

Syllabification of Balinese Words Using the Syllabification Algorithm

Gede Bagus Prawira Putra^{a1}, Ngurah Agus Sanjaya ER^{a2}

^aInformatics Department, Udayana University
Bali, Indonesia

¹bagusprawirap27@gmail.com

²agus sanjaya@unud.ac.id

Abstract

This research discusses the syllabification of Balinese words. The syllabification of Balinese words is necessary in determining padalingsa in pupuh. In order to be able to develop the application of technology in the field of pupuh literary, a system that can automatically find the correct division of a Balinese word into syllables is required. Several studies have been conducted on the syllabification Spanish and Sinhala language using rule-based approach [1][2]. Both of the studies achieved good accuracy score. Thus, in this research we apply the rule-based syllabification approach on Balinese words. The data used in this study were 257 Balinese words in which the system managed to correctly divide the word as many as 244 and obtained an accuracy score of 94.94%.

Keywords: Syllabification Algorithm, Word, Balinese, Word Syllabification, Rule-based

1. Introduction

Language is a system of symbols used to communicate with others, which includes creativity and a system of rules. With this creativity, humans are able to build meaningful sentences by combining words and also pay attention to the rules in making sentences so that, language is something that will never be stopped because it is a medium of communication [3].

There are various languages used in communication. Each region in Indonesia has different regional languages where each language in a region has different rules. As one example in the area of Bali has a regional language that is Balinese.

Language, literature and religion are very closely related and cannot be separated from Balinese culture. Language can be said as a unifying tool for language because language is used as a medium of communication by the community. In general, the Balinese people use Balinese as a medium of communication in daily life and is an important part of culture because in the life of the people, Balinese has its own position and function [4].

Spelling of the Balinese has its own rules, both in the use of letters, use of punctuation, writing and syllabification of words. Syllabification can be used to find syllables in Balinese words. Syllabification in Balinese can also be implemented in a Balinese literary work called pupuh.

The pupuh is bound by the padalingsa law (rules) consisting of the guru wilang and the guru dingdong where the guru wilang is a provision that binds the number of lines and the number of syllables in each row on each type of pupuh, while the guru dingdong is the one who regulates the fall of vowels at the end syllables in each line [5]. To be able to know padalingsa from the pupuh, it is necessary to seize words in each line of the pupuh. In order to be able to develop the application of technology in the field of literary pupuh, a system that can syllabification Balinese words is needed. To be able to make a system of syllabification into syllables can use the syllabification algorithm. There have been many studies that discuss syllabification algorithm, including [1] [2], but there is no related research on syllabification algorithm for Balinese.

In research [1] developed a syllabification algorithm for Spanish language using a rule-based approach. The rules implemented in the algorithm are generated based on the grammar rules of the Spanish language. In that study, an error rate of 1.6%.

In the research [2] developed Syllabification algorithm Sinhalese is the language of Sri Lanka. The algorithm used is also a rule-based approach. The rules used are also based on Sinhala grammar rules. The accuracy obtained in the research is 99.95%.

Because there is no research that develops syllabification algorithm for Balinese and also seen from previous studies using the syllabification algorithm get high accuracy, therefore this research develops the Balinese syllabification algorithm with a rule-based approach. Where the rules for syllabification the Balinese language words are taken from a book entitled "PEDOMAN UMUM EJAAN BAHASA BALI DENGAN HURUF LATIN" compiled by Ida Ayu Mirah Purwati, Ni Putu Ekatini Negari, and I Wayan Sudiarta [6].

2. Research Methods

2.1. Data Collection

The data used are Balinese words both basic words and affixed words. The data used were 257 where the data was obtained on the website page id.wiktionary.org. The data taken is a Balinese word that has not been labeled or whose syllabification are unknown, so labeling is done manually. There will be 2 .txt files where one saves the words to be tested and the other saves the decapitation of those words.

2.2. Balinese Words Syllabification Rule

Balinese has its own rules for syllabification. The result of syllabification is the syllable of the word. The syllables in the Balinese can consist of 6 categories according to figure 1 which consists of one vowel letter, preceded by vowels then followed by consonants, preceded by consonants and then followed by vowels, preceded by consonants then followed by a vowel then consonant, followed by consonant then followed by consonant then vowel, and followed by consonant then followed by consonant then vowel then consonant.

1. V	a -dep
2. VK	en -dep
3. KV	di -ngeh
4. KVK	sam -pi
5. KKV	pra -gat
6. KKVK	klang -sah
Ket:	
V -Vokal	K-Konsonan

Figure 1. Example of a Balinese Syllable Category

Syllabification of Balinese words can be done with the following rules:

- a. If in the middle of the word there are two consecutive vowels, syllabification is done between the two vowels. **Example: si-ap 'chicken'**
- b. If there are consonants in between the vowels in the middle of the word, including consonants, the syllabification is done before the consonants. **Example: ja-gut 'chin'**
- c. If in the middle of a word there are two consonants in a row, syllabification is carried out between the two consonants. **Example: sam-pi 'cow'**

