



THAI GRAPHEME-TO-PHONEME USING PROBABILISTIC GLR PARSER

Pongthai Tarsaku, Virach Sornlertlamvanich and Rachod Thongprasirt

Information R&D Division, National Electronics and Computer Technology Center (NECTEC)
Gypsum Metropolitan Bldg., 22nd Fl., 539/2 Sri Ayudhaya Rd., Rajthevi, Bangkok 10400,
Thailand.
ptarsaku@notes.nectec.or.th

Abstract

Many difficulties in the Thai language such as the absence of boundary word, linking syllables in pronunciation, and homographs are challenging us in developing a Thai Grapheme-to-Phoneme (G2P) converter. Presently there are a couple Thai G2P systems which are proposed in ruled-based and decision-tree approach. The rule-based approach has a drawback in the limitation of employing the context. The decision-tree approach is somehow able to capture the local context for making the decision. On the contrary, the Probabilistic Generalized LR (PGLR) approach is reported that both the global and local context are efficiently captured in the probabilistic model. In this paper, we implement a Thai G2P system based on the PGLR approach. The result of experiment shows 90.44% of word accuracy in case of ignoring vowels length and 72.87% of word accuracy in case of exact match evaluation. These results are superior to those of rule-based and decision-tree approaches.

1. Introduction

A Grapheme-to-Phoneme (G2P) module is a routine that converts an input word sequence into their corresponding phonetic transcription. It is one of the essential routines in developing a text-to-speech (TTS) module.

In order to implement a Thai G2P system, many language specific problems i.e. homograph, non-word boundary, tonal language, etc., need to be solved.

Various approaches, i.e. dictionary-based, rule-based, and statistical-based approaches have already been proposed for Thai G2P system [1], [2], [3], [5], [6], [9].

Luksaneeyanawin [3] and Mittapiyanurak [5] proposed a dictionary-based Thai G2P system for using in the text-to-speech (TTS) system. This approach requires a large dictionary, and it can not deal with unregistered word problem. To overcome the problem, Narupiyakul [6] and Khanya [2] proposed a rule-based approach for Thai G2P system. In their system, they prepared a set of rules of syllable construction and implemented with a finite state machine for detecting syllable structures. Wiboon [9] proposed a rule-based approach using regular expression to deal with the unregistered word problem. However, the rule-based approach can not deal with the ambiguity in syllabification of the Thai language. Chotimongkol [1] proposed a decision tree approach, but the problem of inversion of phonemes across syllable boundaries is left unsolved.

PGLR parsing is a technique invented for incorporating probabilistic value with the parsed trees of a sentence [7]. The PGLR has the advantage of context-sensitivity based on the

GLR parsing framework [8]. The PGLR can efficiently capture the information from both the global and local context during the parse. This paper is organized in 5 sections. Section 2 describes Thai syllable structure and the difficulties in determining word pronunciation. Section 3 explains the architecture of G2P implemented on PGLR framework. Section 4 shows the experimental results.

2. Thai Pronunciation System

2.1. Thai Syllabic Representation

A basic Thai-pronunciation unit is a syllable that can be represented in the form of $C_i V (C_f) (T)$, where C_i , V , C_f , and T denote an initial consonant, a vowel, a final consonant and a tonal marker, respectively. The subunits inside the parenthesis, i.e. C_f and T , are optional. Details of each subunit are described in the followings:

1. Initial consonant (C_i): There are 44 consonants in Thai and all of them can be an initial consonant of a syllable (two of them, “จ” and “ก”, are not often used in modern Thai). However, only 21 phonemes are needed to represent all these consonants as shown in Table 1.

Table 1: Phonetic symbol of Thai consonants.

Consonant	Phoneme		Consonant	Phoneme	
	Initial	Final		Initial	Final
ก	/k/	/k/	บ	/b/	/p/
ข, ค, ฉ	/kh/	/k/	ป	/p/	/p/
ง	/ng/	/ng/	ผ, พ, ภ	/ph/	/p/
จ	/c/	/t/	ฝ, ฟ	/f/	/p/
ฉ, ช, ฅ	/ch/	/t/	ม	/m/	/m/
ซ, ศ, ษ, ส	/s/	/t/	ร	/r/	/n/
ญ, ย	/j/	/j/	ล, ฬ	/l/	/n/
ฎ, ฏ	/d/	/t/	ว	/w/	/w/
ฏ, ฐ	/t/	/t/	ห, ฮ	/h/	-
ฐ, ฉ, ฌ, ฉ, ฌ, ฐ	/th/	/t/	อ	/ʔ/	-
ณ, น	/n/	/n/			

