Journal of Research and Innovation in Language
ISSN (Online): 2685-3906, ISSN (Print): 2685-0818
DOI: https://doi.org/10.31849/reila.14975
Vol. 6, No. 2, August 2024, pp. 125-142

Developing an Android Application for Analyzing Indonesian Syntax: A Rule and Probability-Based POS Tagging Approach

Herpindo Herpindo [1*], Ristiyani Ristiyani [2], Miftahula Rizqin Nikmatullah [3] & Ratih Ngestrini [4]
[1,3] Universitas Tidar, Magelang, Indonesia
[2] Universitas Muria Kudus, Kudus, Indonesia
[4] Utrecht University, Utrecht, Netherlands
herpindo@untidar.ac.id

ARTICLE HISTORY

Received : 2023-07-05
Revised : 2023-07-16
Accepted : 2024-06-29

KEYWORDS

Android Applications
Digital Language Learning
Grammatical Corpus
Indonesian Syntax
Lexical Categories
POS Tagging

ABSTRACT

The investigation of the grammatical level of syntax, particularly Indonesian, focuses exclusively on sentence formation and does not include the corpus. This renders Indonesian grammatical corpus data less relevant in corpus-based grammatical investigations. The study offers a detailed overview of the utilization of Android applications based on POS Tagging data. The method in this study was qualitative focusing on the development of an application that utilizes Rule and Probability-based POS Tagging data from Leipzig Indonesian Mix_2013 to determine the categories, functions, and roles of Indonesian syntax with lexical categories including V (copula, existence, and equative) as the predicate potential on function. The application was designed to be compatible with the Android system by Integrating POS tagging into the System Development Life Cycle (SDLC), enabling wider accessibility to a larger user base. The result of this research introduces a program designed as a tool to search syntactic categories in Indonesian. The program uses a sequential search technique, which is a linear search method, to make it easier for users to find specific syntactic functions. By applying syntactic categories and functions using POS Tagging data from the Leipzig Indonesian Mix_2013 corpus, the study achieved significant insights into the roles of Indonesian syntax. POS Tagging based on the generated rules and probabilities achieves an accuracy rate of 92.53% for category tags and functions.

1. Introduction

Language in the digital era is closely related to the demands of the language community. The language that was traditionally presented conventionally has now undergone a paradigm shift towards language digitalization. The process of digital language transformation from a simple linguistic form to a more complex domain using applications has also been utilized by micro-linguistic studies that examine structures ranging from phonemes, morphemes, words, phrases to sentences. The demand for digitalization in the micro-linguistic domain by the Language Agency has already begun with the launch of the SIPEBI application, which involves linguistic contributors accessing language resources using the KBBI (Indonesian Dictionary) database. The features available in SIPEBI are more inclined towards language error analysis (Ramliyana et al., 2022).

Specifically, the utilization of applications that refer to the grammar of the Indonesian language, which has been developed, still lacks an adequate design. Several factors and linguistic paradigms, as well as micro-linguistic studies, tend to focus only on the internal aspects of language and do not consider contextual language phenomena.

The development of digital-based applications is still dominated by applied linguistic studies, especially language learning (Syamsiah, 2017). Teaching Indonesian grammar is also a strategic plan in the digital field, considering the rapid competition of languages in the context of the fourth industrial revolution. The strategic importance of teaching Indonesian grammar in the digital realm is driven by the competitive nature of languages in the fourth industrial revolution (Changpueng & Patpong, 2021; Rianto, 2020). The advent of digital technology has had a profound impact on language learning, prompting

applied linguistic studies to harness digital platforms for creating innovative tools and applications. Indonesian grammar holds a significant position in this context due to its widespread use, its role as a lingua franca in Southeast Asia, and Indonesia's growing economic influence. By integrating Indonesian grammar into digital language learning applications, applied linguistic studies meet the rising demand for language skills in an interconnected world, equipping learners with valuable linguistic abilities for diverse professional domains and cross-cultural interactions.

In the modern age, the progress of digital technology has profoundly impacted various aspects of our lives, including the advancement of digital-based applications. One field that has greatly benefited from these advancements is applied linguistic studies, with a notable focus on language learning. Among the many languages being taught and studied through digital platforms, the instruction of Indonesian grammar has emerged as a strategic pursuit within the digital realm. This emphasis on teaching Indonesian grammar is motivated by various factors, including the escalating competition among languages in the context of the fourth industrial revolution (Safara et al., 2019). With the widespread utilization of digital platforms and the growing interconnectedness among people worldwide, there is an increasing demand for language learning applications that cater to diverse linguistic needs. Applied linguistic studies have acknowledged this demand and responded by utilizing digital technologies to create innovative tools and applications for language learning. These digital-based applications offer a wide range of interactive and immersive experiences, enhancing the engagement, accessibility, and effectiveness of language learning for learners (Ho, 2023; Olivia & Shaklein, 2019).

Amidst this landscape of language learning applications, the teaching of Indonesian grammar occupies a significant position. Indonesian, as the official language of Indonesia, is spoken by over 270 million individuals in the country and serves as a lingua franca across various Southeast Asian regions. Furthermore, due to Indonesia's strategic geographical location and its increasing economic prominence, there has been a steady rise in global interest in learning the Indonesian language.

Acknowledging the potential of digital platforms in facilitating language learning, applied linguistic studies have embraced the development of digital-based applications dedicated explicitly to teaching Indonesian grammar. These applications incorporate interactive exercises, multimedia resources, and gamified learning techniques to provide learners with a comprehensive understanding of Indonesian grammar rules, structures, and usage. By leveraging digital technologies, these applications can offer personalized learning experiences, adapting to the individual needs of learners and providing them with targeted feedback and guidance.

Accoding to (Ratnawati et al., 2021) the emphasis on teaching Indonesian grammar in the digital domain is further reinforced by the rapid competition among languages in the context of the fourth industrial revolution. As this revolution unfolds, characterized by advancements in automation, artificial intelligence, and digital connectivity, the significance of languages in various professional fields becomes increasingly pronounced. Businesses, organizations, and individuals are increasingly recognize the importance of language skills to effectively navigate the globalized and digitally connected world.

Indonesian, with its growing economic significance and regional influence, is no exception to this trend. Proficiency in Indonesian grammar equips learners with the linguistic tools necessary to engage in international trade, diplomatic relations, tourism, and cultural exchanges. By incorporating Indonesian grammar into digital language learning applications, applied linguistic studies contribute to meeting the needs and demands of learners who seek to acquire language skills that are relevant and valuable in the present and future job markets (Sulistyo et al., 2021).

In conclusion, the development of digital-based applications in Applied Linguistic studies, particularly language learning, is flourishing in today's digital era. Within this context, the instruction of Indonesian grammar holds a prominent position due to its strategic importance in the face of the fourth industrial revolution. These digital applications provide learners with engaging and effective opportunities to master Indonesian grammar, equipping them with valuable language skills for success in an increasingly interconnected and competitive world.

2. Literature Review

The role of grammar applications that provide digital information needed by researchers, teachers, and academics needs to be realized to support language research and language teaching activities at the educational unit level. The development of Indonesian grammar applications is crucial at present, considering that several aspects in Indonesian syntax have not been extensively studied and introduced in Indonesian language. The use of corpus data as the basis for creating rules and probabilities in Indonesian syntax as well as regional languages has been carried out by Herpindo et al., (2022) through the stages of collecting categories, functions, and roles in Indonesian syntax. The rule-based data is derived from categories, while the probability is based on semantic roles. The use of programming data for studying categories and functions has also been conducted in research on regional languages Dewi & Ubaidi (2020); Pramudita et al., (2016); and Widhiyanti & Harjoko, (2013).

The development of an application related to language and the Indonesian language is described in the research conducted by Bustomi et al., (2018). The

study focuses on the development of an Indonesian language learning application aimed at introducing the basics of the Indonesian language. The objective of this research is to introduce the fundamentals of the Indonesian language using multimedia. The application is built using Adobe Flash CS 3. The result of developing this application utilizes an attractive interface, allowing the application to display text, audio, and images while providing direct interaction in the learning process. The application also features an appealing user interface, allowing teachers and students, to utilize the application due to its attractive media support freely.

Several challenges have been encountered in developing this application, such as the lack of a specific foundation on the basics of the Indonesian language that the learners may not well understand. The application primarily focuses on an attractive user interface rather than providing an in-depth analysis of the fundamentals of the Indonesian language, which is equally important in the teaching and learning process. Additionally, the application is not yet registered on the Play Store or App Store.