- d. If in the middle of a word there are three or more consonants, syllabification is done between the first consonants. **Example: kam-plang 'slap'**
- e. The prefix and suffix affixes, including those that have changed shape, syllabification can be done at the turn of a line. **Example: ka-tulis 'written'**

2.3. Syllabification Algorithm for Balinese

Syllabification algorithm developed in this study uses a rule-based approach. The rules that are implemented in the Balinese syllabification algorithm are taken from a book entitled "PEDOMAN UMUM EJAAN BAHASA BALI DENGAN HURUF LATIN" compiled by Ida Ayu Mirah Purwati, Ni Putu Ekadini Negari, and I Wayan Sudiarta as written in sub CHAPTER 2.2 [6]. Following is the syllabification algorithm for Balinese:

- a. Check the characters in the word from the smallest to the largest index.
- b. Initialize symbols representing diphthong letters (\$ = ng, & = ny).
- c. Check letters that cannot be separated / diphthong (ng and ny).
- d. Change diphthong letters with one symbol (\$, &, or another).
- e. Create an index of words containing symbols, namely:
 - 1. If diphthongs are in front, index after the diphthong letters.
 - 2. If diphthongs are in the middle, index before and after diphthong letters.
 - 3. If diphthongs are located behind, index before the diphthong letters.
- f. Check vowels and consonants:
 - 1. if the i-th consonant and the i + 1 consonant are indexed after the i-th letter.
 - 2. If the i letter is vowel and the i letter is + 1 vowel, give an index after the i-th letter.
 - 3. If the i letter is vowel and the i letter is + 1 consonant and the i letter is + 2 vowels, give an index after the i letter.
 - 4. If the consonant i and the i + 1 consonant and the i + 2 consonant, give an index after the i-th letter.
- g. Make beheading according to the specified index.
- h. Change the symbol back to diphthong letters.

3. Result and Discussion

The implementation of the system uses as much test data as the data set has, which is 257 Balinese words. The system syllabification every word contained in the test data. The results will be compared with the syllabification label from the test data which has been done manually syllabification process manually. If the result of the syllabification of the system is the same as the test syllabification label, the system will be considered correct in syllabification. The accuracy of the system is calculated by comparing the amount of true test data with the total of all test data or if it is formulated in a mathematical model it will be like the formula (1)

$$Accuracy = \frac{Db}{N} \times 100\% \tag{1}$$

Explanation:

N = Overall total test data

Db = Test data is correct

The test results show that the algorithm can syllabification almost entire words. Table 1 shows the test results of the Balinese Syllabification Algorithm

Table 1. System Testing Results

Amount of test data	Correct	Incorrect	Accuracy
257	244	13	94,94%

Examples of Balinese words used in testing can be seen in table 2

Table 2. Examples of Test Data and Syllabification Results

No	Word	Result
----	------	--------

1	Ancruk	An-cruk
2	Angklung	Ang-klung
3	Andus	An-dus
4	Baat	Ba-at
5	Badak	Ba-dak
6	Bakat	Ba-kat
7	banjar	Ban-jar
8	Baong	Ba-ong
9	Bapa	Ba-pa
10	Barong	Ba-rong
11	Basing	Ba-sang
12	Basange	Ba-sa-nge
13	Bates	Ba-tes
14	Batis	Ba-tis
15	Bawak	Ba-wak
16	Be	Be
17	Bebekel	Be-be-ke
18	Bedik	Bed-ik
19	Bek	Bek
20	belog	Bel-og

4. Conclusion

Syllabification algorithm is an algorithm that is used to syllabification words. Rules for syllabification vary depending on the language used. In general, syllabification algorithm uses a rule (rule-based) approach and the rule used is the rule in the fragment of the word itself. By using the syllabification algorithm and using the rules of syllabification Balinese words, the system proved successful in syllabification words with quite good results. The accuracy obtained by the system is 94.94% with the amount of data that was successfully beheaded with a large amount of 244 words from 257 test data.

References

- [1] H. Cuayahuitl, "A syllabification algorithm for spanish," in International Conference on Intelligent Text Processing and Computational Linguistics, pp. 412–415, Springer, 2004.
- [2] R. Weerasinghe, A. Wasala, and K. Gamage, "A rule based syllabification algorithm for sinhala," in International Conference on Natural Language Processing, pp. 438–449, Springer, 2005.
- [3] N. Dhieni, L. Fridani, A. Muis, and G. Yarmi, "Metode pengembangan bahasa," 2014.
- [4] I. K. Sudarsana, "Menumbuhkan minat belajar bahasa bali pada kalangan remaja," Prosiding Sembada 2017, 2017.
- [5] I. G. A. M. Putra, P. Aji, and M. H. Syahbani, "Aplikasi pembelajaran nyanyian pupuh tradisional bali berbasis android di sekolah dasar," ePro-ceedings of Applied Science, vol. 1, no. 2, 2015.
- [6] I. W. S. Ida Ayu Mirah Purwati, Ni Putu Ekatini Negari, PEDOMAN UMUM EJAAN BAHASA BALI DENGAN HURUF LATIN. Balai Bahasa Provinsi Bali, 2013.