

Comparison Of Shift Reduce Parsing and Left Corner Parsing Algorithm in Sentence Structure Ambiguity Checker

Reyhan Navind Shaquille^{a,1}, Novi Yusliani^{b,2}, Mastura Diana Marieska^{*b,3}

^{a,b} Artificial Intelligence Laboratory, Faculty of Computer Science, Sriwijaya University
¹rnshaquile@gmail.com; ²novi_yusliani@unsri.ac.id; ³mastura.diana@ilkom.unsri.ac.id
^{*}Corresponding Author

ARTICLE INFO

Article history

Received 1 Juni 2021

Revised 5 Juni 2021

Accepted 14 Juni 2021

Keywords

Structural Ambiguity

Shift Reduce Parsing

Left Corner Parsing

Parsing

ABSTRACT

Indonesian language is the official language of the Republic of Indonesia and the language of the Indonesian nation's unity. Although it is commonly used, ambiguity could occur in the use of Indonesian. Ambiguity can cause misunderstandings in interpreting a word or sentence. Structural ambiguity is a type of ambiguity that occurs when the structure of words in a sentence can be given more than one grammatical structure. Left Corner Parsing and Shift Reduce Parsing are parsing methods used to identify sentence structure ambiguity. This research involves pre-processing, namely case folding, tokenizing and Part Of Speech Tagging. This study uses 90 testing data labelled with facts, 30 ambiguous sentences and 60 unambiguous sentences. Based on the results of checking the ambiguity of the sentence structure, the Shift Reduce Parsing algorithm produces an accuracy of 71%, precision 70.6%, recall 59%, and f-measure 58.2%. Meanwhile, Left Corner Parsing produces an accuracy value of 70%, precision 68.7%, recall 57.5%, and f-measure 55.8%.

1. Introduction

Indonesian is the official language of the Republic of Indonesia and the language of the Indonesian nation's unity. Where the language is used daily in interacting and communicating. Although it is commonly use, ambiguity could occur in the use of Indonesian. Ambiguity can lead to misunderstandings in interpreting a word or sentence[1]. According to the KBBI, ambiguity is the possibility of having more than one meaning in a word, word combination, or sentence. A word or sentence is said to be ambiguous when it has more than one meaning that has the potential to cause confusion in spoken or written speech[2]. Structural ambiguity is a type of ambiguity that occurs when the order of words in a sentence can be given more than one grammatical structure and each has a different meaning[3]. A sentence is categorized as ambiguous, if one sentence can produce more than one sentence pattern[1].

Sentence structure is a structural construction that contains a number of rules or rules as it should be in accordance with the grammar of a language. The existence of various variations of sentence structure construction is a human creativity in language, thinking, and reasoning. Variations in sentence structure can be identified based on the basic sentence patterns so that there is an arrangement of sentence structure elements that changes position. The structure of the sentence elements can be rearranged based on basic sentence patterns and in accordance with the grammar of the language [4].

2. Literature Study

In this section contains the theoretical foundation and some research that has been done by previous researchers. This was made to strengthen the reasoning and rationality of the involvement of several variables in this study. It also functions as a scientific opinion that is integrated with the results of a literature review to build a researcher's mindset in relation to the problem being studied.

Research conducted by [5] used Shift Reduce Parsing to perform ambiguous sentence patterns with a success rate of 83%. The only result of ambiguous sentence recognition is determined by the defined word class, the word class of each word in the phrase used to process the parsing reduction shift.

Research conducted by [6] used Left Corner Parsing to examine the grammar and meaning of words. The results of the study explain that Left Corner Parsing can parse English sentences into their constituent words. Left Corner Parsing can distinguish the meaning of two or more words whose morphology is the same as long as it is not at the beginning of the sentence.

Research conducted by [7] used Left Corner Parsing to apply it to the Indonesian to Bima language translation application with an accuracy rate of success reaching 68% for respondents who strongly agree. Left Corner Parsing is quite well implemented in translating sentences.

Research conducted by [8] used the Shift Reduce Parsing Algorithm to decipher words in the Komerang Rasuan language sentence based on Indonesian rules. The results of the study explain the results of syntax checking using Shift Reduce Parsing by 84%.