The development of an Android-based application, which was carried out by Yunus (2020), is motivated by the lack of interest in learning the Indonesian language among junior high school students due to the influence of gadgets. This application is built using Eclipse Juno software, SQLite, and implemented in the rapidly evolving Android mobile operating system. Indiana and Ramadhani (2019) conducted research that utilizes technological advancements to generate interest in learning the Javanese language. The development of this application harnesses Javanese language lexicon information targeting Javanese language learners and teachers. The application's design also presents the Javanese language in a polite form (Jawa Krama) and includes the appearance of affixes such as—ipun, dipun, and—aken. The implementation of this application aims to assist in Javanese language learning, provide practice exercises, and display Javanese language lexicons and proverbs.

The design of language applications has also incorporated algorithmic data, as seen in the research conducted by Khairul et al., (2018), who developed a Javanese language dictionary application using the RAITA algorithm. The purpose of creating this application is to digitally enhance Javanese language learning and its implementation in educational units.

The application of Indonesian language learning for speakers of other languages (BIPA) has also embraced technological applications. The digital-based implementation of BIPA learning has been carried out by Prasetyo, (2019); Purwono & Asteria, (2021); Rahmawati et al., (2020); Rizki & Meditanala, (2018) utilizing interactive BIPA materials and teaching resources, including grammar instruction, to introduce them to non-native speakers. Regarding using Android applications in these studies, the focus has primarily been on enhancing writing skills. BIPA learners need competence and proficiency in understanding Indonesian grammar to ensure that their writing performance is accompanied by strong grammatical abilities.

The utilization of a language error detection system, specifically covering punctuation marks and capital letters, has been carried out by Ilmiyah & Qoiriah (2021). This detection system adheres to the guidelines provided by PUEBI (Pedoman Umum Ejaan Bahasa Indonesia), which governs the rules of proper and correct Indonesian language usage. In their research, the Booyer-Moore algorithm was employed, which has proven effective in word or string search algorithms and was implemented as a search feature. The results demonstrated the effectiveness of the Booyer-Moore algorithm in error detection, with an average precision score of 0.969, an average recall score of 0.976, and an average accuracy score of 0.917.

The research conducted by Ilmiyah & Qoiriah (2021) has not been implemented in an Android-based application design. Several aspects need to be considered regarding the error detection system research that refers to PUEBI. This research has limited correction features as it cannot fully represent the rules of grammar and, from the user's perspective, it has a smaller user base compared to Android-based applications.

The implementation of POS tagging data in this research has implications for related studies, such as Römer (2011) research on the utilization of corpus linguistics in second language teaching. The primary goal of this article is to provide a comprehensive understanding of how corpora can be applied in language instruction. The article examines two approaches: indirect utilization for syllabus and instructional material design, and direct implementation in the language classroom. The article emphasizes the importance of integrating general and specialized language corpora in these applications.

Additionally, it underscores the necessity for further investigation in the field of applied corpus linguistics. However, a significant aspect that remains unaddressed in this study is the absence of utilizing specific corpora for instructional purposes. Cheng et al., (2003) conducted a study aiming to integrate Corpus Linguistics into an undergraduate English language program, despite the program being packed with courses. The authors justify the combination of two existing courses, Information Technology and Discourse Analysis, to introduce corpus-based language study and data-driven learning to students in a more systematic and meaningful manner. The study provides an overview of the course content, showcases research examples conducted by students, and assesses feedback from both students and instructors regarding this innovative teaching and learning approach.

The results suggest that incorporating corpus linguistics into the curriculum is feasible and valuable. Both students and instructors report that it enhances the original course's worth and offers them new roles in the learning process. Kaya's studies (2022), address the underutilization of corpora in language teaching and assessment, despite their extensive use in linguistic research. The narrative review aims to bridge this gap by providing valuable insights for practitioners and researchers. It explores the benefits and drawbacks of data-driven learning and the use of corpora in foreign language instruction, particularly in academic writing. The paper has three main objectives: (1) to explain data-driven learning and its potential impact on learning, (2) to show how learner corpora can support EFL learners in academic writing, and (3) to offer insights into the indirect application of corpora in teaching and assessing academic writing in L2. Through a comprehensive review of relevant corpus studies and their applications, this review achieves its goals.

Syahroni & Wahab (2019) developed an application utilizing the Natural Language Toolkit (NLTK) to identify Indonesian sentences' syntactic category and function. The application was modified to recognize six phrase levels: NP (Noun Phrase), VP (Verb Phrase), PP (Preposition Phrase), AP (Adjective Phrase), ADVP (Adverb Phrase), and NUMP (Numeral Phrase), achieving an accuracy rate of 93.2%. Furthermore, the study assigned syntactic functions to these six phrase levels in Indonesian sentences, including Subject, Predicate, Object, Modifier, and Complement. The accuracy rate for clause identification reached 91.12%. However, the research remains incomplete as it does not address the syntactic roles of Agent (A) and Patient (P).

The research gap in question arises from the literature review mentioned earlier, which did not extensively exploit language corpora within the realm of syntax, as it solely concentrated on the application of syntactic categories. As a result, this study aims to address this gap by employing Android-based POS Tagging data, thereby enhancing the investigation of corpus-based syntax and computation. This approach ensures a more comprehensive analysis, filling the existing research void.

2.1 Syntax Categories (Lexical Categories)

The lexical concept in syntactic studies refers to the patterns that fill syntactic functions within a clause or sentence. According to Sugiarti et al., (2021), the concept of syntactic categories in this context involves the lexical identities of Nouns (N), Verbs (V), Adjectives (Adj), Numerals (Num), and Prepositions (Prep). Jia & Liang (2020) categorize lexical groups in syntax as rules for forming functional patterns. Van Valin (2001) and Baker (2004) suggest that the essential aspect of lexical items is verbs as the essential property.

Kridalaksana (1988) provides a perspective on syntactic categories as groups or classes of words obtained by connecting lexical items with other elements in syntactic constructions. In line with this, in the Indonesian language, the main components in syntax consist of four categories: 1) verbs (V), 2) nouns (N), 3) adjectives (Adj), and 4) adverbials (Adv) (Raymondra & Bukhori, 2021).

In the study of Indonesian grammatical principles, the detailed examination of verbs in complex structures, especially copulative verbs, has been somewhat neglected. Unlike the Indo-Germanic languages, where copula verbs are essential and frequently appear in nominative slots to support syntactic functions, Indonesian does not require copula verbs in the same manner. Consequently, copulative verbs rarely appear in nominative positions within Indonesian syntax. Despite extensive research on various aspects of Indonesian grammar, the role and elaboration of verbs, particularly copulative verbs, in complex structures, has not received significant attention. This highlights the need for more focused research to enhance our understanding of Indonesian grammar.

In Indonesian, copulative verbs, which connect subjects and predicates, are not commonly used in nominative slots that support syntactic functions. This differs from Indo-Germanic languages like English, German, or Spanish, where copula verbs are essential for constructing complete sentences. In Indonesian, the absence of a copula verb does not hinder grammatical correctness. The non-obligatory nature of copulative verbs in Indonesian may account for the limited emphasis on their study. However, further investigation and analysis of copulative verbs in Indonesian could provide valuable insights into the language's syntax and sentence structure (Auni & Manan, 2023).

In essence, the Indonesian language encompasses sentences with copula predicates or what are referred to as other similar structures. This copula is a unique type of verb and is regarded as obligatory, as noted by (Bradley et al., 2022). The role and formulation of the copula in Indonesian sentences can be better understood through specific examples and rules, which are illustrated in (see Table 1.1). These examples highlight how the copula functions within various sentence structures, demonstrating its essential role in conveying meaning and maintaining grammatical integrity in the language. By examining these formulations, one can gain a deeper appreciation for the intricacies of Indonesian syntax and the importance of the copula in achieving clear and coherent communication.

Table 1.1 *Function of Copula in Indonesian*

Function	Sentence	English Translation
Copula +Pron	Yang saya maksud ***adalah*** ini	What I mean **is** this.
Copula + N	Ini ***adalah*** masalah mereka sendiri	This **is** their own problem.
Copula + Adj	Warna bajunya ***adalah*** hijau	The color of the shirt **is** green.
Copula + V	Perjuangan rakyat sedunia ***adalah*** saling membantu	The struggle of the people worldwide **is** to help each other.
Copula +F yang	Yang kupilih ***adalah*** yang terbaik diantara kesemua itu	What I choose **is** the best among all of them.
Kopula + Prep	Tujuan perubahan tanah ***adalah*** untuk memberikantanah kepada kamu tani yangtidak memiliki tanah	The purpose of land transformation **is** to provide land to farmers who do not have land.

