Utilising Natural Language Processing to Assist ESL Learners in Understanding Parts of Speech

Satwika Nindya Kirana, Yash Munnalal Gupta

Abstract: *The field of Natural Language Processing (NLP) is growing rapidly, as is the number of companies investing in this technology. NLP is changing the way we learn and teach languages. However, it has not been used in ways that benefit ESL (English as a Second Language) educators. It is implied that there is a gap in the application and the use of an NLP tool that focuses on English Part of Speech (POS) analysis to aid the English teaching and learning process. Herein, in this paper, we discuss the prospect of utilising POS analyser to accommodate ESL educators in teaching English POS. The tool development is divided into two sections: 1) the development of a POS analyser; and 2) the implementation of an interface to make the tool become a user-friendly application. We use SpaCy, an NLP opensource library, for the English POS analysis. It offers both statistical and neural network models. It also comes with pre-trained models that can predict the POS. This paper also provides Graphical User Interface (GUI) tool that can be used to create effective and engaging English language teaching materials for learners. GUI tool is created using python programming language. Thus, we first review the NLP-based applications for ESL education, followed by an introduction and overview of our simple POS analysis tool, which is customizable. In the future, we intend to evaluate our tool with the help of ESL educators who are not computer scientists or linguists. The python script used to develop tool is provided at Github: https://github.com/yashmgupta/literate-robot*

Keywords: *English as a Second Language (ESL), Natural Language Processing (NLP), Part of Speech (POS), Syntax*

I. INTRODUCTION

English grammar is a set of rules that describe the structure of expressions in the English language. This includes the structure of words, phrases, clauses, and sentences, right up to the structure of whole texts. One way to begin grammar instruction is through parts of speech (POS) categories. Nouns, pronouns, adjectives, verbs, adverbs, prepositions, conjunctions, and interjections are the main categories. Each POS performs a different function within a sentence [1]. An understanding of these will help to increase writing fluency for English as a Second Language (ESL) learners.

Satwika Nindya Kirana, Business Management and Languages, Faculty of Management Science, Silpakorn University, Petchaburi, 76120, Thailand. Email: nindyakirana_s@silpakorn.edu
Dr. Yash Munnalal Gupta*, Department of Biology, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand. Email: yashmunnalalg@nu.ac.th

It is also a great way to improve ESL learners' ability to converse with others in English. The study of POS is referred to as syntax, and is one factor in determining the grammatical correctness of a sentence. ESL learners, especially the beginners, often struggle with what POS are, and how they are used in the English language [2]. An important part of teaching English as a second language is ensuring that the learners understand the importance of grammar structure, and POS is one of the foundations in it. The importance of teaching parts of speech to ESL learners has been well-documented in the literature. Muchtar [3] claimed that by mastering nine types of POS, learners can more easily master the four language skills of listening, reading, writing, and speaking. In the same line, Haslam [4] argued that it is important for English educators and learners to comprehend the characteristics of each POS. A study conducted by Ruiz Quizhpe [5] also shows that identifying the vocabulary POS can help learners improve their English skills, especially in reading. Thus, it is important that learners are taught the different types of POS, how to use them in a sentence structure, and how they are combined to create a coherent message. Although many English educators do identify POS in sentences and even teach them during class, there is a lack of understanding and knowledge within many ESL learners as to how these POS are used in a sentence and their function and place within a sentence [6]. Therefore, this paper provides a tool to introduce English POS analysis to the ESL learners with ease, and how it can be used by ESL educators to help their learners learn English grammar more easily.

II. LITERATURE REVIEW

A. ESL Learners' Difficulties in Understanding Part of Speech

POS has been found as one of the most important aspects when learning a second language [7, 8]. Teaching POS is one of the most basic, yet challenging tasks in teaching English as a second language. Some difficulties have always been faced by ESL learners in understanding POS due to different grammatical systems between their first language (L1) and the English language. Several studies have been conducted to ESL learners with different first languages, and the results consistently show that their L1 grammar rules interfere their understanding of English grammar, including the POS [9-11]. There are some strategies that can be used to teach POS in more effective ways [12-14]. However, the success of the strategy applied depends on many factors, such as the capability of educators and the learning ability of learners. Consequently, a teaching tool often needs to be used for teaching POS for ESL learners.

