KyotoEBMT System Description for the 1st Workshop on Asian Translation

John Richardson     Fabien Cromières     Toshiaki Nakazawa     Sadao Kurohashi
Graduate School of Informatics, Kyoto University

Language Knowledge Engineering Lab.

Kyoto University

nakazawa@pa.jst.jp
john@nlp.ist.i.kyoto-u.ac.jp
fabien@pa.jst.jp
kuro@i.kyoto-u.ac.jp

Web Interface of Translation

Translation with Lattice Rules

Each path in this lattice corresponds to different choices of
insertion position for X2, morphological forms of “be”, and the
optional insertion of “at”.

• designed to handle an arbitrary number of non-terminals
• able to handle ambiguities of translation hypotheses
  ➢ which target word is going to be used
  ➢ which will be the final position of each non-terminal

Conclusion and Future Perspective

KyotoEBMT system
- source code available under a GPL license at
http://nlp.ist.i.kyoto-u.ac.jp/kyotoebmt/
- uses both source and target dependency analysis
- online example retrieving
- availability of full translation examples at run time

Future perspectives
- use of the input parse forest instead of k-best parses
- parsing the example parallel corpus as forests
- use a target-side tree language model as a decoding/reranking feature
- online tuning of weights
- target-side structural features

WAT2014 Official Results

Dependency Parsers
Ja: KNP [Kawahara and Kurohashi, 2006]
En: NLParser [Charniak and Johnson, 2005] with rules
Zh: SIP [Shen et al., 2012]

Reranking Features
20-best parses with its parsing scores
7-gram language model with Modified Kneser-Ney smoothing
Recurrent Neural Network Language Model with hidden layer size 200

Findings
Team-based rankings of HUMAN score were 2nd, 3rd, 5th and 4th
Reranking works well for all the directions other than J→C
need more investigations

WAT2014, Tokyo Japan, 4th October 2014