Analyzing Interlanguage of Non-English Majors by Proposed Semantic Criteria

DING Yuanrong & SUI Mingcai
Northeast Normal University

Abstract
The research conducted here is an attempt in the analysis of interlanguage from semantic perspective. The study is based on the erroneous expressions in compositions of 118 college non-English major freshmen randomly selected from an independent college in Jilin province. In the process of error analysis, five new semantic criteria, i.e., whether meaning can be figured out, whether meaning is acceptable, whether meaning is ambiguous, whether meaning is appropriate, and whether meaning is well-formed, have been proposed in terms of meaning—the very core of any language. The result of the research suggests that the five semantic criteria should be more convincing in the analysis of learners’ language system in that they may compensate for deficiency of traditional analytical criteria which focus too much on “forms”.

Key words: semantic criteria; erroneous expressions; interlanguage; non-English major freshmen

1. Introduction
Interlanguage (IL), the unique language system of learners, has been focused intensively with the publication of Corder’s *The Significance of Learner’s Errors* (1967), Selinker’s *Interlanguage* (1972), and his *Rediscovering Interlanguage* (1992), because the learner errors are no longer viewed as a failure in the process of learning and the learner language is believed to have their own characteristics besides the characteristics both of the native language (NL) and the target language (TL).

In the following decades, researchers have probed into the nature of the theory more thoroughly and more systematically. Ellis (1999) believes that learner’s language is permeable, dynamic and systematic. Brown (2002) concludes that learners are
indeed creatively operating on a second language—constructing, either consciously or subconsciously, a system for understanding and producing utterances in the language. He also argues that learners are processing language on the basis of knowledge of their own interlanguage, which, as a system lying between two languages, ought not to have the value judgments of either language placed upon it. The researchers became concerned more about the continuum as a whole, not simply the dismembered linguistic features in contrast with the native language or the target language.

Consequently, the focus of research has been gradually converted into the study of the sources of learners’ errors as well as inner mechanism of IL itself. Some researchers (Cook, 1993; Larsen-freeman, 1991) made efforts to develop the practical value of the theory by further studying its developmental sequences; some were devoted to the study of particular IL features such as variability or fossilization (Ellis, 1985; Gregg, 1990; etc.). Methodologically, more focus has been converted into the identification and classification of errors, separating error description from error explanation (Ellis, 2000) and taxonomies of learner’s errors (Dulay, Burt & Krashen, in Keith Johnson & Helen Johnson, 2001).

Though certain achievements have been obtained, it is a pity that most former research on interlanguage overvalued the significance of form but ignored meaning, the very core of any piece of utterance and interpretation, which serves as an equally if not more essential component of learners’ language system. Even though some researchers may notice the key role of semantic aspects, their research did not escape the norm of discrete studies which failed to study meaning and form as a unity. The consequence is that researchers went too far on the path of forms and the hairsplitting scrutiny in form in place of meaning can not retrieve the misleading fate which only makes the nature of errors more veiled and undistinguishable. Just like mere shape without soul would never accomplish a masterpiece, mere form without meaning can never manifest the true properties of interlanguage.

Based on the concerns of reclassifying and analyzing interlanguage in terms of meaning, this paper aims to explore new semantic criteria in light of enlightenments from former theories and methodologies. The classification procedure of the new criteria will also be illustrated with examples and verified with data to show the detailed application of such criteria.

2. Proposal of Five New Semantic Criteria

2.1 Enlightenment from Former Research

Generally speaking, interlanguage theory and semantics have been considered to be two independent areas of research. In the field of semantics, researchers have mainly focused on making sense of meanings in a particular language, while the main concern of interlanguage is the study of the continuum between learners’ native language and the target language. The fact is that when interlanguage is treated as a unique language system, the value of meaning as an access to the system stands out.

According to former research, some researchers have detected the significance of meaning in terms of the effect of learner errors in communication. One famous inquiry is from Littlewood (2000), who has discussed the effects of second language learners’ speech in communication with native speakers of the language and listed several aspects that affect
the seriousness of errors in communication, which is suggestive to the establishment of the semantic criteria which may rate the seriousness of errors that hinder proper comprehension.

