

A Corpus-based Study on Chinese Students' Writing Errors

Pan Jichun, Wang Junsong

Foreign Languages School, Northwestern Polytechnical University, Xi'an, Shaanxi, China

ABSTRACT: A careful study of the errors will reveal the common characteristics shared by the language learners. Based on the theory of Error Analysis and Corpus-based Error Analysis, and with the great help of computer technology, this paper is trying to make an empirical study on common writing errors committed by Chinese non-English major students and put forward some practical approaches to the error correction, which is believed to be of great help to both the language teachers and the language learners.

KEYWORD: Corpus-based error analysis; Writing errors; Error correction

1 INTRODUCTION

Among the five basic language skills---listening, speaking, reading, writing and translation, writing is the most difficult one for both native and non-native learners, especially for those non-native non-English majors. A survey shows that among the English skills badly needed in the future in China, the ability of writing scientific reports or essays was considered to be the most important. However, the present situation of Chinese EFL learners' writing proficiency does not seem to meet the need. Although much effort has been made to improve their writing proficiency, it is still far from being satisfied. It is the fact that even if they have acquired a large vocabulary and have learned much about English grammar, they are still weak in communicative competence, in this case, writing. In Chinese non-English major students' English writings, some of the sentences are really difficult to understand, "Chinglish" expressions can be found in almost every piece of essay and linguistic errors are also very common to see. The large number of language errors shows the problems in our present teaching of the basic language skills.

On the other hand, it is well known that one cannot learn a language without committing errors, as Dulay and Burt charmingly put it "You cannot learn without goofing." [1] Errors seem to be an unavoidable phenomenon in the process of learning a foreign language. Thus the analysis of learners' errors is among the first ways through which researchers try to investigate foreign language acquisition. If the major problems of the language learners were found out and the corresponding remedial teaching methods were put forward, students would surely make less language errors in

their writings and their writing proficiency would be greatly improved. Hence, detecting the language errors, analyzing them and offering the effective solutions will be most effective means to improve both our teaching and the students' writing proficiency.

Hence, a study was designed to analyze the errors made by the Chinese non-English major college students in their writings. Based on the theory of Error Analysis (EA) and Corpus-based Error Analysis (CEA), and with the great help of computer technology, this study aims to find out the main types of linguistic errors and reveal the common characteristics of errors shared by the language learners, which is believed to be of great help to both the language teachers and the language learners.

2 THEORETICAL BACKGROUND

The study of learners' errors has been driven mainly by two theories. They are Contrastive Analysis (CA) and Error Analysis (EA). When CA lost its initial appeals, EA superseded CA as the dominant approach to studying the learners' language. In the early 1990s, with the development of computer technology, EA was reinvented in a new form: Corpus-based Error Analysis (CEA), which is more capable of dealing with large-sized research data and overcomes some inevitable weaknesses in traditional Error Analysis. The following part of the paper is dealing with an overview of the theoretical foundation of the study: EA and CEA.

2.1 Error Analysis (EA)

Error analysis is a branch of Applied Linguistics

which plays an important part in the study on second and foreign language learning. Error analysis is the study and analysis of the errors made by second and foreign language learners. It is carried out in order to find out how well someone knows a language and how a person learns a language, and obtain information on common difficulties in language learning, so as to provide an aid in teaching or in the preparation of teaching materials. [2] Error analysis examines the actual errors produced by the learners in L2. It views both first and second language acquisition as a process involving the active participation of the learners. This approach is based on cognitive psychology which sees errors as a clue to what is happening in the mind of the learners. In this approach, errors are seen as a natural phenomenon that must occur when learning a first or second language before correct language rules are completely internalized. Much of the work in the field of EA is attributed to Corder. In 1967, he first suggested that a better understanding of language learning would come from a more systematic investigation of learners' errors by discovering the 'built-in syllabus' of the language learners. [3] Many of the efforts of the following decade were in fact directed to discovering the natural sequences of EFL learning. It not only helps us gain some insights into the process of L2 learning, but also throws some light on the strategies the learners employ in the learning process.

2.2 Corpus-based Error Analysis (CEA)

"A corpus is a collection of linguistic data, either written texts or a transcription of a recorded speech, which can be used as a starting-point for linguistic description or a means of testing hypotheses about a language". [4] Corpus linguistics is the result of the interdisciplinary development between computer science and linguistics. It is a new way of thinking, a new method to gain a deeper understanding of the nature of language and a new discipline of applied linguistics which provides more objective views on language study. It is a new research method in language study in the linguistic field which greatly depends on the use of computer and concordance software. With the aid of corpus, it is possible for researchers to collect, observe and analyze linguistic data and find out the similarities and differences between second language learner and native language speaker or between learners at different language proficiency levels in order to facilitate language research.