One of the effective ways to teach POS is by using visual to identify POS in different types of sentences. Studies conducted by Cabrera, Castillo [15] and Purgina, Mozgovoy [16] show that visuals help ESL students understand English grammar more easily. Hence, we attempted to create a visual tool to assist educators in demonstrating POS in English sentences. The tool uses natural language processing (NLP) to extract information from sentences. This tool's output can be used in the English teaching and learning process.

This study was carried out with a view to discovering how educators can integrate NLP tools into their lessons. Although NLP tools are often viewed as being useful for language learning [17], they are not widely used by educators in the classroom setting, yet. This is due to several factors, including the fact that many NLP tools are complex and difficult to use [18], and may not be suitable for integration into an ESL class setting. In order to address this problem, we present a simple tool which can be used by educators to teach parts of speech. This tool can be customised, and is based on an existing NLP algorithm which has been proved to be effective at identifying words that belong to different POS.

In addition, many of the available NLP tools require educators to be familiar with, or to have background knowledge on NLP tools, in order to use them effectively [19]. Therefore, we propose using a new method which uses simple syntax rules and allows educators to customize their lessons based on the learners needs without necessarily needing to have any knowledge of NLP tools.

B. Natural Language Processing and English Syntax

English syntax is a part of natural language processing, which is an area of artificial intelligence. It describes the rules that define the structure of English sentences. NLP is a field that has seen a lot of growth in recent years, but much of the work done falls short of its potential for teaching languages. The promise of NLP is that it can be used to simplify the teaching and learning of English syntax. [20, 21] One specific area where NLP can help teachers is with simplifying the analysis of syntax for non-native English speakers. NLP is used in a variety of applications, including machine translation and data mining. A more basic use of NLP is to help English educators improve their teaching methods. One way this can be accomplished is by analysing the syntax of the written form of a student's assignment.

There are several tools used for natural language processing in Python, including Natural Language Toolkit (NLTK) [22] and SpaCy [23]. These libraries can be used not only to build applications like chatbots and speech recognition systems, but they are also useful for analysing text. NLTK provides a rich suite of methods and objects that can be used to build NLP applications. In particular, NLTK is ideal for text analytics and text mining applications. However, NLTK can be cumbersome for English educators who wants a quick way to set their project without getting bogged down in scripting. On the other hand, SpaCy has fewer libraries but is faster and easier to use. SpaCy is a more modern toolkit that is better suited for developing applications.

The GUI tool created in this paper attempts to do this by focusing on English syntax analysis, which saves time and helps educators focus on what they need to teach rather than

spending time understanding the technicalities of syntax analysis. GUI application uses SpaCy for syntax analysis and Streamlit (https://streamlit.io) for visualization. SpaCy comes with a command line interface for training models and performing various NLP analysis tasks from the terminal or scripts. The GUI application Streamlit incorporates all these capabilities into one easy-to-use application.

English syntax structure is widely used in natural language processing as a type of computer grammar. Consequently, the English syntax structure and other similar grammars have been used to help computers analyse and process natural language. However, only few tools have been created that focus on this specific use case, and those that have been created tend to focus only on the syntactic elements of English syntax structure while ignoring the semantic factors. The GUI tool uses pretrained machine-learning algorithms to predict POS tags in text and then outputs a parse tree.

The parse tree produced can be used by ESL educators to teach learners how syntactic parsing works. Another possible application of POS taggers in educational field is automatic feedback on student writing assignments. Learners can upload their writing works to the tool, and then receive feedback on their use of POS. By implementing such solution, the feedback given will be more precise and objective than human evaluation. The next direction is POS exercise scoring in the classroom. In this case, the POS analyser can be used to score written or oral language exercises. The fourth possible application is to teach ESL learners the use of POS tags. This task is usually difficult, tricky, and time-consuming for educators. With the help of the POS analyser, it can be made much easier. It also provides a more efficient language practice.

