

Report on the activities of the SIGLEX-MWE Section in 2020

*Carlos Ramisch, Agata Savary, Paul Cook, Jelena Mitrović, Petya Osenova,
Carla Parra Escartín, Shiva Taslimipoor, Ashwini Vaidya*

January 8, 2021

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 2020.

SIGLEX elections

In 2019 the [SIGLEX constitution](#) was amended. The SIGLEX board now has fewer members (still including the MWE Section representative) and their term is 2 instead of 3 years. A call for nominations for the new board was sent out in November 2019 and elections were organized in February-March 2020. The position of the MWE Section representative only received one nomination, so no election was run. Carlos Ramisch from Aix Marseille University in France thus became the elected representative of the Section for 2020-2022.

Standing committee

Since August 2018, the SIGLEX-MWE Section's [standing committee](#) (SC) consists of one elected officer (the MWE Section representative) and four nominated officers. The duration of the term of an SC nominated officer is 2 years, and term periods for different SC nominated officers should alternate, so as to ensure a transfer of competences within the SC.

On July 31, 2020, a call for new SC officers was announced, and some potential candidates were also approached personally by the Section's Representative. As a result, 4 expressions of interest were submitted via the online form by active members of the community who regularly publish research on multiword expressions. After discussion within the SC, a ranking was submitted to the SIGLEX board, who validated the nomination of Shiva Taslimipoor (University of Cambridge, UK) and Paul Cook (University of New Brunswick, Canada) on the basis of various balance factors including gender, scientific excellence, diversity and multilingualism. The term of the two newly nominated members is from September 2020 to August 2022.

A term overlap strategy was adopted this year to cope with the postponing of MWE-LEX 2020 due to the global covid-19 pandemic. From September to December, the SC exceptionally had 6 nominated officers, so that the two outgoing officers were still involved in the Section's decisions until the 2020 workshop took place in December. We thank the outgoing officers Jelena Mitrović (Passau University, Germany) and Carla Parra Escartín (Iconic Translation, Ireland) for their valuable contribution.

Membership

According to the Section's constitution, SIGLEX-MWE members are those who have registered to SIGLEX and indicated their membership to the MWE Section in the registration form. In 2020, SIGLEX renewed its website including the [members directory](#), with direct impact on the Section's members list. SIGLEX members were asked to re-register using the new registration form. On December 29, 2020 the SIGLEX directory indicates that 227 members re-registered as MWE Section members, that is, 53 registrations less with respect to the "280 registered identities, some of which are obsolete or duplicated" reported in 2019. On the one hand, we may have lost some members who did not re-register, but on the other hand, the directory now contains a more up-to-date list of members, much easier to manage, deduplicate, contact, etc.

The Section also maintains a mailing list (multiword-expressions@lists.sourceforge.net) whose members' list does not fully overlap with the SIGLEX-MWE members. Unfortunately, since 2016 Sourceforge no longer allows the list administrators to view or change the names and addresses of their members. We are currently studying alternatives to migrate this mailing list to a more manageable platform and homogenise it with the SIGLEX members directory.

Communication infrastructure

In parallel with improvements in member management, we are working on new tools for communicating within the Section and with the general public. Given the 2-year term of SC officers, and to ensure continuity irrespective of the composition of the SC at any given time, we created an email siglex-mwe-board@googlegroups.com for the Section's SC. This single alias can be used both for internal communication among SC officers and for anyone who wishes to contact the Section's SC.

We have also introduced the Section's advisory board, composed of all former Section representatives and former SC nominated officers. The advisory committee can also be contacted using an email alias siglex-mwe-advisory@googlegroups.com.

In terms of external communication, we have created two Twitter accounts: [@siglex_mwe](#), for the Section; and [@mwe_workshop](#) for the workshop series. The creation of a new website and mailing list for the Section is work in progress and should be launched in 2021.

MWE-LEX 2020 workshop

The Section joined forces with the [ELEXIS](#) community to organise a joint event, the Joint Workshop on Multiword Expressions and Electronic Lexicons ([MWE-LEX 2020](#)). It was co-located with the COLING 2020 conference, initially planned for September 2020 in Barcelona, Spain, and postponed to December 2020 online due to the covid-19 pandemic. The workshop took place online on December 13, 2020 and was the 16th edition of the yearly Workshop on Multiword Expressions ([MWE](#)). The event also hosted edition 1.2 of the

PARSEME shared task (see below).

