

Analysis of the preposition error types in Corpus for English Majors (CEM)

Shuochen Wang¹

¹Graduate School of Language Education, Takushoku University

tonywang531@gmail.com

Abstract

This paper will focus on the analysis of the prepositional error types found in the Corpus for English Majors (referred to as CEM). CEM is a 2007 learner corpus of Chinese learners of English Major containing approximately one million words. It contains written essays questions of TEM-4 (Test for English Majors) and TEM-8 including translation works. TEM-4 is administered at the end of the second year and TEM-8 at the end of the fourth year in their undergraduate programs. All the errors made by learners are marked in xml style designed by the publisher.

My method of approach is to analyze the number of occurrences of prepositional error types in the corpus. I have classified the prepositional errors into 3 levels: addition, deletion and modification; simple, complex and idiomatic; finally the misused preposition itself. The results show the most dominant preposition error type is the modification error type, which is the confusion of the preposition. In addition, most of the errors of the one word preposition type. The top 10 preposition students struggled with the most are: *in, to, on, of, for, from, with, by, at* and *as*.

Keywords

Corpus for English Majors, learner corpus, prepositions, error analysis, corpus linguistics.

Introduction

Prepositions are among the most difficult items to master in a second language. Yet so far there has been few studies on which type of preposition errors the Chinese learners make using a corpus approach. The advantage of a corpus approach is its large data size as the representative of the whole population. By analyzing the occurrences of preposition errors in a one million words corpus, we can study

what kind of preposition errors students make in a manner which the previous SLA studies could not achieve before.

1 Objectives

In this paper I aim to find out what type of prepositional errors Chinese university students make. By identifying the prepositions learners struggle the most, we can give insights and suggestions on how to teach preposition better.

2 Method of Analysis

The method of analysis I have conducted for this paper included the following steps.

1. Reconstruct correct prepositional usages for all instances
2. Classify the errors by types
3. Count the frequency of errors for each type

My error classifications contain 3 levels.

The first level refers to addition, deletion and modification. The second level refers to simple, complex and idiomatic. The third level refers to the misused preposition itself.

3 Results

Although preposition errors are not the top error in the corpus, it is still one of the top errors, therefore they are still significant in terms of errors.

Table 1: Top Errors Present in the Corpus

Word choice	19,532	23%
Spelling	12,771	15%
Tense	6,823	8%
Noun	5,535	6%
Preposition	3,507	4%
Total	85,882	

For the first level, the majority of the errors are of modification errors. This means most students made the mistake of confusing two prepositions. No correction refers to instances where I do not agree with the marker's decision: I do not believe the sentence is mistaken or it is more appropriate to fix other parts of speech other than the preposition.

Table 2: Distribution of First Level Classifications

Addition	519	15%
Deletion	207	6%
Modification	2,569	76%
No correction	77	2%
Total	3,372	

For the second level, the majority of errors are errors of a simple preposition. Complex prepositions are few and not significant. However, some students seemed to struggle with idiomatic expressions.

Table 3: Distribution of Second level classifications

Simple	3,024	90%
Complex	57	2%
Idiomatic	214	6%
No correction	77	2%
Total	3,372	

For the third level, the top 10 prepositions students struggled with the most are: *in, to, on, of, for, from, with, by, at* and *as*.

Table 4: Distribution of third level classifications

In	533	16%
To	485	14%
On	404	12%
Of	342	10%
For	245	7%
From	189	6%
With	148	4%
By	142	4%
At	114	3%
As	84	2%

4 Conclusion

The results agree with other studies in the following ways: preposition errors are significant, most errors are the confusion between two prepositions (Rozovskaya, 2010). In addition, most preposition errors are of single word preposition errors. I have also identified the top 10 prepositions; however whether this is a unique distribution of the Chinese students or universal for all learners is open for future research.

References

- Corder, S. P. (1967). The significance of learners' errors. *Int. Rev. Appl. Linguist.*, 5, 161-169.
- Corpus for English Majors*. (2008). Shanghai Foreign Language Education Press. ISBN 978-7-900717-38-2/H·24.
- Dulay H, M. Burt., & S. Krashen. (1982). *Language two*. Oxford University Press.
- Gamon, M., Gao, J., Brockett, C., Klementiev, A., Dolan, W., D, Belenko., & Vanderwende, L. (2008). Using Contextual Speller Techniques and Language Modeling for ESL Error Correction. *Proceedings of IJCNLP*.
- James, C. (1998). *Errors in Language Learning and Use*. London: Addison Wesley Longman."
- Leech, G. (1998). "Learner corpora: what they are and what can be done with them". *Learner English on Computer*. Ed. S. Granger. London: Longman. xiv-xx.
- Rozovskaya, A., Roth, D. (2010). Annotating ESL Errors: Challenges and Rewards. *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*. 28-36.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sudhakaran, B. (2015). Acquisition of English Prepositions among Malaysian Learners: A Case Study. *International Journal of English Linguistics*; Vol. 5. No. 3.
- Tetreault, J., & Chodorow, M. (2008). Native Judgments of Non-Native Usage: Experiments in Preposition Error Detection. *COLING Workshop on Human Judgments in Computational Linguistics*, Manchester, UK.