



PART-OF-SPEECH TAGGING ACCURACY IN ESL LEARNER CORPORA: AN ERROR-BASED COMPUTATIONAL STUDY

M. Ihsan Haris
ihsanharis388@gmail.com

MPhil Scholar, Department of English, NUML University,
Faisalabad Campus, Faisalabad, Punjab, Pakistan.

Dr. Aftab Akram
aakrum@numl.edu.pk

Lecturer, Department of English, NUML University,
Faisalabad Campus, Faisalabad, Punjab, Pakistan

Abstract

This study investigates the accuracy of automated part-of-speech (POS) tagging in English as a Second Language (ESL) learner writing using an error-based computational approach. A corpus of fifteen learner essays, totaling approximately 2,870 tokens, was analyzed using the TagAnt POS tagging tool, and the automated outputs were meticulously compared with manually verified correct tags. The analysis focused on identifying common tagging errors, classifying them into categories, and examining patterns in relation to interlanguage development. Findings indicate that verb-related errors, particularly tense misclassifications and subject-verb agreement issues, are the most prevalent, whereas noun, adjective, and adverb errors occur less frequently. These systematic deviations reflect learners' transitional interlanguage systems rather than random mistakes, offering insights into developmental stages of second language acquisition. The study achieved an overall tagging accuracy of 97.8%, demonstrating the general reliability of automated tools while revealing their limitations when processing non-native writing. By linking computational error analysis with Interlanguage Theory, the research provides theoretical contributions to understanding learner language development and practical guidance for educators, curriculum designers, and computational linguists. The results emphasize the importance of targeted pedagogical interventions, refinement of automated tagging systems, and the application of corpus-based computational methods to improve the assessment and support of ESL learners. This study bridges the gap between linguistic theory and computational practice, highlighting the role of automated analysis in enhancing both educational outcomes and the development of language technologies.

Keywords: *Automated Error Analysis, ESL Learner Corpus, Interlanguage Theory, Part-of-Speech Tagging*

Corresponding Author: M. Ihsan Haris (MPhil Scholar, Department of English, NUML University, Faisalabad Campus, Faisalabad, Punjab, Pakistan.)
Email: ihsanharis388@gmail.com

1. Introduction

English is widely used as a second language in many countries, and learning it correctly is important for clear communication. ESL students often make mistakes in grammar, vocabulary, and sentence structure. These mistakes are not random. They show the level of a learner's language development. Studying these errors helps us understand how students learn English and what kind of difficulties they face. Today, computer tools make it easier to study learner errors because they can analyze texts quickly and handle a large amount of data. This study focuses on part-of-speech tagging errors in ESL learner essays using a computational method.

The main purpose of this study is to fill a gap in research related to error-based computational analysis of ESL writing. Many previous studies have examined learner errors manually, which takes a lot of time and effort. However, fewer studies have used automated tools to identify and analyze errors in part-of-speech tagging. By using a computational approach, this research provides clear information about tagging accuracy, common error types, and patterns in learner writing. It also checks how reliable automated POS tagging tools are when they are used with learner language instead of native speaker texts.

This study uses an evaluative error-analysis design and quantitative corpus analysis. The data for this research consists of fifteen essays written by ESL learners. These essays were processed using TagAnt, which is an automated POS tagging tool. After tagging, the results were compared with manually checked correct tags to find mistakes made by the tool. The tagging errors were then classified into different categories, and the overall accuracy of the tool was calculated using percentage.

The study is guided by Interlanguage Theory, which explains that learners develop their own temporary language system while learning English. Because of this developing system, learners make systematic errors. By connecting tagging errors with interlanguage development, this study combines computational analysis with language learning theory. In this way, the research helps in understanding ESL learner writing and also evaluates how effective automated tagging tools are for analyzing learner data.

1.1. Research Questions

1. What types of part-of-speech tagging errors are most common in ESL learner writing when analyzed computationally?
2. Why do automated POS taggers produce errors in learner language, and how do these errors relate to interlanguage development?
3. How can computational analysis of tagging errors improve our understanding of ESL learner writing accuracy?

1.2. Research Objectives

1. To find out the common part-of-speech tagging errors in ESL learner essays using automated tools.
2. To examine the reasons for errors produced by automated POS taggers in learner writing.
3. To analyze the accuracy of automated tagging and identify patterns in learner language.

