

Report on the activities of the SIGLEX-MWE Section in 2023

Voula Giouli, Archna Bhatia, Kilian Evang, Marcos Garcia, Lifeng Han

January 3, 2024

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 2023.

SIGLEX elections — Section representative

Following the [SIGLEX constitution](#), in 2023 there were no elections in SIGLEX. Next elections are planned for 2024, when a new Section representative will be elected along with other SIGLEX board members

The other members of the standing committee are not elected but nominated (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 July 28, 2023, a call for new standing committee officers was announced on the MWE members mailing list, and some potential candidates were also approached personally by the Section representative. 2 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 standing committee, both candidacies were submitted to the SIGLEX board, who validated the nomination of A. Seza Dođruöz (Ghent University, Belgium) and Alexandre Rademaker (IBM Research and FGV EMAP, Brazil) on the basis of various balance factors including scientific excellence, diversity, seniority, linguistics/computer background and multilingualism. The term of the two newly nominated members is from October 2023 to September 2025.

We thank the outgoing officers Archna Bhatia (Florida Institute for Human and Machine Cognition, USA) and Marcos Garcia (Universidade de Santiago de Compostela, Spain) for their valuable contribution. Both have 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. On December 31, 2024 the [SIGLEX directory](#) indicated that 366 members registered as MWE Section members (336 members on Dec 31, 2023). These

numbers indicate that the MWE Section keeps growing. 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.

MWE 2023 workshop

The [19th MWE 2023 workshop](#) took place as a hybrid one-day workshop at EACL 2023, in Dubrovnik, Croatia, on May 6. Since the Clinical NLP workshop was accepted at ACL 2023 (Toronto, Canada), we organised a joint track and invited a panel; we also invited two members of the Clinical NLP community to give keynote speech at the MWE 2023 workshop. In our proposal, we had opted for either EACL2023 (Dubrovnik, Croatia) or ACL (Toronto, Canada) – in this order – taking into account the results of a post-workshop survey sent after MWE 2021 (details in the [2021 yearly report](#)) and also considering that EMNLP would be the least attractive venue for the targeted audience of our workshop. The goal was to favour networking, facilitating logistics in case of online or hybrid modes, being geographically close to a large part of the community (Europe). Pursuing the MWE Section's tradition of synergies with other communities, we organised a special track on [MWEs in clinical NLP](#) jointly with the [Clinical NLP workshop](#) for shared papers/poster presentations.

The call for papers mentioned the following topics: *MWE processing and identification in specialised languages and domains*, *MWE processing to enhance end-user applications*, *MWE identification and interpretation in pre-trained language models*, and *MWE processing in low-resource languages*. The focus on MWE processing in specialised languages was materialised in a panel discussion jointly organised with the [Clinical NLP workshop](#) on *MWEs in Knowledge-intensive domains*.

For the first time, besides the official Start submission platform, papers with reviews from ACL Rolling Review (ARR) were also considered. In total, 21 submissions of original research papers (10 long and 11 short) were received and 14 of them were selected for presentation, with an overall acceptance rate of 66.67% for the archival submissions.

The workshop's program included:

- Oral and poster presentations of 7 long papers, 7 short papers, plus 2 presentations (1 oral and 1 poster) of non-archival papers (published elsewhere);
- Special Track on MWEs in Clinical NLP in which 2 of the accepted papers were presented;
- 2 invited talks, one by [Leo Wanner](#) (ICREA and Universitat Pompeu Fabra, and another one jointly by [Asma Ben Abacha](#) (Microsoft) and [Goran Nenadic](#) (University of Manchester), featuring thus both academia and the industry;
- a discussion panel jointly organised with the [Clinical NLP Workshop](#) (co-located with ACL 2023) featuring [Asma Ben Abacha](#) (Microsoft), [Goran Nenadic](#) (University of Manchester), [Stefan Schulz](#) (Medical University of Graz, Austria and Averbis GmbH, Germany), and [Kirk Roberts](#) (UTHealth Houston) to discuss *Multiword Expressions in Knowledge-intensive Domains: Clinical Text as a Case Study*;
- a community discussion, open to all.

As of 2023, the MWE workshop organising committee decided to confer a best paper award based on the reviews of accepted papers. The criteria for selecting the best paper are scientific significance, originality, technical quality and clarity of presentation. This year, two papers were selected - presented under the main workshop program and the Special Track.

Participation in the workshop can only be roughly estimated since registration to EACL 2023 was not specific to a workshop. The workshop had an attendance of 20-40 participants onsite and around 16-24 participants online, with the upper range figures occurring during keynote (Leo Wanner) and special track keynote (Goran and Asma) speeches and during the panel session. Slides and posters are available on the website for most presentations, and we also share the community discussion [slides](#).

Proceedings were published in the [ACL Anthology](#).

MWE-related initiatives

In the past, members of the Section were involved in the organisation of the [PARSEME shared tasks](#), co-located with the annual MWE workshop. The shared tasks were synchronised with the release of (training and evaluation) corpora annotated for verbal MWEs. In 2023, the Section did not organise any shared task, since the [PARSEME network](#) has worked towards disconnecting the corpus development, maintenance and releases from the organisation of shared tasks. [PARSEME corpus release 1.3](#) was presented as a paper in the workshop.

After its kick-off meeting in Brussels in September 2022, [UniDive](#) Cost Action CA21167, an interdisciplinary scientific network devoted to universality, diversity and idiosyncrasy in language technology was started. The main objective of the Action is to reconcile language diversity with rapid progress in language technology by embracing both inter- and intra-language diversity, i.e. a diversity understood both in terms of the differences among the existing languages and of the variety of linguistic phenomena exhibited within a language. UniDive gathers more than 250 interdisciplinary experts (linguists, computational linguists, computer scientists, psycholinguists, and industrials) from almost 40 COST countries (36 of which are COST countries) representing dozens of languages from many different language genera. MWEs lie at the heart of UniDive, along the axes of corpus annotation, representation in lexica and their efficient processing. The goal is to harmonise annotation practice between what has been achieved in PARSEME and Universal Dependencies and to expand PARSEME MWE annotation to non-verbal ones.

MWE 2024 workshop

The Organizers of LREC-COLING 2024 invited proposals for workshops to be held in conjunction with the joint conference in Lingotto Conference Centre – Turin (Italy). Being a conference with focus on computational linguistics and resources, traditionally attended by a large part of the linguistics and NLP community alike, LREC-COLING was the best venue for the 20th edition of the MWE workshop. This year, we seek synergies with the Universal Dependencies (UD) community. This is a timely collaboration because the two communities clearly have shared interests, i.e., in developing guidelines, data-sets, and tools that can be applied to a wide range of typologically diverse languages, raising

fundamental questions about tokenization, lemmatization, and morphological decomposition of tokens. These issues are also central to the recently started [UniDive COST action \(CA21167\)](#) who accepted to be co-organiser of the event.

The proposal for the [MWE-UD 2024](#) workshop was submitted by the organising committee members (Kilian Evang, Voula Giouli, Gosse Bouma, Archana Bhatia, Marcos Garcia, Lifeng Han, and Joakim Nivre) and was accepted as a hybrid one-day workshop at LREC-COLING 2024, in Turin, Italy (May 25, 2024). The workshop is organised and sponsored by the Special Interest Group on the Lexicon ([SIGLEX](#)) of the Association for Computational Linguistics ([ACL](#)), SIGLEX's Multiword Expressions Section ([SIGLEX-MWE](#)), Universal Dependencies ([UD](#)) and [UniDive](#) Cost Action CA21167. The 1st call for papers was released in December 2023.