III. DEVELOPMENT PROCESSES OF POS ANALYSER

The tool development falls into two parts: the first was to develop a part-of-speech (POS) tagger, and the second was to implement an interface that incorporates all these capabilities into one easy-to-use application. SpaCy NLP python library was employed to tag POS. SpaCy has pre-trained statistical models and word vectors used for POS analysis. The model used for development is available from the following link: https://github.com/explosion/spacy models/releases/tag/en_core_web_sm-3.2.0

Specifically, the main command line interface was implemented in Python, while the additional graphical user interface (GUI) was developed using Streamlit. Streamlit is an open-source Python library that makes it easy to build custom web-apps for machine learning and data science. The code snippets of POS application are deposited on GitHub and made available for public use (https://github.com/yashmgupta/literate-robot).

The Fig. 1 (a) illustrates SpaCy implementation and demonstrates how the POS analyser happens in practice. Fig. 1 (b) illustrates the execution of the Streamlit library, which was used to create a web application for POS visualization.

13

When an input is given in the form of English sentence, the POS analyser splits the sentence and tags the words with predicted POS. Furthermore, the information produced by the POS analyser can also be used for further research works on NLP and machine learning.
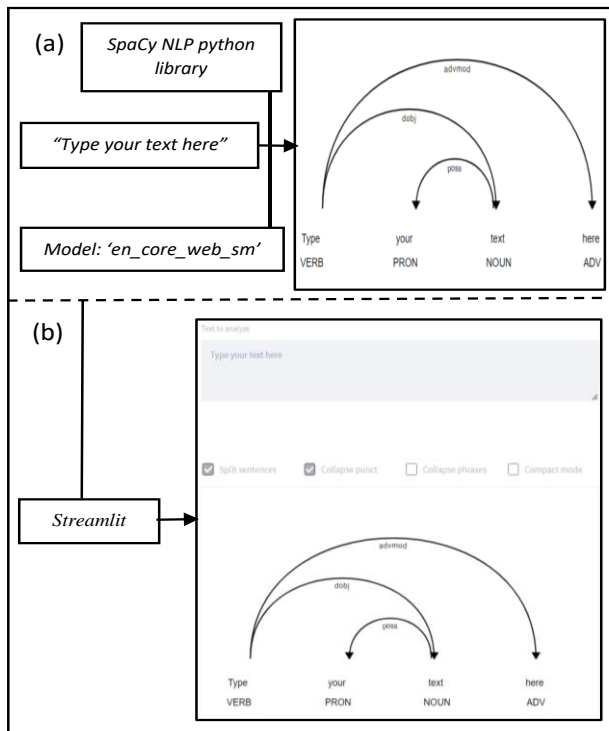

Fig.1. (a) POS analysis using 'en_core_web_sm' model from SpaCy library. (b) POS visualization using Streamlit

IV. FUTURE PROSPECTIVE POS ANALYSER

In this paper, we have presented a simple POS analyser that can be used by ESL educators and learners to assist and assess test their grammatical knowledge. In the future, we will enhance this tool with the following features: (1) Text to speech: Text to speech feature will help learners to listen to English words and their correct pronunciation. (2) Addition of new features for teaching grammar based on SpaCy and NLP. (3) Speech to text: Speech to text feature will be added to improve ESL learners' English pronunciation and to assist them in English spelling.

V. CONCLUSION

In summary, we have created a simple tool to help ESL educators incorporate POS identification into their lessons. This tool is easy to use and will help both educators and learners to get a better understanding of English POS without requiring them to have a background knowledge in NLP. This simple tool can also be easily modified by the educators to meet their needs in teaching English to ESL learners and can be used for various English teaching and learning purposes.

