Grapheme-to-Phoneme for Thai

Pongthai Tarsaku
Content

- Introduction
- Grapheme-to-Phoneme in TTS system
- Problems in Thai
- PGLR Approach
- Experiment & Results & Discussion
- Conclusion
Introduction

- Grapheme-to-Phoneme (G2P) is a module in TTS system.
- Grapheme-to-Phoneme approaches.
  - Dictionary base.
  - Rule base.
  - Statistical base.
- Probabilistic Generalized LR (PGLR) parser is statistical base approach.
G2P in TTS system

Text Segmentation

Grapheme-to-Phoneme

/п@:4/п@:p1/п@:н0/т@:н3/т@:и2/м@:и0/ж@:м2/ч@:м0/н@:м0/

Prosody Generation

Speech Signal Synthesis

Speech Waveform
Problems in Thai (1)

- Ambiguity in grapheme-phoneme mapping.
  - “มณฑา” is pronounced as /mon0/tha:0/
  - “มณฑป” is pronounced as /mon0/dop1/

- Homograph.
  - “เพลา” (axe) is pronounced as /phlaw0/
  - “เพลา” (time) is pronounced as /phe:0/la:0/

- Vowel’s length
  - “น้ำ” is phonologically pronounced as /nam3/ but usually pronounced as /na:m3/
Problems in Thai (2)

- Linking syllable pronunciation.
  - “วิท” in “วิทยา” is pronounced as /wit2/tha2/

- Ambiguity in consonantal functionality.
  - “อัฐิ” is pronounced as /?at1/thi1/

- Word boundary.
  - “ตากลม” can be segmented into “ตา” (round eye) and “กลม” (to expose wind) which are pronounced /ta:0/klom0/ and /ta:k1/lom0/ respectively.
PGLR Approach

- PGLR : Probabilistic Generalize LR parsing.
- PGLR has advantage in context-sensitivity.
- PGLR is able to capture two levels of context.
  - Global context - over structures from the CFG rules.
  - Local n-gram context.
Context-Free Grammar Rules

- A CFG rule is prepared for Thai syllable construction.
- A set of CFG rules is grouped by Thai vowel unit.
  (21 groups and 3 special groups)
- CFG rules are able to cover both monosyllable and polysyllable.

<Grp_1> → <init> 1
<Grp_1> → <init> 1 <final>
<Grp_2> → 1 <init>
<Grp_2> → 1 <init> <final>
Thai Grapheme-to Phoneme system

PGLR parser → Most probable parse tree → G-P Mapping → Toneme Generation

PGLR Table

CFG Rules

G-P Table

สมชาย

/som/chaj/

/som4/chaj0/
Grapheme-Phoneme Mapping

Example.
Experiment

Database
- LEXiTRON: The Thai electronic dictionary is used for training and testing.
- ~23000 Thai words with pronunciation.

Training
- Four-fifth of database is used for training.

Testing
- One-fifth of database is used for testing.
## Result

<table>
<thead>
<tr>
<th>Model</th>
<th>Conversion (word) accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact match</td>
</tr>
<tr>
<td><strong>PGLR</strong></td>
<td>72.87</td>
</tr>
<tr>
<td>Rule-based</td>
<td>67.14</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>68.76</td>
</tr>
</tbody>
</table>
Discussion

- Vowel’s length problem is dominant (90.44 - 72.87).
- Half of all errors (~5%) come from linking syllable problem.
- To improve accuracy, more training data is required.
Conclusion

- PGLR approach has advantage in context-sensitivity (both global and local context).
- The efficiency of PGLR parser depends on carefully writing in CFG rules.
- This approach can be applied in syllable segmentation framework or soundex conversion framework.
Thank you
Tone in Thai

- There are 5 tone levels (Tonemes) in Thai.
  - mid-level : 0
  - low-level : 1
  - falling-level : 2
  - high-level : 3
  - rising-level : 4

- Toneme is depended on consonant class, syllable type and tone marker.
## Tonemic Generation

<table>
<thead>
<tr>
<th>Consonant Class</th>
<th>Syllable type</th>
<th>Tone Markers and Tonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>unmarked</td>
</tr>
<tr>
<td><strong>High Class</strong></td>
<td>Live Syllable</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Dead Syllable</td>
<td>1</td>
</tr>
<tr>
<td><strong>Middle Class</strong></td>
<td>Live Syllable</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dead Syllable</td>
<td>1</td>
</tr>
<tr>
<td><strong>Low Class</strong></td>
<td>Live Syllable</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dead Syllable</td>
<td>1</td>
</tr>
</tbody>
</table>

**Notes:**
- The final consonant is n, ng, m, or j.
- Long vowel with no final consonant (z).
- Short vowel with final consonant k, t, or p.
- Long vowel with final consonant k, t, or p.
Parse Tree Selection

The selected parse tree is used for training.

The selected parse tree is used for training.