Comparing Native and Nonnative Teachers' Evaluations of Error Seriousness

Norbert Schmitt

Minatogawa Women's College

There is a widely held belief that Japanese teachers of English place much more emphasis on grammatical accuracy in error correction than do their Assistant English Teacher (AET) colleagues. To test the validity of this belief, a survey instrument was designed which asked both groups to evaluate a variety of student composition errors for seriousness. Both groups of teachers were then asked to state the criteria they used in their error judgments. The results showed that the Japanese teachers did indeed tend to judge grammatical errors more harshly than their native-speaking counterparts, and some explicitly used grammatical accuracy as their main criterion in grading papers. Most AETs noted comprehensibility as the primary basis for their judgments. Interestingly, despite their harsher appraisal of grammatical errors, the majority of Japanese teachers also reported using comprehensibility as their most important criterion.

学生の間違いに対する日本人教師とＡＥＴの評価の対比

日本人英語教師はＡＥＴに比べて、間違いの指摘の面で文法の正確さをより強調すると広く信じられているようだ。この考えの妥当性を調べるために、日本人教師とＡＥＴ両グループにどんな間違いがより深刻だと思われるかを尋ね、調査票を作成した。そして両グループの教師にどんな基準で間違いを指摘したかを述べてもらうことにした。事実、結果は日本人教師はネイティブ・スピーカーよりも厳しく文法的誤りを評価していた。また、その中の何人かは作文を審査する時に文法の正確さを最も重要な基準とするとはっきり述べた。これに比し、ほとんどのＡＥＴは、解りやすさを主要因とした。興味深いことに日本人教師は文法面で厳しく作文を審査したにもかかわらず、彼等の大多数が判断の基準はＡＥＴと同じように解りやすさが一番重要な基準だと述べた。

1. Introduction

In a bid by the Japanese government to improve English classes in Japan, an increasing number of native-speaking assistant English teachers (AETs) have been introduced into the Japanese classroom. In the resulting interaction between the AETs and resident Japanese teachers of English, a number of differences in teaching emphases and priorities have surfaced. Possibly one of the most discussed differences has been how to deal with errors. Almost all teachers would agree that there are some errors that need to be corrected. The question becomes, which errors and when? In light of the fact that several

studies have shown that errors do not usually prevent comprehension (Chastain, 1980, 1981; Guntermann, 1978; Olsson, 1973; Piazza, 1980), there seems to be an impression among the AETs that the Japanese teachers place a disproportionately high emphasis on formal accuracy.

This paper will explore whether AET and Japanese teachers do in fact approach errors differently, focusing specifically on how seriously AET and Japanese teachers judge various kinds of written errors. To achieve this, two particular areas need to be addressed: error classification and error evaluation. In order to compare AET and Japanese teachers' judgments of errors in a principled way, it is first necessary to categorize the errors to be judged. Also, any report on error judgment should include an attempt to isolate some of the criteria teachers use when evaluating those errors.

2. Classification of Errors

The most common way errors have been classified is according to categories such as phonology, lexis, semantics, and syntax. These categories are useful in a general discussion, but the tendency of errors to cross category boundaries may limit their use when more precise definitions are desirable. The following example illustrates how the traditional categories can become blurred, even when attempting to define errors limited to a single word.

* He dribed the horse yesterday.

The word "dribed" consists of three simultaneous errors. Should we focus on the misspelling of "b" for "v," the incorrect morphological form "drived" for "drove," the lexical misuse of "drove" for "rode," or a combination of the three? At the sentence or discourse level, errors can become even more convoluted. Clearly, many errors may prove rather complex for these descriptions.

A different approach, called a Surface Strategy Taxonomy, was developed by Dulay, Burt, and Krashen (1982). In it, errors are classified as one of the following: (a) error of omission (necessary item is omitted); (b) error of addition (prohibited item is present); (c) error of substitution (incorrect item is present in place of a correct item); (d) error of misordering (items are in an incorrect order). This system reflects the view that language consists of incremental units strung together serially. It may have serious difficulties classifying errors occurring at the sentence or discourse level in one of the four specific categories (Lennon, 1991).

Burt (1975) made a distinction which recognized that errors may impinge upon more than one linguistic component at one time. She differentiated

between global errors (ones which effect overall sentence organization) and local errors (ones which are limited to a single part of the sentence). She found that global errors tend to seriously hinder communication, while local errors do not.

