METER-WSD

**Why WSD?**
- Meteor and Meteor-NEXT map words with identical surface forms or having the same stem, WordNet synonyms and paraphrases but all variants are treated as semantically equivalent ⇒ synonyms and paraphrases of different senses mapped during evaluation
- **METER-WSD:** identify the correct set of synonyms/paraphrases for a word/phrase in context

Alignment-based MultiWord Expression (MWE) identification
- WMT’14 data lemmatised, PoS tagged and word aligned
- candidate MWE: a sequence of words in the reference that is aligned
- Meteor-WSD: identify the correct set of synonyms/paraphrases for a word/phrase in context
- **M** • Meteor-WSD: identify the correct set of synonyms/paraphrases for a word/phrase in context
- **A** • Alignment-based sense selection in M • **E** • METER-WSD

Alignment-based Disambiguation
- meteor and meteor-next map words with identical surface forms or having the same stem, wordnet synonyms and paraphrases but all variants are treated as semantically equivalent ⇒ synonyms and paraphrases of different senses mapped during evaluation
- **METER-WSD:** identify the correct set of synonyms/paraphrases for a word/phrase in context

### Results
- at segment-level (metric: Kendall’s τ, data: WMT’14)
- at system-level (metric: Pearson’s coefficient, data: WMT’14)

### RATATOUILLE

**A metric combination**
- combine Meteor-WSD and nine other metrics: PER, WER, CDER, TER, GTM, BLEU, Meteor 1.5, RIBES 1.03.1 and BEER 1.0
- each metric gives a score at segment-level
- RATATOUILLE is the result of the log-linear combination of each metric’s score

**Tuning**
- the weight for each metric score is tuned using a similar approach to PRO (Hopkins and May, 2011; Guzmán et al., 2014)
- a pairwise approach: candidate translation pairs are classified into 2 categories (correctly or incorrectly ordered)
- training examples: all translation pairs or only translation pairs of significant different quality separated by at least 3 ranks in the human judgments
- at segment-level, experiments show a slightly better correlation when using only translation pairs of significant different quality

### Results and importance of Meteor-WSD
- at segment-level (metric: Kendall’s τ, data: WMT’14)
- at system-level, experiments show a slightly better correlation when using only translation pairs of significant different quality

**CONCLUSION AND FUTURE WORK**
- context-dependent sense selection helps Meteor establish better correspondences between hypotheses and references and improves the performance of the RATATOUILLE metric in almost all language pairs
- in the future, we intend to perform context-based filtering of pivot paraphrases and replace Meteor by Meteor-WSD in BEER to improve its correlation with human judgments