Research interests
Last Updated vendredi, 01 septembre 2017

Some researches I am interesting in:

- Terminology extraction

- Term alignment from comparable corpora

- Multi-word expressions: multi-word terms, term variants, collocations

- Sentiment analysis: fine-grained sentiment analysis

Terminology extraction

Terminology extraction identifies and collects candidate terms that occur in a monolingual corpus of a specialised domain. Some questions raise in terminology extraction:

- the specialised corpus: the delimitation of the specialised domain, its size, a collection of native texts, text genres, communicative intentions in specialised domains;

- the candidate terms: single terms and multi-word terms;

- the method to identify the candidate terms in texts;

- the measures and methods to rank, to filter, to cluster the candidate terms;

- the definition of a unified framework to evaluate term extraction tools.

Two tools dedicated to terminology extraction in the public domain:

- TermSuite: a tool suite for multilingual terminology extraction from comparable corpora. This tool suite offers a user-friendly graphical interface for designing UIMA-based tool chains whose components:

- form a functional architecture,

- manage 7 languages of 5 different families: English, French, German, Spanish, Latvian, Chinese and Russian.

- support standardised file formats,

- extract single- and multi- word terms languages by languages

- and align them by pairs of languages.
TermSuite has been developed within TTC project funded from the European Community's Seventh Framework Programme.

- ACABIT is a terminology extraction that proposes as output a list of multi-word term (MWT) candidates. It is available for French and English. It has been adapted to Japanese, Malagasy, Spanish and Italian.

Term alignment from comparable corpora

Given two corpora of a specialised domain in two languages, term alignment is about finding for a term in a source language, its translations in a target languages. Some questions that are raised with bilingual terminology extraction from comparable corpora:

- to ensure a good comparable corpora

- the vector-based method to align terms from comparable corpora: adopting the vector model, how to obtain the best configuration of parameters to obtain the context the most representative of a meaning of a term, the best ranking of the candidate translations

- the adaptation of the vector model to compute the context of multi-word terms

- a unified framework for evaluation term alignment from comparable corpora

Multi-word expressions

- Multi-word terms

- Linguistic specification and semantics of multi-word terms

- Compound term segmentation

- Term variants

- Typology of the linguistic operations leading to a term variant

- Inference of semantic links between terms through term variation

- Diachronic variants

- Multilingual variants

- Collocations

Sentiment analysis
Fine-grained sentiment analysis extract subjectivity textual segments and label them with several features: axiological polarity, discursive role (judgement (moral/ethic), assessment (intellect/ pragmatic/esthetic/affect), agreement and disagreement), enunciative strategy, speaker engagement (does he assume his subjectivity or does he try to hide it?) according to linguistic theories on sentiment language. Fine-grained sentiment analysis require lexical and semantic resources and an sentiment grammar.
Some questions raise in sentiment mining:

- Automatic lexical enhancement of sentiments resources

- Sentiment topic detection

- multilingual and multimodal sentiment analysis

The tool Aposis performs fine-grained sentiment analysis in French . It has been developed within the national project Blogoscopie (2006-2008) funded by ANR.