1.3. Significance of the Study

This study is important because it shows how error-based computational methods can be applied to ESL learner corpora to identify patterns in grammar, vocabulary, and sentence structure efficiently. The findings help teachers design targeted strategies to address common learner mistakes and guide curriculum developers in creating materials aligned with learners' needs. It also informs software developers about limitations in automated POS tagging tools and ways to improve accuracy for non-native texts. The study supports Interlanguage Theory by showing that many errors reflect systematic stages of language learning. Linking tagging inaccuracies to learner development bridges theory and practice, providing insights into second-language acquisition. The research combines quantitative and qualitative analysis, strengthening corpus-based studies and offering evidence for improved teaching and computational methods. Overall, it benefits educators, researchers, and developers by enhancing understanding of learner language and improving instructional and technological tools.

1.4. Statement of the Problem

Errors in ESL learner writing remain a challenge for both teachers and automated language tools. Manual error analysis is time-consuming and limited in scope, while automated POS taggers often misclassify words because they are trained on native-speaker corpora. Learner texts frequently contain incorrect verb forms, missing articles, and unusual word order, which lead to tagging errors and affect analysis. This study addresses this issue by using an error-based computational approach to systematically identify and categorize tagging inaccuracies. By comparing automated tags with manually verified correct tags, the research evaluates the reliability of POS tools and links errors to interlanguage development. The findings provide insights for improving teaching practices and designing more robust tagging systems capable of handling learner language.

2. Literature Review

Learner errors have long been considered an important source of evidence in second language acquisition research. According to Interlanguage Theory, learners create a developing linguistic system that contains systematic deviations from the target language. These deviations are not random but reflect stages of learning and rule formation. Studying such errors helps researchers understand language development and learning difficulties.

In recent years, computational tools have allowed researchers to examine these errors more systematically in large learner corpora (Selinker, 1972; Ellis, 2008).

Part of speech tagging plays a central role in corpus based linguistic analysis. It assigns grammatical labels such as noun, verb, adjective, and adverb to each word in a text. Accurate tagging allows researchers to study grammatical patterns, frequency distributions, and syntactic structures. In native corpora, automated taggers achieve high levels of accuracy. However, learner corpora present additional challenges because grammatical errors interfere with tagging performance (Jurafsky & Martin, 2020).

Traditional POS tagging systems are trained on standard native speaker data. As a result, they expect grammatically correct sentence structures. When learner writing includes incorrect verb forms, missing articles, or unusual word order, the tagging system may misclassify words. These misclassifications affect further linguistic analysis and may distort research findings. Therefore, evaluating tagging accuracy in ESL learner corpora is necessary before relying on automated results (Tetreault & Chodorow, 2008).

Error analysis as a research method emerged from the work of Corder, who argued that learner errors are meaningful and reveal learning strategies. This perspective shifted attention from simply correcting mistakes to understanding them. In an error based computational study, errors are not only identified but also quantified using automated tools. This combination of qualitative insight and quantitative measurement strengthens research reliability. It allows researchers to move beyond description toward systematic evaluation (Corder, 1967; Ellis, 2008).

Corpus linguistics provides the methodological foundation for computational analysis of language. By collecting authentic texts into a corpus, researchers can analyze real language use rather than isolated sentences. In ESL research, learner corpora have been developed to examine patterns of grammatical errors and vocabulary use. Computational tagging tools make it possible to process large numbers of essays efficiently. However, the accuracy of these tools with learner language remains an important research concern (Biber et al., 1998; Granger, 2002).

Automated tagging tools such as spaCy and TagAnt are widely used in linguistic research. These tools rely on statistical and machine learning models to assign grammatical categories. They are designed to work quickly and handle large amounts of text. Despite their efficiency, they are not specifically trained on learner language data. This limitation makes error based computational evaluation essential when applying them to ESL writing (Anthony, 2023; Manning et al., 2014).

Several studies have reported that verbs are among the most frequently mis-tagged categories in learner corpora. Errors in tense, agreement, and verb form create ambiguity for automated systems. Determiners and prepositions also cause difficulties due to omission or misuse by learners. These tagging errors can lead to incorrect frequency counts

and misleading conclusions about learner proficiency. Therefore, systematic comparison between automated tags and correct tags is necessary (Heilman et al., 2014).

