



# Rules for Orthographic Word Parsing of the Philippines' Cebuano-Visayan Language Using Context-Free Grammars

Roseclaremath A Caroro, Misamis University, Ozamiz, Philippines

 <https://orcid.org/0000-0002-2351-531X>

Rolysent K. Paredes, Misamis University, Ozamiz, Philippines

 <https://orcid.org/0000-0002-5041-9358>

Jerry M. Lumasag, Misamis University, Ozamiz, Philippines

## ABSTRACT

Syllabication is essential in the preprocessing stage of speech systems. In the context of the Philippines' Cebuano-Visayan language's syllabication rules, the existing rules do not include hyphenated words although the hyphen defines the syllable boundary in a word. Hence, this study created grammar rules for hyphenated words which include sequences of a hyphen between vowel-consonant, consonant-consonant, vowel-vowel, and consonant-vowel. The test was done for the enhanced grammar rules for Cebuano-Visayan syllabication with 1,465 representative hyphenated and non-hyphenated words of varying lengths. The result further implies that the syllabication analysis for hyphenated words showed that hyphens improve the naturalness and intelligibility in the utterance of the words, thereby enhancing the understanding and comprehension of the Cebuano-Visayan discourse.

## KEYWORDS

Cebuano-Visayan Language, Context-free Grammar, Orthographic Syllabication, Parsing Algorithm, Rule-based Syllabication

## INTRODUCTION

The Cebuano-Visayan dialect of the greater central Philippines belongs to a Malayo-Polynesian based language (Reid, 2017), being the second strongest and most spoken language in the country, particularly in Visayas and Mindanao. Its syllabication used the vowel as the marker for the number of syllables of the word which spells as it sounds with slight difference considering its meaning (Caroro et al., 2015), similar to that of the Filipino language (Fajardo & Kim, 2015) and some other dialects in other areas of the country. The dialect has 15 consonants and three (3) vowels. However, the advent of colonization in the country including the greater central area recognized the vowels e and o. Thus, the present Cebuano-Visayan orthography adopted the alphabet consisting of 5 vowels and 15 consonants (Fajardo & Kim, 2015).

A current study on the Cebuano-Visayan language defined 11 syllabication rules (Garcia et al., 2015). These syllabication rules generated a text-to-speech system for the language (Garcia et al., 2015) and a data-driven prosodic text-to-speech system (Caroro et al., 2015). Thus, the syllabication is significantly important in e-learning and speech systems.

DOI: 10.4018/IJSSCI.2020040103

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Syllabication is an essential process for determining syllable boundaries and for retrieving similar phones using either its phonetic or orthographic structure (Ramli et al., 2015; Räsänen et al., 2018). Also, the syllable structure varies in languages and approaches which could be orthographic or phonetic (Krisnawati & Mahastama, 2018). Furthermore, various languages require different models of automatically generating syllables, such as rule-based syllabication (Devi et al., 2017; Kovač & Marković, 2018), data-driven (Haverals et al., 2019), nearest neighbor rule (Suyanto et al., 2016), and finite-state transducers (Krisnawati & Mahastama, 2018), among others. Rule-based syllabication, a reliable approach in generating syllables, follows certain rules in deriving the syllables of the word (Krisnawati & Mahastama, 2018), such as that of Asian scripts (Neto et al., 2015).

Part of orthographic learning is the word syllabication which the Cebuano-Visayan language already defined rules based on its orthographic sequences. However, the existing syllabication rules do not include the production of syllables containing hyphenated Cebuano-Visayan words. Improper syllabication without considering the presence of the hyphen in the word will result in the incorrect utterance of the word or incorrect generation of speech synthesis if applied to automated speech systems. In this instance, the incorrect utterance due to incorrect syllabication will lead to misunderstanding on the side of the non-speakers of the language. Besides, the incorrect syllabication rules for hyphenated words if applied to systems providing language tutorials will produce unreliable tutorials to the end-users. Hence, this study created additional grammar rules to include the syllabication of hyphenated words. These grammar rules include parsing the given words with a hyphen between a vowel and a consonant, a consonant and a consonant, a vowel and a vowel and a consonant and a vowel sequence. The syllabication rules for hyphenated words are necessary to preserve the consistency of syllabication analysis (Devi et al., 2017). The syllabication rules play a very vital role in most speech systems wherein the speech synthesis uses syllables. Furthermore, most speech systems obtained speech synthesis through a concatenative approach, thereby concatenating speech signals based on the syllable sequences generated by the system.