In the late twentieth century, Error Analysis based on learner corpora which was initiated by Granger provides a brand-new perspective into the aspect. The essential characteristics of CEA are: "It is empirical, analyzing the actual patterns of use in natural texts. It utilizes a large and principled collection of natural texts, known as a "corpus", as

the basis for analysis. It makes extensive use of computers for analysis, using both automatic and interactive techniques. It depends on both quantitative and qualitative analytical techniques". [5] Corpus-based Error Analysis makes it possible for research workers not only to analyze what is wrong but also to describe what is right. Linguists can observe the language produced by EFL learners in contrast to that uttered by native speakers.

Gui Shichun [6] summarizes the advantages of CEA as follows:

It can collect and store a great quantity of linguistic data and the users can extract the data according to one's own use, beneficial to all kinds of researchers.

It avoids subject and groundless conclusion by using qualitative analytical techniques and statistical inferences.

It can be used to do vertical researches to some extent for observing the development of learners' language.

An error-tagged corpus could provide more information about learners' language and suggestions about teaching.

By taking advantage of learner corpora and specific software tools, the distribution of errors, frequency of errors and sorts of errors can be obtained quickly and accurately. However, CEA analysis should be seen as a complementary approach to the traditional approaches rather than the single correct approach.

3 RESEARCH DESIGN

3.1 Corpus Building

The data of the present study is based on the compositions written by the non-English major students in NWPU. During the English teaching and study, the teacher assigned the students four writing tasks each term and students were required to complete them online (www. pigai. org---a correcting network providing online service to correct 160 students writings automatically based on the corpus and cloud computing technology). 160 freshmen from ten different departments submitted their essays and the writing topics are argumentative, descriptive and narrative in genre with the length varying from 120 words to 200 words each. The teacher thus collected 1000 online essays as the research samples and built up a mini writing corpus for study.

3.2 Methodology and Instrument

To identify errors means to recognize and locate errors, which is a challenging task. To achieve the purpose, error tagging is an indispensable procedure for any corpus-based EA study. In this study, AnnoTool software is applied to finish this task.

After tagging the error types in AnnoTool, the researcher used another function of it---insertion, to insert these tags into the compositions.

In this study, both qualitative approach and quantitative approach were employed. The former one was used to describe the types of the errors while the latter approach was to deal with the number and the frequency of writing errors occurred in the materials.

4 RESULTS AND DISCUSSIONS

4.1 Classification of writing errors

The errors found in the present study are classified according to linguistic description and content and organization the writings as well. After the absorption of the taxonomies by Duly, Burt and Krashen [7], students' writing errors were first classified into three general levels: linguistic errors, discourse errors and pragmatic errors. In addition, another level is added in this study, the level of "idiomatic English" which is named "Chinese English" or "Chinglish".

4.1.1 Linguistic Errors

Linguistic errors refer to grammatical errors. Errors at this level are divided into the following subcategories: morphological errors, lexical errors, syntactical errors, and cohesive and coherent errors.

4.1.1.1 Morphological errors

Morphological errors mainly involve misspelling, misuse of the plural forms, omission of third person singular ending and errors of capitalization and punctuation.

4.1.1.2 Lexical errors

Lexical errors in this study mainly refer to the semantic or conceptual errors in lexis. It mainly involves two types, namely, malformation and coinage and collocation errors.

4.1.1.3 Syntactical errors

Syntactical errors include errors in the use of structure words including articles, prepositions, conjunctions, auxiliary verb "be" and pronouns, errors in sentence structure and errors in tense, voice and mood.

4.1.1.4 Cohesive and coherent errors

Cohesion refers to the "cohesive ties", a term created by Halliday and Hasan which involves such content as reference, substitution, ellipsis, conjunction and lexical cohesion. [8]

4.1.2 Discourse Errors

Compared with the linguistic errors which are relatively overt and can be identified easily, there are still some errors which are covert and difficult to identify and are commonly reflected in idea

production and organization of the writings. These errors are categorized in the name of discourse errors. According to study, it is found that the discourse errors mainly lie in idea coherence and information ordering.

4.1.3 Chinglish

Chinese English can be easily witnessed in students' writings. It means that Chinese student applies his mother tongue rules to those of English, and with the inference of Chinese thing mode and the specific culture, he produces the "deformed" English that deviate from standard English.

4.2 Quantitative Analysis

After classifying all the errors, the classified errors were then encoded and put into computer. With the help of Excel, some statistical analysis to the writing errors has been carried out. The results obtained from the quantitative analysis are the primary focus of this paper as shown in the following two tables.

Table 1.Types of writing errors

Types of errors	linguistic errors	discourse errors	Chinese English	total
Distribution of errors				
<i>Number of errors</i>	1743	582	334	2659
<i>Proportions</i>	65.5%	22%	12.5%	100%

Table 2 Sub-branches of linguistic errors

Types of errors	Lexical errors	Syntactical errors	Morphological errors	Cohesive and coherent errors
Proportions				
<i>Number of errors</i>	37%	25%	21%	19%

5 PEDAGOGICAL IMPLICATIONS

After learning the different types of errors which students commonly share in their writings, one might ask what we could actually do in our teaching and learning practice to improve our students' oral competence. Hence, in the following, the paper is going to present some suggestions to treat this thorny question.