Quantitative corpus analysis strengthens error based computational studies by allowing researchers to calculate measurable accuracy rates. Accuracy is usually determined by dividing the number of correct tags by the total number of tags and multiplying by one hundred. This numerical measurement provides objective evidence about tool performance. It also allows comparison across studies and datasets. Such statistical evaluation adds transparency and reliability to learner corpus research (Brezina, 2018).

Interlanguage Theory provides the theoretical framework for understanding why tagging errors occur in learner writing. Since learners operate within a transitional linguistic system, their grammatical forms may not match standard English categories. Automated taggers interpret words according to native norms, which may not align with learner usage. As a result, computational errors often reflect deeper developmental patterns. Linking tagging inaccuracies to interlanguage development creates a meaningful connection between technology and theory (Selinker, 1972; Odlin, 1989).

Error categorization is another important aspect of computational research. Instead of simply counting mistakes, researchers classify them into types such as verb errors, noun errors, or adjective misclassification. This categorization reveals which grammatical areas present the greatest difficulty for both learners and tagging tools. It also supports more focused pedagogical recommendations. Systematic classification strengthens the analytical depth of an error based computational study (Bitchener & Ferris, 2012).

Recent developments in natural language processing have improved tagging algorithms through machine learning and neural networks. These systems learn from large annotated datasets and predict word categories based on contextual patterns. Although accuracy has improved in standard corpora, learner corpora remain underrepresented in training data. As a result, further evaluation of automated tools in ESL contexts is necessary. Error based computational analysis contributes to this ongoing development (Jurafsky & Martin, 2020).

Comparative evaluation between manual and automated tagging is considered a reliable method for measuring tool accuracy. In this process, human corrected tags are treated as the standard reference. Automated output is then compared against this reference to identify discrepancies. This comparison provides both numerical accuracy rates and qualitative insights into tagging weaknesses. Such evaluation is essential in ensuring that computational findings are trustworthy (Flowerdew, 2015).

Educational implications of computational error analysis are significant. Teachers can use tagging results to identify common grammatical problems among students. Curriculum designers can adjust learning materials to address frequently misused

structures. Additionally, automated analysis supports faster feedback on learner writing. By combining computational efficiency with theoretical understanding, research can directly inform classroom practice (Reppen, 2010; Granger, 2003).

Despite these advances, gaps remain in the literature. Many studies focus on general tagging performance rather than detailed error based evaluation in small learner datasets. Few studies systematically examine tagging errors in a controlled set of essays and categorize them carefully. There is also limited research connecting tagging inaccuracies directly with interlanguage theory. Addressing these gaps strengthens both applied linguistics and computational research (Biber et al., 1998).

In conclusion, previous research highlights the importance of POS tagging, error analysis, and corpus based methods in studying learner language. However, the specific area of error based computational evaluation of tagging accuracy in ESL learner essays remains underdeveloped. By combining quantitative corpus analysis with theoretical insights from Interlanguage Theory, the present study contributes a focused evaluation of automated tagging tools. This approach ensures that computational methods are critically examined rather than assumed to be accurate (Ellis, 2008; Flowerdew, 2015).

3. Research Design

This study uses an evaluative error-analysis design. The purpose of this design is to evaluate the accuracy of an automatic part-of-speech tagging tool when it is applied to ESL learner writing. The study focuses on identifying and measuring tagging errors produced by the software. It does not only describe errors but also evaluates how accurately the tool performs. The research is guided by Interlanguage Theory, which explains that learner errors are systematic and reflect stages of language development. Therefore, tagging errors are examined both as computational mistakes and as part of learner language patterns.

4. Research Methodology

This research follows a quantitative corpus analysis method. The data consists of fifteen essays written by ESL learners. These essays were selected as the learner corpus for analysis. First, the essays were processed using TagAnt, an automatic POS tagging tool. The software assigned grammatical labels such as noun, verb, adjective, and adverb to each word in the essays. The tagged files were saved for further examination.

After tagging, the automated tags were manually compared with the correct tags. Differences between the automatic and correct tags were identified as tagging errors. The errors were then classified into grammatical categories such as verb errors, noun errors, adjective errors, and others.

