A Corpus Based Study of the Errors Committed by Pakistani Learners of English at Graduation Level

Muhammad Tayyib Ijaz
G. C. University, Quaid-e-Azam Road, Faisalabad, Pakistan
Tel: +92-300-7288600 E-mail:tayyabijaz@gmail.com

Muhammad Asim Mahmood
G. C. University, Faisalabad, Pakistan
E-mail:masimrai@gmail.com

Aisha Ameer
G. C. University, Quaid-e-Azam Road, Faisalabad, Pakistan
Tel: +92-314-4050493 E-mail:duafatimah@gmail.com

Abstract

Man is bound to make mistakes. Language learning is totally a process of trial and error. As language errors are the topic of this paper, let us provisionally define language errors first as an unsuccessful bit of language. Though errors were considered a sin in the past, now the linguists stress the importance of learners' errors. These errors are helpul to teachers, researchers, and learners themselves (Corder 1967). Error analysis is the process that determines the incidences, nature, causes and consequences of the unsuccessful language i.e. errors. Keeping in mind the recent trends in language teaching, studying learners' output is important to define the parameters of learning second language.

The present corpus based work intends to find errors made by Pakistani learners of English at graduation level, when their production is in written form. The learner corpus used for this research Pakistani component of ICLE (International Centre for Learner English). It contains 0.2 million words in total, written in the form of argumentative essays, by the learners of English at the graduation level and is POS tagged using CLAWS7 POS tagger.

Keywords: Error Analysis, Learner Errors, Corpus Based Error Analysis, Learner Corpora, Pakistani English

1. Introduction

As the famous saying goes, "To err is human", in the process of language development, this is also the case. When the utterances produced by learners are examined and compared with target language norms, they are often found to be full of errors. In the present research, the term "error" refers to a systematic deviation from a selected norm (after Burt et al. 1982) or set of norms. Since it is inevitable that learners commit errors, students should be enabled to benefit from various forms of feedback on those errors so that they do not commit them repeatedly and continuously. But most of the times, errors of the learners are not dealt in a technical way. It almost seems as if they are dismissed as a matter of no particular importance, as possibly annoying, distracting, but an inevitable by-product of learning a language. The present research paper explains the errors committed by Pakistani learners of English, and their causes, based on a corpus of errors collected from the Pakistani graduate students. This study focuses on the errors committed, the classification of those errors and the causes of those errors. It would also suggest some remedial measures in this regard. Although a lot of work on Error Analysis has been done, the present research paper is an effort of contribution in this area in Pakistani perspective. One of its aims is to find out the problem areas in language learning and language acquisition as faced by Pakistani Language learners at various levels.

There may be many reasons behind the errors. First language interference, carelessness on the part of both the teacher and the learner, overgeneralization or incomplete application of the rules, etc. might be the possible causes, but they are a natural phenomenon, not to be viewed as a stumbling block in the process of learning, rather to be viewed as a sign of progress in the process of learning a second language.

Errors are always inevitable in the process of language learning. Like other learners, Pakistani learners also face problems in the learning of English and commit mistakes and errors. But it is a fact that there is no error analysis practically in our classrooms. It is only the 'Error Detection' which is limited to finding the error and underlining or encircling it. One of its reasons is that there is no repository of errors. There is no way of comparing the errors. It is the intuitive judgment of the teacher that helps to detect the errors. As errors are inevitable, and there are many factors behind those errors, this paper tries to find out those problems the Pakistani learners of English face

at graduation level.

This Error Analysis would help determine the reasons of the errors which would ultimately help in the process of learning English as a foreign language. The present study was delimited to the corpus data collected from Pakistani learners of English at graduation level. The genre was argumentative essays. The length of the essay was from 500 up to 1000 words. Data collection was not bound to any social, political, or economic background and mixed sampling was done.

2. Review of Literature

Having its origin in 70's, error analysis turned to be a scientific method to judge and analyze the errors of the learners. It owes a lot to the work of Corder (1967), Richards (1970) and Selinker (1972). They identified various aspects of second language learners' language system, an independent language system that was neither their first language nor the target language, but another one, the third language, that Selinker calls 'interlanguage' (Selinker 1972). According to Corder (1967) learners' errors are systematic and not random. Richards (1970) categorized the learners' errors into three types: Interference errors—caused by L1 transfer, Intralingual errors—caused by overgeneralized application of rules, and developmental errors—caused by developing a faulty hypothesis.

Error analysis has a unique place in the field of ESL/EFL/SLA. It gives information about how a language is learned. The errors represent the internal construct of the learners. It further explains about the amount of language knowledge the learner has. Error Analysis is crucial in the field of SLA because it gives a direct approach to the observation of actual learner output ad helps researchers explain how learning progresses.