Research conducted by [9] used the Left Corner Parsing Algorithm to check the grammar in Indonesian sentences. Left Corner Parsing is very dependent on the data dictionary used. If the input word is not found in the data dictionary, the sentence pattern will not be recognized or become an incorrect pattern.

3. Methodology

The data used in this research is an Indonesian word dictionary which contains a collection of Indonesian words and word classes taken from the NLP_ITB package. The test data is secondary data, which consists of 90 single sentences labeled ambiguous sentences and unambiguous sentences. Testing data will go through text preprocessing. Then the ambiguity checking is done using Shift reduce parsing and left corner parsing. All stages are illustrated in Figure 1.

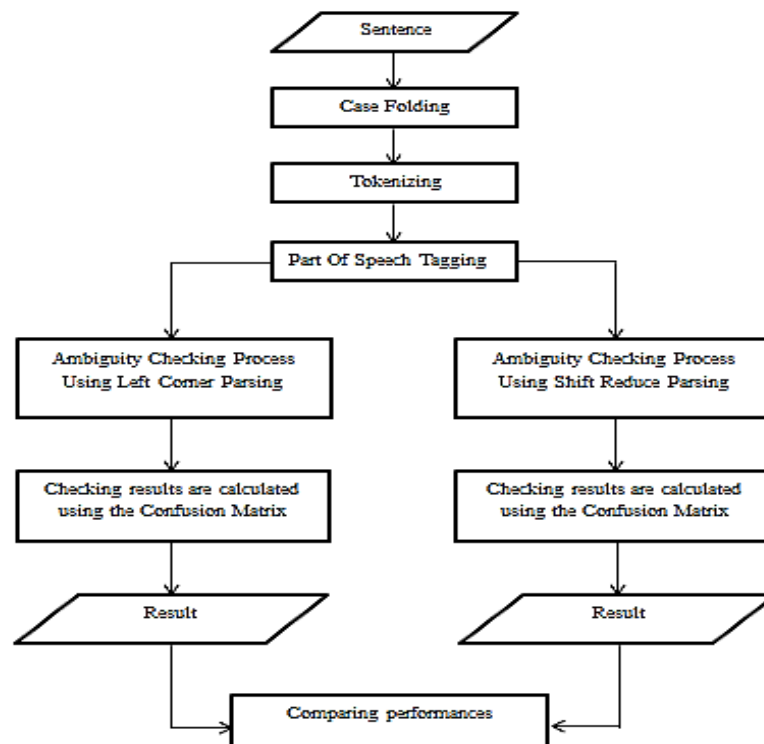


Fig. 1. Framework Diagram

3.1. Preprocessing

Preprocessing is a process to prepare raw data before other processes are carried out [10]. In this case, preprocessing includes case folding, tokenizing, and part of speech tagging. The case folding process will change all the letters in the document to lowercase. Next, the sentence will be broken down into words or known as tokenizing. The last step is part of speech tagging, in which is assigning word classes to each word, using the NLP_ITB library.

3.2. Shift Reduce Parsing

Shift Reduce Parsing is a parsing technique that belongs to the Bottom-Up parsing category. Shift-reduce table is used to guide the parse. Shift Reduce Parsing is used as a token and forms a production line to build a parse tree. Shift Reduce Parsing uses a stack to keep the order of each token[5].

The steps involved in the Shift Reduce Parsing process are as follows[11]:

1. Shift, which adds one word to the stack. The Shift action only moves (shifts) words, in this case per word until they form one sentence from the input sentence.
2. Reduce, which removes or removes some elements from the stack and replaces them with new elements.

Fig 2. illustrated the use of Shift Reducing Parsing in checking an ambiguous sentence, which is resulted more than one grammatical structure.