As an additional reference, Sneddon et al., (2010) provide an illustration of copula verbs in Indonesian language using a tree diagram. Copula verbs in this context are correlated with linking verbs. This can be observed in the following tree diagram.

Figure 2.1 Copula Verb Tree Diagram

2.2 Function

As fillers of entities in clauses or sentences, copula verbs play a role in determining functional functions and positions. The Syntactic function is closely related to the subject and predicate. As lexical roles, the functions in a sentence specifically consist of subject (S), predicate (P), object (O), complement (Pel), and adverbial (Ket) (Moeliono et al, 2017).

Falk (2011) provides an additional insight into the functional elements of a sentence. He argues that in language, a clause or sentence must have at least an S (subject) and a P (predicate). It is worth noting that P, which is the lexical role of the verb, serves as the core or nucleus of a sentence or clause, making S and P obligatory elements (Tallerman, 2014).

The functions mentioned above are derived concepts from lexical categories that are related to function. The positions to be filled by these syntactic functions are realized in Indonesian language as constituents. According to Chomsky (2021) in the perspective of transformative grammar, constituents are labeled in active sentences following the syntactic pattern of NP1-V-NP2, where NP as S occupies the position before VP as the subject. The realization of this transformative grammar pattern can be seen in the modified form of the following Indonesian sentence in Figure 2.2

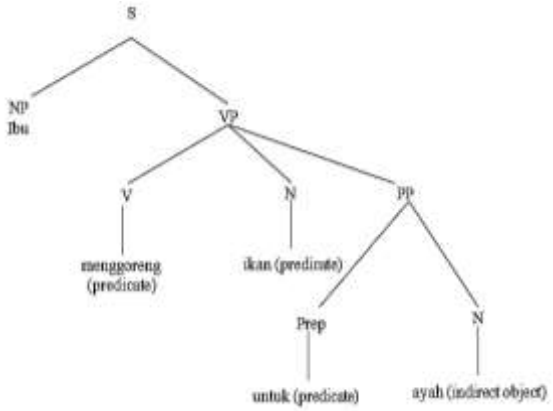

Figure 2.2 Diagram of Sentences in Transformative Grammar Function

According to transformational grammar, syntactic functions are related to rules that have various meanings. These rules serve as a means for categories in syntax to reconstruct different levels of syntactic elements (Chomsky, 2020). They provide guidance or

instructions for transforming from a categorical to a functional level, as follows:

1) Rule of Categorization: This rule assigns categories or labels to the constituents of a sentence, such as Noun Phrases (NP), Verb Phrases (VP), etc.
2) Rule of Phrase Structure: This rule determines the hierarchical structure of phrases and how they combine to form larger constituents within a sentence.
3) Rule of Transformations: These rules specify the operations or transformations that can be applied to the syntactic structure of a sentence to derive different surface forms or structures. Transformations include movement, deletion, insertion, and other operations that modify the syntactic arrangement of constituents.
4) Rule of Functional Interpretation: This rule assigns functional roles to the constituents of a sentence, such as subject, object, complement, modifier, etc. It determines how the syntactic categories relate to their semantic or functional roles in the sentence.

These rules guide the transformational process from categorical representation to functional interpretation, allowing for the construction of different syntactic structures with varying meanings. They form the basis for analyzing and generating sentences within the framework of transformational grammar.

According to Sneddon et al. (2010), they have a perspective regarding the rules proposed by Chomsky. Sneddon et al. (2010) introduce the concept of "function fillers" within categories. They add that copulative verbs, which do not fit into Chomsky's transformative grammar, can occupy the predicate position in syntactic functions, thus completing the components of the form filler rule. This is illustrated in the following table:

Table 2.2 Copulative Verb Distribution Syntax Function in Indonesian

No	Relation	Copula	Verb		
1	Copula with NP Predicate	[1]ø [2]*adalah/*ialah/*merupakan [3]*adalah/*ialah/*merupakan	[1]Copula' (x, [pred'] [2]Copula' (x,y) [3]Copula' (x,y)	[1]x= attributivelly, y=atributive [2]x,y=referen [3]x,y=referen	[1]Budi ø guru [2]Budi adalah guru [3]Anak Umar adalah mahasiswa Budi
2	Copula with AdJ P Predicate	[1]ø	[1]Copula' (x, [pred']	[1]x=attributivelly y=atribute	[1]Ani ø cantik
3	Copula Clause with PP	[1]ada/berada [2]ø	Copula' (x,y) [2]Copula' (x, [pred']	x,y=referen [2]x=attributivelly y=atribute	[2]Budi (ada/berada)

The position of a syntactic category in a sentence is determined by its syntactic function. The form filler, in this context, includes not only concrete verb categories but also copulative, equative, and existential verbs, as noted by Verhaar & Alip (1996). These verbs play essential roles in sentence construction by linking subjects to complements, establishing equality, or asserting existence. The detailed table below illustrates these relationships, providing a comprehensive overview of how different form fillers function within Indonesian syntax. This table helps to understand the diverse roles that verbs can play beyond simple actions.

Table 2.3 Relationship between Function and Category in Indonesian

Functions	Categories
Subject	Noun, Noun Phrase, Nominal Clause
Predicate	Verb, Equative, Copulative, and Existential
Object	Noun, Noun Phrase, Nominal Clause
Complement	mandatory presence
Adjunct	arbitrary

As an implication of developing a grammar system for syntactic categories and functions in this research and to assist in learning Indonesian language, particularly Indonesian grammar, digitalization is necessary to facilitate the identification and study of these syntactic categories and functions

The previous research conducted by Herpindo et al., (2022) with POS Tagging data on the Function and Syntactic Category serves as a foundation for level of 92.53% in terms of syntactic categories and functions. This accuracy level can be observed in the graph.

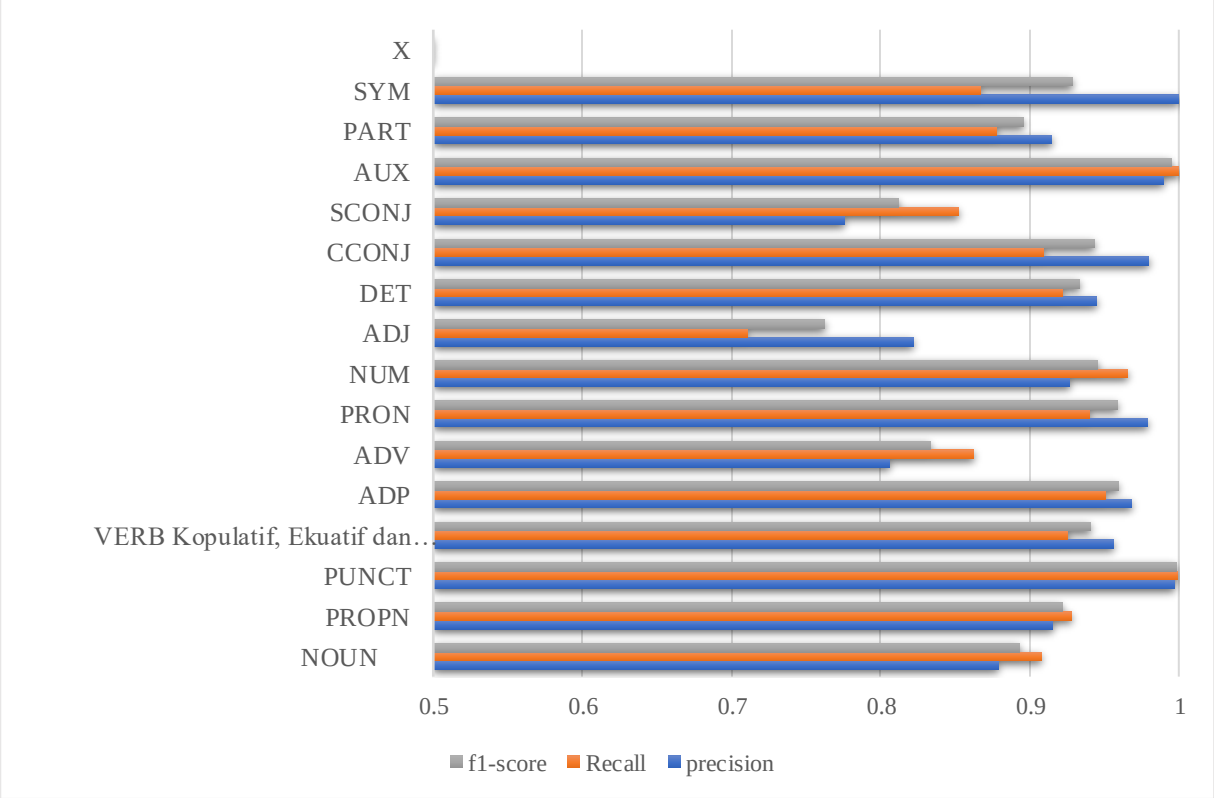

Figure 2.3 Graph of Accuracy for Syntactic Categories and Functions in Indonesian Language using POS Tagging

3. Method

The method used in this study was qualitative, utilizing POS Tagging data for Syntactic Categories and Functions in the Indonesian language with a model developed by Herpindo et al., (2022). The researchers employed a Rule-Based approach with an accuracy rate of 92.53% for the tagging results. The data used in this study were statistical computations extracted from the Leipzig Indonesian Mix 2013 corpus. The corpus data required language engineering (Carroll, 2013). The created rules were then implemented into the application program as the foundation for developing the syntactic category and function application for the Indonesian language.

