Report on the activities
of the SIGLEX-MWE Section in 2021

Carlos Ramisch, Archna Bhatia, Paul Cook, Marcos Garcia, Petya Osenova, Shiva Taslimipoor, Ashwini Vaidya

February 4, 2022

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 2021.

SIGLEX elections

In 2021 there were no elections in SIGLEX. Next elections are planned for the beginning of 2022, when a new Section representative will be elected along with other SIGLEX board members. Nominated members of the standing committee are not elected (see below).

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 August 11, 2021, a call for new SC officers was announced, and some potential candidates were also approached personally by the Section representative. As a result, 6 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 Archna Bhatia (Florida Institute for Human and Machine Cognition, USA) and Marcos Garcia (Universidade de Santiago de Compostela, Spain) on the basis of various balance factors including gender, scientific excellence, diversity, seniority, linguistics/computer background and multilingualism. The term of the two newly nominated members is from October 2021 to September 2023.

We thank the outgoing officers Ashwini Vaidya (Indian Institute of Technology, Delhi, India) and Petya Osenova (Sofia University and Bulgarian Academy of Sciences, Bulgaria) for their valuable contribution. Both have now joined the MWE Section's advisory committee.

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 2021,

some information was added to the SIGLEX website and registration form to indicate how the members' data is used and how to request its suppression, making it GDPR-compliant.

On December 31, 2021 the [SIGLEX directory](#) indicated that 290 members registered as MWE Section members (227 members on Dec 29, 2020). It should be noted that the number of members might appear slightly lower on the public SIGLEX directory, as some members opt out of displaying their identity on the web. These numbers indicate that the MWE Section keeps growing, exceeding the number of members reported before the migration from SourceForge.

Communication infrastructure

One of the most important changes in 2021 was the creation of a fresh new website and URL for the MWE Section. We thank Elias Ponvert for having donated the *multiword.org* domain to the Section, and the ACL for funding and managing the domain. The website's new content and layout was developed and populated by the standing committee members in January 2021. It uses more recent technology (github, jekyll, markdown pages) and is open-source. The new website was revised by the advisory board and announced on the mailing list on January 28, 2021. Any member of the MWE community can now make pull requests to edit its content on https://github.com/multiword/multiword.github.io.

In March 2021 and after several announcements, we shut down the historical MWE mailing list (multiword-expressions@lists.sourceforge.net) due to sourceforge's restrictions. A new mailing list was created using Google groups siglex-mwe-members@googlegroups.com. We manually added all members who (a) registered in the SIGLEX members directory, (b) indicated their affiliation to the MWE Section, and (c) did not opt out of the SIGLEX mailing lists. This list is manually updated by the Section representative as soon as there are updates or new members in the SIGLEX directory (usually takes 2-3 days). The list is moderated by the Section's standing committee, following the [ACL anti-harassment policy](#).

This mailing list completes our new email infrastructure, which also features two mailing lists created in 2020 for the Section's standing committee (siglex-mwe-board@googlegroups.com) and for the advisory board (siglex-mwe-advisory@googlegroups.com).

Also in March 2021, we unified our Twitter accounts under a unified web handler [@multiword](#). We thank Elias Ponvert for having transferred ownership of this Twitter account to the Section. Under this new handler, we post announcements and news for the MWE workshop series, shared tasks, Section news, etc. The previous handler [@mwe_workshop](#) was kept active for archival purposes but a message pointing to the new account was posted.

MWE 2021 workshop

The [MWE 2021 workshop](#) took place online on August 6, as part of [ACL-IJCNLP 2021](#). The workshop's [program](#) included:

- live presentations of 3 short papers, 4 long papers, and 1 ACL Findings long paper

- 1 [invited talk](#) by Vered Shwartz
- a joint discussion with the [workshop of online abuse and harms](#)
- a community discussion, open to all

The workshop had an attendance of 15-20 participants online simultaneously on Zoom. Slides are available on the website for all presentations and the [community discussion](#). Our invited speaker kindly shared a publicly available [recorded video](#) of her invited talk. Proceedings were published in the [ACL Anthology](#).