AntConc was used to examine tag frequencies, search for specific tag patterns, and assist in counting the total number of tags. Finally, the accuracy rate was calculated by dividing the number of correct tags by the total number of tags and multiplying by 100 to

get the percentage of accuracy. This process ensures that the evaluation of the tagging tool is clear, systematic, and measurable.

5. Theoretical Framework

This study is based on Interlanguage Theory. This theory explains that second language learners develop their own temporary language system while learning a new language. This system is not the same as their first language and not fully the same as the target language. It is a developing system that changes over time as learners gain more knowledge and experience.

According to Interlanguage Theory, learner errors are not random mistakes. They are systematic and show the learner's current stage of language development. These errors happen because learners are testing rules, forming hypotheses, and gradually improving their understanding of English grammar. Therefore, errors are natural and important part of language learning.

In this research, tagging errors produced by the automatic POS tagging tool are examined in relation to learner language. Some tagging errors may happen because learner sentences contain grammatical deviations that reflect interlanguage development. By using Interlanguage Theory, the study does not only evaluate the performance of the tagging tool but also considers how learner language characteristics may influence tagging accuracy.

This theoretical framework helps connect computational analysis with second language acquisition theory. It provides a clear explanation for why learner texts may create challenges for automatic tagging systems and supports the analysis of error patterns found in the study.

6. Data Analysis

This study analyzes the accuracy of automated part-of-speech (POS) tagging in ESL learner essays using a computational approach. The primary objective is to evaluate the extent to which a POS tagging tool correctly assigns grammatical categories to words in learner texts and to identify systematic patterns of tagging errors. Understanding these patterns is crucial because errors in automated tagging may not only indicate limitations of the software but also reflect underlying characteristics of learner language, particularly as predicted by Interlanguage Theory.

The dataset comprises fifteen essays written by ESL learners, each ranging from approximately 175 to 210 tokens, totaling around 2,870 tokens. These essays were selected to provide a representative sample of learner language at an intermediate level. All essays were processed using the TagAnt tool, a widely recognized POS tagging software in corpus linguistics, and the output was manually verified to identify tagging errors. Manual verification involved checking each token individually against its correct grammatical

category, ensuring an accurate reference for calculating error frequencies and tagging accuracy.

The analysis is guided by Interlanguage Theory, which posits that second-language learners develop a dynamic linguistic system that is distinct from both their first language (L1) and the target language (L2). Errors in learner language are not random; they are systematic and often reveal the learner's stage of language development. Accordingly, POS tagging errors in learner essays may arise not solely from software limitations but also from deviations inherent in interlanguage development.

6.1. Corpus Description

The learner corpus in this study consists of fifteen essays written by ESL learners on various topics, including school experiences, hobbies, personal routines, and social observations. Each essay is relatively short, with 175–210 words, which makes the dataset manageable for detailed analysis while still providing a broad range of linguistic examples. Table 1 presents an overview of the essays and their approximate token counts.

Essays	Topic	Tokens
1	First day at school	210
2	Favorite hobby	185
3	Zoo visit	190
4	Best friend	175
5	Daily routine	200
6	Speech competition	195
7	Favorite food	180
8	Important life lesson	185
9	Favorite teacher	210
10	Trip to Murree	200
11	Importance of education	190
12	School library	185
13	Social media	180
14	Time management	185
15	Future Aim	20

Table No 1

Total tokens ≈ 2,870

These essays represent authentic learner writing, containing natural deviations from standard English grammar. Typical ESL errors observed include verb tense mistakes,

subject–verb agreement issues, missing articles, misused prepositions, and occasional lexical misapplications. These deviations provide a valuable opportunity to examine how automated POS taggers handle non-standard language and to identify systematic patterns of misclassification. For example, one essay describing a school trip contained the sentence “We goes to the zoo yesterday and see many animals”, which illustrates both tense and agreement errors.

