We propose a method for extending our current fuzzy matching framework with semantic information and SMT pretranslation functionality:

- Use of fuzzy metrics based on lexical semantics and on semantic roles from PropBank/NomBank
- Integration of fuzzy matches with SMT by pretranslating matching parts using word alignment or parse tree alignment
- Use of semantic roles during parse tree alignment

The method has been partially implemented and tested, for English-Dutch.

**Fuzzy matching framework**

**Origin:** general-purpose similarity metrics, metrics for MT evaluation, ...

**Type:** linguistically (un)aware metrics

**Combined metrics:** regression trees with match scores as features

**Evaluation procedure:** predict usability of translation of match

**Integration of fuzzy matches with SMT**

We are testing two types of alignment for determining matching parts for pretranslation.

**Word alignment**

They stress the need to bring the countries of this region closer, politically, to the EU.

They stress the need to bring the countries of this region closer, politically, to the EU.

Similar to method of Koehn and Senellart (2010)

**Parse tree alignment**

Similar to method of Zhechev and van Genabith (2010)

**Semantics-based fuzzy matching**

We are testing fuzzy metrics which use two types of semantics. These metrics can also be applied for evaluating the translation of a fuzzy match.

**Lexical semantics**

METEOR → Synonyms from Dutch part of EuroWordNet → Paraphrases from English-Dutch phrase pairs: Parex (Denkowski and Lavie 2010; Banard and Callison-Burch 2005)

**Semantic roles**

MEANT (Lo and Wu 2011) SR metrics in Asiya toolkit (Giménez and Márquez 2010)

**Evaluation metric:** Levenshtein, final word share, Parer match, Percent match, TER

**Correlation of fuzzy metrics with TER (Vanallemeersch and Vandeghinste 2015):**

To do: manual evaluation by translators

**Semantic tree alignment**

We are testing two alignment strategies which make use of semantic roles in order to approach the problem of diverging syntactic structures.

**Role alignment**

Alignment of PropBank/NomBank roles:

- Semantic role labels
- Lexical translation probabilities

**Semantics features in aligner**

Role labels as features in discriminative tree aligner Lingua-Align (Tiedemann 2010)

**Sentence alignment**

They stress the need to bring the countries of this region closer, politically, to the EU.

They stress the need to bring the countries of this region closer, politically, to the EU.

Similar to method of Zhechev and van Genabith (2010)

**SR systems**

English: System of LTH (Johansson and Nuges 2008) for PropBank/NomBank

Swirl (Asiya toolkit)

Dutch: System trained on crosslingual projections from English to Dutch (Vanallemeech 2012)

**References**


Intelligence Research.


This research is funded by the Flemish government agency IWT (project 130401, SCATE). See http://www.ccl.kuleuven.be/scate.