky found that the most typical of global mistakes are those which confuse the relationship among clauses such as the use of connectors, distinction between coordinate and relative clause constructions, parallel structure in reduced coordinate clauses, and tense continuity across clauses among others. Piazza (1980) in examining the French tolerance for grammatical errors committed by Americans concluded that "incorrect word order is relatively not irritating but can be a problem for comprehension" (p.424). In addition, in a comprehensive study involving Iowa State University faculty members, it was found that " respondents
judged as least acceptable those errors which, for the most part, are global and/or are relatively rare violations for native speakers, e.g., word order, it-deletion, tense, relative clause errors, and word choice" (Vann, Meyer, and Lorenz, 1984, p.432). Kresovich (1988) in a study of errors committed by Japanese students supports the general conclusion of Vann, Meyer, and Lorenz (1984), stating that the more an error impedes understanding of meaning, the less it is tolerated. According to his findings, since a word order error causes significant confusion for the reader of the L2 writings, it is judged less tolerable. But errors such as comma splice, wrong word choice, wrong preposition, and article mistakes, which do not detract comprehensibility, are judged as more tolerable.

Relatively little attention has been paid to the effects of lexical item selection on native speakers comprehension. In a study of language errors considered the most serious in intermediate Spanish classes, Chastain (1980 a) concluded that "comprehension is most severely limited by word usage, the use of wrong word or the addition or omission of words"(p. 212). This finding is also supported by Khalil (1985), who did a study on errors made by Arab EFL learners. Khalil suggests that semantically deviant utterances, i.e., errors in lexis and collocation, are judged by native speakers of English less intelligible and are interpreted with less accuracy than are grammatically deviant utterances.
The degree of consistency or inconsistency in native speakers' judgement of L2 deviant utterances has been another fascinating aspect of communicative error evaluation in the last two decades. In their study Vann, Meyer, and Lorenz (1984) had 164 respondents rank the relative gravity of 12 typical ESL written errors. Results indicated that most respondents did not judge all errors as equally grievous; rather, their judgements generated a hierarchy of errors. Furthermore, the results showed that 97 percent of the respondents were consistent at least half of the time. According to the authors, the pattern of response in their study also suggested a connection between consistency and severity of judgement; that is, "those respondents who were most consistent likely to be less tolerant in their judgement" (p. 433). In a similar study, Kresovich (1988) had 43 teachers of ESL (one British, sixteen American, and twenty six Japanese) judge the acceptability of specific error types produced by Japanese students. Contrary to the claim made by some studies (e.g., James 1977, Hughes and Lascaratou 1982), the results of the study did not show a great difference in the error perceptions of non-native and native speaking teachers in general. Kresovich suggests that an explanation could be the particularly advanced English language proficiency of the non-native speaking teachers in the study.

There are also a number of studies that report major inconsistencies in native speakers' judgements. A case in point is the work done by Chastain (1981 b). He asked his Spanish
informants to rate paragraphs from student compositions as comprehensible and acceptable, comprehensible but unacceptable, or non-comprehensible. He found that "of the total possible number of instances of errors thirty-four percent of the NPF [Noun Phrase Form], forty-five percent of the NPW [Noun Phrase Word] errors, twenty-three percent of the VPF [Verb Phrase Word] errors, and forty-seven percent of the VPW [Verb Phrase Word] errors did not attract enough attention to be underlined by the native speaker evaluators" (cited by Ludwig 1982, p. 293).

The Purposes of the Study

The following study demonstrates the communicative values of major grammatical errors as judged by a panel of native speakers of English. It further indicates how the members of the panel differ in their perception of the gravity of the learners' errors. It finally illustrates how non-native speakers, in this case Iranian students, perform differently in two different modes of English writing.

The Subjects for the Study

The group of subjects for this study was comprised of 20 Iranian students. The students majored in diverse fields of science and engineering. Eleven of the students were completing their graduate studies in either a Master's or Doctoral program and the remaining nine students were undergraduate students in some kind of science major. In addition to having taken regular
English courses in their high school years in Iran, eighteen of the students had taken one or two semesters of ESL courses in the U.S. before entering their academic programs. Fourteen of these students had taken the TOEFL four years before the present study and averaged a score of 476. All the students in the sample claimed that they used English for academic purposes and, obviously, for communicating with non-Iranians. At the time of the study, the students had lived in the U.S. an average of five years.