The joint session with the WOAH workshop was a more or less informal panel discussion and took place on Gather.town. It focused on the tasks and resources that would be necessary to integrate MWEs and automatic online hate processing.

The community discussion included a presentation of [SemEval 2022 task 2](#), focusing on MWEs, and various announcements (PMWE series, EACL 2021 games, twitter, new website, mailing lists...). Another topic of the community discussion was the low number of submissions in 2021: in spite of a large (and growing) community and program committee, we only received 19 submissions for MWE 2021. A discussion followed on possible reasons for this, including registration fees (no workshop-only fees, lack of visibility at submission time), lack of publicity, lack of prestige for workshops with respect to conferences, timeline (our previous workshop, MWE-LEX 2020, took place only 5 months before the 2021 deadline), initially planned location in Thailand, uncertainty about travel conditions, overall lack of motivation for attending online events, and MWE processing being perceived as a very difficult task in NLP. The participants also discussed the pros and cons of co-locating online workshops with *ACL events. A summary of the arguments was sent to the Section's members mailing list on August 13, 2021.

Post-workshop survey

After the workshop, the standing committee launched an online survey to help understand what is happening. We received 47 responses from August 20 to October 10, 2021. These responses were analysed by the standing committee members. Although it is difficult to summarise the diversity and richness of the feedback, some trends could be observed:

- 91.5% of respondents were aware of MWE 2021.
- 40.1% of respondents evoked a personal situation preventing them from submitting.
- Other reasons for not submitting mentioned by some participants included:
    - Travel uncertainty and registration and travel costs;
    - Lack of motivation for online events;
    - Lack of time to prepare their papers, or no current MWE research.
- 48.9% of respondents considered attending the workshop.
- 25.5% of respondents did not attend because they need to have a paper to be funded.
- Other reasons for not attending mentioned by some participants included:
    - Registration costs;
    - Lack of motivation for online events;

- - ○ Uninteresting program and timezone or date incompatibility.
  - When asked about the preferred format for MWE 2022, respondents replied as follows:
    - 55.3% Online event, independent of *ACL conference, free attendance;
    - 34% Hybrid event, co-located with *ACL conference, paid attendance;
    - 23.4% On-site event, co-located with *ACL conference, paid attendance;
    - 17% Online event, co-located with *ACL conference, paid attendance.
  - What respondents value the most about MWE workshops is networking.
    - Synergies, high-quality papers, new ideas, shared tasks were also mentioned.

These results were taken into account when choosing the venue and format of MWE 2022.

MWE-related initiatives

In 2021, the Section did not organise any shared task. The PARSEME network is currently working towards disconnecting the corpus development, maintenance and releases from PARSEME shared tasks. The first regular release of the corpus is planned for 2022.

The standing committee endorsed and gave feedback on the organisation of SemEval 2022 task 2, on Multilingual Idiomaticity Detection and Sentence Embedding.

Some members of the section participated at the EACL 2021 language diversity panel and games. The event included a game in which the meanings of word-by-word English translations of idiomatic MWEs in other languages had to be guessed by participants.

Some members of the section have also participated in the online UNLID Dagstuhl seminar, which joined experts in MWEs, universal dependencies and linguistic typology. This event continues the MWE community's efforts to make synergies with the UD and typology communities, especially in the development of annotated corpora.

MWE 2022 workshop

The next edition of the MWE workshop will take place on June 25, 2022, co-located with LREC 2022, possibly hybrid or on-site in Marseilles, France. This venue was chosen for favouring networking, facilitating logistics in case of online or hybrid modes, being geographically close to a large part of the community (Europe), and being a conference with focus on resources, traditionally attended by a large part of the linguistics community.

The call for papers mentions three main topics: *MWE processing in low-resource languages*, *MWE identification and interpretation in pre-trained language models*, and *MWE processing to enhance end-user applications*. The latter was also emphasised in 2021, being reconducted this year given its importance to connect the field to other NLP applications. The focus on low-resource languages will be materialised by a joint session with the 2022 workshop of the Special Interest Group in Under-resourced Languages (SIGUL). We expect that this session can create a common ground for possible synergies in the intersection between MWE research and inclusive multilingual language technology.