Some initial consonants are written pairly and act as an initial consonant. This type of initial consonants is called ‘double consonant’. There are four categories of double consonant:

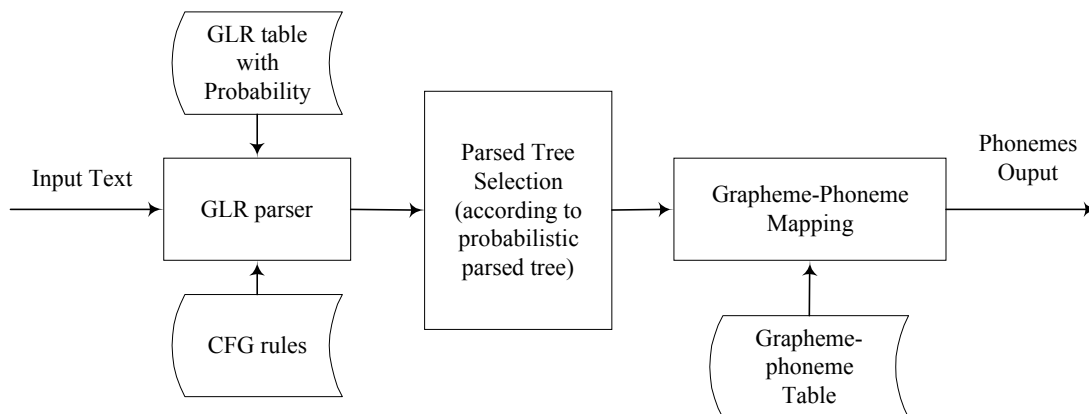


Figure 1: Thai G2P System: the PGLR approach.

3.2 Context-Free Grammar Rules

Syllable structures are described in CFG rules for preparing the LR parsing table. The CFG rules are designed according to the vowel structure. For example, the CFG rules for the syllable structure. For example, [initial consonant] [vowel] [Final consonant] to parse the words “การ”, “นาห”, and “ชาร” is represented as:

$$\langle \text{Syl_Type} \rangle \rightarrow \langle \text{ini-cons} \rangle Vx \langle \text{fin-cons} \rangle,$$

where $\langle \text{Syl_Type} \rangle$, $\langle \text{ini-cons} \rangle$ and $\langle \text{fin-cons} \rangle$ are non-terminal nodes representing syllable type, initial consonant, and final consonant, respectively and Vx is a terminal node representing a vowel.

Since a Thai word can be either monosyllable or polysyllable, followings are the sample of CFG rules for word formation from syllable units:

$$\begin{aligned} \langle \text{Word} \rangle &\rightarrow \langle \text{Syllable} \rangle \\ \langle \text{Syllable} \rangle &\rightarrow \langle \text{Syllable} \rangle \langle \text{Syl_Type} \rangle \\ \langle \text{Syllable} \rangle &\rightarrow \langle \text{Syl_Type} \rangle. \end{aligned}$$

In case that some combinations of characters are not allowed due to the spelling regulation, additional restricted rules are prepared to reduce the acceptable parsed candidates. For example, the final consonant “ง” can not follow some vowels, e.g. อ, ใ, ใ, ใ, ใ, ใ, and ใ. This kind of restriction is introduced to restrict the invalid syllable sequences.

In order to handle the linking-syllable pronunciation problem, the possible final consonants for pronunciation as a linking-syllable are defined with a new non-terminal type $\langle \text{FinConwithLinking} \rangle$ composed in a syllable structure such as

$$\langle \text{Syl_TypewithLinking} \rangle \rightarrow \langle \text{InitCon} \rangle \gamma \langle \text{FinConwithLinking} \rangle$$

corresponding to the original rule:

$$\langle \text{Syl_Type} \rangle \rightarrow \langle \text{InitCon} \rangle \gamma \langle \text{FinCon} \rangle.$$

Since some final consonants, e.g. “ง”, “จ”, and “น”, can not be pronounced as a linking-syllable, the new non-terminal $\langle \text{FinConwithLinking} \rangle$ is defined as a different non-terminal from the ordinary final consonant $\langle \text{FinCon} \rangle$.