Another beneficial inspiration is from Gass & Selinker (2001): “Crucial to the success of any conversation is the ability to understand and to be understood.” Comprehensibility serves here as the bottom line for a basic comprehension. Relevant inquiries are also closely connected with EG (Error Gravity). James (2001) suggested that “comprehensibility” be used as one of the criteria of EG to cover both the success of the text producer and that of the text receiver.

In The Study of Second Language Acquisition (1994), Ellis has raised several questions in identifying errors. One question concerns whether the analysis should examine only deviations in correctness or also deviations in appropriateness. The former involves rules of usage and the latter involves rules of language use. Both contribute to the identification of errors in the present research.

Corder’s research (1981) also came into our view. He acknowledged the importance of “interpretation” of meaning and proposed an elaborate procedure for identifying errors, providing a clear distinction between overt and covert errors. The procedure can be seen in Figure 1:

\[\text{Figure 1. The Procedure for Description of Idiosyncratic Dialects (Corder, 1981: 23)}\]
As the figure shows, learners’ language has been distinguished according to whether the meaning makes sense in certain circumstance. However, in the procedure, further distinction concerning interpretation is rather rough and obscure. Corder’s procedure hasn’t escaped from the traditional norms of comparison either from CA or from EA perspective, which fails to analyze from a more detailed semantic angle.

2.2 Classification Procedure of New Semantic Criteria
Based on former research and semantic characteristics of the errors, five semantic criteria for error classification have been proposed in the light of whether meaning can be figured out, whether meaning is acceptable, whether meaning is ambiguous, whether meaning is appropriate, and whether meaning is well-formed. The classification procedure of the five semantic criteria can be demonstrated in Figure 2 as the following:

![Figure 2. Criteria for Error Classification in terms of Meaning](image)

Figure 2 shows how the five criteria work in the process of classifying freshmen’s errors in their compositions. In the process, we would first see whether the meaning could be figured out. If the meaning of an erroneous expression could not be figured out, the expression belongs to the category of incomprehensible expressions; otherwise it could be further classified according to the following criteria. For example, “Smoking, should to be forbided.” is not correct. If it is analyzed in
terms of the five criteria, the first question to be asked is that “Can it be figured out?” —Yes. Then, “Is the meaning acceptable?” —Yes. It is not ambiguous and appropriate. However, it is not well-formed. Therefore, it is a badly-formed expression. The teacher’s next job might be finding out in which way he/she could possibly inform the student of the correct form if necessary.

3. Research Design

The research consists of three parts. Firstly, subjects and materials have to be chosen. Secondly, authentic interlanguage data are to be collected. Finally, the results of data analysis and the advantages of new criteria are to be further discussed in the end of the paper.

3.1 Subjects and Materials

To obtain necessary interlanguage data, the subjects are firstly randomly selected from the non-English majors in an independent college—College of Humanities and Sciences of Northeast Normal University. They are freshmen from three classes of different departments, namely, Chinese Department, Human Resources (HR) Department and the Department of Business Administration (BA).

Then, the English level and personal data of the subjects have been investigated by a questionnaire (see Appendix I), from which their average score in the entrance examination (85.5 points) has been obtained. The fact showed that the average score is below the pass score (90 points, 150 points in total), indicating in general that the subjects are of an intermediate or relatively lower intermediate English level.

After that, the subjects have been assigned a composition entitled “a letter to a friend / a teacher / my parents, etc.”, which is proposed due to its adaptation to different levels of the freshmen in question. It can be a letter to anyone they want to address. These freshmen have been informed that the written task needed to be accomplished in class within 30 minutes and the body of the letter was not expected to be less than 80 words (see Appendix II). With 123 non-English major freshmen taking part in the research, 118 valid samples were successfully collected.

3.2 Data Collection

As it is, errors in interlanguage are not easy to identify and calculate due to the complexity of their sources. To catch a general view of the interlanguage and the features of the system, an effort was made by observing certain principles and methods.

A. The subject was inquired about his/her intention when we came across some difficulty in interpreting the intended meaning in the expressions or in doubting why such errors occurred. The method is believed effective in understanding the inner thought of the subjects since without the enquiries, we find it difficult to learn about the sources or the meaning of the errors.