6.2. Tagging Method and Manual Verification

All essays were processed using TagAnt, a widely used POS tagging tool in corpus linguistics. TagAnt assigns grammatical categories to each word based on standard English usage. After automated tagging, the output was manually compared to the correct grammatical tags for each word. Discrepancies between the automated and correct tags were marked as tagging errors. Manual verification involved examining every token individually and identifying errors such as:

- Misclassification of verbs (e.g., tense, agreement)
- Misclassification of nouns or pronouns
- Errors in adjectives, adverbs, and determiners
- Lexical errors where the expected POS differed from the automated tag

This step ensures the reliability of the analysis by providing an accurate reference for calculating error frequencies and tagging accuracy

6.3. Sample Tagging Errors

The analysis of the fifteen essays revealed that most tagging errors occurred in verb forms, particularly in tense and agreement. Other errors included lexical misclassification and confusion in noun and adjective tagging. Table 2 presents as

Word	Automated Tag	Correct Tag	Error Type
Waked	VBD	VBD (correct, manually checked)	–
Goes	VBZ	VBD	tense error
Wear	VB	VBD	tense error
Participate	VBP	VBD	tense error
Brings	VBP	VBZ	agreement error
Smell	VBP	VBZ	agreement error
Have	VBP	VBZ	agreement error
Fights	VBZ	VBP	agreement error
Goodbye	NN	UH	lexical misclassification
Effect	VB	NN	lexical error

Share	NN	VB	POS confusion
-------	----	----	---------------

Table No 2

These examples demonstrate the primary challenges automated POS taggers face when processing ESL learner texts. Verb forms are particularly problematic due to incorrect tense usage, missing auxiliary verbs, or subject–verb agreement errors.

6.4. Categorization of Tagging Errors

After identifying individual errors, the tagging inaccuracies were classified into broader categories to understand patterns. The following categories were used:

1. **Verb Tense Errors** – Instances where the automated tag misclassified the tense of verbs in past, present, or future contexts.
2. **Subject–Verb Agreement Errors** – Cases where verbs did not match the grammatical number of the subject, resulting in tagging discrepancies.
3. **Lexical Category Errors** – Words that were misclassified as a different POS due to non-standard usage or ambiguity.
4. **Structural/Missing Verb Errors** – Instances where sentences had incomplete verb structures, leading to misclassification.

Table 3 shows the frequency of each error type in the corpus.

Error Type	Frequency
Verb tense errors	31
Subject–verb agreement errors	18
Lexical category errors	9
Structural/missing verb errors	6
Total Errors	64

Table No 3

Observation: Verb tense errors represent the majority of tagging errors, highlighting the sensitivity of automated taggers to non-standard verb forms in ESL writing.

6.5. POS Categories Affected by Errors

The errors were also analyzed according to the POS category they affected. Table 4 summarizes the findings.

POS Category	Error Frequency	Observation
Verb (VB, VBD, VBZ, VBP, VBG)	49	High frequency due to tense and agreement issues
Noun (NN, NNS, NNP, NNPS)	9	Moderate frequency, often with plural or proper noun misclassification
Adjective (JJ, JJR, JJS)	3	Low frequency, occasional misclassification

Adverb (RB, RBR, RBS)	2	Very low frequency, minor misclassification
Other	1	Miscellaneous lexical errors

Interpretation: Verb morphology is the main source of errors, which aligns with patterns expected in ESL learner writing. Nouns and adjectives are less problematic, while adverbs and other categories rarely cause tagging issues.

6.6. POS Tag Frequency Distribution

Using AntConc, the frequency of POS tags across all essays was calculated. Table 5 presents the approximate counts for major POS categories.

POS Tag	Frequency
NNP (Proper Noun)	120
VB (Base Verb)	340
VBD (Past Verb)	260
VBZ (3rd person singular)	180
VBP (Non-3rd person singular present)	140
JJ (Adjective)	210
RB (Adverb)	160
IN (Preposition)	280
DT (Determiner)	250
PRP/PRP\$ (Pronouns)	300
NNP (Proper Noun)	120
VB (Base Verb)	340
VBD (Past Verb)	260

Table No 5

Observation: Verbs and nouns are among the most frequent POS categories in the learner corpus. This explains why verb misclassification has a significant impact on tagging accuracy.

Tagging Accuracy Calculation

The accuracy of the POS tagging system was calculated using the formula:

$$\text{Accuracy} = (\text{Correct Tags} \div \text{Total Tags}) \times 100\%$$

From the corpus:

- Total tokens = 2,870
 - Correctly tagged tokens = 2,806
 - Errors = 64
- $$\text{Accuracy} = (2806 \div 2,870) \times 100 \approx 97.8\%$$

Interpretation: The overall tagging accuracy is high. However, detailed analysis shows that most errors are concentrated in verb forms, which may affect downstream grammatical analysis if uncorrected.