Lennon (1991) proposed combining the global/local concept with breadth of error to derive a two-component classification system. The "extent" of error is the linguistic unit which the error permeates. This might be a morpheme, word, phrase, clause, sentence, or discourse. Often, the error cannot be discerned by looking only at the unit in which it exists; the listener/reader may have to check the sentence or discourse context to identify it. How much language the reader/listener must examine to determine if an error has occurred is referred to as the "domain" of the error.

From these descriptions, we can develop a list of possible error categorizations (Table 1). Domain will always be at an equal or higher rank than extent, never at a lower rank, because the amount of language necessary to determine the complete error can never be less than the error itself.

Table 1
Extent/Domain Categories

	Extent						
Domain	Morpheme	Word	Phrase	Clause	Sentence	Discourse	Extralingual
Word	M/W	W/W					
Phrase	M/P	W/P	P/P				
Clause	M/C	W/C	P/C	C/C			
Sentence	M/S	W/S	P/S	C/S	S/S		
Discourse	M/D	W/D	P/D	C/D	S/D	D/D	
Extralingual	M/E	W/E	P/E	C/E	S/E	D/E	E/E

A few examples will help to illustrate the extent/domain concept. The following error examples come from student book summaries of the novel *The Outsiders* (Hinton, 1989) and were used in the survey instrument. In the error "hitted," the morpheme "ed" is the problem, and that fact is discernible by looking at the word, therefore it is a Morpheme/Word error. On the other hand, in the sentence "The church was safety," the morpheme "ty" is incorrect, but we have to look at the complete sentence to ascertain that, making it a Morpheme/Sentence error. Spelling errors are usually Word/Word errors as the word is incorrect, and that is usually obvious without looking further afield, that is, "fighting" misspelled as "fithing." Let us try a more global example at the discourse level. "Ponyboy and Johnny like watching movies.

He went to watch a movie." The second sentence is fine if it stands alone, but in a discourse context the pronoun "he" is erroneous since the previous referent is ambiguous. This error is classified as Word/Discourse. A person would be committing a Word/Extralingual Context error by saying "They went south," but pointing or gesturing to the north.

This system may prove a principled way to describe errors since it includes the notion that errors often reach beyond their immediate component into the surrounding linguistic environment.

3. Evaluation of Errors

When evaluating errors it might be assumed that comprehensibility is a prime consideration. However, Chastain (1980) found that his subjects rated about 50 percent of erroneous forms as comprehensible, but unacceptable. From this data, he concluded that many errors are considered unacceptable for reasons other than lack of comprehension. Some possible reasons appear in Ludwig's (1982) survey of native-speaker error judgment studies. In it, she identifies five recurring factors affecting judgments of correctness: comprehension, irritation, acceptability, communication strategies, and personality. There are conflicting reports on the relative tolerance of errors by native and nonnative speaking subjects. The prevailing view seems to be that nonnatives are less tolerant of nonnative errors than native speakers (Galloway, 1980; Sheory, 1985).

For example, Santos (1988) found more severe judgments of composition errors by nonnative speaking professors than by native-speaking ones. In fact, Sheory (1985) concluded that tolerance of errors increases as language proficiency increases.

By way of contrast, Birdsong and Kassen (1988) concluded that as people increase in language proficiency they become harsher in their judgments of error seriousness. They found that French-speaking teachers of French were harsher judges than English-speaking ones, and in general, teachers judged errors more harshly than students. A study by Ervin (1978) reported similar results in a Russian context. A recent study by Kobayashi (1992) found that native English speakers were stricter about grammaticality when judging ESL compositions than were native Japanese speakers.

In order to gather empirical information on this unresolved subject in a Japanese context, and to explore assumptions made about AET and Japanese teachers' evaluations of errors, a survey study was designed focusing on the following questions:

1. Do Japanese teachers judge errors more harshly than AETs?

2. Which categories of errors are judged more harshly by each group?

On the methodological side, a decision was made to utilize Lennon's (1991) Extent/Domain distinction as the way to classify errors. This raised a third research question:

3. Would the Extent/Domain classification system prove to be a viable way to describe errors in this study?

4. Procedure

From the 27 possible error categories (Table 1), ten were selected which would allow a representative and manageable sampling of errors along the global/local continuum. The phrase and clause categories were collapsed into a subsentential (SS) category. The extralingual classifications, being more appropriate to spoken communication, were ignored. The final ten categories were Morpheme/Word (M/W), Morpheme/Sentence (M/S), Word/Word (W/W), Word/Subsentential (W/SS), Word/Sentence (W/S), Word/Discourse (W/D), Subsentential/Subsentential (SS/SS), Subsentential/Sentence (SS/S), Sentence/Sentence (S/S), and Sentence/Discourse (S/D).