B. In the process of data analysis, if the errors with an identical exterior result from the same reason in a text, it could only be calculated once. Otherwise, it could be regarded as different errors owing to different reasons. This principle has also been mentioned in
some other research (Ellis, 1994).

There are 16,392 words with 1,166 errors in all. Detailed frequencies of error occurrence of each class are demonstrated in Table 1.

**Table 1. Error Frequency in Three Classes**

<table>
<thead>
<tr>
<th>Class 1 (Students from Chinese department)</th>
<th>Class 2 (Students from HR department)</th>
<th>Class 3 (Students from BA department)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td>Total error number</td>
<td>314</td>
<td>443</td>
</tr>
<tr>
<td>Total word number</td>
<td>4384</td>
<td>6290</td>
</tr>
<tr>
<td>Error frequency</td>
<td>0.0716</td>
<td>0.0704</td>
</tr>
</tbody>
</table>

The figures illustrated above provide us with a general view of the error frequencies of the non-English major freshmen in study. The average error frequency of Class 1 is 0.0716; the average error frequency of Class 2 is 0.0704; and the average error frequency of Class 3 is 0.0715. Therefore the average error frequency for the three classes is 0.0711, which may suggest a general status of errors in the interlanguage of the compositions.

4. Results and Discussion

4.1 Applying the Proposed Semantic Criteria

As illustrated in Figure 2, the analysis process is a step-by-step procedure. To better present the classification procedure, a more detailed list of examples are demonstrated in the following, with the stereotypical errors underlined and some of the possible improved forms suggested in the parentheses.

(1) **Whether meaning can be figured out**

A. The meaning is not figured out.

   e. g. I west for you what our meet on the university all the time. (I wish we could meet on the university someday.)

   You attended yourself, I will … (Take care of yourself, …)

B. The meaning can be figured out.

   All the other errors except errors in (1) A.

(2) **Whether meaning is acceptable**

A. The meaning is not acceptable.

   e. g. I hope will have a long time to go home, but I have many classes in a day all the time.

   (some time left over at home)

   Because the wind is very happy and the sands flys in the sky. (heavy)

B. The meaning is acceptable

   All the other errors except errors in (1) A and (2) A.

   e. g. Yesterday evening you have finished speaking an excellent speech, Your smiles on your face from begin to end. (Yesterday evening you gave an excellent speech, with smile
on the face from beginning to end.

In last winter, I have dreamed the snow what is my first time in my life. (Last winter, for the first time in my life, I saw the snow in a dream.)

(3) **Whether meaning is ambiguous**

A. The meaning is ambiguous.

  e. g. How are you still at home? (How are you? / Are you still at home?)

  …, I *could* go to Beijing *as possible as I can*. ( will; as soon as possible/ if possible )

B. The meaning is not ambiguous.

All the other errors except errors in (1) A, (2) A and (3) A.

  e. g. But we is the best friend as the brother.

    The students *is* friendly for me. (*are; to *)

(4) **Whether meaning is appropriate**

A. The meaning is not appropriate.

  e. g. I *can’t say how much*, let’s stop! ( I won’t say much, … )

    They love me *as long as you love me*. ( just as )

B. The meaning is appropriate.

All the other errors except errors in (1) A, (2) A, (3) A and (4) A.

  e. g. I’m fine, the school-life *very comfortable, too.*

    Classmates here are all nice to me. ( All the classmates here are nice to me. )

(5) **Whether meaning is well-formed**

A. The meaning is not well-formed.

  e. g. Smoking, should *to be forbid*. 

    I have’nt wrote…

B. The meaning is well-formed.

Good Expression/No Error.

As can be seen, the first four error types are related more or less with the proper interpretation of the intended meaning, while the last error type is simply resulted from bad form, which seems to shed the lightest influence in meaning interpretation.

4.2 Results

To verify the feasibility of the proposed criteria, the statistical result of data analysis, namely, the number and the percentage of new error types have been presented in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Incomprehensible expressions</th>
<th>Wrong expressions</th>
<th>Ambiguous expressions</th>
<th>Inappropriate expressions</th>
<th>Badly-formed expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Number</td>
<td>12</td>
<td>21</td>
<td>26</td>
<td>170</td>
<td>937</td>
</tr>
<tr>
<td>Error Percentage</td>
<td>1.03%</td>
<td>1.80%</td>
<td>2.23%</td>
<td>14.58%</td>
<td>80.36%</td>
</tr>
</tbody>
</table>

In this table, inappropriate expressions and badly-formed expressions, which separately take up 14.58% and 80.36% of the total erroneous expressions, account for an
overwhelming majority of the errors in the compositions, indicating a clear meaning intention but an inappropriate or incorrect convey, which may indicate the insufficiency of target language intake and misrepresentation of meaning in the process of construction.