6.7. Patterns of Errors

Verb Errors

- **Past Tense Mistags:** Tokens like *goes*, *wear*, *participate*, and *announce* were often tagged incorrectly due to learners using incorrect forms of verbs.
- **Agreement Errors:** Learners sometimes wrote verbs that do not match the subject, e.g., *have aim* instead of *has aim*. Automated tagging misclassified these.
- **Infinitive/Participle Confusion:** Errors occurred with verbs in non-finite forms, such as *speak*, *reading*, *writing*, and *play*, where the context required a different verb form.

Noun Errors

- Proper nouns were occasionally misclassified, particularly names of people (Ali, Ahmed) and places (Murree).
- Plural nouns were sometimes tagged as singular forms due to learner errors or unusual sentence structures.

Lexical Category Errors

- Some common words were tagged incorrectly due to non-standard usage in learner essays. For example, *goodbye* was tagged as a noun (NN) instead of an interjection (UH).

Structural/Missing Verb Errors

- A few sentences lacked auxiliary verbs or contained incomplete structures. These instances caused misclassification, particularly with modal verbs (can, should, may) or copular constructions (is happy, was tired).

6.8. Linking Errors to Interlanguage Development

Interlanguage Theory provides a framework for understanding why tagging errors appear frequently in learner texts:

1. Learners produce systematic deviations that are not random, such as incorrect tense or agreement patterns.
2. These deviations affect automated tagging because standard POS taggers are trained on native-speaker texts.
3. Tagging errors often reflect the learner's current stage of language development rather than a failure of the tagging system.

Example:

In Essay 1, the token *goes* was used in a past narrative context, but the automated system tagged it as present tense (VBZ). This reflects the learner's difficulty with past tense conjugation, which is common in early interlanguage development.

6.9. Summary of Findings

- The automated POS tagging system achieved a high overall accuracy of approximately **97.8%** across 15 essays.
- Most tagging errors were related to **verbs**, particularly tense and subject–verb agreement.
- Noun and adjective misclassifications were less frequent, and adverbs rarely caused errors.
- Lexical misclassification occurred in specific contexts, particularly with non-standard word use in learner writing.
- Errors observed align with predictions from **Interlanguage Theory**, highlighting systematic patterns in learner language rather than random mistakes
- The distribution of errors indicates that computational tools perform best with standard grammatical structures and are challenged by learner deviations.

6.10. Conclusion of Data Analysis

The analysis of fifteen ESL learner essays demonstrates that automated POS tagging tools can achieve high overall accuracy but face consistent challenges with learner deviations, particularly verb tense and agreement. The errors observed are systematic and reflect the learners' interlanguage development, which aligns with theoretical expectations. By carefully categorizing and quantifying errors, this study provides a detailed evaluation of tagging accuracy in learner corpora and highlights areas for pedagogical and computational improvement.

This comprehensive error-based computational analysis forms the foundation for connecting learner language characteristics with automated linguistic tools, providing both theoretical and practical insights for corpus-based ESL research.

7. Discussion

The findings of this study reveal that automated POS tagging tools, while generally accurate, face systematic challenges when applied to ESL learner corpora. With an overall accuracy of 97.8%, TagAnt performed well in tagging standard structures but struggled with deviations commonly found in learner writing. Verb forms were the most frequent source of errors, particularly tense misclassification and subject–verb agreement errors. This aligns with previous research (Heilman et al., 2014; Jurafsky & Martin, 2020) showing that verbs are highly sensitive to learner errors and pose difficulties for statistical tagging models trained on native-speaker data.

Noun and adjective errors occurred less frequently, suggesting that these categories are less impacted by interlanguage deviations, although pluralization and proper noun recognition occasionally caused misclassification. Lexical errors, though rare, highlight the impact of unconventional word use in learner texts on automated tools. Overall, these patterns underscore that while tagging tools are robust, ESL learner language presents systematic challenges due to interlanguage characteristics rather than random mistakes.

The frequency distribution of POS tags further supports these observations. Verbs and nouns dominate the corpus, which explains the disproportionate impact of verb misclassification on overall tagging accuracy. In addition, manual verification confirmed that the errors were largely predictable and aligned with interlanguage development patterns, such as overgeneralization of tense rules and missing auxiliary verbs. The combination of quantitative accuracy calculation and qualitative error categorization strengthens the validity of these findings.

