



Multilingual, Open-text Unified Syntax-independent SEmantics

Reporting

Project Information

MOUSSE

Grant agreement ID: 726487

[Project website](#) 

DOI

[10.3030/726487](https://doi.org/10.3030/726487) 

Project closed

EC signature date

27 February 2017

Start date

1 June 2017

End date

31 May 2023

Funded under

EXCELLENT SCIENCE - European Research Council (ERC)

Total cost

€ 1 497 250,00

EU contribution

€ 1 497 250,00

Coordinated by

UNIVERSITA DEGLI STUDI DI
ROMA LA SAPIENZA



Italy

This project is featured in...



**Artificial intelligence:
expanding scientific
boundaries and
enhancing innovation**

Periodic Reporting for period 4 - MOUSSE (Multilingual, Open-text Unified Syntax-independent SEmantics)

Reporting period: 2021-12-01 to 2023-05-31

Summary of the context and overall objectives of the project



The exponential growth of the Web is resulting in vast amounts of online content. However, the information expressed therein is not at easy reach: what we typically browse is only an infinitesimal part of the Web. And even if we had time to read all the Web we could not understand it, as most of it is written in languages we do not speak. Computers, instead, have the power to process the entire Web. But, in order to "read" it, that is perform machine reading, they still have to face the hard problem of Natural Language Understanding, i.e. automatically making sense of human language. To tackle this long-lasting challenge in Natural Language Processing (NLP), the task of semantic parsing has recently gained popularity. This aims at creating structured representations of meaning for an input text. However, current semantic parsers require supervision, binding them to the language of interest and hindering their extension to multiple languages. Here we propose a research program to investigate radically new directions for enabling multilingual semantic parsing, without the heavy requirement of annotating training data for each new language. The key intuitions of our proposal are treating multilinguality as a resource rather than an obstacle and integrating the knowledge-based paradigm which allows supervision in the machine learning sense to be accompanied with efficacious use of lexical knowledge resources. Key joint goals of the project were multilingual Word Sense Disambiguation, multi-inventory and multilingual Semantic Role Labeling and multilingual Semantic Parsing, thanks to the development and exploitation of a novel formalism, the BabelNet Meaning Representation (BMR), which enables language-independent, semantically-grounded representations of text. These tasks have also been shown to be mutually beneficial, progressively enriching less-resourced languages and contributing towards leveling the playing field for all languages. Enabling Natural Language Understanding across languages will have an impact on NLP and other areas of AI, plus a societal impact on language learners. An important benefit for the society will be the ability for automatic systems to "explain" their understanding of text in an intelligible way.

Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far ✓

The MOUSSE project aims at advancing the frontiers in the field of Natural Language Understanding (NLU). The team carried out innovative research work along four main research vectors, with the key aim of unifying the various tasks:

- Multi-inventory and Multilingual Word Sense Disambiguation (WSD)

Thanks to novel neural models, we put forward innovative techniques that scale robustly across languages and that can integrate neuro-symbolic information within the network, making them more interpretable, and within the loss functions. We also investigated generating definitions which explain the meaning of words in context, proposing a new task for multilingual/cross-lingual Word-in-Context disambiguation and casting WSD as a Question Answering task. We also pioneered silver-data creation in WSD by proposing new frameworks for acquiring large amounts of sense-tagged sentences across languages.

- Multi-inventory and Multilingual Semantic Role Labeling (SRL)

For the first time, we introduced a multilingual and multi-inventory approach to SRL, work that received an outstanding paper award at NAACL 2021. This changed the field's landscape by: 1) reducing the gap from high- to low-resource languages and 2) bringing together the different inventories of predicates and roles in the literature, also across languages. We also brought together two different task "styles", that is, dependency- and span-based. We presented a novel multilingual multi-inventory resource, UnitedD-SRL, and provided APIs and software, including InVeRo-SRL. Lately, we explored definition modeling to empower Semantic Role Labeling and analyzed the behavior of multilingual SRL via probing.

- Language-independent, semantically-grounded Semantic Parsing

We created meaning representations at the sentence level that are independent of the language (current representations are instead bound to English or few other languages). To achieve this goal, we proposed VerbAtlas, a novel unified resource which enables state-of-the-art multilingual semantic role labelling, SyntagNet, the first large-scale language-independent resource of semantic collocations, and multilingual sense embeddings. We enabled cross-lingual and multilingual Semantic Parsing with a novel model for cross-lingual Abstract Meaning Representation (AMR) parsing and graph generation in a seq2seq fashion. To move away from English AMR, we proposed SPRING, a new seq2seq model. After only 2 years it is the reference approach to the task. Last but not least, we proposed a novel formalism, BabelNet Meaning Representation (BMR), together with a truly semantic parsing aimed at overcoming the current issue of explicit sentence representations like AMR and enabling for the first time language-independent semantically-grounded representations.

- Related areas, including Machine Translation, LLMs, Multilingual Named Entity Recognition, Entity Disambiguation, Relation Extraction

We put forward multi-genre, fine-grained Named Entity Recognition. We proposed a novel approach to Entity Disambiguation based on extractive disambiguation, also empowered with textual definitions. We concluded the project by showing the relevance of lexical bias in Machine Translation with a novel benchmark for MT evaluation (work recipient of the best resource paper award at ACL 2022).

Close to the end of the project, we organized in Rome a Workshop on Ten Years of BabelNet and Multilingual Neuro-Symbolic Natural Language Understanding, with talks from the MOUSSE team and colleagues all around the world, was a big success, with guess from all over the world: Simon Krek, Anna Rogers, Bonnie Webber, Mark Steedman, Luke Zettlemoyer, Ed Hovy, Hinrich Schutze, Iryna Gurevich, Nathan Schneider, Ekaterina Shutova, Rico Sennrich, Alexander Koller, Daniel Hershcovich, Johan Bos, Steven Schockaert, Thierry Declerck, Jan Hajic, Carla Marello. As a result of one of the brainstorming session held during this workshop, a position paper on the hype on superhuman performance of LLMs in NLU was presented at ACL 2023 (recipient of an outstanding paper award).

Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the project so far)

- 1) It is now possible to perform state-of-the-art multilingual Word Sense Disambiguation with arbitrary inventories and without needing manually created training data, thanks to novel approaches to high-quality silver data creation.
- 2) Knowledge-based approaches have been shown to rival neural supervised approaches thanks to the integration of lexical-semantic syntagmatic information.
- 3) It is now possible to perform Semantic Role Labeling in arbitrary languages, thanks to the availability of VerbAtlas, a novel verb resource which overcomes the issues of PropBank and related resources (scalability, language specificity, human readability) in the literature and encodes the semantics of verb predicates and their arguments in a language-independent manner. It is also possible to perform Semantic Role Labeling with multiple inventories, an outcome enabled by the project.
- 4) Semantic parsing can now be carried out multilingually and in a truly semantic, language-independent fashion, thanks to the BabelNet Meaning Representation and neuro-symbolic semantic parsers.