5.1 Proper attitude toward error and error correction

People's attitudes toward errors are quite different. With the development of applied linguistics, there have grown two kinds of attitudes toward learners' errors, the behaviorist attitude and the mentalist attitude. Which attitude should we take in language teaching? In our opinion, it is absolutely wrong to go into either extreme. But as far as writing teaching is concerned, the mentalist viewpoint that errors provide valuable evidence of learning problems and thus supply the teachers with information on which they can base their remedial teaching [9], is more

objective and will be adopted more frequently than behaviorist viewpoint because we know that errors are the inevitable product of learning. They are indicators that learning is taking place, and also evidence that the mysterious language acquisition device is working. Learners' errors are seen as an indispensable part of the learning process because learners are encouraged to explore the target language.

5.2 *Strategies of error correction*

When deciding to give students feed back about the effects of their errors they made, teachers should take some techniques and procedures that need to be designed and tested. During the accurate reproduction stage some correction techniques can be employed to achieve the aim. [10]

5.2.1 *Self-correction*

Students prefer to put their own errors right by themselves rather than be corrected by others so as to keep face. So, it is advisable to give a chance for students to re-experience the language and retest their hypotheses against the language. What the teacher needs to do is simply to mark the error when it occurs. Surely, learners can take responsibility for the treatment of minor production-centered errors but this must also be balanced by the teacher's focus on major process-centered errors.

5.2.2 *Peer-correction*

If the student may not know what to do with the errors, or may even make another error, the teacher can ask students to cooperate and help with each other. This is called peer-correction. There are some good reasons in encouraging the use of peer-correction. First, we should stress the value of communication between and among students. Secondly, most of our students take English courses where a significant number of their fellows will also be non-native speakers of English. So, it is clearly important that students get used to the necessity to understand and be understood by other non-native speakers. Thirdly, it is useful for students themselves to get feedback on exactly how much of what they have produced has actually been comprehensible to the members of their audience in the actual face-to-face communication. Fourthly, by peer-correction, both learners are involved in the learning and thinking about the language. Fifthly, peer-correction helps learners cooperate and make them less dependent on teachers. Sixthly, peer-correction is useful when students work in pairs and groups, when teacher's help is not often available.

5.2.3 *Teacher Correction*

In the most of the time, it is the teacher's responsibility to take charge of correction. Burt and Kipasky suggest that global errors, such as the wrong

use of connectors, unclear distinctions between coordinate and relative clause constructions, ill balanced parallel structure and inconsistent tense, are much more severe than local errors such as noun formation and articles, etc. Thus, the teacher may mainly focus on such errors and the error correction in such way will be especially appropriate when the teacher sees that a majority of class is having the same problem. When in this situation, the teacher must realize that this point has not yet been generally learned, even though the teacher might think that he or she has taught it, and it obviously needs teaching again in a different way.

6 CONCLUSION

It has to be admitted that due to the limitation of the time and corpus building, what is investigated in this paper is by no means complete. More sophisticated research in this area would be expected in the future. However, it is hoped that the analysis to the common writing errors and the practical approach listed in this paper will be of some help to the English teachers who are wondering how to deal with the students' errors and to the Chinese EFL learners who are wondering at the cross-road, suffering from not knowing the proper ways to improve their writing competence.

REFERENCES

- [1] Dulat, H.C. and Burt, M.K., "You cannot learn without goofing an analysis of children's second language 'errors'". *Error Analysis Perspectives on Second Language Acquisition*. ed. by Richards, J.C. Longman Group Limited, 1974, pp.95-123.
- [2] Richards. J.C., "Social factors, interlanguage and language learning. *Error Analysis: Perspectives on Second Language Acquisition*. ed. by Richards, J.C. Longman Group Limited, 1974, pp.64-94.
- [3] Corder, S.P. (1967)., "The Significance of Learners' Errors", *International Review of Applied Linguistics* Vol. 5 No.4, 167-178
- [4] Crystal, D., "Stylistic profiling". *English Corpus Linguistics: Studies in Honor of Jan Swartvik*. London: Longman, 1991. pp. 28-37.
- [5] Douglas, B., Susan, C. & Randi. R., *Corpus Linguistics*. Beijing: Foreign Language Teaching and Research Press, 2000.
- [6] Gui, Shichun, "A cognitive model of courpus-based analysis of Chinese learners' errors of English". *Modern Foreign Languages*, 2004, Vol 27, pp129-139.
- [7] Dulay, H., Burt, M. and Krashen, S.D., *Language Two*. Newbury House, Rowley, MA, 1982.
- [8] Halliday, M.A.K. and Hason, R., *Cohesion in English*. London: Longman, 1976, pp.31-292.
- [9] Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- [10] Burt, M. and Kiparsky, C., *Global and Local Mistakes in New Frontiers in Second Language Learning*, eds. by J. Schumann and N. Stenson Rowley. Mass. Newbury House, 1974.