



# COST

**PARSEME**

**Parsing and Multi-Word Expressions**

**IC1207**

**Start date: 08/03/2013**

**End date: 07/03/2017**

**Year: 1**

**Agata Savary**

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 **knowledge-based** and **data-driven** methods.
- **Novelty:**
  - First highly cross-lingual, cross-theoretical, and cross-methodological forum for MWEs in parsing.
  - 23 languages, 9 language families, 7 dialects



# Working groups

- 1. Lexicon/Grammar Interface**
- 2. Parsing Techniques for Multi-Word Expressions**
- 3. Hybrid Parsing of Multi-Word Expressions**
- 4. Annotating Multi-Word Expressions in Treebanks**



# Future Plan and Challenges (1/2)

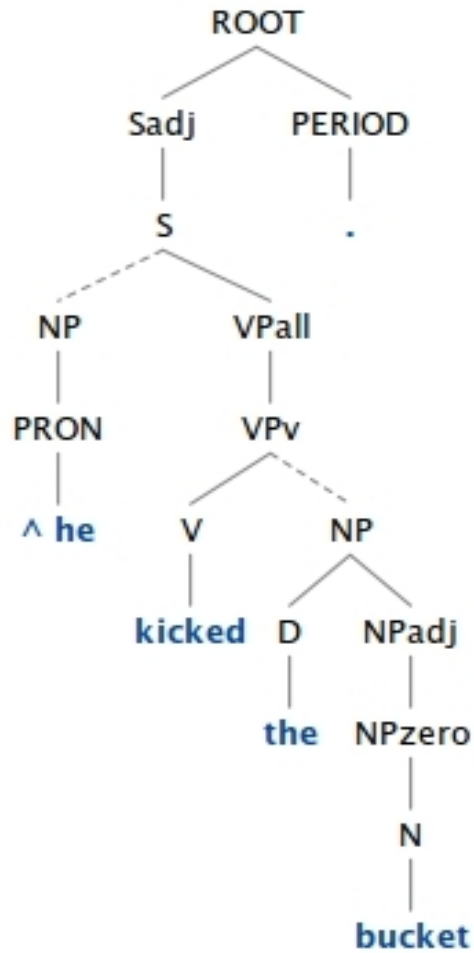
- **Outreach:**
  - **Extending** the number of partners, countries and languages
  - Public **website** addressed to **a diverse public**
  - Establishing a balanced collaboration with the MWE **community outside Europe** (*MWE Workshop*)
- **Networking:**
  - Efficient organization of the Working Groups
  - Creating internal communication and management tools
  - Organizing 2 plenary scientific sessions and 9 STSMs



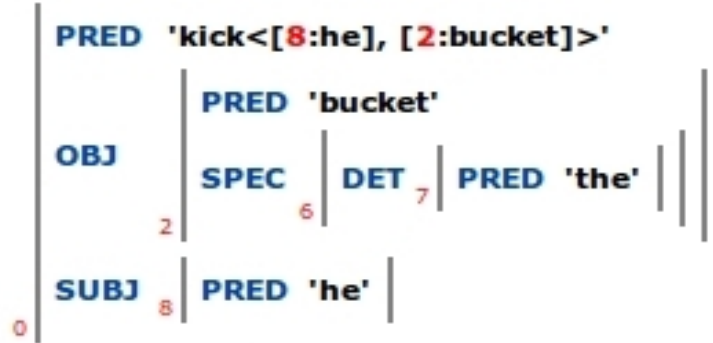
# Future Plan and Challenges (2/2)

- **Scientific program:**
  - Contrastive studies of the **linguistic properties of MWEs** in different languages.
  - Contrastive studies of the state-of-the-art **parsing frameworks** wrt. MWEs.
  - Contrastive studies of the state of the art in **annotating MWEs** in treebanks.
  - First steps towards **extending existing resources and tools** with MWEs.
  - Preparing **evaluation testbeds**.

## C-structure



## F-structure



Thank you