Based on a combination of both qualitative and quantitative analyses, the research conducted here may have some theoretical and pedagogical implications and would to some extent facilitate English learning in composition.

In the first place, we may penetrate the hierarchical difficulty these non-English major freshmen may come across in their English learning since the criteria are proposed according to the seriousness these erroneous expressions hinder direct comprehension. For example, the incomprehensible expressions, which occur especially when the task level assigned is much higher than the students’ English level, may suggest the most serious difficulty in both interpretation and expression of the intended meaning. In firm support of the learning-centered theory (Sui & Wei, 1998), this finding may urge the EFL teacher to design task and treat errors in a more discerning and proper way.

The study reveals that the output of EFL learning is not a replica of form input, but a process involving choices among alternatives with vast semantic, psychological and cognitive basis. Dai & Shu (1999) have noticed that: “In authentic communicative situations, learners would pay more attention to expressing meaning.” As mentioned previously, erroneous expressions occur when the students fail to find ways to express their intended meaning. Besides the ignorance of rules and cultures of the target language, simplistic imitation of Chinese habitual expressions, wrong perception of the Chinese language and “unclear thinking” also contribute to such errors. This finding matches well with some of the former research. Slobin (1993) has argued that adult learner retain a scale of conceptual complexity, based on their own cognitive development, and at first search the target language for the grammatical marking of those notions which represent some primordial core of basicness or simplicity. Robinson (2005) also proposes the cognition hypothesis which claims that pedagogic tasks be sequenced for learners on the basis of their cognitive complexity.

In the past, the occurrence of errors is believed to be due to the transfer of the native language or bad mastery of the target language since language transfer and overgeneralization have been believed to be two major sources of errors. In contrast to traditional views, the findings show that badly-formed expressions, which take up 80.36% of the total erroneous expressions, stem from insufficiency of target language intake and failure to establish connection between language input and output since more learner’s factors, the production mechanism as well as interpretation mechanism have been intensively focused. Therefore, the priority to the improvement of students’ English level is to help them build effective mappings from meaning to its corresponding forms as well as enlarge their English intake.

4.3 Discussion
The reanalyzed data present a totally different perspective in error classification since they focus on meaning and the relationship between meaning and form. As it is, the five semantic criteria attempt to state the erroneous expressions from the perspective
of meaning, with certain structural rules observed in the interpretation of the intended meanings, since meaning and form, as two separate aspects of language, are indispensable to each other. As Schlesinger (1995) has put it, language has rules stating how meanings are expressed by linguistic constructions; or, put differently, grammars describe (often complex and indirect) mappings from cognitive space into syntactic structures.

Interestingly, in the process of the analysis, we found that besides the consideration of semantic difficulties which hinder proper expression, some errors from the former four criteria also result from the inappropriateness of forms. Since form might convey specific meaning, wrong form might also result in wrong interpretations in meaning. The complexity of the error sources could to some extent account for the seriousness of such errors.

It is worth mentioning that the advantages of the five semantic criteria stand out since they are competent in the classifying and interpreting the erroneous expressions in compositions of the non-English major freshmen. In contrast with traditional taxonomies such as the linguistic classification from Richard et al. (2000) and Selinker’s five central processes (1972), four main advantages emerged.

Firstly, different from any traditional analytical criteria, the five semantic criteria may provide a brand-new perspective focusing on meaning in interlanguage research. The new procedure regarded form and meaning as a unity in classifying errors, which made the study probe more easily into the nature of interlanguage, while traditional taxonomies focused more on form instead of meaning or totally ignored meaning, which may lead to irreversible intricacy on trifles and on the contrary inhibits good description and classification of errors.

