Subject: First CFP: 18th Workshop on Multiword Expressions (MWE 2022)

[apologies for cross-postings]

****

First Call for Papers
18th Workshop on Multiword Expressions (MWE 2022)
Organized and sponsored by SIGLEX, the Special Interest Group on the Lexicon of the ACL

Full-day workshop colocated with LREC 2022 | Marseille, France | June 25, 2022

Submission deadline: **April 8, 2022**

MWE 2022 website: https://multiword.org/mwe2022/

****

Description:

Multiword expressions (MWEs) are word combinations which exhibit lexical, syntactic, semantic, pragmatic and/or statistical idiosyncrasies (Baldwin & Kim 2010), such as *by and large*, *hot dog*, *pay a visit* and *pull one's leg*. The notion encompasses closely related phenomena: idioms, compounds, light-verb constructions, phrasal verbs, rhetorical figures, collocations, institutionalized phrases, etc. Their behavior is often unpredictable; for example, their meaning often does not result from the direct combination of the meanings of their parts. Given their irregular nature, MWEs often pose complex problems in linguistic modeling (e.g. annotation), NLP tasks (e.g. parsing), and end-user applications (e.g. natural language understanding and MT), hence still representing an open issue for computational linguistics (Constant et al. 2017).

For almost two decades, modeling and processing MWEs for NLP has been the topic of the MWE workshop organized by the MWE section of SIGLEX in conjunction with major NLP conferences since 2003. Impressive progress has been made in the field, but our understanding of MWEs still requires much research considering its need and usefulness in NLP applications. For this 18th edition of the workshop, we identified three topics on which contributions are particularly encouraged:

- **MWE processing in low-resource languages:** The PARSEME shared tasks (Ramisch et al. 2020; 2018; Savary et al. 2017), among others, have fostered significant progress in MWE identification, providing datasets that include low-resource languages, evaluation measures and tools that now allow fully integrating MWE identification into end-user applications. A few efforts have recently explored methods for automatic interpretation of MWEs (Bhatia et al. 2018; 2017). Pursuing similar efforts on understanding MWEs in low-resource languages is beneficial. There are some recent efforts on processing of MWEs in low-resource languages (Liu & Wang 2020; Kumar et al. 2017; Wei et al. 2015). Resource creation and sharing should be pursued in parallel to the development of methods able to capitalize on small datasets.
- **MWE identification and interpretation in pre-trained language models:** Most current MWE processing is limited to their identification and detection using pre-trained language models (Taslimipoor et al. 2020), but we lack understanding about how MWEs are represented and dealt with therein (Nedumpozhimana & Kelleher 2021; Garcia et al.

2021, Fakharian & Cook 2021). Now that NLP has shifted towards end-to-end neural models like BERT, capable of solving complex end-user tasks with little or no intermediary linguistic symbols, questions arise about the extent to which MWEs should be implicitly or explicitly modeled in such models (Shwartz & Dagan 2019).

- **MWE processing to enhance end-user applications:** As underlined by the MWE 2021 call for papers, MWEs gained particular attention in end-user applications, including MT (Zaninello & Birch 2020), simplification (Kochmar et al. 2020), language learning and assessment (Paquot et al. 2019; Christiansen & Arnon 2017), social media mining (Maisto et al. 2017), and abusive language detection (Zampieri et al. 2020; Caselli et al. 2020). We believe that it is crucial to extend and deepen these first attempts to integrate and evaluate MWE technology in these and further end-user applications.

Through this workshop, we would like to bring together and encourage researchers in various NLP subfields to submit MWE-related research, so that approaches that deal with processing of MWEs including processing for low-resource languages and for various applications can benefit from each other. We also intend to consolidate the converging effects of previous joint workshops LAW-MWE-CxG 2018, MWE-WN 2019 and MWE-LEX 2020, and the joint MWE-WOAH panel in 2021, extending our scope to MWEs in e-lexicons and WordNets, MWE annotation, as well as grammatical constructions. Correspondingly, we call for papers on research related (but not limited) to MWEs and constructions in:

- Computationally-applicable theoretical work in psycholinguistics and corpus linguistics
- Annotation and representation in resources such as corpora, treebanks, e-lexicons, and WordNets
- Processing in syntactic and semantic frameworks (e.g. CCG, CxG, HPSG, LFG, TAG, UD, etc.)
- Discovery and identification methods
- Interpretation of MWEs and understanding of text containing them
- Language acquisition, language learning, and non-standard language (e.g. tweets, speech)
- Evaluation of annotation and processing techniques
- Retrospective comparative analyses from the PARSEME shared tasks
- Processing for end-user applications (e.g. MT, NLU, summarisation, language learning, etc.)
- Implicit and explicit representation in pre-trained language models and end-user applications
- Evaluation and probing of pre-trained language models and end-user applications
- Resources and tools (e.g. lexicons, identifiers) and their integration into end-user applications
- Theoretical and computational linguistic description and modeling in low-resource languages
- Annotation guidelines and methods in low-resource languages (expert, crowdsourcing, automatic)
- Adaptation and transfer of annotations and related resources to low-resource languages
- Processing in low-resource languages (supervised, semi-supervised, and unsupervised methods for identification, discovery, and interpretation)
- Evaluation of annotations and processing techniques for low-resource languages
- Processing for end-user applications in low-resource languages

Joint Session with SIGUL 2022 Workshop:

Pursuing the MWE Section's tradition of synergies with other communities, we will organize a joint session with the workshop of the [Special Interest Group on Under-resourced Languages](#) (SIGUL 2022). The goal is to foster future synergies that could address scientific challenges in the creation of resources, models and applications to deal with multiword expressions and related phenomena in low-resource scenarios, in accordance with one of our special topics in MWE 2022. The session format is currently under discussion. Submissions describing research on MWEs in under-resource languages, especially introducing new datasets or new tools and resources, are welcome.

Submission Details:

The workshop invites two types of submissions:

1. **Archival submissions** present substantially original research. Submissions will follow the [LREC stylesheet](#). They can be **long papers** (8 content pages + references) or **short papers** (4 content pages + references). The decisions as to **oral or poster presentations** will be taken by the PC chairs, with no distinction in the proceedings. Submission will be **double-blind**.
2. **Non-archival submissions** of abstracts will also be considered for presentation, but not included in the proceedings. Abstracts will go through a light reviewing process.

All papers should be submitted via the workshop's START submission page, available soon. Please choose the appropriate submission format (archival/non-archival).

Identify, Describe and Share your LRs:

- Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about "Sharing LRs" (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.
- As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2022 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

Important Dates:
All deadlines are at 23:59 UTC-12 (Anywhere on Earth).

Paper Submission Deadline: **April 8, 2022**
Notification of Acceptance: May   3, 2022
Camera-ready Papers Deadline: May 23, 2022
Workshop: June 25, 2022

Organizing Committee:
Program chairs: Archna Bhatia, Paul Cook and Shiva Taslimipoor
Publication chairs: Marcos Garcia
Communication chair: Carlos Ramisch

Contact:
For any inquiries regarding the workshop, please send an email to the Organizing Committee at
mweworkshop2022@gmail.com .