Furthermore, consonant type specific rules are also prepared for solving the ambiguity in consonant functionality problem.

These rules depend on the syllable structure. For example, the rule:

$\langle \text{SylTypeSpecial} \rangle \rightarrow \langle \text{InitCon} \rangle \text{Vowel1} \langle \text{FinCons} \rangle \text{Vowel2}$ is prepared for parsing the word such as “อช” (ash). In this case, $\langle \text{InitCon} \rangle$, Vowel1 , $\langle \text{FinCons} \rangle$, and Vowel2 correspond to the characters “อ”, “ช”, “อ”, and “ช”, respectively.

3.3 Training of the PGLR parsing table

The probabilistic GLR parsing table is a LR table in which each action is assigned a probability gained from parsing a bracketed training set [7]. LEXiTRON is a Thai electronic dictionary containing 23,000 entries with pronunciation provided [10]. A four-fifth of the LEXiTRON is randomly selected for training. All possible parsed trees for each word are generated by the GLR parser. Only those with the correct parsed trees (phoneme sequence) comparing with the word pronunciation in LEXiTRON, are selected. As a result, the correct parsed tree for each word is listed up for training.

4. Experimental results

The remained one-fifth of the LEXiTRON, is used to test the PGLR model. Table 3 shows the results of the PGLR model comparing with the rule-based [9] and the decision-tree models [1]. The evaluation shows the grapheme-to-phoneme conversion accuracy in both case of exact match and ignorance of vowel length match. The exact match accuracy indicates that the output of the model has exactly the same phoneme sequence as the one given in LEXiTRON. The ignorance of vowel length match accuracy shows the model accuracy in yielding the phoneme sequence by ignoring the length of vowel which is less important than the phoneme sequence. It is expected that the vowel length may be determined by other means.

The PGLR approach yields the best result in comparing with those of the rule-based and decision-tree approaches. However, about half of the errors are the words that include linking syllables. This problem is also not trivial for native speakers either when the word is unseen.



Table 3: Grapheme-to-phoneme conversion accuracy

Model	Conversion (word) accuracy(%)	
	Exact match	Ignorance of Vow. Length
PGLR	72.87	90.44
Rule-based	67.14	83.81
Decision Tree	68.76	86.94

5. Conclusions

Theoretically, PGLR can capture both local and global context which is reported the high performance in syntactic parsing for sentences. Our research has shown the best result in applying PGLR in G2P problem. This concludes that the context for probabilistic decision in PGLR is good enough for a relatively small structure such as words (phoneme sequence) as well.

6. References

- [1] Chotimongkol, A. "Statistically Trained Orthographic to Sound Models for Thai," Proceeding of the 6th International Conference on Spoken Language Processing," pp. 533-554, Oct. 2000.
- [2] Khamya, A. "SATTS: Syllable Analysis for Text-to-Speech System," Proceeding of the Fourth Symposium on Natural Language Processing 2000 (SLNP'2000), pp. 336-340, May 2000.
- [3] Luksaneeyanawin, S. "A Thai Text to Speech System," Proceeding of the Conference of the Regional Workshops on Computer Processing of Asian Languages. Asian Institute of Technology, pp. 305-315, 1990.
- [4] Luksaneeyanawin, S. "Speech Computing and Technology in Thailand," NLP in Thailand, pp. 276-321, 1993.
- [5] Mittapiyanurak, P., Hansakunbuntheung, C. and Teprasit, V. and Sornlertlamvanich, V. "Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach," Proceedings of NECTEC Annual Conference, Bangkok Thailand, pp. 483-495, 2000.
- [6] Narupiyakul, L. "The Phonetic Transcription of Thai Word," Proceeding of the International Symposium on Intelligent Signal Processing and Communication Systems, pp. 789-792, 1999.
- [7] Sornlertlamvanich, V. " Probabilistic Language Modeling for Generalized LR Parsing," Technical Report, Department of Computer Science, Tokyo Institute of Technology, September, 1998.
- [8] Tomita, M., "Generalized LR Parsing," Kluwer Academic Publishers, 1991.
- [9] Wiboon, T. "The Engineering Training Report at NECTEC," Engineering Faculty, Khasetsart University, 1999.
- [10] <http://www.links.nectec.or.th/lexit/>.