Moreover, linking errors to Interlanguage Theory provides explanatory power. Learner errors are not random; they reflect systematic hypotheses and experimentation with English grammar. The tagging errors captured by TagAnt reveal learners' evolving linguistic system, demonstrating that computational tools can indirectly provide insights into language development. This confirms the theoretical argument that analyzing learner corpora with automated tools can illuminate both linguistic accuracy and interlanguage dynamics.

Finally, this study shows the potential of computational methods in second language research. Automated tagging facilitates rapid analysis of multiple essays, allowing researchers to quantify error patterns efficiently. However, the limitations of tools trained on native-speaker data suggest a need for careful evaluation and potential adaptation for learner language. These findings provide a foundation for targeted pedagogical interventions and inform the design of more robust tagging systems for ESL contexts.

8. Conclusion

The first research question was about identifying the most common types of part-of-speech tagging errors in ESL learner writing. The analysis of the learner corpus revealed that verb-related tagging errors occurred most frequently. In particular, verb tense misclassification and subject–verb agreement problems accounted for the majority of tagging inaccuracies observed in the data. These errors often appeared when learners used incorrect verb forms in past or present contexts, which caused the automated tagging tool to assign incorrect grammatical labels. In comparison, noun, adjective, and adverb misclassifications occurred less frequently. This pattern shows that verb morphology remains the most problematic grammatical area for both ESL learners and automated POS tagging systems.

The second research question was about the reasons automated POS taggers produce errors in learner language and how these errors relate to interlanguage development. The findings indicate that tagging errors occur mainly because automated tagging systems are trained on grammatically correct native-speaker texts. When learner writing contains grammatical deviations such as missing auxiliary verbs, incorrect verb forms, or unusual sentence structures, the tagging tool interprets these forms according to standard grammatical rules and may assign incorrect tags. These deviations are closely related to interlanguage development, where learners create a transitional linguistic system while learning the target language. As a result, tagging inaccuracies often reflect predictable patterns in learner language rather than random mistakes.

The third research question was about how computational analysis of tagging errors can contribute to a better understanding of ESL learner writing accuracy. The results demonstrate that computational analysis provides an efficient and systematic way to examine learner language patterns. By comparing automated POS tags with manually corrected tags, the study was able to measure tagging accuracy and identify grammatical areas where learners face difficulties. This approach allows researchers to process learner texts quickly while maintaining objective analysis of grammatical structures. Consequently, computational tagging analysis helps researchers gain deeper insight into learner writing and identify systematic patterns of language development.

The first research objective of this study was to identify common part-of-speech tagging errors in ESL learner essays using automated tools. This objective was successfully achieved through the analysis of the learner corpus processed with the TagAnt tagging tool. The results showed that verb tense errors and subject-verb agreement errors were the most common tagging problems. Other types of errors, such as lexical misclassification and noun-related tagging issues, occurred less frequently. The identification and categorization of these errors provide valuable information about the grammatical structures that are most challenging for both ESL learners and automated tagging systems.

The second research objective was to examine the reasons for errors produced by automated POS tagging tools when analyzing learner writing. The analysis showed that tagging errors mainly occur because automated systems are designed to process standard grammatical structures. Learner writing often includes deviations from these structures due to the developing nature of interlanguage. As a result, automated tools may misclassify certain words, particularly verbs, when they appear in incorrect grammatical forms. The findings therefore highlight the relationship between learner language characteristics and the performance of computational tagging systems.

The third research objective was to analyze the overall accuracy of automated POS tagging and identify patterns in learner language. The results show that the tagging system

achieved a high overall accuracy rate of approximately **97.8%**, indicating that automated tools can perform reliably when analyzing learner corpora. However, the detailed error analysis also revealed that certain grammatical categories, particularly verbs, remain challenging for automated tagging systems. These findings suggest that automated POS tagging tools are useful for corpus-based ESL research, but manual verification is still necessary to ensure accurate interpretation of learner language patterns.

8.1. Findings

The analysis of the learner corpus shows that automated part-of-speech tagging tools can achieve a high level of accuracy when applied to ESL learner writing. The tagging system used in this study produced an overall accuracy rate of approximately **97.8%** across fifteen learner essays containing about **2,870 tokens**. This result indicates that automated tagging tools such as TagAnt are generally reliable for analyzing learner corpora. However, a detailed examination of tagging errors shows that certain grammatical categories remain more challenging for automated systems, particularly when learner language deviates from standard English structures.