The Sources of Data

For the purposes of this study each of the twenty students in the sample was asked to write two free compositions in two different modes, one in the expository mode and one in the imaginative mode. However, due to their busy academic schedules, eight of them wrote only one composition. Therefore, the total number of compositions was 32, of which eighteen were written in the expository mode and fourteen in the imaginative mode. The compositions ranged in length from 157 words to 869 words, with a mean length of 428 words. The students' compositions were longer in the imaginative mode (Mean=483 words) than in the expository mode (Mean=385 words).

The topic of the compositions was chosen by the students themselves. In order to keep the writing environment consistent for all the students, they were asked to meet in a classroom and were given up two hours to finish the two compositions. They
were not permitted to consult any grammar or usage books while they were writing; however, they were allowed to use dictionaries.

The Analysis of Data

Upon recognition of errors in all the compositions, a frequency count was made of the total number of errors committed. The frequency count indicated that the students had made a total of 1387 errors in the thirty-two compositions. These errors were dispersed among fifty-one categories.

Since a vast number of errors were related to other components of language other than structure, a good number of errors were deleted from the corpus under study. Thus, in the final analysis, only twenty-one grammatical categories remained to be examined. These twenty-one categories were classified into two categories: Major Group and Minor Group. Having rank ordered the twenty-one categories based on the relative frequencies from the highest to the lowest, the category lying exactly on the 50th percentile (the median) divided the data into two groups. The ten categories above the middle ground comprised the major categories, while the ten categories falling below the middle ground constituted the minor categories.

In this study, because of the excessive amount of data, only the major categories of errors were considered. These categories, which consisted of 702 grammatical errors, are displayed in the following table:
## (Table 1)

**Major Error Types in Compositions in Both Modes**

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Order of Rank</th>
<th>Absolute Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles</td>
<td>1</td>
<td>133</td>
<td>14.4</td>
</tr>
<tr>
<td>Prepositions</td>
<td>*2</td>
<td>86</td>
<td>9.3</td>
</tr>
<tr>
<td>Unclear Antecedents</td>
<td>*3</td>
<td>86</td>
<td>9.3</td>
</tr>
<tr>
<td>Incorrect and Confusing Tenses</td>
<td>4</td>
<td>82</td>
<td>8.9</td>
</tr>
<tr>
<td>Number</td>
<td>5</td>
<td>80</td>
<td>8.6</td>
</tr>
<tr>
<td>Conjunction/Transition</td>
<td>6</td>
<td>76</td>
<td>8.2</td>
</tr>
<tr>
<td>Adjectives</td>
<td>7</td>
<td>43</td>
<td>5.3</td>
</tr>
<tr>
<td>Subjects and Predicates</td>
<td>8</td>
<td>42</td>
<td>4.5</td>
</tr>
<tr>
<td>Verb Phrases</td>
<td>9</td>
<td>38</td>
<td>4.1</td>
</tr>
<tr>
<td>Pronouns</td>
<td>10</td>
<td>36</td>
<td>3.9</td>
</tr>
</tbody>
</table>

* Categories 2 and 3 were equally recurrent. However, in order to keep the total number of error categories consistent, they were rank ordered differently.

The total number of grammatical errors committed in the compositions written in an expository mode was 537 errors of which 495 errors fell under the "Major" groups, while the total number of structural errors made in the compositions written in an imaginative mode was 399 errors of which 324 errors fell under the "Major" categories.

In general, this study found that this group of students, despite their extensive exposure to expository writings at school, not only made more errors when writing in an expository mode, but they also made more "Major" errors in their expository writing than in their imaginative writing.
Although in the present study no particular investigation was conducted to empirically identify possible factors responsible for this discrepancy, perhaps the difference in the number of errors in the two types of compositions could be attributed to, among other factors, the students' previous experience in writing in an imaginative mode in English or in Persian, as some of the participants in the study indicated. The difference could also be attributed to their extensive reading experience in the imaginative mode.

According to the findings of this study, it also seems that the length of compositions is not responsible for an increase or decrease in the number of errors in the students' writing samples.