Stack	Input	Action
	r + v + n + a + t + n	Shift r
r	+ v + n + a + t + n	Reduce S → r
S	v + n + a + t + n	Shift + v
S + v	+ n + a + t + n	Reduce P → v
S + P	+ n + a + t + n	Shift + n
S + P + n	+ a + t + n	Shift + a
S + P + n + a	+ t + n	Reduce O → n + a
S + P + O	+ t + n	Shift + t
S + P + O + t	+ n	Shift + n
S + P + O + t + n		Reduce K → t + n
S + P + O + K		

Stack	Input	Action
	r + v + n + a + t + n	Shift r
r	+ v + n + a + t + n	Reduce S → r
S	+ v + n + a + t + n	Shift + v
S + v	+ n + a + t + n	Shift + n
S + v + n	+ a + t + n	Shift + a
S + v + n + a	+ t + n	Reduce V → v + n + a
S + V	+ t + n	Reduce P → V
S + P	+ t + n	Shift + t
S + P + t	+ n	Shift + n
S + P + t + n		Reduce K → t + n
S + P + K		

Fig. 2. Checking Ambiguity Using Shift Reduce Parsing

3.3. Left Corner Parsing

Left Corner Parsing is a parsing strategy that uses data in Bottom-Up Parsing and predictions from Top-Down[7]. The stages in the Left Corner Parsing (LCP) process are as follows, Top-Down Parsing performs a search from the root/peak to the leaf. Top-Down works by deciphering a sentence from the largest constituent to the smallest. Constituents are sentence-forming elements that can stand alone. Bottom-Up Parsing works by taking one word from a sentence, to be assembled into a larger constituent. This is done continuously until the constituent formed is a complete sentence. Therefore, Left Corner Parsing process starts bottom-up and ends top-down [12]. Fig 3. illustrated the use of Left Corner Parsing for a sentence.

