

IC1207 COST Action PARSEME

PARSIng and Multi-word Expressions

Towards linguistic precision and computational efficiency
in natural language processing

7th General Meeting
26–27 September 2016
Dubrovnik, Croatia

This meeting

7th general meeting

Dubrovnik, Croatia

26–28 September 2016 (Monday-Wednesday)

Croatian Language Technologies Society

University of Zagreb

final meeting of the action



Feedback from **Dr. Ralph Stübner**, COST Scientific Officer for PARSEME

IC1207 COST Action **PARSEME** in September 2016



- scientific network, over 200 members
- 31 COST countries
- 2 non-COST institutions
- 6 general meetings:
(Warsaw, Athens, Frankfurt, Valletta, Iasi, Struga)
- 30 short-term missions
- 4 workshops (Gothenburg, Málaga, Iasi, Skopje)
- 2 training schools (Prague, La Rochelle)
- 5 national spin-off projects

Duration

4 years: 8 March 2013 – 7 March 2017

Year 4: past activities

2nd training school

La Rochelle, France

27 June – 1 July 2016

University of La Rochelle

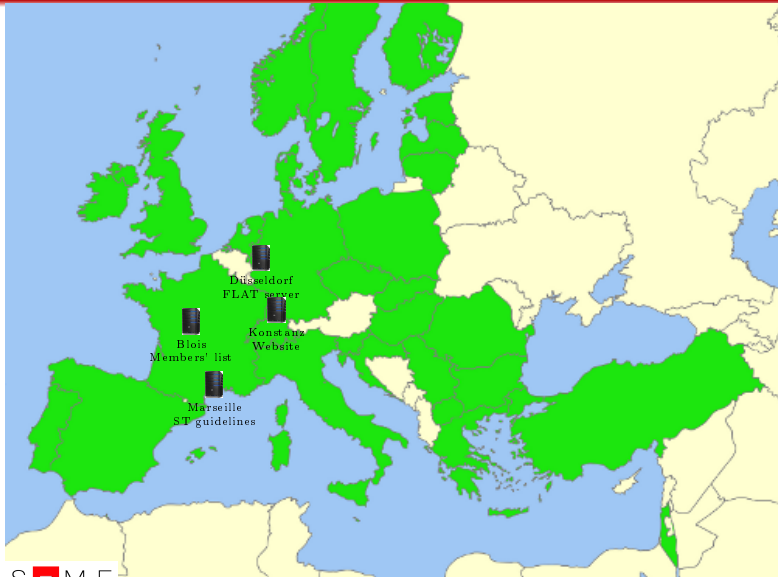
6 trainers, 40 trainees from 21 countries,
3 courses, posters, group projects



Year 4: current activities

- **WG4 survey** on MWE annotation in treebanks – ongoing (TLT 2015, LREC 2016)
- **WG3 survey** (*MWE discovery, parsing and translation*) - close to submission,
- **WG2 book** (*MWE representation and parsing*) - chapter proposals reviewed,
- **WG1 book** (*MWEs from a multilingual perspective*) - close to publication,
- **WG1 MWE template** – ongoing, publication planned,
- **WG1/WG3 survey** on multilingual MWE resources - call published,
- A MWE-dedicated **book series** accepted by LangSci (see below),
- **Shared task** on automatic detection of verbal MWEs (see below),
- call for STSMs: last deadline **early November**.

"PARSEME-SHARE" infrastructure



Action extension

Extension request

- Extension of the duration until **April 2017** and of the budget by **20,000 EUR** requested,
- Objective: covering the shared task final workshop,
- COST Head of the Science Operation and Head of Finance – in favor of the extension,
- COST Executive Board – decides end October.

Final workshop of the PARSEME shared task

Workshop proposal for **EACL, Valencia, Spain, 3–4 April 2017**,
Joint event: 13th MWE workshop & PARSEME ST workshop,
Proposal to submit by 30 September.

Follow-up initiatives

Going on with PARSEME annual meetings unfunded?

Follow-up projects

- Prague: *Translation, Parallel Data and MWEs* (poster A12),
- Athens: *SentOp: Expression of Sentiments and Opinions* (poster A13),
- Athens/Belgrade: *European culinary culture* (poster B13).

SIGLEX-MWE section

- Integrating PARSEME outcomes into a broader international community,
- [▶ MWE workshop](#) – annual meeting of the community,
- New initiatives:
 - shared tasks,
 - data repositories (contact with Mona Diab, USA – Arabic, English, Spanish, Farsi, and Chinese),
- Call for members - see [▶ SIGLEX](#) .

Book series on MWEs

- Editor: Language Science Press – open access,
- Title: *Phraseology & Multiword Expressions* (or vice versa),
- Scope: theoretical, computational, and empirical approaches to phraseology and MWEs,
- Editors: Manfred, Agata, Mike, Victoria, Yannick,
- Editorial board: PARSEME SC and more,
- Proposal **accepted** last week,
- First volumes:
 - WG1 & WG2 books,
 - MWE workshop selected papers.

Shared task on automatic detection of verbal MWEs

- Objectives:
 - achieve a universal VMWE typology and annotation guidelines,
 - boost the development of MWE-aware NLP tools.
- Challenge: highly **multilingual** participation
- **21 languages**, 4 language groups (Germanic, Romance, Slavic, others),
- Very active **core group** of 10 people,
- Community of about **40 annotators**,
- The methodology is **not** English-centered .

Shared task on automatic detection of verbal MWEs

- Progress so far:
 - two **pilot annotation** phases complete (200 sentences per language),
 - annotation **guidelines** (v6) ready,
 - annotation **tool** selected – **FLAT** – and server set up,
 - intensive support from FLAT's main developer, **Maarten van Gompel**, Radboud University, NL.
- Future work:
 - completing the **platinum standard** (a version agreed on by all annotators) compatible with guidelines v6 (within PARSEME),
 - main **annotation**, 3,500-4,000 MWEs per language – Sept–Dec 2016 (within PARSEME),
 - **evaluation** of systems – Jan-Feb 2017 (open worldwide),
 - final **workshop** planned for 3–4 April 2017 (EACL).

Shared task on automatic detection of verbal MWEs

- Expected outcomes:
 - **universal** annotation **guidelines** for 21 languages,
 - room for **language-specific** categories and properties,
 - **corpora** annotated uniformly (under an **open license**),
 - **tools** for VMWE identification,
 - shared task **methodology** for as many languages (with few systems per language).

Call for participants

- You may **join** your language annotators' team,
- You may present a **system** for the shared task.

Come today the **Annotathon** to learn more!

Organization matters

- sign the **attendance list** each day,
- be ready for a **group photo**,
- send poster files in **A4** format to `parseme@nlp.ipipan.waw.pl` by **next week**,
- send **publications** having directly benefited from PARSEME to `parseme@nlp.ipipan.waw.pl`,
- appointment for the **guided tour** and the **gala dinner**: today 17:15,
- appointment for the **tour on Wednesday**: ?.