To integrate rule-based POS Tagging and probability in Android app development, the first step was to compile the necessary POS Tagging data and resources, such as training data and language rules. Next, the program implemented the input text processing process, which involved entering text and parsing it into individual words for further processing. The rule-based POS Tagging module was created by defining a set of grammatical rules that determine the POS tag of each word. These rules included frequent patterns in the target language, such as specific word endings indicating their purpose in a sentence. Another technique involved training a probability model using SDLC training data (Pinciroli et al., 2022).

Integrating POS tagging into the System Development Life Cycle (SDLC) entails gradually introducing natural language processing techniques to improve text processing and analysis capabilities. This integration occurs at several stages of the SDLC, including planning, analysis, design, implementation, testing, deployment, and maintenance. The necessity for text processing inside the system is first established, followed by a study of text data to identify linguistic elements useful for POS tagging. During design, the system architecture is created to handle POS tagging components, and algorithms are defined. Implementation entails creating or integrating POS tagging algorithms into the system. Testing assures the correctness and performance of POS tagging, while deployment comprises integrating the system with POS tagging capabilities into the target environment (Tekinerdogan et al., 2016).

The SDLC (System Development Life Cycle) method was used to process the Leipzig Indonesian Mix 2013 corpus data tagged using a Rule-Based approach. The Rule-Based data for syntactic categories and functions were processed by following a sequential software development flow, including stages of design, coding, and support.

The stages in this research began with requirement analysis, system design, and implementation. The next stage was the analysis of the Indonesian language application system for syntactic category and function, ensuring its accessibility. The final stage involved the implementation and testing of the application to ensure that it meets the specifications and functions properly.

4. Results

The tool used in this research is a program for searching syntactic categories in the Indonesian language, utilizing the sequential search technique, which is a linear search method. This technique allows users of the application to easily remember and find syntactic functions.

This research introduces a program designed as a tool for searching syntactic categories in the Indonesian language. The program utilizes the sequential search technique, a linear search method, to facilitate users in conveniently locating specific syntactic functions.

The sequential search technique involves systematically examining each element in each list or dataset until the desired element is found or the end of the list is reached. In the context of this research, the program leverages this approach to enable users to search for syntactic categories in Indonesian. Users can input their desired syntactic function, and the program will sequentially search through the available data to identify and retrieve the corresponding category.

By incorporating the sequential search technique, the program offers an effective and user-friendly means for individuals to navigate and retrieve information about syntactic categories in the Indonesian language. This allows users to easily access and remember the specific syntactic functions they are interested in, enhancing their comprehension and utilization of Indonesian syntax. The integration of the sequential search technique into the program provides an efficient and user-friendly way for individuals to explore and retrieve information on syntactic categories in the Indonesian language. The program's interface allows users to input specific syntactic functions they are interested in, such as subject, object, or modifier.

Once the user submits their query, the program performs a sequential search through the available data, examining each entry one by one until it finds a match or reaches the end of the list. This approach ensures that all relevant syntactic categories associated with the input functions are identified and displayed to the user.

By using this program, individuals can easily access and retrieve information about the syntactic categories they are interested in. This enhances their understanding of Indonesian syntax, as they can quickly find the relevant categories for different syntactic functions. It also facilitates the utilization of Indonesian syntax in their own language production or analysis tasks, as they have a convenient tool for reference and guidance.

Overall, the incorporation of the sequential search technique in the program streamlines the process of exploring and comprehending syntactic categories in Indonesian, enabling users to improve their comprehension and usage of the language's syntax.

The design used in this application is a use case diagram, which is the most effective tool for stimulating users. In the use case diagram, the user is depicted as an actor representing the overall scope of the system in a concise manner. Below is an image that illustrates the application of syntactic category and function using the POS Tagging data from the Leipzig Indonesian Mix_2013 corpus. The input and output data of this application consist of the rule-based POS Tagging data, which can be seen in the following table:

Table 4.1 Syntactic Tag of Categories

No	Tag	Categories	Indonesia
1.	N	Noun	*Kata benda*
2.	ADJ	Adjective	*Kata sifat*
3.	NUM	Number	*Angka*
4.	ADV	Adverb	*Keterangan*
5.	V	Verb (Copula, Equative, and Existence)	*Kata kerja*
6.	CCONJ	Conjunction	*Kata penghubung*
7.	ADP	Preposition	*Preposisi*
8.	AUX	Auxiliary	*Kata kerja bantu*
9.	INTJ	Interjection	*Kata seru*
10.	PART	Particle	*Partikel*
11.	PRON	Pronoun	*Pronomina*
12	PUNCT	Punctuation	*Tanda baca*
13.	SYM	Symbol	*Simbol*
14.	X	Others	*Lainnya*

The realization of the speech category tag serves as the foundation for determining the form to fill in the empty slots in syntactic functions. The results of POS Tagging for syntactic categories are then used to create rules, which leads to achieving an accuracy level that represents the ratio of correct predictions to the total data. Precision, on the other hand, represents the ratio of true positive predictions to the overall positive predictions. Recall is the ratio of true positive predictions to the total accurate positive data. The F1-score of the syntax is incorporated into a form filler system, which fills the syntactic functions according to the concepts proposed by Chomsky, (2020); Haryono et al., (2018); Howes & Gibson, (2021); Van Valin, (2001) These concepts emphasize the part-of-speech category.

The POS Tagging results for syntactic categories are then used to create rules, which in turn yield accuracy rate (Accuracy) that represents the ratio of correct predictions to the total data. Precision is the ratio of true positive predictions to the overall positive predictions. The recall is the ratio of true positive predictions to the total accurate positive data. F1-score is a measure that combines precision and recall, providing a balanced evaluation of the model's performance. It is calculated as the harmonic mean of precision and recall. The F1-score is a commonly used metric for assessing the overall effectiveness of a model's predictions.

The POS Tagging results for syntactic categories are used to create rules that yield an accuracy rate (Accuracy), representing the ratio of correct predictions to the total data. Precision is the ratio of accurate positive predictions to the overall positive predictions, indicating the correctness of positive predictions. Recall, or sensitivity, is the ratio of true positive predictions to the total actual positive data, measuring the ability to identify all relevant instances. The F1-score is a weighted average of precision and recall, providing a single metric that balances both concerns, especially useful when the data distribution is imbalanced.

In the table below (see table 4.2), the syntax for determining syntactic categories in Indonesian. And to clarify the accuracy level of table 4.2, the following describes the accuracy of syntactic categories with Probability-Base POS Tagging (see figure 4.3).

