

# IC1207 COST Action PARSEME

## PARSIng and Multi-word Expressions

Towards linguistic precision and computational efficiency  
in natural language processing

6th General Meeting  
7–8 April 2016  
Struga, FYR Macedonia

# IC1207 COST Action PARSEME in April 2016



scientific network

31 COST countries

2 non-COST institutions

6 general meetings:

(Warsaw, Athens, Frankfurt,  
Valletta, Iasi, Struga)

26 short-time missions

4 workshops (Gothenburg, Málaga,  
Iasi, Skopje)

1 training school (Prague)

## Duration

4 years: 8 March 2013 – 7 March 2017

## Preceding events

### Month 30 Progress Report

- submitted to COST late 2015,
- review: "overall excellent results".

### PARSEME/ENeL workshop on MWE e-lexicons

#### Skopje, FYR Macedonia

5-6 April 2016 (Tuesday-Wednesday)

Ss Cyril and Methodius University in Skopje

Co-organized with the IS1305 COST action **ENeL**

23 participants (10 from ENeL, 13 from PARSEME)



## Intensive activity between meetings

- **PARSEME dissemination paper**: LTC 2015,
- **2 WG4 surveys** (MWE annotation in treebanks) published: TLT 2015, LREC 2016
- **WG1 survey** (MWE lexicons): LREC 2016,
- **WG1 book** (MWEs from a multilingual perspective) - close to publication,
- **WG3 survey** (MWE discovery, parsing and translation) – close to submission,
- **WG2 book** (MWE representation and parsing) - call published (deadline in **May**)
- A MWE-dedicated **series** planned for the Language Science Press (application in 2016)
- **Shared task** on automatic detection of verbal MWEs

# Year 4

## Budget

- Year 4: 1 May 2016 – 7 March 2016,
- Budget: **31% decrease** compared to year 3,
- 8th general meeting (February 2017) **canceled**,
- September meeting in Dubrovnik is our **final meeting**.

## Action extension and follow-up

- Extension of the duration until **April 2017** and of the budget by **20,000 EUR** requested,
- Objective: covering the shared task final workshop,
- PARSEME **follow-up initiatives** will be promoted in poster selection and in STSM calls.

# Future events

## 2nd training school

**La Rochelle, France**

**27 June – 1 July 2016**

University of La Rochelle



## 7th general meeting

**Dubrovnik, Croatia**

**26–28 September 2016** (Monday-Wednesday)

Croatian Language Technologies Society

University of Zagreb

**final meeting** of the action



# 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
- **19 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

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

# Shared task on automatic detection of verbal MWEs

- Progress so far:
  - **teams** structured, roles defined
  - **two pilot annotation phases** complete (200 sentences per language)
  - first **inter-annotator agreement** results
  - annotation **guidelines** drafted and gradually enhanced
  - annotation **format** defined
  - annotation **tools** pre-selected (see hands-on session on Friday)
  - joint **paper** submitted
- Future work:
  - main **annotation** – Apr–Sept 2016 (within PARSEME)
  - system **evaluation** – winter 2016/2017 (open worldwide)
  - final **workshop** planned for April 2017 (EACL)



# Shared task on automatic detection of verbal MWEs

- Expected outcomes:
  - **unified** annotation **guidelines** for 19 languages
  - room for **language-specific** categories and properties
  - **corpora** annotated uniformly (under open license)
  - **tools** for VMWE identification
  - shared task **methodology** for as many languages (with few systems per language)
- Links with WGs:
  - WG1 – VMWE typology
  - WG4 – VMWE annotation guidelines
  - WG3 – VMWE identification methods and their evaluation
  - WG2 – VMWE identification with symbolic grammars

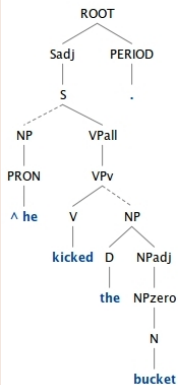
## Organization matters

- sign the **attendance list** each day (cf. **reimbursement**),
- **60 EUR / 3700 DIN** paid at registration for meals (2 lunches and a dinner) will be reimbursed,
- fill in the questionnaire on MWEs from Latin or Ancient Greek,
- be ready for a **group photo**,
- submit reimbursement claim very soon (end of budget year on 30 April, possible STSMs in April), but ...
- ... 2nd budget installment for year 3 not yet received; **reimbursement delays** expected,
- send poster files in **A4** format to `parseme@nlp.ipipan.waw.pl` by **15 April**,
- send **publications** having directly benefited from PARSEME to `parseme@nlp.ipipan.waw.pl`,

# Questions?

Thank you

## C-structure



## F-structure

