

# **CSE-431**

Name: GM Mohaiminuzzaman Apurbo

ID: 20301100

Section: 01

## **Paper Title**

Recognising the English Language using Context Free Grammar with PyFormlang

## **Paper Link**

[Recognising the English Language using Context Free Grammar with PyFormlang  
| IEEE Conference Publication | IEEE Xplore](#)

## **1. Summary**

### **1.1 Motivation**

The goal of this work is to develop an English language recognizer using context-free grammar (CFG) and the Pyformlang package. The preprocessing and parsing of sentences to ascertain their validity in the English language is the main topic of this study. This method advances the field of natural language processing (NLP) by enabling machines to comprehend and analyze complicated human languages. The article offers a comprehensive collection of English language components and rules, as well as an overview of Pyformlang's capabilities and benefits. According to reports, the proposed approach has an accuracy of 84.90%.

### **1.2 Contribution**

In this paper, the authors make several contributions. Firstly, they demonstrate the usage of the Pyformlang library for English Language recognition, which is helpful for beginners to understand the parsing of sentences using formal language concepts [2a]. They provide a detailed set of possible rules for the English Language [2b]. The authors also discuss the advantages of Pyformlang over other libraries and its main functionalities [2c]. They lay out the rules for lexical parsing of the English language to ensure reusability of the process [1a]. Additionally, the authors define a custom English CFG with parts-of-speech (POS) tags as constituents using Pyformlang library [5]. The CFG has a start symbol, terminals and non-terminals, and productions, and is used for parsing the English language [1b]. This approach has an accuracy of 84.90% [1c].

### **1.3 Methodology**

The methodology of this paper involves using context-free grammar (CFG) and the Pyformlang library to develop an English language recognizer. The first step is to pre-process and parse a statement to check if it is a legitimate sentence or not. The CFG is defined with a start symbol, terminals, and non-terminals, and the Pyformlang library is used for implementation. The paper provides a detailed set of rules for lexical parsing of the English language, ensuring reusability of the process. The implementation involves using the Pyformlang and nltk libraries for defining the CFG and parsing the sentences. The recognizer is tested on a custom dataset and achieves an accuracy of 84.90%. The process of parsing involves analyzing the sentence with respect to its grammatical constituents and syntactic relations. The semantics of the sentence are not captured in this phase. Overall, the methodology involves using CFG and Pyformlang to parse English sentences and check for their validity based on the defined grammar.

### **1.4 Conclusion**

The English language recognizer, which was developed with the Pyformlang package and a customized Context-Free Grammar (CFG), has an accuracy of 84.90% for parsing English sentences, according to the paper's conclusion. The efficiency of the CFG in identifying syntactically correct phrases is demonstrated by the metrics used to evaluate the recognizer, including accuracy score, precision score, recall score, and f1\_score. The paper also discusses Pyformlang's possible uses in a number of computer science domains and emphasizes the benefits of utilizing CFGs in language recognition. To assure reusability, the study also offers a comprehensive set of rules for the lexical parsing of the English language. All things considered, the study advances knowledge about sentence parsing using formal languages and functions as a useful teaching aid for beginners.

## **2. Limitations**

### **2.1 First Limitation:**

Complexity of CFG: Context-free grammars can become quite complex, especially for languages with highly ambiguous and nuanced syntactic structures like English. Creating an accurate and comprehensive CFG for English requires careful consideration of all possible sentence structures, which can be time-consuming and prone to human error. Additionally, as the size of the CFG increases, the parsing process becomes computationally expensive and can lead to slower performance.

### **2.2 Second Limitation:**

Lack of semantic analysis: While CFG and Pyformlang are effective in syntactic parsing and checking the grammatical correctness of sentences, they do not consider the semantic meaning of the sentences. Semantic analysis, which involves understanding the meaning and intent behind the language, is crucial for more advanced natural language processing tasks like sentiment analysis, question answering, and text summarization. Without incorporating semantic analysis, the CFG-based approach may not be suitable for applications that require deeper understanding of the language.

## **3. Synthesis**

This paper discusses the use of context-free grammar (CFG) and the Pyformlang library for English language recognition through natural language processing (NLP). The authors highlight the importance of NLP in enabling machines to understand complex human languages. They explain that the first step in NLP is pre-processing and parsing a statement to determine if it is a valid sentence. To achieve this, they utilize automata theory and CFG, presenting a custom English CFG with parts-of-speech (POS) tags as constituents. The Pyformlang library is used for implementation, providing operations for working with CFGs. The paper

outlines the advantages of Pyformlang over other libraries and provides the set of rules for lexical parsing of the English language. Through their approach, they achieve an accuracy of 84.90% in English language recognition. The paper also mentions previous research on English language parsing and provides a detailed set of possible rules for the English language. Overall, the paper aims to demonstrate the usage of Pyformlang for English language recognition using CFG and provide a practical learning tool for understanding the concepts of formal languages.