Table 4.2 Input Data for Syntactic Categories in Indonesian Language Application

By Class	Precision	Recall	F1-score	Suport
NOUN	0.8789	0.9076	0.8930	2488
PROPN	0.9155	0.9280	0.9217	2207
PUNCT	0.9970	0.9994	0.9982	1645
VERB	0.9562	0.9256	0.9407	1250
ADP	0.9682	0.9506	0.9593	1153
ADV	0.8061	0.8624	0.8333	487
PRON	0.9791	0.9398	0.9590	498
NUM	0.9269	0.9656	0.9458	407
ADJ	0.8219	0.7109	0.7624	422
DET	0.9449	0.9220	0.9333	372
CCONJ	0.9799	0.9093	0.9433	375
SCONJ	0.7756	0.8521	0.8121	142
AUX	0.9897	1.0000	0.9948	96
PART	0.9149	0.8776	0.8958	49
SYM	1.0000	0.8667	0.9286	30
X	0.0000	0.0000	0.0000	1
micro avg	0.9253	0.9253	0.9253	11622
macro avg	0.8659	0.8511	0.8576	11622
weighted avg	0.9260	0.9253	0.9253	11622
samples avg	0.9253	0.9253	0.9253	11622

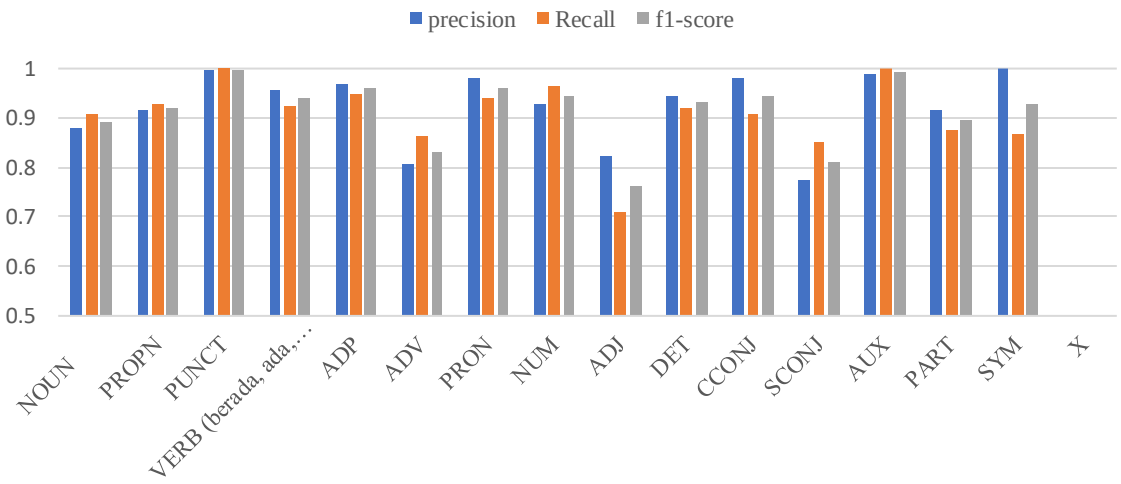

Figure 4.3 Graph of Syntactic Categories with *Probability-Based PoS Tagging*

Below, explain the Syntax for Determining Categories (see Figure 4.4). In processing the input, there are three columns: *KALIMAT* column, AGENT column, and PATIENT column, with 155 rows of sentences. For each row, this function applies a prediction function to determine the category tag for each word in the sentence. The output of this function is an input table that has been updated with an additional column named CATEGORY. The CATEGORY column contains each word and its corresponding category tag.

```
def kategori_f(x):
  for i in range(len(x)):
   sentence = Sentence(x['KALIMAT_k'].iloc[i]) tag_pos.predict(sentence)
   x.loc[i,'Kategori']=sentence.to_tagged_string()
```

Figure 4.4 Syntax for Determining Categories

```
def spok_f(x):
  for i in range(len(x)):
    tree = x['Frasa'].iloc[i]
    tree = nltk.tree.Tree.fromstring(str(tree))
    ptree = ParentedTree.convert(tree)
    for subtree in ptree.subtrees():
      if subtree.label() == 'VP':
        if subtree.left_sibling() is not None:
          if type(subtree.left_sibling()) is str:
            x.loc[i,'Subjek'] = str(subtree.left_sibling())
            x.loc[i,'Predikat'] = str(subtree.leaves()[0])
            x.loc[i,'Objek'] = str(subtree.leaves()[1:])
          else:
            if subtree.left_sibling().label() == 'NP':
              x.loc[i,'Subjek'] = str(subtree.left_sibling())
            elif subtree.left sibling().label() == 'NUMP':
              x.loc[i,'Subjek'] = str(subtree.left_sibling())
            elif subtree.left_sibling().label() == 'PP':
              x.loc[i,'Keterangan'] = str(subtree.left sibling())
            else:
              x.loc[i,'Predikat'] = str(subtree.leaves())
      if subtree.label() == 'VP':
        if subtree.right_sibling() is not None:
          if type(subtree.right_sibling()) is str:
            x.loc[i,'Predikat'] = str(subtree.leaves()[0])
            x.loc[i,'Objek'] = str(subtree.leaves()[1:])
            x.loc[i,'Keterangan'] = str(subtree.right_sibling())
          else:
            if subtree.right_sibling().label() == 'NP':
              x.loc[i,'Objek'] = str(subtree.right_sibling())
            elif subtree.right sibling().label() == 'NUMP':
              x.loc[i,'Objek'] = str(subtree.right_sibling())
            elif subtree.right sibling().label() == 'PP':
              x.loc[i,'Keterangan'] = str(subtree.right_sibling())
            else:
              x.loc[i,'Predikat'] = str(subtree.leaves())
      if ((subtree.right_sibling() is None) and (subtree.left_sibling() is None
)):
        x.loc[i,'Subjek'] = str(subtree.leaves()[0])
        x.loc[i,'Predikat'] = str(subtree.leaves()[1])
```

Figure 4.5 Syntax for Determining Functions

The function above (see Figure 4.5) processes the input in the form of a table. The table has five columns: *KALIMAT* column, *AGEN* column, *PASIEN* column, *KATEGORI* column, and *FRASA* column, with 155 rows. For each row, the function applies the prediction of function tags. The output of this function is an updated input table with four additional columns: *SUBJEK* (subject), *PREDIKAT* (predicate), *OBJEK* (object), and *KETERANGAN* (complement). These columns contain the words that occupy specific functions within a sentence.

The function in this context takes a table as input. This table consists of five columns: *KALIMAT* column, *AGEN* column, *PASIEN* column, *KATEGORI* column, and *FRASA* column. It contains a total of 155 rows, representing different sentences or linguistic units. For each row in the table, the function performs the prediction of function tags. These function tags indicate the specific syntactic roles or functions that words or phrases in a sentence may occupy, such as subject, predicate, object, or complement.

After applying the function, the input table is updated to include four additional columns: *SUBJEK* (subject), *PREDIKAT* (predicate), *OBJEK* (object), and *KETERANGAN* (complement). These columns serve as placeholders to store the words or phrases that fulfill the respective syntactic functions within a sentence. The function assigns the appropriate words or phrases from the input table to these columns based on the predicted function tags.

The function generates the updated table with the additional columns, providing a clearer representation of the sentences' syntactic structure. This information can be valuable for further analysis, understanding the relationships between different elements in a sentence, and extracting meaningful linguistic insights.

Syntactic categories can occupy grammatical syntactic functions, which are then defined through rule-based stages. These stages involve creating rules and are represented in the syntactic function determination diagram, which will be implemented in the input determination of the Indonesian language syntactic category and function application.

The implementation of the Indonesian language syntactic category and function application with POS Tagging data from the Leipzig Indonesian Mix 2013 corpus is then tested. The testing of input data POS Tagging is organized through the use case diagram of the user application, as shown in the following image (see figure 4.6).

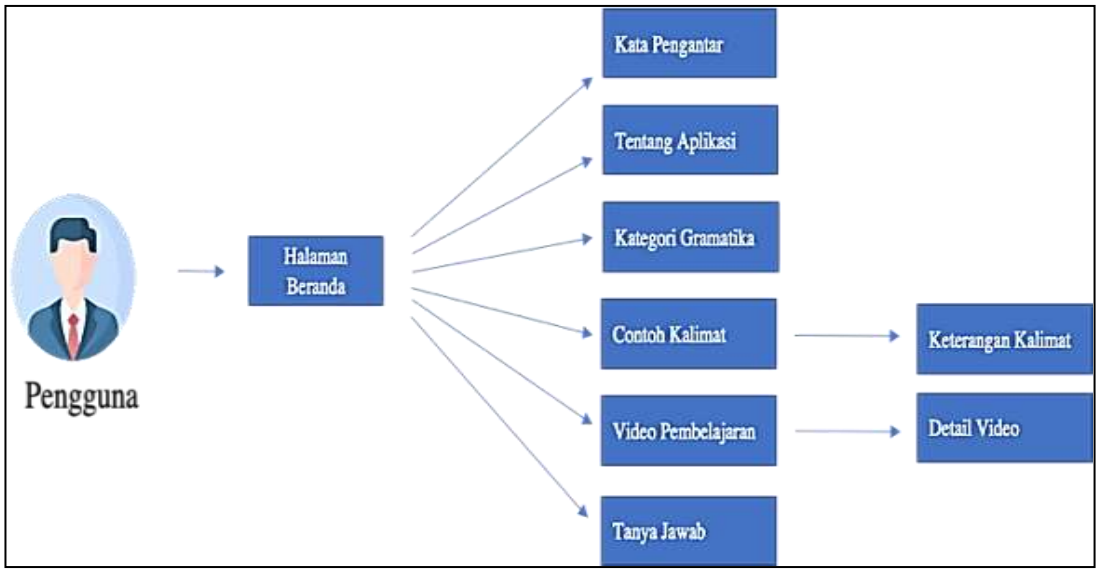

Figure 4.6 Blog Diagram of Indonesian Language Syntactic Categories and Functions App Users

The use case diagram of the Indonesian language category application user and syntax function is used to show the function flow of the application user use case. This application user flow is managed by the admin who will run interactions with users. Interaction with users is seen in the application admin blog diagram and can be seen in the following image (see figure 4.7).