Fourteen book summaries were collected from students enrolled in a pre-college intensive English program at Temple University in Osaka, Japan. Summaries of the same book were taken so that sentences containing individual errors could eventually be formed into a cohesive discourse. Approximately 60 error-bearing sentences were extracted from the student summaries and presented to four native-speaking raters who had been trained in the simplified Extent/Domain error classification system. Sentences in which the error classification was agreed upon by at least three of the four raters were put into a pool from which three examples for each of the ten categories were chosen.[1]

The final survey instrument (see Appendix) was created by arranging the 30 error-bearing sentences into sequence and adding supplementary contextual information in brackets to make the resulting summary cohesive. A seven-point Likert scale was attached to each erroneous sentence. The respondents were asked to indicate the seriousness of the error contained in each sentence by circling a value on the Likert scale and then, when finished, to answer the following question, "On what basis did you judge the seriousness of the errors?"[2]

Thirty-eight surveys were collected, twenty from AETs and eighteen from Japanese teachers. Most respondents were males teaching at the high school or college level. The average teaching experience was six years for AETs and 12.1 years for Japanese teachers.

The ratings for the three sentences in each error category were averaged to achieve one rating per respondent per error category. A mean could then be derived for each of the ten categories. Next, an analysis of variance procedure (ANOVA) was used to establish whether any error categories contained statistically significant differences between the teacher groups.

5. Results

Each error category contained from 51 to 60 ratings (three erroneous sentences per category x 17-20 respondents). Table 2 illustrates the mean of these ratings for each category.

Table 2
Error Category Means

	Japanese (n = 18)		AET (n = 20)	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Morpheme/Word	3.64	1.03	2.80	.96
Morpheme/Sentence	3.83	.88	2.28	.88
Word/Word	3.07	1.20	1.83	.74
Word/Subsentential	3.94	1.10	3.38	1.04
Word/Sentence	4.01	.78	3.61	.90
Word/Discourse	3.44	1.30	3.98	.87
Subsentential/Subsentential	4.08	.96	3.00	.81
Subsentential/Sentence	4.23	.97	3.87	.96
Sentence/Sentence*	3.93	1.16	4.50	.98
Sentence/Discourse	2.60	1.06	2.54	1.59

*Only 17 subjects.

The mean ($\bar{X}$) of the Japanese ratings was higher than the mean AET rating for every category of error except Word/Discourse and Sentence/Sentence. Table 3, illustrating the ANOVA analysis results, shows that the difference was significant in four of the categories: Morpheme/Word, Morpheme/Sentence, Word/Word, and Subsentence/Subsentence.

6. Discussion

An examination of the means in Table 2 supports the position that nonnative teachers are harsher on errors than native teachers, at least when dealing with story summaries. In eight out of 10 categories, Japanese teachers judged the errors as being more serious than did the AET teachers. The difference between the average of the Japanese means (3.79) and the AET means (3.25) is quite striking, although only four of the 10 comparisons are significant.

Table 3
Univariate *F*-Tests With (1,35) D.F.

Variable	Hypothesis SS	Error SS	Hypothesis MS	Error MS	*F*	Significance of *F*
M/W	6.74	37.33	6.74	1.06	6.32	.017*
M/S	23.49	28.93	23.49	.82	28.41	.000*
W/W	16.57	33.69	16.57	.96	17.22	.000*
W/SS	2.44	42.18	2.44	1.20	2.02	.163
W/S	1.64	27.36	1.64	.78	2.09	.156
W/D	2.79	45.38	2.79	1.29	2.15	.151
SS/SS	11.47	29.32	11.47	.83	13.70	.001*
SS/S	1.30	35.24	1.30	1.00	1.29	.263
S/S	3.28	42.28	3.28	1.20	2.72	.108
S/D	.08	70.71	.08	2.02	.04	.836

* $p < .05$

However, the four significant figures do point in a similar direction. The significant categories are mainly clustered at the "local" end of the error hierarchy. It is here that the smaller components of language (morphemes, words) and their corresponding rules of use are prominent. Morpheme/Word errors typically consist of the incorrect use of verbal inflection, such as *-ing* or *-ed*. Morpheme/Sentence errors are likely to be lack of agreement between subject and verb, or misuse of the inflections which change cognates into different parts of speech. Word/Word errors are usually spelling mistakes. These are all areas normally associated with grammatical accuracy. In these categories at least, Japanese teachers do grade more severely, and seem to be more concerned with formal accuracy than their native-speaking counterparts.

