8.3 Thai Grapheme-to-Phoneme by Probabilistic GLR Approach

Virach Sornlertlamvanich, TCL, NICT, Thailand

Applying a common writing system to a language benefits us on both reading a text written in an unfamiliar language script, and machine processing in text conversion procedure. There are several efforts based on Romanization, transliteration, or transcription approaches proposed to represent a Thai text by a universal alphabetic system. Though some ambiguities occur in Romanization, we introduce the approach to Grapheme-to-Phoneme (G2P) conversion to avoid the use of diacritical marks to promote the legibility and interoperability. Because of the greater number of characters in the Thai writing system, some characters are inevitably used to share in representing the same sounds.

Luksaneeyanawin (1993) and Mittapiyanurak, et al. (2000) proposed a dictionary-based Thai G2P system for using in the text-to-speech (TTS) system. This approach requires a large and up-to-date dictionary, and the research did not consider how to deal with the unregistered word problem. To overcome the problem, Narupiyakul (1999) and Khamya (2000) proposed a rule-based approach for Thai G2P system. In their system, they prepared a set of rules for syllable construction and implemented with a finite state machine for detecting syllable structures. Wiboon (1999) proposed a rule-based approach using regular expression to deal with the unregistered word problem. However, the rule-based approach still cannot deal with the ambiguity in homograph and syllabification of the Thai language. Chotimongkol (2000) proposed a decision-tree based approach, but the problem of inversion of phonemes across syllable boundaries is left unsolved.

We proposed the Probabilistic Generalized LR (PGLR) approach which is reported to be able to capture both the global and local context efficiently in creating a probabilistic model. The experiment of G2P conversion for the Thai language has shown that the result of PGLR approach is superior to those of rule-based and decision-tree based approaches.

8.3.1 Thai Pronunciation System

8.3.1.1 Thai Syllabic Representation

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

8.3.1.2 Difficulties in Thai pronunciation system

Ambiguity in grapheme to phoneme mapping: Some graphemes can be mapped to multiple phonemes depending on the context.

Homograph: The words that are pronounced differently according to the semantic context. In order to pronounce a homograph word, the sentential context is required.

Vowel length: The problem occurs when a vowel is pronounced as a short (long) vowel according to its grapheme, but by any reasons it is pronounced as a long (short) vowel instead.

Linking syllable pronunciation: The problem occurs in a polysyllabic word where the final consonant of the forthcoming syllable is explicitly pronounced with /a/ vowel as an additional syllable.

Ambiguity in consonantal functionality: In some Thai words, a consonant can be both the final consonant of forthcoming syllable and the initial consonant for the next syllable.

Word boundary: Unlike some other languages such as English, Thai has no word boundary. Therefore, different segmentation can yield different words and pronounce in different syllable units.

8.3.2 Grapheme-to-Phoneme System

8.3.2.1 PGLR Approach

PGLR is based on GLR parsing, modeled to incorporate the probabilistic value of the parsed trees of a sentence. The trained probabilistic value of a parsed tree is distributed in shift and reduce actions, normalized by the state and look-ahead in the LR table. The PGLR has the advantage of context-sensitivity based on the GLR parsing framework. The PGLR can efficiently capture the information from both the global and local context during the parse.

8.3.2.2 Context-Free Grammar for Syllable Structure

Syllable structures are described in CFG rules for preparing the LR parsing table. The rules are grouped according to the vowel structure, e.g. [*initial consonant*] [*vowel*] [*final consonant*] is expressed as:

<Syl_Type> → *<ini-cons> Vx <fin-cons>*

where *<Syl_Type>*, *<ini-cons>* and *<fin-cons>* 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:

<Word> → *<Syllable>*
<Syllable> → *<Syllable> <Syl_Type>*
<Syllable> → *<Syl_Type>*

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 "ร" cannot follow some vowels, e.g. " ิ ", " ี ", " ุ ", " ู ", " ึ ", and " ื ".

In order to handle the linking-syllable pronunciation problem, the possible final consonants to be pronounced as a linking-syllable are defined with a new non-terminal type <*FinConwithLinking*> composed in a syllable structure such as

<Syl_TypewithLinking> → *<InitCon>* า *<FinConwithLinking>*

corresponding to the original rule:

<Syl_Type> → *<InitCon>* า *<FinCon>*

Since some final consonants, e.g. "ง" ,"ด", and "บ", cannot be pronounced as a linking-syllable, the new non-terminal *<FinConwithoutLinking>* is defined as a different non-terminal from the ordinary final consonant *<FinCon>*.

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:

<SylTypeSpecial> → *<InitCon> Vowel1 <FinCons> Vowel2*

is prepared for parsing the word such as “อัฐิ” (ash). In this case, *<InitCon>*, *Vowel1*, *<FinCons>*, and *Vowel2* correspond to the characters “อ”, “ ั ”, “ฐ”, and “ ิ ”, respectively.

8.3.2.3 Training of the PGLR parsing table

A four-fifth of the LEXiTRON, a Thai electronic dictionary containing 23,000 entries with correct word pronunciation, 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 to generate the training set.

8.3.3 Experimental results

The remained one-fifth of the LEXiTRON, is used to test the PGLR model comparing with the rule-based and the decision-tree models. The accuracy of ‘with vowel length’ indicates that the output of the model has exactly the same phoneme sequence as the one given in LEXiTRON. The accuracy of ‘without vowel length’ shows the model accuracy in yielding the phoneme sequence by ignoring the length of vowel.

Table 1.1 Accuracy of grapheme-to-phoneme conversion.

Model	Word conversion accuracy (%)	
	With vowel length	Without vowel length
PGLR	72.87	90.44
Rule-based	67.14	83.81
Decision Tree	68.76	86.94

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.

8.3.4 Conclusion

Based on the syllable structure rule written in CFG manner, after training the model by the pronunciation of word entry of a dictionary, the PGLR approach can outperform both rule-based and decision tree based approaches in G2P problem. This concludes that the context for building a probabilistic model in PGLR is considerable for a relatively small structure of a word (namely sequence of phoneme) as well.

Bibliography

Chotimongkol, A. (2000). Statistically Trained Orthographic to Sound Models for Thai, *Proceedings of the 6th International Conference on Spoken Language Processing*, pp. 533-554.

Khamya, A. (2000). SATTS: Syllable Analysis for Text-to-Speech System, *Proceedings of the Fourth Symposium on Natural Language Processing 2000 (SLNP'2000)*, pp. 336-340.

Luksaneeyanawin, S. (1993). Speech Computing and Technology in Thailand, *NLP in Thailand*, pp. 276-321.

Mittapiyanurak, P., Hansakunbuntheung, C. and Tepprasit, V. and Sornlertlamvanich, V. (2000). Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach, *Proceedings of NECTEC Annual Conference,* Bangkok Thailand, pp. 483-495.

Narupiyakul, L. (1999). The Phonetic Transcription of Thai Word, *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 789-792.

Sornlertlamvanich, V. (1998). Probabilistic Language Modeling for Generalized LR Parsing, *Technical Report,* Department of Computer Science, Tokyo Institute of Technology.

Tarsaku, P., Sornlertlamvanich, V. and Thongpresirt, R. (2001). Thai Grapheme-to-Phoneme Using Probabilistic GLR Parser, *Proceedings of Eurospeech 2001*, Aalborg, Denmark.

Wiboon, T. (1999). The Engineering Training Report at NECTEC, Engineering Faculty, Khasetsart University.