The Communicative Value of Major Errors

Given the current emphasis on communicative language teaching, researchers and language educators in recent years have placed much emphasis on the gravity of errors produced by the L2 learners as they attempt to speak or write in the second language.

One of the primary objectives of this study was also to discover which errors, particularly which high frequency errors, most impeded the accurate comprehension of the students' writings. To this end, a panel of three native-speaker judges consisting of a grammar teacher (Evaluator 1), an ESL professor (Evaluator 2) and an administrator (Evaluator 3) was hired to
evaluate the effects of the students' errors on their comprehension of the students' intended messages in the thirty-two compositions on a five-point scale from (1) "highly excusable" errors to (5) "highly serious" errors.

First, the grammar teacher evaluated the errors. Based on her judgement, the categories in Table 2 were the ones that most hindered the comprehension of the students' intended messages.
(Table 2)

Error Categories Impeding Communication the Most

<table>
<thead>
<tr>
<th>Rank Order</th>
<th>Error Type</th>
<th>Number of Occurrence</th>
<th>Communicative Value</th>
<th>Index of Grav.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unclear Antecedent</td>
<td>86</td>
<td>0 8 60 11 7</td>
<td>18.33</td>
</tr>
<tr>
<td>2</td>
<td>Conjunctions/Transitions</td>
<td>76</td>
<td>20 27 19 8</td>
<td>2 11.53</td>
</tr>
<tr>
<td>3</td>
<td>Articles</td>
<td>133</td>
<td>124 8 1 0 0</td>
<td>9.53</td>
</tr>
<tr>
<td>4</td>
<td>Incorrect or Confusing Tenses</td>
<td>82</td>
<td>36 35 10 0</td>
<td>1 9.40</td>
</tr>
<tr>
<td>5</td>
<td>Prepositions</td>
<td>86</td>
<td>66 15 5 0 0</td>
<td>7.40</td>
</tr>
<tr>
<td>6</td>
<td>Number</td>
<td>80</td>
<td>63 16 0 0 1</td>
<td>6.67</td>
</tr>
<tr>
<td>7</td>
<td>Adjectives</td>
<td>43</td>
<td>17 13 8 4 1</td>
<td>5.87</td>
</tr>
<tr>
<td>8</td>
<td>Verb Phrases</td>
<td>38</td>
<td>14 14 7 2 1</td>
<td>5.07</td>
</tr>
<tr>
<td>9</td>
<td>Subjects and Predicates</td>
<td>42</td>
<td>20 15 4 3 0</td>
<td>4.93</td>
</tr>
<tr>
<td>10</td>
<td>Pronouns</td>
<td>36</td>
<td>11 15 8 2 0</td>
<td>4.87</td>
</tr>
</tbody>
</table>

A close examination of the data presented in Tables 1 and 2 indicates that, although the error categories in the two tables do not present the same hierarchical order, they do present the same error categories. In other words, it can be concluded that in general the most recurrent errors (the "Major" errors) are also the errors that most impede normal comprehension by the native speaker. More importantly, it can also be concluded that the degree of gravity is not necessarily correlated with the level of frequency because, as clearly demonstrated in Table
3, the majority of the most frequent errors do not establish the same rank order as the most serious errors.

(Table 3)

No One to One Correspondence Between Occurrence and Gravity of Errors

<table>
<thead>
<tr>
<th>Order of Rank</th>
<th>Most Recurrent Errors</th>
<th>Most Serious Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Articles</td>
<td>Unclear Antecedent</td>
</tr>
<tr>
<td>2</td>
<td>Preposition</td>
<td>Conjunctions/Transitions</td>
</tr>
<tr>
<td>3</td>
<td>Unclear Antecedent</td>
<td>Articles</td>
</tr>
<tr>
<td>4</td>
<td>Incorrect and Confusing Tenses</td>
<td>Incorrect and Confusing Tenses</td>
</tr>
<tr>
<td>5</td>
<td>Number</td>
<td>Prepositions</td>
</tr>
<tr>
<td>6</td>
<td>Conjunctions/Transitions</td>
<td>Number</td>
</tr>
<tr>
<td>7</td>
<td>Adjectives</td>
<td>Adjectives</td>
</tr>
<tr>
<td>8</td>
<td>Subjects and Predicates</td>
<td>Verb Phrases</td>
</tr>
<tr>
<td>9</td>
<td>Verb Phrases</td>
<td>Subjects and Predicates</td>
</tr>
<tr>
<td>10</td>
<td>Pronouns</td>
<td>Pronoun</td>
</tr>
</tbody>
</table>