At the other end of the scale, Sentence/Discourse errors, (where the sentence itself is correct, but is misplaced in discourse) were consistently more difficult to evaluate for both teacher groups. There were respondents from both groups who saw the confusion these sentences caused in the story flow and rated them severely; conversely, other respondents from both groups rated these same sentences as either correct or not serious. This led to very inconsistent results within each group. A possible explanation for this lies in the way the survey instrument was constructed. Although respondents were explicitly instructed that some sentences might not fit well with the rest of the composition, it is likely that some viewed the sentences attached to the Likert scales as discrete, isolated entities, instead of as part of a cohesive summary. This would explain the correct/not serious judgments. Additionally, during the initial error categorization stage, the four native-speaking raters also had

difficulties coming to agreement on errors in this category. Clearly, the problem of categorizing the most global types of errors has not been resolved.

The respondents' comments on how they judged the seriousness of errors were interesting on several counts. They seemed to indicate that comprehensibility was the overriding criterion for judging the gravity of error. Out of the 20 AET responses, 19 included some indication that obstruction of meaning was a primary factor when dealing with errors. In addition, no AET indicated that grammatical correctness was their main basis for error evaluation. These responses were expected, but somewhat surprisingly, most of the Japanese teachers stated similar views. Ten out of the 14 Japanese teachers who responded to the question also indicated that the ability to transfer meaning was more important than grammatical accuracy, although three teachers specifically mentioned their preference for grammar. Many expressed the feeling that the Japanese school system places too much stress on grammatical accuracy, making it difficult for students to improve in communicative fluency.

Comparing these criteria comments with the error seriousness ratings highlights a discrepancy between professed beliefs and actual error correction practice. One possible reason for this discrepancy is the tendency for teachers to judge students' work according to language aspects the teacher knows best. Following this, English teachers who emerge from the Japanese school system grammatically competent but lacking in confidence to actually use English in a meaningful way, would stress grammar in their grading, even if they are aware of the importance of meaning. Speculation aside, the error seriousness criteria comments can be taken as an indication that although Japanese teachers evaluate formal errors more severely than native speakers, most are also very conscious of the importance of comprehensibility.

Another recurring point concerns "mental effort" as a criterion for judging the severity of errors. Four respondents stated that the amount of time or number of readings necessary to understand the meaning of a sentence was the primary basis they used in evaluation. This suggests a possible direction for new research: Can time required to make an error evaluation be used as a measure of (a) comprehensibility and (b) error seriousness?

As to the viability of the Extent/Discourse classification system, there are signs that it is useful. The survey included a wide variety of learner errors, some of them quite complex. The Extent/Discourse system seemed better able to describe this variety than any of the other systems discussed. It still has serious difficulties describing the most global error areas, but this appears to

be a universal weakness. It may be that broad global errors are simply too complex to be easily fitted into convenient categories.

7. Summary

The results of this study indicate that Japanese teachers put a greater emphasis on formal accuracy than AET teachers. They judge grammatical errors more gravely overall than their native-speaking counterparts, and a proportion of them explicitly cited formal accuracy as their prime criterion in evaluating errors. However, since most Japanese teachers indicated comprehensibility as the most important measure for assessing errors, we may see a shift away from this formal emphasis in the future, especially if external factors, particularly modifications in the college entrance examinations, are conducive to change.

Norbert Schmitt is a lecturer at Minatogawa Women's College. His main research interest is second language vocabulary acquisition. He is currently researching the vocabulary learning strategies of Japanese students.

Notes

[1] Because there were no Sentence/Discourse errors (in which a sentence is grammatically correct but out of place in the discourse) in the student summaries, these sentences were contrived.

[2] Khalil (1985) stresses the need for authentic, contextualized language data in error studies. To obtain a variety of error types, it was necessary to use errors from several students. However, care was taken to contextualize the errors by embedding them in a single discourse. Also, respondents read a synopsis of the story before they began rating the errors in the summary.

References

Birdsong, D. & Kassen M. A. (1988). Teachers' and students' evaluations of foreign language errors: A meeting of the minds? *Modern Language Journal, 72,* 1-12.

Burt, M. (1975). Error analysis in the adult EFL classroom. *TESOL Quarterly, 9,* 53-63.

Chastain, K. (1980). Native speaker reaction to instructor-identified student second-language errors. *Modern Language Journal, 64,* 210-215.

Chastain, K. (1981). Native speaker evaluation of student composition errors. *Modern Language Journal, 65,* 288-294.

Dulay, H., Burt M., & Krashen, S. (1982). *Language two.* New York: Oxford University Press.

