METEOR-WSD
Improved Sense Matching in MT Evaluation

Marianna Apidianaki[1] Benjamin Marie [1,2]

[1] LIMSI-CNRS, Orsay, France

[2] Lingua et Machina, Le Chesnay, France

SSST'9 Workshop
June 4, 2015

Lexical variation in MT

- METEOR (Banerjee and Lavie, 2005)
  - mapping of words with the same stem or belonging to the same WordNet synset
- METEOR-NEXT (Denkowski and Lavie, 2010)
  - semantic mapping extended to languages other than English and to longer text segments using pivot paraphrases (Bannard and Callison-Burch, 2005)

Lexical variation in MT

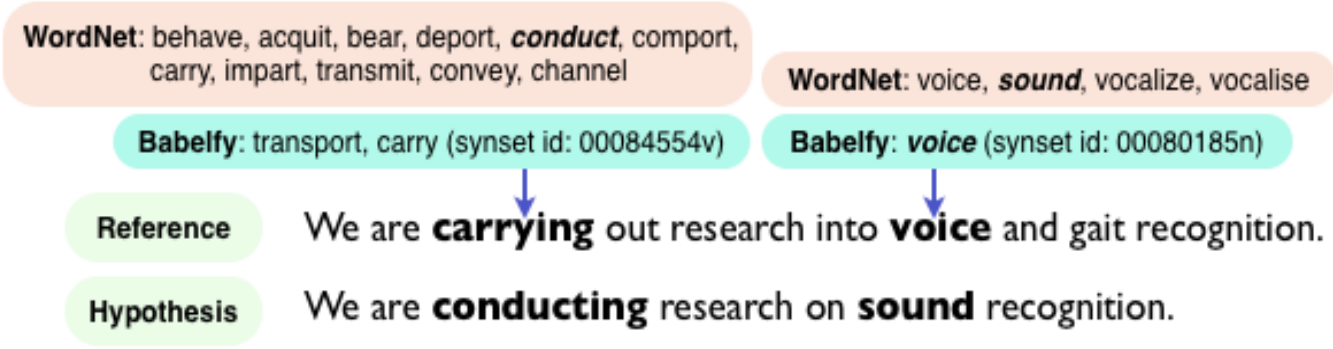
- METEOR (Banerjee and Lavie, 2005)
  - mapping of words with the same stem or belonging to the same WordNet synset
- METEOR-NEXT (Denkowski and Lavie, 2010)
  - semantic mapping extended to languages other than English and to longer text segments using pivot paraphrases (Bannard and Callison-Burch, 2005)

Increased correlation with human judgments of translation quality!

Sense matching without WSD: synonyms/paraphrases corresponding to different senses are treated as semantically equivalent

What WSD can do?

- identify the correct synset (subset of paraphrases) for a word (phrase) in context
- avoid erroenous matchings between text segments carrying different senses
- help METEOR establish better sense correspondences



Data sets and tools

- English translations of news texts from the languages of the WMT'14 Metrics Shared Task: French, Hindi, German, Czech, Russian

- WSD tool: in this first experiment we use Babelfy (Moro et al., 2014)
  - graph-based WSD by exploiting the structure of the multilingual network BabelNet (Navigli and Ponzetto, 2012)

Results

Segment-level Kendall's tau correlations between METEOR and human judgments

METEOR configuration	fr-en	de-en	hi-en	cs-en	ru-en
METEOR	.406	.334	.420	**.282**	.329
METEOR-WSD	**.410**	**.335**	**.422**	.278	**.331**

For more details and questions,
see you at the poster session!