## LITERATURE REVIEW

The paper includes varied literature that significantly presents useful information needed in realizing the goal of the study. The literature includes the concept of syllabication, the applications of the syllabication algorithm to speech systems, the brief discussion of one of the instances of the Malayo-Polynesian language, the Cebuano-Visayan dialect, and the Cebuano-Visayan syllabication process.

### The Syllabication Process

Syllabication is the first process in developing speech systems, particularly text-to-speech (TTS), which supports the naturalness and intelligibility of the TTS system (Ramli et al., 2015). Hence, proper syllabication is necessary particularly in showing various features of the language's articulation and pronunciation as syllable-based synthesizers produce more speech synthesis (Pradhan et al., 2015). Some syllabication processes are through the rule-based approach, although there are other approaches to segmenting the words.

The scope of rule-based approaches include data mining such as association rule mining (Wang et al., 2019), formal rules with fuzzy logic (Jin et al., 2019), and natural language processing (Nimaiti & Izumi, 2014; Kovač & Marković, 2018; Krisnawati & Mahastama, 2018; Ravishankar & Shriram, 2018; Rajendran & Kumar, 2019; Zerrouki et al., 2019). Rule-based approach to natural language processing spans its scope among topics such as translation system (Nimaiti & Izumi, 2014), text classifications (Ravishankar & Shriram, 2018), text-to-speech systems (Rajendran & Kumar, 2019; Zerrouki et al., 2019), auto-syllabication systems (Kovač & Marković, 2018; Krisnawati & Mahastama, 2018), among others. A rule-based syllabication algorithm refers to segmenting words based on the splitting rules required in particular language orthography (Neto et al., 2015; Ramli et al., 2015; Devi et al., 2017; Räsänen et al., 2018). Besides, the syllabication rules defined the syllabication process

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/rules-for-orthographic-word-parsing-of-the-philippines-cebuano-visayan-language-using-context-free-grammars/252214](http://www.igi-global.com/article/rules-for-orthographic-word-parsing-of-the-philippines-cebuano-visayan-language-using-context-free-grammars/252214)

## Related Content

---

### SOM-Based Class Discovery for Emotion Detection Based on DEAP Dataset

Aladdin Ayesh, Miguel Arevalillo-Herráez and Pablo Arnau-González (2018). *International Journal of Software Science and Computational Intelligence* (pp. 15-26). [www.irma-international.org/article/som-based-class-discovery-for-emotion-detection-based-on-deap-dataset/199014](http://www.irma-international.org/article/som-based-class-discovery-for-emotion-detection-based-on-deap-dataset/199014)

### A Least-Laxity-First Scheduling Algorithm of Variable Time Slice for Periodic Tasks

Shaohua Teng, Wei Zhang, Haibin Zhu, Xiufen Fu, Jiangyi Su and Baoliang Cui (2010). *International Journal of Software Science and Computational Intelligence* (pp. 86-104). [www.irma-international.org/article/least-laxity-first-scheduling-algorithm/43899](http://www.irma-international.org/article/least-laxity-first-scheduling-algorithm/43899)

### Missing Data Estimation Using Rough Sets

Tshilidzi Marwala (2009). *Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques* (pp. 94-116). [www.irma-international.org/chapter/missing-data-estimation-using-rough/6797](http://www.irma-international.org/chapter/missing-data-estimation-using-rough/6797)

### A Novel Convolutional Neural Network Based Localization System for Monocular Images

Chen Sun, Chunping Li and Yan Zhu (2019). *International Journal of Software Science and Computational Intelligence* (pp. 38-50). [www.irma-international.org/article/a-novel-convolutional-neural-network-based-localization-system-for-monocular-images/233522](http://www.irma-international.org/article/a-novel-convolutional-neural-network-based-localization-system-for-monocular-images/233522)

### Data Mining Applications in the Electrical Industry

Rubén Jaramillo Vacío, Carlos Alberto Ochoa Ortiz Zezzatti and Armando Ríos (2012). *Logistics Management and Optimization through Hybrid Artificial Intelligence Systems* (pp. 380-402). [www.irma-international.org/chapter/data-mining-applications-electrical-industry/64930](http://www.irma-international.org/chapter/data-mining-applications-electrical-industry/64930)