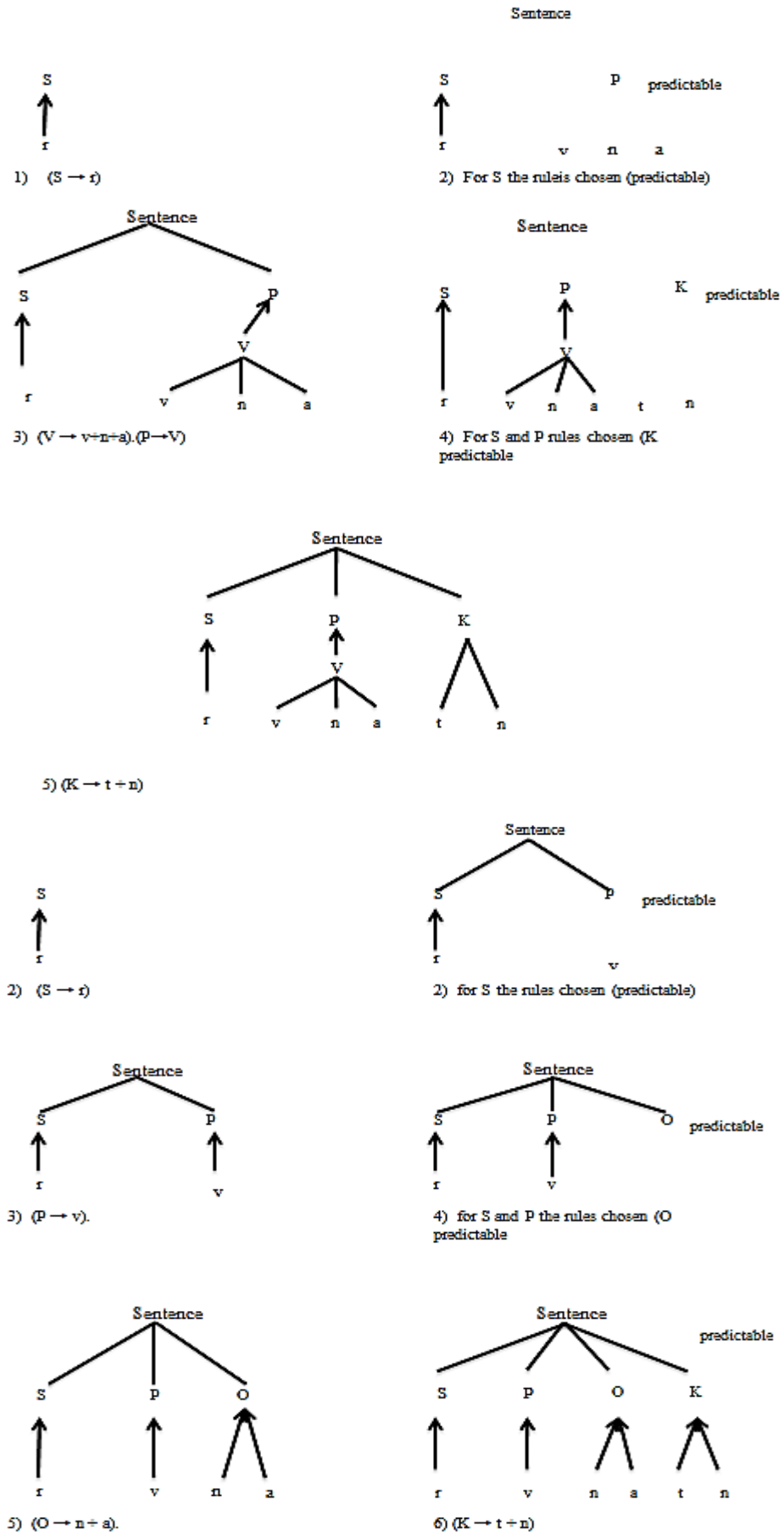


Fig. 3. Checking Ambiguity Using Left Corner Parsing

4. Result and Discussion

The test is carried out using the Shift Reduce Parsing and Left Corner Parsing methods to produce a Confusion Matrix table. From the results of the table, the accuracy, precision, recall, and f-measure values will be obtained. The data used in this study are 90 single sentences that have been labeled ambiguous and unambiguous sentence facts.

Table 1. List of Test Results

No	Sentence	Category	Shift Reduce Parsing	Left Corner Parsing
1	Ibu Dini dirawat di rumahsakit	Unambiguous	Unambiguous	Unambiguous
2	Ali membeli sepeda baru seperti Diko	Ambiguous	Ambiguous	Ambiguous
3	Ibu membeli sayuran di pasar	Unambiguous	Unambiguous	Unambiguous
...
90	Ibu membaca dongeng baru ke cucunya	Ambiguous	Ambiguous	Ambiguous

Furthermore, the process of calculating the percentage of accuracy shown in tables 2 and 3 are the results of checking using Shift Reduce Parsing and Left Corner Parsing for each class.

Table 2. Evaluation Table in Confusion Matrix

Shift Reduce Parsing			
Class	Precision	Recall	F-Measure
Ambigu	0.70	0.23	0.35
Tidak Ambigu	0.7125	0.95	0.8143
Accuracy	0.71 = 71%		

Table 3. Evaluation Table in Confusion Matrix

Left Corner Parsing			
Class	Precision	Recall	F-Measure
Ambigu	0.67	0.2	0.3076
Tidak Ambigu	0.7037	0.95	0.8085
Accuracy	0.7 = 70%		

The analysis of the research results explains the comparison of measurement results with Shift Reduce Parsing and Left Corner Parsing which is shown in Fig. 4 comparison chart of system checks.