Ervin, G. (1978). A study of the use and acceptability of target language communication strategies employed by American students of Russian. *Dissertation Abstracts International, 38,* 6579A-80A.

Galloway, V. (1980). Perception of the communication effects of American students of Spanish. *Modern Language Journal, 64*, 428-433.

Guntermann, G. (1978). A study of the frequency and communicative effects of errors in Spanish. *Modern Language Journal, 62*, 249-253.

Hinton, S. E. (1989). *The Outsiders*. New York; Dell.

Khalil, A. (1985). Communicative error evaluation: Native speakers' evaluation and interpretation of writer errors of oral EFL learners. *TESOL Quarterly, 19*, 335-351.

Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly, 26*, 81-112.

Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics, 12*, 180-195.

Ludwig, J. (1982). Native speaker judgments of second language learners' efforts at communication: A review. *Modern Language Journal, 66*, 274-283.

Olsson, M. (1973). The effects of different types of errors in the communicative situation. In Jan Svartvik (Ed.), *Errata: Papers in Error Analysis*. (pp. 153-160). Lund, Sweden: Gleerup.

Piazza, L. (1980). French tolerance for grammatical errors made by Americans. *Modern Language Journal, 64*, 422-427.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*, 69-90.

Sheory, R. (1985, March). *Goof gravity in ESL: Native vs. nonnative perceptions*. Paper presented at the Nineteenth Annual TESOL Convention, New York, New York.

Appendix

Error-Bearing Sentences Extracted from Student Summaries of *The Outsiders* (Hinton, 1989).

[Continuity is maintained by author's summary in brackets.]

HOW SERIOUS ARE THE MISTAKES IN THE FOLLOWING SENTENCES?

Indicate your opinion by circling one number on each scale.

CIRCLE ONE NUMBER: Not at all serious Very serious

1 2 3 4 5 6 7

[This 1 to 7 scale appeared to the right of each error-bearing sentence.]

The character in this story are Darry, Sodapop, Ponyboy, Dally, Johnny, and Two-Bit.

He is one main character, Ponyboy.

Ponyboy is the youngest brother in his family, but this doesn't mean that has parents.

His parents died in an accident sevral years ago.

[Ponyboy's older brothers are Sodapop and Darry.]

I love Soda better than Darry.

[They belong to a gang called the Greasers.]

They have companions having group consciousness like a gang each other.

The Greasers are poor, but they who are interested in many things are always searching for something to satisfy them but never find it.
[The Greasers have many problems]
They have some problems except money problems.
[The rival gang is the rich kid gang called the Socs.]
The Greasers have long hair. On the other hand, the Socs have cutted hair. One night, his brother told him that he was going to marry Sandy.
[Johnny is Ponyboy's best friend in the Greasers gang.]
Ponyboy and Johnny like watching movies. He went to watch a movie.
[There was a fight by the theater]
They saw fithing between the Socs and a Mexican man.
[Two other Greaser friends, Dally and Two-Bit, met them at the theater.]
Dally has the strongest and meanest of the gang.
They met two girls there whom they are Socs girls.
[Cherry is one of the Soc girls.]
Dally talked to Cherry, but she didn't have interest to Dally.
[Dally became angry with Cherry.]
Johnny's character is not brisk, but he told Dally to leave.
[After Dally left, Ponyboy, Johnny, and Two-Bit talked with the girls. Later, the girls wanted to go home, but they didn't have a car.]
So Two-Bit finally spoke them into letting him drive them home in his car.
After he dropped off the girls, Two-Bit went to home.
They always liked doing their favorite things.
[After talking with Johnny for awhile in the park, Ponyboy hurried home because it was very late.]
When he returned home, his older brother Darry got angree.
[Darry criticized Ponyboy for staying out too late.]
Darry blamed Ponyboy had been out too late.
Darry hitted him.
[Ponyboy ran away from his house. He went back to the park. Johnny was still there, so they started talking again. Meanwhile, a drunken Soc gang drove to the park to attack Ponyboy and Johnny.]
They were drunking.
Johnny was afraid of the Socs, the reason he was attacked by them before.
[The Soc gang held Ponyboy underwater in a fountain. Afraid that Ponyboy would drown, Johnny stabbed the leader of the Socs with a knife.]
Johnny killed Ponyboy.
[When they saw their leader dead, the other Socs ran away. Johnny pulled Ponyboy out of the fountain.]
Ponyboy was okay though looked so deadly.
Johnny was scared because he kills someone.
[Fearing the police, they left town.]
They got on the train at this night.
They went to the church on the hill.
The church was safety.