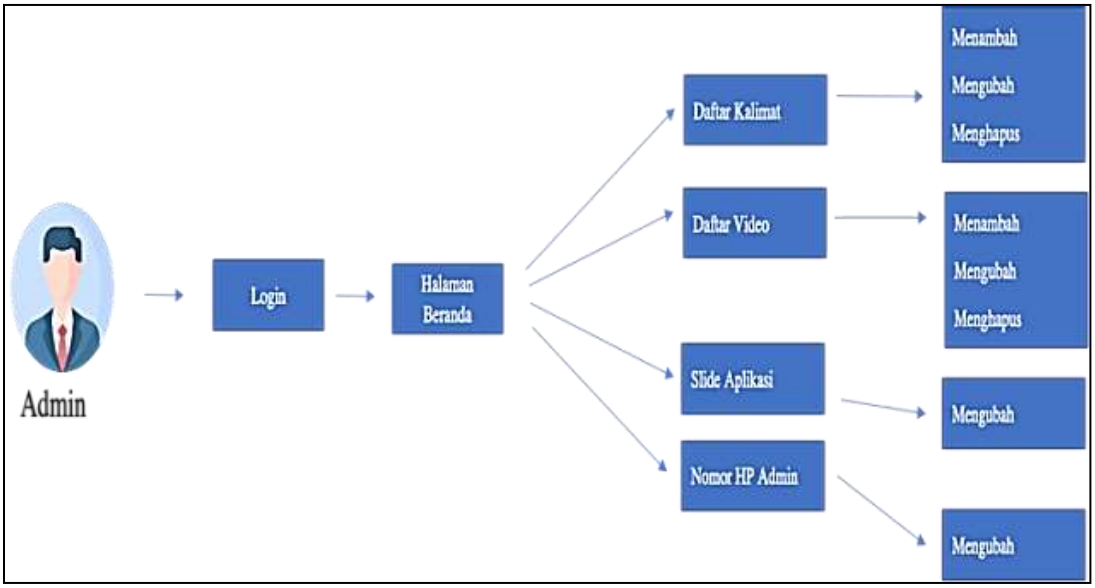

Figure 4.7 Blog Diagram Admin Application Categories and Syntax Functions of Indonesian Language

4.1 Test Results of Indonesian Syntactic Categories and Functions App

The test results show that the application built has run in accordance with the objectives. The reports generated provide information that is in accordance with the target users and management. Here is a view of the main menu of the application (see Figure 4.8)

Meanwhile, the menu of this application consists of several sub menus that will make it easier and clearer for users to understand Indonesian grammatical categories. The menu contains defines of predicative verbs that can occupy syntactic functions. The information is contained in the grammatical category menu in the following image (see Figure 4.9).

Figure 4.8 Main Menu

Figure 4.9 Indonesian Grammar Category

This application in the grammatical category sub menu will discuss the phenome of language category structures, especially on more complex verbs (core/nucleus) that have not been touched by other grammar applications. In this sub menu, users will be given information related to other verbs that have the potential to become predicates functionally. These verbs consist of copulative, equative, and existential verbs and their usage in Indonesian contexts. The pattern in this grammatical category is based on the principles proposed by Van Valin (2001) and Kridalaksana (1988) in the description of the table that has been included in the application as follows (see table 4.3).

Table 4.3 Function and Category Fillers in Indonesian Language Syntactics on the Application

No	Categories	Function
1.	NV (Copula, Equative, Existence, ø)	S, P (Copula, Equative, Existence, ø) SP
2.	NV (Copula, Equative, Existence, ø) N	S, P (Copula, Equative, Existence, ø) O
3.	NVN	S, P (Copula, Equative, Existence, ø), O, K
4.	NVN	S, P (Copula, Equative, Existence, ø) O, Pel

Sentences in Indonesian that are entered according to syntactic categories will occupy various functions. The clause or sentence will occupy the function of the subject in the form of a noun (N) or noun phrase (FN) element that precedes the predicate, the predicate as an element in the syntactic category that can be filled with verbs (V), other verb categories such as copulative, equative and existential as additional supplements in this application, the object function occupied by a noun (N) or noun phrase (FN) that fulfills a verb (V) with the rule of being in the transitive active position, and behind the predicate, next is the adverb function which in the syntactic category and function can be filled with verbs (V). (V) with the rule of being in the transitive active position, and is behind the predicate, next is the adverbial function which in the category and syntactic function in this application is arbitrary (whichever you like), the complementary function which can be filled by Verbs (V), Nouns (N), Noun Phrases (FN), Adjective Phrases (FAdj), Numeral Phrases (FNum) which describe the top of the verbal predicate. Here is a specific view of the app that explains the syntax category links and their functions (see Figure 5.5).

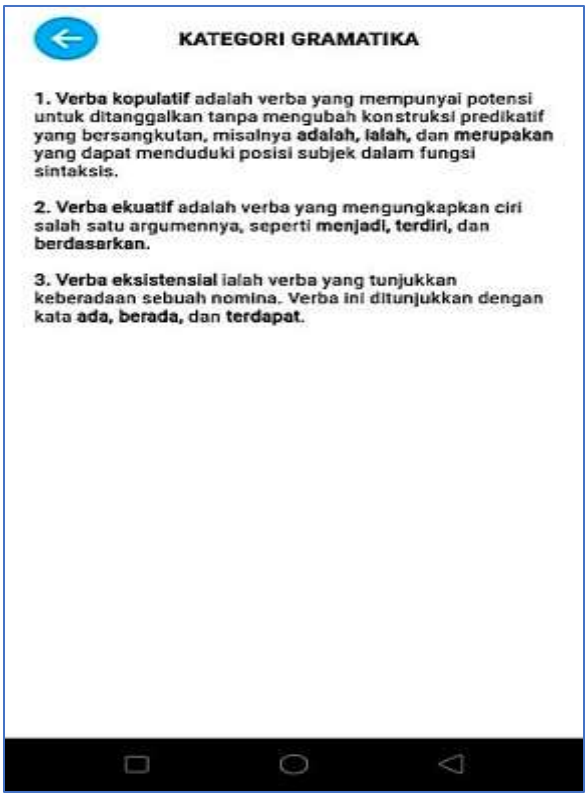

Figure 5.5 Category and Function Relationships

Users of the Indonesian language syntactic category and function application can not only view example sentences but also information about the grammatical categories and functions in Indonesian. The user interface provides detailed information regarding the placement of elements within clauses or sentences. The application also presents the functional components, making it easier to determine the position and role of these elements. Additionally, it includes copulative, equative, and existential verbs as additional components that provide insights into the position of the predicate.

5. Discussion

This study's discussion of the development of category and function of Indonesian syntax is based on POS Tagging data. This research effort used a sequential search strategy to create a program for searching syntactic categories in the Indonesian language.

In this study, a specific program was created to help users identify and retrieve syntactic categories in the Indonesian language. The adoption of a sequential search methodology, which is known for its linear search method, ensures a simple and user-friendly approach. This technique tries to empower users by making syntactic functions easier to recall and locate (Dong & Liu, 2013).

This study proposes a novel program aimed to be a useful tool for searching and retrieving syntactic categories in the Indonesian language. This program is poised to transform how users manage the complexity of Indonesian syntax (Shamsujjoha et al., 2021). It provides a user-friendly experience that can greatly contribute to linguistic study and language learning by applying the sequential search methodology, a linear search strategy.

It is worth noting the program's emphasis on a user-centric approach similar with (Flemming et al., 2004) work in object-oriented development in CAD. The sequential search technique, with its step-by-step evaluation of syntactic categories, is designed to provide consumers with an easy and memorable experience. This method recognizes the inherent intricacies of linguistic structures and seeks to make the process of discovering certain syntactic functions within the Indonesian language easier. This ease of use is a significant step toward making linguistics and language study more accessible (Ortin & Cueva, 2004).