As displayed in the above table, among the ten categories of errors, only three categories--namely those of "Incorrect and Confusing tenses," "adjectives," and the category of mistakes with "Pronouns"--correspond to the same rank order and the remaining seven categories do not.
Discrepancy in the Native Speakers' Judgements of Error Gravity

The literature of modern linguistics is replete with references to the power of the native speaker's linguistic intuition to judge the accuracy and appropriateness of language use. Chomsky (1965), just to use one example from the literature, states: "...There is no way to avoid the traditional assumption that speaker-hearer's linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, linguistic theory, or operational test..." (p. 21). The problem is that when it comes to judging the communicative value of errors, all native speakers of English do not necessarily perceive the learner's errors with the same degree of tolerance. Native speakers' judgements of error gravity are often influenced by, among other factors, their educational backgrounds, formal knowledge of English grammar, the particular dialects they speak, and their professional backgrounds.

In this study a five-way comparison was made to determine how different the grammar teacher and the other two evaluators judge the seriousness of the students' grammatical deviances. A selected sample including 239 of the most frequent and troublesome errors whose communicative values were first determined by the grammar teacher was given to the other two members of the panel for reevaluation. The three evaluations were done on three different occasions so that the evaluators
would not influence one another's judgement. The five-way comparison was conducted with respect to the following questions:

- a) How many times did the grammar teacher and the other two evaluators agree on the value of errors?
- b) How many times did the grammar teacher agree with the ESL professor on the value of errors?
- c) How many times did the grammar teacher agree with the administrator on the value of errors?
- d) How many times did the three of them disagree with one another on the value of errors?
- e) How many times did the ESL professor and the administrator agree with each other on the value of error?

As the data in Table 4 indicates, it was found that only 9.21% of the time the grammar teacher (Evaluator 1) was in harmony with the other two evaluators. 8.37% of the time, the grammar teacher and the ESL professor (Evaluator 2) were consistent with each other; 10.46% of the time the grammar teacher and the administrator (Evaluator 3) agreed with one another; interestingly 40.59% of the time the ESL professor and the administrator assigned the same value to the errors, and finally 31.37% of the time the three evaluators disagreed in their judgements of the errors under study.
It is significant to note that in 39 cases (16.31%) the second and the third evaluator judged the item/phrases which were previously marked as unacceptable by the first evaluator (the grammar teacher) as acceptable (Value "0" in column 4 of Table 4). In general, it seems that there was more agreement on communicative value of errors between the second and the third evaluator (40.59%) than there was between the grammar teacher and the other two evaluators (9.21%). This fact becomes even clearer when it is found that 31.37% of the time the three evaluators disagreed in their judgements of the seriousness of errors.

It was beyond the scope of this study to explore the reasons behind the observed differences in gravity judgement of errors. Perhaps, it would be logical to state that since the grammar teacher always taught English grammar, she generally tended to be...
stricter in her judgement of errors than the other two error assessor.

**Conclusions**

The findings in this study can be summarized as follows:

1- In general, students committed more errors when writing in an expository mode than when writing in an imaginative mode.

2- Variation in the length of writings could not be held responsible for variation in the number of errors in the students' writings.

3- The frequency of error-type occurrence differed in the expository and imaginative writings.

4- As far as communication is concerned, mistakes with unclear antecedents were most impeding. After that came conjunctions/transitions, articles, incorrect tenses, prepositions, number, adjectives, verb phrases, subjects/predicates and pronouns.

5- Although the most recurrent errors and the most serious errors did not present the same hierarchical order, they did present the same error categories.

6- The degree of error gravity was not necessarily correlated to the number of times an error recurred.

7- Native speakers often cannot reach a consensus on the gravity of errors. It seems that native speakers' educational, professional backgrounds, dialects and formal knowledge of English grammar are among the possible factors that influence their judgment of error gravity.
References


