

PARSEME Parsing and multi-word expressions IC1207

Start date: 08/03/2013

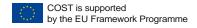
End date: 07/03/2017

Year: 1

Agata Savary

MC chair

Université François Rabelais Tours / France



Scientific context and objectives (1/2)

Background / Problem statement:

- Natural Language Processing (NLP): "understanding" and processing human texts by a computer (information extraction, machine translation, question answering, automatic text summarization, sentiment and opinion mining, human-machine dialogue, etc.)
- Multi-Word Expressions (MWEs): sequences of words with unpredicted properties (to count somebody in, to take a haircut, to kick the bucket)

Brief reminder of MoU objectives:

- To put **multilingualism** in focus of linguistic and technological studies.
- To establish a long-lasting collaboration of Natural Language Processing (NLP) experts within a cross-lingual, cross-theoretical and cross-methodological research network.
- To bridge the gap between **linguistic precision** and **computational efficiency** in NLP applications.



Scientific context and objectives (2/2)

Research directions:

- **Contrastive studies** of MWE properties and treatment in different languages and frameworks.
- Extending pre-existing language resources and tools (lexicons, grammars, treebanks, parsers) with MWEs.
- New formalisms and best practices for cost-saving lexicon, grammar and treebank production methodologies.
- Crossing borders between symbolic and probabilistic methods.

Innovation:

- First highly cross-lingual, cross-theoretical, and cross-methodological forum for MWEs in parsing
- MWE properties, criteria, linguistic tests and best practices are defined and studied for many languages at a time, as opposed to the English-dominated state-of-the-art NLP.



Working groups

- 1. Lexicon/Grammar Interface
- 2. Parsing Techniques for Multi-Word Expressions
- 3. Statistical, Hybrid and Multilingual Processing of MWEs
- 4. Annotating Multi-Word Expressions in Treebanks

Results vs. Objectives

- Extension to 30 countries (from 20)
- 2 general meetings:
 - 16-18 September 2013, Warsaw, Poland
 - 10-11 March 2014 Athens, Greece
- Co-organizing an established workshop at a top conference
- Website, 6 mailing lists, 2 Wiki spaces
- Dynamic list of members
- 4 national spin-off project proposals



First general meeting Warsaw, 16-18 September 2013

- 45 participants
- 15 presentations
- Discussions
- Defining the scientific program of the WGs

Impact:

- 4 national spin-off project proposals followed.



Second general meeting Athens, 10-11 March 2014



- 82 participants (twice as much as in Warsaw!)
- 44 posters
- Brainstorming sessions

Challenges

- Year 1 deviation: Extension of the WG3 scope to multilingual applications such as Machine Translation
- Year 2 challenges:
 - Defining common vocabularies, best practices and challenging cases
 - High-impact publications of state-of-the-art surveys
 - Two general meetings (Haifa, Malta)
 - Training School
 - More STSMs
 - Motivating MC members for WG participation
 - Action's logo



IC1207 – Action Data

Activity (No.)	Year 1	Year 2	Year 3	Year 4
MC/WG Meetings	2 / 2*4			
STSMs	8+1			
Training Schools	0			
Workshops or Conferences	1			
Joint Publications	26			
Participating COST Countries	30			
Participating non-COST-countries / institutions	2/2			
No. of individual participants (MC +WGs)	114			
Share ESRs (WGs)	44%			
Share female researchers (WGs)	49%			