Error analysis is also important because it directly provides methodology to study the learner language and errors, and the study of errors has a significant role in the evolution of language teaching. The previous formal methods considered errors as the evidence of mislearning, and utmost effort was done to avoid them as the impeccable learning was the objective. But now the scenario is totally changed. Now the errors are considered to be the "eternal unavoidable companion". In fact study of errors proves advantageous in language teaching pedagogies. It assists in all aspects, from syllabus to material.

The term corpus, basically derived from Latin, as it was first used for body, refers to the large body/collection of texts. Oxford English Dictionary defines corpus as "the body of written or spoken material upon which a linguistic analysis is based". Francis (1982) calls it to be 'a collection of texts assumed to be representative of a given language, to be used for linguistic analyses. Corpus have been used in three major areas: lexicographical studies, dialectological studies and grammar creation (Francis 1982). Corpus linguistics proved to be a revolutionary agent in the field of linguistics as it deal with the real data and enables linguist to work and find real results, irrespective of any other factor. McEnery, Xiao and Tono (2006:7) consider corpus linguistics as "a whole system of methods and principles of how to apply corpora in language studies and teaching/learning".

Around the globe, the learner corpora has gained much weight. ICLE claims to have the largest collection of learner corpora from the learners of English as second/foreign language. The date for the present research was also collected under the ICLE guidelines as there was no previous work on learner corpora in Pakistan.

3. Methodology

3.1 Corpus Description

The International Corpus of Learner English (ICLE) is the generic name of the corpus that was used for the present research. ICLE is the name of the consortium of many institutions from different countries with a collaborative work, with the corpus of almost 30 million words, containing the writings of learners of English from 21 different mother tongues background. This corpus can be manipulated for research purposes in many ways. The sub-corpus used for the present research contains 0.2 million words written in the form of argumentative essays by the advanced level learners of English. Along with data, the ICLE also provides an Error Tagging Manual (Dagneaux et al., 1996) which gives an in detail description of the errors. Mainly there are seven types of errors.

The data collected was in the form of plain text, and was POS tagged using CLAWS 7 POS Tagger. It was further processed by using Wordsmith 5.0, a concordance software that gives a variety of options to manipulate data.

3.2 Error Categories

The error tagging manual (V 1.2) categorizes the errors in following groups.
1. Formal errors
2. Grammatical errors, i.e. errors that break general rules of English grammar
3. Lexico-grammar errors, i.e. errors where the morpho-syntactic properties of a word have been violated

4. Lexical errors, i.e. errors involving the semantic properties of single words and phrases

5. Word Redundant, Word Missing and Word Order errors

6. Punctuation errors

7. Style errors

3.3 Process of Analysis

Corder (1974) suggests three steps in the process of error analysis:

Recognition

Description

Explanation

In the whole process of error analysis, he further emphasizes the importance of data analysis and data collection. As this is a corpus based work, the most important and essential task is data analysis in order to draw conclusions. Keeping these guidelines in the mind, the following route was mapped to reach the results.

Collection of the corpus

Recognition of errors

Description of errors

Explanation of errors

After the process of recognition and description of the errors, the analysis of the nature of errors and statistical results were formulated.

4. Results and Discussion

On the basis of the results of data analysis it can be said that tenses are the most problematic area of the learners at graduation level. Among all the plethora of the errors 19.06 % errors (709 errors) are regarding the correct use of tenses.

In case of errors of tenses, learners are found unable to use proper tense according to the demand of the situation. Most of the errors are in the form of combination of the present or future with the past tense in the same sentence.

For example,

They said that dreams are needed if one is unable to get what one wants.   (PAGF 1014)

There are errors regarding the proper use of verb in the simple present formation.

For example,

She perform her duties in different manner.	(PAGF 1020)
Usually he get out of temper.	(PAUF1008)
Education play an important role in the progress.	(PAGF1005)
It play important role for betterment.	(PAGF1022)

Some of the examples in the category of verb errors show that the learners have learnt the rules but the application of those rules is either wrong or at improper place. The addition of s/es morpheme with the base form of verb to construct simple present tense for 3rd person was misused frequently.

For example,

They *sits* all the time in front of it.	(PAGF1074)
They *seems* happy.	(PAGF1084)
They *provides* true information.	(PAPL1001)
You *feels* the financial pressure of trying to pay bills.	(PAVL1005)

The students seem to commit the errors because of ignorance of rules restrictions and lack of practice in grammar exercises. The use of s/es morpheme with 2nd/3rd person plural pronoun shows that the students are well aware of the rule of using s/es morpheme in the tenses, but they are not applying the rule at the proper place. Similarly the omission of s/es morpheme while making sentences of 3rd person/singular subject in the present tense formation also shows that they are either oblivious of the rule, or they have learnt it but they are unable to use it at proper place.