On the one hand, it facilitates linguistic scholars' search for syntactic categories. It can help with more effective data processing and the creation of Indonesian-specific linguistic theories. The program, on the other hand, is a godsend for language learners. Students and others interested in learning Indonesian will find it extremely useful for better understanding sentence structures and grasping grammatical complexities.

Comparisons were made with research conducted by Tyers and Howell (2021). The research proves that there are limitations to Mayan speech tagging. There is a shortcoming in the research material object, namely the corpus size. The corpus size in this study is relatively small for a research model so that it can limit the results in generalizability. In addition, this study shows that improving the annotation of K'iche data can

improve part-of-speech tagging performance for related languages such as Uspanteko (Tyers & Howell, 2021). But of course, the limitations of this research require a larger and more diverse data set to improve accuracy and the application of the model developed in the research. Therefore, this research brings renewal by developing an application utilizing Rule and Probability-Base POS Tagging data from the Leipzig Indonesian Mix_2013 corpus to determine the category and syntactic function of Indonesian with lexical category V (copula, existence, and equative) as a potential predicate in the function.

Looking ahead, the prospects for this initiative appear excellent. Future research and development could include improving the tool's ability to cover more languages, increasing its utility. Additionally, improving the program's search algorithms and user interface can make it more user-friendly and robust. To ensure the effectiveness and continual improvement of the development process, user research and feedback should be implemented.

The creation of a program specialized in finding syntactic categories in Indonesian, using a user-friendly sequential search technique, has the potential to simplify and transform the way we interact with linguistic complexities. It serves the dual function of assisting linguistic research and language learning, making it a valuable tool. With a commitment to continued development and growth, this software has the potential to affect not only the study of Indonesian, but also to create a precedent for similar tools in other languages, enhancing our understanding of linguistics (Papenhausen & Mueller, 2018).

The sequential search strategy examines each element in a list or dataset in a methodical and straightforward manner until the required element is discovered or the end of the list is reached. This approach, used in the study, is a critical component of the program's architecture, focused on the retrieval of syntactic categories in Indonesian.

The implementation of the sequential search methodology by the application indicates a dedication to a user-friendly and systematic approach. Users have the option of typing in their desired syntactic function, and the application subsequently searches through the available data in a methodical manner. This design approach not only streamlines the procedure but also assures that users may quickly find the syntactic category they want.

The sequential search approach's incorporation into the application greatly improves its accessibility and usefulness. This study is consistent with a larger trend in technology to prioritize user experience. It implies that syntactic categories in Indonesian may be investigated and comprehended more easily by linguistic scholars and language learners alike. This method bridges the gap between advanced linguists and individuals who are fresh to the topic or language.

This study's use of the sequential search strategy is not confined to the Indonesian language. Its systematic approach allows it to be adapted to other languages, expanding the program's potential uses. It establishes a precedent for comparable tools in diverse linguistic situations by offering a framework for rapid search and retrieval of syntactic categories, making language analysis, and learning more accessible globally.

The researchers' introduction of the sequential search strategy into the software marks an important addition to the field of language technology. It simultaneously enables linguists in their research and learners in their language studies by reducing the process of looking for syntactic categories in the Indonesian language. Furthermore, this technique has the potential to alter the way we approach linguistic analysis and language acquisition, with far-reaching ramifications beyond the Indonesian language. As technology advances, the seamless integration of linguistic research and language learning tools with systematic search approaches is a potential future trend.

In category research, a corpus with post-tagging enables researchers to determine the grammatical category of each word in a sentence, such as subject, predicate, object, and so on. This facilitates syntactic analysis and a better grasp of sentence structure.

For syntactic function research, a corpus with post-tagging can help researchers determine each word's syntactic role in a sentence, such as subject, direct object, indirect object, or complement. This enables for a more in-depth investigation of how words interact within a sentence.

Meanwhile, in syntactic role research, a corpus with post-tagging enables researchers to track the specific roles that words play in a given sentence context, such as time markers, location markers, or inter-clause connection markers. This helps us grasp how syntax contributes to the creation of meaning in language.

6. Conclusions

The development and implementation of the Indonesian grammar category and syntactic function application mark a significant advancement in digital linguistic tools, particularly for Indonesian language studies. This application addresses a critical need by incorporating grammatical categories such as lexical verbs, including copulative, equative, and existential verbs, and providing extensive information on sentence parts within clauses or sentences, thus aiding academics in Indonesian syntax studies. The application demonstrates high accuracy, with a 92.53% success rate in POS tagging based on rules and probabilities, significantly enhancing linguistic analysis and facilitating efficient research. Additionally, it serves as a valuable teaching resource for Indonesian language instructors at various

educational levels, allowing them to demonstrate and explain complex grammar concepts effectively. Despite the limitations of sample coverage and methodology, this application provides a comprehensive tool for both research and education. Future development should focus on incorporating syntactic role components, which will enrich the application's capabilities and enable more in-depth analyses. Ultimately, this application not only fills a gap in digital grammar resources but also holds immense potential for advancing linguistic research and improving language instruction.

References

Auni, L., & Manan, A. (2023). A contrastive analysis of morphological and syntactic aspects of English and Indonesian adjectives. *Studies in English Language and Education*, *10*(1), 403–423. https://doi.org/10.24815/siele.v10i1.27401

Baker, M. C. (2004). *Lexical categories*. Cambridge University Press.

Bradley, T.-D., Terilla, J., & Vlassopoulos, Y. (2022). An enriched category theory of language: From syntax to semantics. *La Matematica*, *1*, 551–580. https://doi.org/10.1007/s44007-022-00021-2

Bustomi, T., Kridalaksana, H., & Fachruddin, F. (2018). Aplikasi media pembelajaran Bahasa Indonesia berbasis multimedia. *Sebatik*, *10*(1), 1-7. https://doi.org/10.46984/sebatik.v10i1.58

Carroll, J. (2013). *The Routledge handbook of corpus linguistics*. Routledge. https://doi.org/10.4324/9780367076399

Changpueng, P., & Patpong, P. (2021). Genre analysis of minutes of meetings conducted in English by Thai engineers. *Indonesian Journal of Applied Linguistics*, *11*(1), 134-145. https://doi.org/10.17509/ijal.v11i1.34584

Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System*, *31*(2), 173–186. https://doi.org/10.1016/S0346-251X(03)00019-8

Chomsky, N. (2020). Remarks on nominalization. In *Oxford University Press* (pp. 25–28). https://doi.org/10.1093/oso/9780198865544.003.0002

Chomsky, N. (2021). Linguistics then and now: Some personal reflections. *Annual Review of Linguistics*, *7*(1), 1–11. https://doi.org/10.1146/annurev-linguistics-081720-111352

Dewi, R., & Ubaidi, A. (2020). Pos tagging Bahasa Madura dengan menggunakan algoritma Brill tagger. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(6), 1121-1128. https://doi.org/10.25126/jtiik.2020722449

Dong, C., & Liu, X. (2013). Development of Android application for language studies. *IERI Procedia*, *4*, 8–16. https://doi.org/10.1016/j.ieri.2013.11.003

Falk, Y. (2011). *Lexical-functional grammar*. Oxford University Press.

Flemming, U., Erhan, H., & Özkaya, I. (2004). Object-oriented application development in CAD: A graduate course. *Automation in Construction*, *13*(2), 147–158. https://doi.org/10.1016/j.autcon.2003.09.009

Haryono, H., Lelono, B., & Kholifah, A. N. (2018). Typography, morphology, and syntax characteristics of texting. *Lingua Cultura*, *12*(2), 179-185. https://doi.org/10.21512/lc.v12i2.3976

Herpindo, H., Wijayanti, A., & Shalima, I. (2022). Kategori, fungsi, dan peran sintaksis bahasa Indonesia dengan PoS tagging berbasis rule dan probability. *Kembara: Jurnal Keilmuwan Bahasa, Sastra Dan Pengajarannya*, *8*(1), 51–65. https://doi.org/10.22219/kembara.v8i1.18602

Ho, J. (2023). #FleeingWuhan: Legitimation and delegitimation strategies in hostile online discourse. *Applied Linguistics*, *44*(3), 391–419. https://doi.org/10.1093/applin/amac061

Howes, C., & Gibson, H. (2021). Dynamic syntax. *Journal of Logic, Language and Information*, *30*(2), 263–276. https://doi.org/10.1007/s10849-021-09334-x

Ilmiyah, M., & Qoiriah, A. (2021). Sistem deteksi kesalahan tanda baca dan huruf kapital pada karya tulis ilmiah berbahasa Indonesia menggunakan algoritma Boyer-Moore. *Journal of Informatics and Computer Science (JINACS)*, *2*(03), 185-193. https://doi.org/10.26740/jinacs.v2n03.p185-193

Indiana, B. D., & Ramadhani, I. (2019). Aplikasi pembelajaran Bahasa Jawa berbasis Android. *CAHAYAtech*, *8*(1), 40-57. https://doi.org/10.47047/ct.v8i1.18

Jia, H., & Liang, J. (2020). Lexical category bias across interpreting types: Implications for synergy between cognitive constraints and language representations. *Lingua*, *239*, 102809. https://doi.org/10.1016/j.lingua.2020.102809

Kaya, Ö. F. (2022). Using corpora for language teaching and assessment in L2 writing: A narrative review. *Focus on ELT Journal*, *4*(3), 46–62. https://doi.org/10.14744/felt.2022.4.3.4

Khairul, K., Haryati, S., & Yusman, Y. (2018). Aplikasi kamus Bahasa Jawa Indonesia dengan algoritma Raita berbasis Android. *Jurnal Teknologi Informasi Dan Pendidikan, 11*(1). https://doi.org/10.24036/tip.v11i1.102

Kridalaksana, H. (1988). *Beberapa prinsip perpaduan leksem dalam Bahasa Indonesia*. Kanisius.

