

Report on the activities of the SIGLEX-MWE Section in 2017

Agata Savary, Francis Bond, Styliani Markantonatou, Johanna Monti, Carlos Ramisch

1 February 2018

According to the constitution of the [SIGLEX-MWE section](#), the Standing Committee of the Section should report yearly on the Section's activities to its members. This document is the report of the Section's activities in 2017.

Reorganisation of the section

In 2017 the section gained new members, notably due to absorbing members of the IC1207 [PARSEME COST](#) action, whose funding came to an end in April 2017. In May 2017, the Section had over 200 registered members and steps were taken towards a more efficient management (which until then had remained in the hands of a single Section's representative at the SIGLEX board). In August 2017, a [Constitution](#) of the Section was drafted by the Section's members and validated by the [SIGLEX board](#). The constitution stipulates, inter alia, that the Section is now coordinated by a Standing Committee, consisting of one elected and four nominated officers. The **nominations** of the 4 officers and their validation by the SIGLEX board took place in August 2017. The new organisation is documented on the Section's [webpage](#).

MWE 2017 workshop

The 13th edition of the annual Workshop on Multiword Expressions ([MWE 2017](#)) was co-located with the EACL 2017 conference in Valencia, Spain on 4 April.

- [Participation](#): over 50 participants (MWE 2017 was one of the biggest, if not the biggest, satellite event of EACL 2017).
- [Invited talk](#): "Exploiting multilingual lexical resources to predict the compositionality of MWEs" by Paul Cook, University of New Brunswick, Canada.
- [Submissions](#): 34 papers (22 short and 12 long) were submitted to the main track; 7 papers (4 short and 3 long) were selected as oral presentations and 14 (11 short and 3 long) as posters; the overall selectivity rate was 62%. 6 system description papers were submitted to the shared task track.
- [Proceedings](#): published in the [ACL Anthology](#).
- [Shared task](#): the workshop was the culminating event of the PARSEME [shared task](#) on automatic identification of verbal MWEs (see below). The shared task organization and results were presented orally, and 6 out of the 7 participating systems were presented as posters.
- [Panel discussion](#) (see [slides](#)): it included MWE-related announcements, the

reorganization of the SIGLEX-MWE section (see above), shared task edition 1.1 (in 2018) and 2.0 (in 2019).

PMWE volume with selected papers from the MWE 2017 workshop

A new book series on "[Phraseology and Multiword Expressions](#)" was established in 2017 at [Language Science Press](#). A proposal of a volume containing extended papers from the MWE 2017 workshop was accepted by this series (see the [call](#)). Until late 2017, all chapters were drafted, reviewed, and revised. The final selection contains 13 chapters and should be published in spring 2018.

PARSEME shared task on automatic identification of verbal MWEs - edition 1.0

This initiative started in the framework of the IC1207 [PARSEME](#) COST action, and since the end of action (in April 2017) is overtaken by the SIGLEX-MWE section.

[Edition 1.0](#), thanks to a considerable effort of 18 language teams, resulted in:

- Universal [guidelines](#), with examples in many languages and room for language-specific specifications.
- The final [corpus](#) of **5.5 million tokens** and **60,000 VMWE annotations** in **18 languages**, distributed under different versions of the Creative Commons license.
- 7 systems for VMWE identification (5 of them were multilingual, all 18 languages were covered) - see the [results](#)
- Project management, with a structure based on [language groups](#) and roles (organizers, language group leaders, language leaders, annotators, etc.).
- Customizable annotation platform [FLAT](#)
- Dedicated tools to verify coherence and silence
- File formats (including [parsemetsv](#)), validators, converters and evaluation tools
- Communication tools: mailing lists, git issue tracker, websites
- Data repositories (gitlab)

PARSEME shared task - edition 1.1

Work on [edition 1.1](#) of the shared task started even before the MWE 2017 workshop in Valencia. Until the end of 2017:

- new language teams were recruited and trained,
- feedback was gathered from annotators concerning the annotations guidelines,
- discussion (via Gitlab issues) took place and about 70 issues were solved and implemented in [version 1.1 of the guidelines](#),

- annotation started for version 1.1 of the corpus.

LAW-MWE-CxG 2018 workshop proposal

In order to achieve better synergies with related initiatives, the Section made contact with the communities of the Construction Grammar and of the Linguistic Annotation Workshop. A joint workshop proposal, entitled "Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions ([LAW-MWE-CxG-2018](#))", was submitted to ACL/COLING/EMNLP/NAACL 2018 and selected to be colocated with [COLING 2018](#) in Santa Fé, USA in August 2018. The nomination of the workshop organisation and PC chairs was handled jointly by the SIGLEX-MWE Standing Committee and [SIGANN](#)'s leaders. The Program Committee gathers experts from the LAW, MWE and CxG communities.