During the data analysis it was found that the spelling mistakes were the ruling party. If the spelling mistakes of each category i.e. addition, omission, replacement, misspelling and capitalization are dealt collectively, these errors reach the count of 2087 which is a colossal figure, and it is evident of an alarming situation. This high toll of spelling errors shows that this area of language is most neglected one.

The errors of capitalization were the most important and prominent one in the spelling errors. The learners even do not know the simple fact that the proper nouns are begun with capital letters. Similarly they seem to be ignorant of the simple rules that the new sentence begins with capital letters. They use small letters where the capital letters must have been used and they use the capital letters where they are not required.

The learners commit the spelling errors because they do not know the exact spellings and it means that in their

courses of language, the spellings are given less importance as compared to the other areas of language learning. The students seem to be following the pronunciation of the word whereas in reality the orthography of some words is totally different from the pronunciation. Lack of proper feedback is also responsible for this grievous situation.

There might be some other reasons behind spelling errors, but lack of concentration while reading seems to be the major cause of spelling errors. It also proves that the students are not taught the spelling rules, and they simply rely on the pronunciation of the words, and try to write as they hear it. This situation may easily be handled with the help of activities/co-activities which would improve the spellings of the learners.

Here are few examples	
The *socity* has taught us…………	(PAAO1012)
It should creat a fair system of tax.	(PAGF1021)
They monitor it every *houre.*	(PAAO1007)

Another high proportion of errors are the errors of prepositions. The learners of English as a foreign language sometimes omit the preposition where it is necessary and sometimes use unnecessary preposition and in the few places they use wrong prepositions. L1 interference seems to be the main cause of these errors. The errors of the sentence fragments came out to be the fourth highest proportion of errors with a percentage of 8.73. The learners seem to be quite ignorant of the rules of using relative clauses. These errors seem to occur because of the false concepts hypothesized, and ignorance of rules restrictions.

Similarly, the errors in articles with sizeable proportions of 51 have also been found to occur as a result of omission of articles and unnecessary or wrong use of articles. It is the 11th highest error category with percentage of 1.37%. The learners are found to be inconsistent in the use of articles. The errors of adjectives mostly occur because of the wrong use of the degree of adjectives, inappropriate adjectives or the use of a noun instead of an adjective, whereas the errors of adverbs are mostly the result of the misplacement of the adverbs.

Errors of punctuation are committed in a bulky magnitude, particularly, the errors in the use of full stops, commas and apostrophe. The students, mostly use "of" to show possession for animate nouns. It means that the rules of punctuation are not taught to them properly and they keep on making errors of punctuation.

Negation errors, though in a very small proportion, occur because the learners seem to be unable to convert simple sentences into negative forms. They have also made use of double negatives or only not without any auxiliary verb.

5. Conclusion

The study has given an in detail account of the errors made by Pakistani learners of English at graduate level. Based on the findings, discussion and examples given, it can be concluded that the Pakistani learners made eight common types of errors in their writings i.e. verb tense errors, spelling, punctuation, capitalization, articles, preposition, subject verb agreement and double negation. Out of all these errors, verb tense errors are at the top, showing that the tenses are hard area for Pakistani learners of English. Ignorance of the rules restriction and lack of practice in grammar exercises seem to be main cause of these errors. Analysis of these errors shows that the learners are aware of the rules but they are unable to apply it at proper place. Furthermore, the lack of practice is prevalent everywhere. The Pakistani learners have less exposure to the target language, having minimum knowledge of language in use. This factor causes them producing more unsuccessful language. Keeping in view the results, it is suggested that more grammatical exercises should be added to the course contents at graduate level. Furthermore, there should be special language courses for all graduate students to enhance their proficiency in English.

References

Corder, S. P. (1967). THE SIGNIFICANCE OF LEARNER'S ERRORS. *Iral-international Review of Applied Linguistics in Language Teaching*. doi:10.1515/iral.1967.5.1-4.161

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

In Allen, J. P., & In Corder, S. P. (1974). *The Edinburgh course in applied linguistics: 3*. (The Edinburgh Course in Applied Linguistics.) London: Oxford university press.

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

Richards, J. C. (1970). *A non-contrastive approach to error analysis: By Jack C. Richards*. Quebec: Centre International de Recherches sur le Bilinguisme.

Selinker, L. (1972). INTERLANGUAGE. *Iral-international Review of Applied Linguistics in Language Teaching*. doi:10.1515/iral.1972.10.1-4.209

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar