PARSME

PARSing and Multi-word Expressions
Towards linguistic precision and computational efficiency in NLP

http://www.cost.eu/COST_Actions/ict/IC1207
It’s time to...

let the cat
out of the bag!
• 8 March 2013 – 7 March 2017
• 30 countries
• 2 non-COST institutions (Stanford U., Federal U. of Rio Grande do Sul)
• 6 dialects from 10 language families
• *interdisciplinary* experts (linguists, computational linguists, computer scientists, psycholinguists, and industrials)
• different *parsing frameworks*: CCG, DG, GG, HPSG, LFG, TAG
• different *methodologies* (symbolic, probabilistic and hybrid parsing) and LT applications (MT, IR, ...)

[Map of Europe with green regions indicating participation in the COST project]
Working groups

• WG1: Lexicon/grammar interface
• WG2: Parsing techniques for MWEs
• WG3: Statistical, Hybrid and Multilingual Processing of MWEs
• WG4: Annotating MWEs in treebanks
WG1: Lexicon/grammar interface

• Better understanding of linguistic properties of MWEs, in particular at the lexical and syntactic level
• Enhancing the usability of MWE lexicons and valence dictionaries in parsing
• Paving the way towards interoperability of lexicons and the reduction of their production cost.
WG2: Parsing techniques for MWEs

• A better understanding of the potential of different linguistic frameworks with respect to parsing MWEs,
• Enhancing parsing efficiency,
• Reducing the cost of grammar production.
WG3: Statistical, Hybrid and Multilingual Processing of MWEs

• Elaborate ways of combining data-driven and knowledge-based methods to yield various hybrid processing models.

• Increase the efficiency and accuracy of hybrid parsing methods.

• Improve our understanding of how these may be applied to the processing of MWEs.

• Make better use of widely accessible unannotated data in order to improve grammars and models based on annotated data.

• Investigate relation between hybrid processing methods and multilingual applications.
WG4: Annotating MWEs in treebanks

• Enhanced MWE-aware methodologies of treebank construction.
• Optimal usability of MWEs in parsing.
What’s in the past

• 5 general meetings: (Warsaw, Athens, Frankfurt, Valletta, Iasi)
• 19 short-time missions (STSMs)
• 3 workshops (Gothenburg, Malaga, Iasi)
• 1 training school (Prague)
What’s in the future

• 3 general meetings
  • 7-8 April 2016, Struga, FYR Macedonia
  • 26-27 September 2016, Dubrovnik, Croatia
• short-time missions (STSMs)
• 1 workshop
  • PARSEME/ENeL workshop on MWE e-lexicons, s
    • 5-6 April 2016, University of Skopje
• 1 training school (La Rochelle, France, 27 June – 1 July)
What’s in the future: Shared TASK

- PARSEME shared task on automatic detection of verbal MWEs

Objectives:
- to cover many languages (18) of (4) different language families
- to boost the development of MWE processing tools (13)
- to make results of tools performing similar tasks comparable (1 workshop)
- to take discontinuous MWEs into account
- to bring MWE detection closer to parsing
PARSEME shared task

• annotating **training** and **evaluation** corpora of about 3,500-4,000 MWEs per language (by combining pre-existing data sets with newly annotated ones)

• annotation layers to decide (MWE layer, possibly also the PoS layer)

• common annotation guidelines (language specificities taken into account)

• selecting one or two text genres (e.g. newspaper texts)

• annotation experts will be mostly from PARSEME

• teams participating in the contest may also be from inside & outside the action
**PARSEME shared task**

- **Organizers:** Veronika Vincze, Antoine Doucet, Agata Savary

- **Calendar**
  - October-November 2015: feedback of the annotation guidelines
  - December 2015: annotation guidelines finalized; annotation tool set up
  - January-June 2016: annotation, development of evaluation measures and tools
  - September 2016: annotated datasets and evaluation tools finalized
  - **October 2016 - spring 2017:** system training, evaluation, paper submissions and reviewing, notifications
  - **2017:** final workshop (April, EACL, Valencia or CoNLL)
Thank you!

• Join us!
  http://typo.uni-konstanz.de/parseme/index.php/contact/how-to-join-us

• Join COST!
  http://www.cost.eu/COST_Actions/all_actions