Moeliono, A. M., dkk. (2017). *Tata bahasa baku Bahasa Indonesia* (Edisi keempat). Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan.

Olivia, N., & Shaklein, V. (2019). Preparation specifics of students-philologists: Modern technologies of literary and linguistic text analysis. *International Journal of Language Education, 3*(2), 91–98. https://doi.org/10.26858/ijole.v3i2.9749

Ortin, F., & Cueva, J. M. (2004). Dynamic adaptation of application aspects. *Journal of Systems and Software, 71*(3), 229–243. https://doi.org/10.1016/S0164-1212(02)00157-7

Papenhausen, E., & Mueller, K. (2018). Coding ants: Optimization of GPU code using ant colony optimization. *Computer Languages, Systems & Structures, 54*, 119–138. https://doi.org/10.1016/j.cl.2018.05.003

Pinciroli, F., Barros Justo, J. L., & Forradellas, R. (2022). Systematic mapping study: On the coverage of aspect-oriented methodologies for the early phases of the software development life cycle. *Journal of King Saud University - Computer and Information Sciences, 34*(6), 2883–2896. https://doi.org/10.1016/J.JKSUCI.2020.10.029

Pramudita, H. R., Utami, E., & Amborowati, A. (2016). Pengaruh part of speech tagging berbasis aturan dan distribusi probabilitas maximum entropy untuk Bahasa Jawa Krama. *Jurnal Buana Informatika, 7*(4), 235-244. https://doi.org/10.24002/jbi.v7i4.764

Prasetyo, E. A. (2019). Aplikasi pembelajaran BIPA (Bahasa Indonesia bagi penutur asing) tingkat dasar berbasis Android. *J-INTECH, 6*(02), 229-234. https://doi.org/10.32664/j-intech.v6i02.256

Purwono, P. Y., & Asteria, P. V. (2021). Pembelajaran BIPA dengan aplikasi AWAN ASA berbasis pengenalan lintas budaya. *Fon: Jurnal Pendidikan Bahasa Dan Sastra Indonesia, 17*(1), 97-107. https://doi.org/10.25134/fjpbsi.v17i1.3892

Rahmawati, I. Y., Aziz, F., & Wahyono, T. (2020). Rancang bangun aplikasi pengenalan materi pengajaran Bahasa Indonesia bagi penutur asing (BIPA) berbasis Android di Universitas Muhammadiyah Ponorogo. In *Konferensi Internasional Pengajaran Bahasa Indonesia bagi Penutur Asing (KIPBIPA) XI*.

Ramliyana, R., Pratiwi, N. K., & Megiati, Y. E. (2022). Analysis of Indonesian language error in writing reports of students' learning results of the Amanah Fitrah Rabbani Foundation using the Sipebi application. *Hortatori: Jurnal Pendidikan Bahasa Dan Sastra Indonesia, 6*(1), 6–16. https://doi.org/10.30998/jh.v6i1.998

Ratnawati, N., Wahyuningtyas, N., Ruja, I. N., Habibi, M. M., Anggraini, R., & The, H. Y. (2021). Developing multimedia-based learning media for basic skill of teaching material in order to equip professional teachers. *International Journal of Emerging Technologies in Learning (IJET), 16*(07), 77. https://doi.org/10.3991/ijet.v16i07.21203

Raymondra, K. A. P., & Bukhori, H. A. (2021). Interferensi sintaksis Bahasa Indonesia terhadap Bahasa Jerman pada Schriftlicher Ausdruck dalam matakuliah B1-Prüfüngsvorbereitung. *JoLLA: Journal of Language, Literature, and Arts, 1*(1). https://doi.org/10.17977/um064v1i12021p25-36

Rianto, A. (2020). A study of language learning strategy use among Indonesian EFL university students. *Register Journal, 13*(2), 231–256. https://doi.org/10.18326/rgt.v13i2.231-256

Rizki , T., & Meditanala, T. (2018). Perancangan media pembelajaran membaca dan menulis Bahasa Indonesia bagi penutur asing melalui aplikasi berbasis Android. In *Seminar Internasional Riksa Bahasa*.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225. https://doi.org/10.1017/S0267190511000055

Safara, A., Zaim, M., & Refnaldi. (2019). Mobile-based learning in digital era: Android application as a media to teach grammar. In *Proceedings of the 1st International Conference on Education Social Sciences and Humanities (ICESSHum 2019)*. https://doi.org/10.2991/icesshum-19.2019.157

Shamsujjoha, Md., Grundy, J., Li, L., Khalajzadeh, H., & Lu, Q. (2021). Developing mobile applications via model driven development: A systematic literature review. *Information and Software Technology, 140*, 106693. https://doi.org/10.1016/j.infsof.2021.106693

Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian reference grammar*. Allen & Unwin.

Sugiarti, R., Budiyono, C., & Sunu. (2021). Fungsi, kategori dan peran sintaksis pada cerita pendek dalam koran Jawa Pos bulan Juli 2016. *Buana Bastra*, *5*(1). https://doi.org/10.36456/bastra.vol5.no1.a3582

Sulistyo, W. D., Khakim, M. N. L., Jauhari, N., & Anggraeni, R. D. (2021). Fun learning history: Explore the history of water sites based on Android. *International Journal of Emerging Technologies in Learning (IJET), 16*(07), 105. https://doi.org/10.3991/ijet.v16i07.21215

Syahroni, A., & Wahab, H. (2019). Aplikasi penentuan kategori dan fungsi sintaksis kalimat Bahasa Indonesia. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, *4*(1). https://doi.org/10.30743/infotekjar.v4i1.1537

Syamsiah, S. (2017). Pengembangan aplikasi multimedia pembelajaran interaktif untuk mata pelajaran Bahasa Indonesia. *SAP (Susunan Artikel Pendidikan)*, *2*(1). https://doi.org/10.30998/sap.v2i1.1723

Tallerman, M. (2014). *Understanding syntax* (4th ed.). Routledge. https://doi.org/10.4324/9781315758084

Tekinerdogan, B., Ali, N., Grundy, J., Mistrik, I., & Soley, R. (2016). Quality concerns in large-scale and complex software-intensive systems. In *Software Quality Assurance: In Large Scale and Complex Software-Intensive Systems* (pp. 1–17). https://doi.org/10.1016/B978-0-12-802301-3.00001-6

Tyers, F., & Howell, N. (2021). A survey of part-of-speech tagging approaches applied to K'iche'. In *Proceedings of the 1st Workshop on Natural Language Processing for Indigenous Languages of the Americas, AmericasNLP 2021*. https://doi.org/10.18653/v1/2021.americasnlp-1.6

Van Valin, R. D., Jr. (2001). *Syntax: Structure, meaning and function*. Cambridge University Press.

Verhaar, J. W., & Alip, B. (1996). *Asas-asas linguistik umum*. Gadjah Mada University Press.

Widhiyanti, K., & Harjoko, A. (2013). POS tagging Bahasa Indonesia dengan HMM dan rule based. *Jurnal Informatika*, *8*(2). https://doi.org/10.21460/inf.2012.82.125

Yunus, Y. H. (2020). Aplikasi pembelajaran Bahasa Indonesia berbasis Android pada sekolah menengah pertama. *Jurnal Teknologi Informasi Indonesia (JTII)*, *5*(2). https://doi.org/10.30869/jtii.v5i2.646