ACKNOWLEDGMENT

REFERENCES

1. Dalby, J. and D. Kewley-Port, Explicit pronunciation training using automatic speech recognition technology. *CALICO journal*: 425-445, 1999.
2. Hafiz, S., et al., Analysis of syntactic errors in English writing: A case study of Jazan University preparatory year students. *Journal of Education and Practice*. **9**(11): 113-120, 2018.
3. Muchtar, N., Nine Parts of Speech in Teaching English to Electrical Engineering Students. 2017.
4. Haslam, M., Parts of Speech. *The TESOL Encyclopedia of English Language Teaching*: 1-8, 2018.
5. Ruiz Quizhpe, R.Y., *Classifying Vocabulary into Parts of Speech to Improve Reading Comprehension*. 2019, Universidad Casa Grande. Departamento de Posgrado.
6. Djurayeva, Y., et al., Current problems and resolutions of teaching english grammar. *Academic research in educational sciences*,(3): 572-576, 2020.
7. Richards, J.C., *Error analysis: Perspectives on second language acquisition*. 2015: Routledge.
8. Cook, V., *Second language learning and language teaching*. 2016: Routledge.
9. Cabaruan, D.J.C. and I.M. Cue, Grammar proficiency and first language interference in learning English among SASTE students of St. Paul University Philippines. *SPUP Research Digest*. **21**(1), 2018.
10. Budiharto, R.A., Native language interference on target language writings of Indonesian EFL students: An exploratory case study. *Indonesian EFL Journal*. **5**(1): 107-116, 2019.
11. Masood, M.H., et al., Interference of L1 (Urdu) in L2 (English) in Pakistan: Teaching English as a second language. *International Journal of Applied Linguistics and English Literature*. **9**(5): 110-118, 2020.
12. Supakorn, P., et al., Strategies for Better Learning of English Grammar: Chinese vs. Thais. *English Language Teaching*. **11**(3): 24-39, 2018.
13. Chu, H.-C., et al., Impacts of concept map-based collaborative mobile gaming on English grammar learning performance and behaviors. *Journal of Educational Technology & Society*. **22**(2): 86-100, 2019.
14. Juniar, R., *A survey of grammar learning strategies used by EFL learners in Indonesia*. 2019, Universitas Islam Indonesia.
15. Cabrera, P., et al., The Impact of Using" Pixton" for Teaching Grammar and Vocabulary in the EFL Ecuadorian Context. *Teaching English with Technology*. **18**(1): 53-76, 2018.
16. Purgina, M., et al., WordBricks: Mobile technology and visual grammar formalism for gamification of natural language grammar acquisition. *Journal of Educational Computing Research*. **58**(1): 126-159, 2020.
17. Meurers, D. and M. Dickinson, Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*. **67**(S1): 66-95, 2017.
18. Kraif, O., et al. *NLP tools for CALL: the simpler, the better*. in *InSTIL/ICALL Symposium 2004*. 2004.
19. Meurers, D., Natural language processing and language learning. *Encyclopedia of applied linguistics*: 4193-4205, 2012.
20. Chowdhary, K., Natural language processing. *Fundamentals of artificial intelligence*: 603-649, 2020.
21. Chopra, A., et al., Natural language processing. *International journal of technology enhancements and emerging engineering research*. **1**(4): 131-134, 2013.
22. Bird, S. and E. Loper. *NLTK: the natural language toolkit*. 2004. Association for Computational Linguistics.
23. Vasiliev, Y., *Natural Language Processing with Python and SpaCy: A Practical Introduction*. 2020: No Starch Press.

AUTHORS PROFILE

Satwika Nindya Kirana is an Indonesian EFL lecturer at Silpakorn University, Thailand. She graduated from Yogyakarta State University with a Bachelor of Education in English in 2014, and Naresuan University with a Master of Arts in English in 2018. She has been teaching English in Indonesia and Thailand for approximately nine years and has been actively conducting research on English education. She has published four papers on English ideologies, English translation, and English idioms. Her research interests include English Language Teaching, Teaching English as a Foreign Language, World Englishes, and Language Ideologies.

Dr. Yash Munnalal Gupta is lecturer at Naresuan University, Thailand. He is detail-oriented researcher with hands-on experience performing computational and laboratory work to advance research goals. Highly proficient in conducting comprehensive literature reviews and computational research support and preparing samples and specimens for testing. Adept at prepping resources, equipment and materials for research. Extensive background in investigating population genetics.