- Participation: the number of participants was not provided by the COLING 2020 organisers, since the conference registration fees and form comprised the whole event, including the workshops. We estimate that there were about 20-25 active participants during the sessions, plus a certain number of intermittent listeners.
- 1 invited talk: A close look at Generatory or: "How We Went beyond Sense Inventories and Learned to Gloss" by [Roberto Navigli](#) from the Sapienza University of Rome (Italy).
- Papers: The **research track** received 25 submissions (14 long and 11 short papers); 13 papers were accepted (7 long and 6 short); all of them were presented orally as videos on Underline, with live Q&A sessions during the event. The overall selectivity rate was 52%, comparable to that of MWE-WN 2019 (54%). The **shared task track** received 6 submissions of system description papers (out of 7 participating teams which submitted 9 system results); all 6 papers were accepted to be presented orally as videos as well. We have called for EMNLP Findings presentations, but received no proposal.
- Proceedings: published in the [ACL Anthology](#).
- Community discussion (see [slides](#)): it included feedback from participants, MWE-related announcements, plans for future shared tasks and MWE workshops. Here are some highlights from the discussion:
 - **MWE workshops**: the synergy with ELEXIS could be further explored: we do not know what would be useful for e-lexicon development, maybe we should focus on semantics and/or definition generation; it would be interesting to explore synergies with representation learning and distributional semantics (e.g. can contextual embeddings be used to discover new MWEs?); focus on resources, their development and use, and in particular lexicons should be encouraged.
 - **Shared task**: the focus on a specific phenomenon (e.g. unseen MWEs) is useful; a weighted Token-based metric could help making evaluation more language independent; the task attracted less participants this year but the problem is far from solved, it could benefit from some extra visibility (e.g. as part of SemEval or hosted on a platform like CodaLab); ideas for extension include machine translation, word sense disambiguation, definition generation.

PARSEME shared task on automatic identification of verbal MWEs - edition 1.2

[Edition 1.2](#) of this shared task, whose culmination took place at the MWE-LEX 2020 workshop, was a follow-up of editions 1.0 and 1.1 (MWE 2017 and LAW-MWE-CxG 2018 workshops, respectively). Edition 1.2 used a new version of the PARSEME multilingual corpus of verbal multiword expressions, with 14 languages (3 new ones), enhanced annotation guidelines, and newly annotated texts. This initiative started in the framework of the IC1207 [PARSEME](#) COST action, and since the end of action (in April 2017) was overtaken by the SIGLEX-MWE section.

Edition 1.2, thanks to a considerable effort of 14 language teams, resulted in:

- Universal [guidelines](#), enhanced with respect to editions 1.0 and 1.1, with examples in many languages and room for language-specific specifications.
- The final [corpus](#) of over **5.5 million tokens** and over **68,500 VMWE annotations** in **14 languages**, annotated according to [unified guidelines](#) and distributed under different versions of the Creative Commons license via the CLARIN/LINDAT infrastructure.
- 9 systems (from 7 teams) for VMWE identification (8 of them were multilingual, and 4 covered all 14 languages) - see the [results](#).
- Novel evaluation scenarios, focusing on VMWEs **unseen** in the training data.
- New development infrastructure for [PARSEME corpora](#).

MWE 2021 workshop

The next edition of the MWE workshop will take place on August 5 or 6, 2021, with a special focus on applications. The event will be co-located with ACL-IJCNLP 2021 planned to happen in Bangkok, Thailand or virtually, depending on the evolution of the covid-19 crisis. Although this edition will not be joint, as it was usually the case since 2017, we are planning a joint session with the [Workshop on Online Abuse and Harm](#), also co-located with ACL-IJCNLP 2021, to foster collaborations and maybe propose a joint event in the future.

Future synergies

The Section's SC contacted the following communities to propose the organisation of joint events in the coming years: Universal Dependencies ([UD](#)), similar languages, varieties and dialects ([VarDial](#)), figurative language ([FigLang](#)), machine translation ([WMT](#), [MUMTTT](#)), meaning representations ([DMR](#)), online abuse and harm ([WOAH](#)).