Multilingual Lexicon

Tokunaga Takenobu

Tokyo Institute of Technology

Goal & Milestones

- A multilingual lexicon of 200,000 entries for the 20 languages (hopefully including Asian languages) that is usable for MT systems by 2015

- Defining specifications of entries (3 yrs)

- Basic 5,000 entries (2 yrs)

- Scaling up to 200,000 entries (5 yrs)

EAGLES applied on Asian Languages (ISLE WS)

- Chinese
- (Formosan)
- Hindi
- Japanese
- Korean
- Thai

Methodology

- Distribute the EAGLES documents

- Test applicability on each language (noun, verb) by comparing existing linguistic resources (lexicon, tagged corpora etc.)

- Comparison and summarization

General Comments

- Clear contract between inflectional languages and others
- Information marked by inflection in the EAGLES framework is often expressed by auxiliaries, particles and affixes in Asian languages.
- Information and its representation should be distinguished clearly

Contributors

Chu-Ren Huang (Chinese)
Elizabeth Zeiton (Formosan)
Paul J.K. Li (Formosan)
Dipti Misra Sharma (Hindi)
Tokunaga Takenobu (Japanese)
Hae-Yun Lee (Korean)
Asanee Kawtrakul (Thai)
Thatsanee Charoenporn (Thai)
Virach Sornlertlamvanich (Thai)