Secondly, the new criteria revealed the production mechanism as well as the interpretation mechanism and provided a hierarchical classification standard; therefore, they may reveal the seriousness of errors which hinder interpretation and meanwhile reflect different levels of difficulty in meaning expression. In contrast, some traditional criteria may fail to notice either mechanism or both and can not distinguish the seriousness of the semantic difficulties. As a result, traditional methodologies would conceal the difference in effects within the same category, though they might clearly distinguish different error features in the classification. For instance, errors in verb phrase and in noun phrase may be rated quite differently in the effect of communication; however, they may be put into the same linguistic category without any distinction according to traditional methodology such as the linguistic classification from Richard et al. (2000).

Thirdly, as suggested by Johnson & Johnson (2001): “For a taxonomy to be effective it should be easy to classify items uniquely under one category or another.” The new criteria provided priority sequence in the classification procedure. The erroneous expression may be analyzed in sequence, i.e., from the first criterion to the last. The advantage is that such sequences may succeed in avoiding such dilemmas like determining why an error should be classified in one way rather than another. For example, according to traditional criteria, “We often going to…” may have different descriptions and explanations. Can we summarize it as overtraining or overgeneralization or a poor mastery of present tense of verbs, or—just a slip? Which should be considered prior than the others? We can not find answers from traditional taxonomies since they often confuse or mislead in such cases.
Fourthly, the new criteria can interpret most errors traditional criteria fail to explain. For instances, the first criterion—whether meaning can be figured out, can be applied pinpoint to incomprehensible utterances like “I west for you what our meet on the university all the time”, while it might be extremely difficult for traditional methodologies to categorize and interpret such errors when the intended meanings are not clear to the reader. By applying the new criteria, students’ avoidance of difficult structures may also be detected since too much communication strategies may add difficulty in understanding. The phenomenon has been explained previously by Littlewood (2000): “Presumably this is not directly a result of the communication strategies, but because the strategies reflect the fact that the learner has special problem in expressing himself.”

In the classification and analysis of the errors in question, two underlining maxims are also observed as follows.

Firstly, the errors are viewed from non-native English-speaking teacher’s maxim. Therefore, some particular transferred errors might be figured out even though some native English speakers might feel it awkward or they couldn’t even understand it. In the sentence: “It understand me that the beauty is creatived distance”, we could figure out that the student actually means that “this made me understand the Chinese proverb ‘geniality would emerge after separation’ (In Chinese, ‘Jùlí chūnhěng měi’).” We believe that it is an “inappropriate expression” whereas a native English speaker might put it in the category of “wrong expression” for he/she might not think the meaning is acceptable.

Secondly, the first criterion is subjective and subject to variation. One can be bewildered sometimes in the figuring out of the meaning. While the other four criteria—whether meaning is acceptable, whether meaning is ambiguous, whether meaning is appropriate and whether meaning is well-formed, are not subjective or not up to the reader's feeling or judgment but objective or up to the fact. It is determined by whether the error indicates the intended meaning.

5. Concluding Remarks

With the purpose of a more effective analysis of learners’ language system, five semantic criteria have been proposed and applied in the research. The advantages and the pedagogical implications of the new criteria emerged after a detailed illustration of the classification procedure. In contrast with former studies which mainly focus on the comparisons between the interrelated systems, the semantic approach shows how meaning functions in interlanguage itself. It urges us to think over the mechanism of production as well as interpretation. Although these five semantic criteria might be rough and needs more effort in scrutiny in the study of interlanguage, they reveal the essence of interlanguage analysis and deserve further study.
References


戴炜栋、束定芳, 1999，对比分析、错误分析和中介语研究中的若干问题——外语教学理论研究之二，《外国语》第5期。

隋铭才、魏立明，1998，学习中心论：问题与思考，《外国语》第4期。
Appendix

I. Questionnaire for Their Personal Information

Name: ______________  Age: _____
Hometown: ______________
Which school did you graduate from?
____________________
When did you start learning English?
____________________
Your English score in the entrance Examination is: ____________

II. The Composition Required

Direction: this is a guided composition which you are required to write in accordance with the following requirements.
Title: A letter to a friend / a teacher / my parents, etc.
Letter format: informal
Style: informal
Number of words required: at least 80 words
Time: 30 minutes
Write the compositions in class.
Do not consult dictionaries.
Do not discuss with others.