A major finding of the study is that **verb-related errors represent the largest proportion of tagging inaccuracies**. Errors involving verb tense and subject–verb agreement occurred most frequently in the dataset. These errors often appeared when learners used incorrect verb forms or when verbs did not agree with the subject in number. Because automated tagging tools rely on standard grammatical patterns, such deviations frequently resulted in incorrect tagging. This finding highlights that verb morphology is the grammatical category most sensitive to learner errors in automated tagging.

The analysis also revealed that **noun-related tagging errors and lexical misclassifications occurred less frequently than verb errors**. In some cases, proper nouns or plural nouns were incorrectly tagged due to unusual sentence structures or inconsistent capitalization in learner writing. Lexical category errors were observed when words were used in non-standard contexts, causing the tagging system to assign incorrect grammatical labels. Nevertheless, these types of errors were relatively limited compared to verb-related tagging problems.

Another important finding is that **tagging errors are closely connected to patterns of interlanguage development**. Many of the grammatical deviations found in the essays reflect the transitional linguistic system that learners develop while acquiring English. Errors such as incorrect verb tense usage, missing auxiliary verbs, and simplified sentence structures are typical features of learner language. Therefore, tagging inaccuracies not only reveal the limitations of automated tagging tools but also provide insight into the developmental stages of ESL learners. These results demonstrate that computational analysis can be a valuable method for identifying systematic patterns in learner writing and understanding the linguistic challenges faced by ESL students.

8.2. Future Recommendations

Future studies should analyze larger and more diverse ESL learner corpora to capture a broader range of grammatical patterns. Including learners of varying proficiency levels, linguistic backgrounds, and writing contexts would improve the generalizability of findings and provide a more comprehensive understanding of tagging accuracy across different learner populations.

Comparative evaluation of multiple automated POS tagging tools is another important direction. Assessing different systems would help determine whether observed errors are tool-specific or common across platforms and identify which tagging tools perform most reliably with learner language. Integrating advanced computational methods, such as machine learning models trained on learner data, could further improve tagging performance for non-standard grammatical forms.

Finally, future research should focus on the pedagogical implications of automated tagging analysis. By identifying systematic learner errors, teachers could design targeted instructional strategies and materials to address common difficulties. Additionally, examining the connection between tagging errors and stages of language development would provide deeper insight into interlanguage patterns, contributing to both second language acquisition research and computational linguistics.

References

- Ali, M., Khan, M., & Alharbi, Y. (2024). *A conditional random field based approach for high-accuracy part-of-speech tagging using language-independent features*. PeerJ Computer Science.
- Andersen, G. (2025). *The importance of context in part-of-speech tagging: enhancing accuracy and understanding*.
- Anthony, L. (2023). *AntConc 4.1: A learner and researcher-friendly corpus analysis toolkit*. Waseda University.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
- Expert Systems with Applications. (2025). *Comparison of deep learning and machine learning models for POS tagging*, 288, 128026.

- Flowerdew, J. (2015). Corpus-based perspectives on second language acquisition and writing. In J. Flowerdew & M. Brock (Eds.), *Academic writing for graduate students* (3rd ed., pp. 15–36). University of Michigan Press.
- Gardazi, N. M., Daud, A., Malik, M. K., et al. (2025). *BERT applications in natural language processing: a review*. *Artificial Intelligence Review*, 58, 166.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). John Benjamins.
- Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2014). Part-of-speech tagging for learner English. *Language Resources and Evaluation*, 48(1), 1–24.
- Ishaq, M., Asim, M., Ahadullah, et al. (2025). *Development of a novel Urdu language part-of-speech tagging system for improved Urdu text classification*. *Social Network Analysis and Mining*, 15, 58.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed., draft). Pearson.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
- Raja, K. R., Chaurasia, P. K., & Midhunchakkaravarthy. (2025). *Evaluating machine learning and deep learning techniques for POS tagging in morphologically rich languages*. *International Journal of Basic and Applied Sciences*.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209–231.
- Tetreault, J., & Chodorow, M. (2008). The ups and downs of preposition error detection in ESL writing. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 865–872.