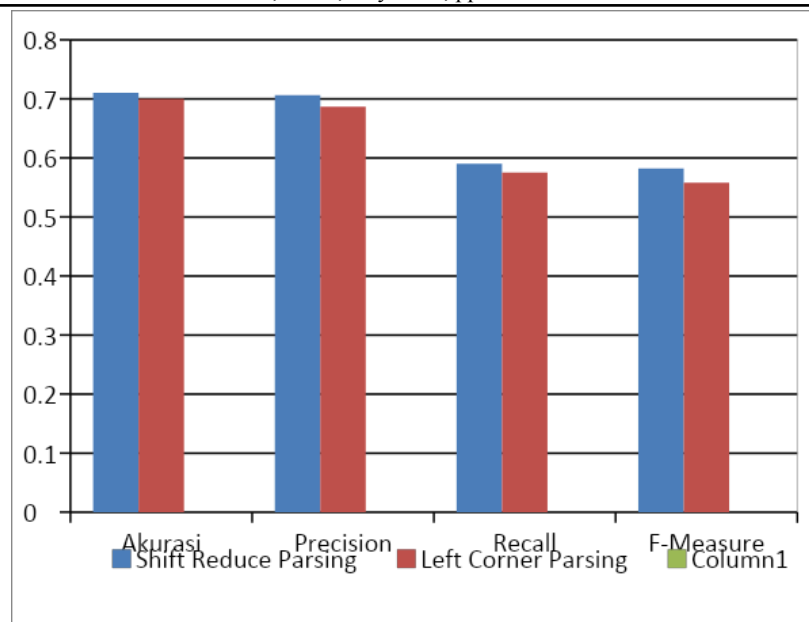


Fig. 4. Comparison Performance

Fig. 4 shows that in terms of accuracy, precision, recall, and f-measure, Shift Reduce Parsing is superior to Left Corner Parsing. We can conclude that the results of checking ambiguity using Shift Reduce Parsing are better than using Left Corner Parsing.

5. Conclusion

Based on the research results, Shift Reduce Parsing algorithm produces an accuracy value of 71%, precision 70.6%, recall 59%, and f-measure 58.2% in checking the ambiguity of sentence structure. While Left Corner Parsing produces an accuracy value of 70%, precision 68.7%, recall 57.5%, and f-measure 55.8%.

Further research is expected to apply other methods or other libraries in order to obtain better results.

References

- [1] Rizki, T.D. & Yusliani, N. 2016. Rancang Bangun Sistem Pengecekan Ambiguitas Kalimat Berbahasa Indonesia Menggunakan Harmony Search Algorithm. Annual Research Seminar, vol.2, N0.(1): 173–176.
- [2] Ambiguitas. 2016. Pada KBBI Daring. Diambil 07 Desember 2020, dari <https://kbbi.kemendikbud.go.id/entri/ambiguitas>
- [3] Lelywiay, C.J.S., Widowati, S. & L, K.M. 2019. Deteksi Pola Ambiguitas Struktural pada Spesifikasi Perangkat Lunak menggunakan Pemrosesan Bahasa Alami. 4: 51–64.
- [4] Liusti, S.A. 2016. Analisis Kalimat Berdasarkan Pola Kalimat Dasar Dan Kalkulus Predikat. Adabiyāt: Jurnal Bahasa dan Sastra, 15(2): 157.
- [5] Fachrurrozi, M., Yusliani, N. & Agustin, M.M. 2014. Identification of Ambiguous Sentence Pattern in Indonesian Using Shift-Reduce Parsing. Proceeding of The 1st International Conference on Computer Science and Engineering, 61–63.
- [6] Suprpto, S. 2000. Pemeriksaan Tata Bahasa dan Makna Kata Dalam Kalimat Bahasa Inggris Menggunakan Algoritma Left Corner Parsing.
- [7] Istiqamah, E. & Soyusiawaty, D. 2017. Penerapan Metode Left Corner Parsing dalam Aplikasi Terjemahan Bahasa Indonesia ke Bahasa Bima. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), 1–8.
- [8] Harnanto, A.D. 2007. Penguraiankata pada kalimat bahasa komering rasuan berdasarkan kaidah bahasa indonesia menggunakan teori automata. I(I): 31–36.
- [9] Rante, O., Paseru, D. & Saputro, I. 2016. Aplikasi Pemeriksaan Tata Bahasa Dalam Kalimat Bahasa Indonesia Menggunakan Algoritma Left Corner Parsing. JURNAL ELEKTRO, 9: 35–44.
- [10] Mujilahwati, S. 2016. Pre-Processing Text Mining Pada Data Twitter. Seminar Nasional Teknologi Informasi dan Komunikasi, 2016(Sentika): 2089–9815.

-
- [11] Shieber, S.M. 1982. Sentence Disambiguation by a Shift-Reduce Parsing Technique.
- [12] Tomuro, N. 2016. Left-Corner Parsing Algorithm for Uni cation Grammars. (February).
- [13] Wicaksono, A.F. 2014. HMM Based Part-of-Speech Tagger for Bahasa Indonesia HMM Based Part-of-Speech Tagger for Bahasa Indonesia. (January 2010).