

AIM-WEST Presentation - LICIA Workshop



LIG (Grenoble), Lattice (Paris), LIF (Marseille)
UFRGS (Porto Alegre), UFSCar (São Carlos)

Funded by CNRS, FAPERGS and FAPESP

September 2016

AIM-WEST Keywords

- Natural Language Processing
- MultiWord Expressions (MWEs)
- Machine Translation (MT)
- Automatic Speech Recognition (ASR)
- Language Technology
- Non Compositionality
- Cognitive Lexical Models

Goal

This project aims to investigate techniques, resources and protocols for *evaluating* and *integrating* models of *MWE processing* into *MT and ASR technology*.

Partners

- LIG (Grenoble, France): Speech translation (machine translation and speech recognition), multilingual systems
- LIF (Marseille, France): Multiword expressions, speech recognition, natural language processing
- LATTICE (Paris, France): Multiword expressions, cognitive models of the lexicon
- UFRGS (Porto-Alegre, Brasil): Multiword expressions, cognitive models of the lexicon
- UFSCar (San Carlos, Brasil): Machine translation, automatic learning of computational-linguistic resources

AIM-WEST Funding

- From jan 2014 - dec 2016
- Budget managed by LIG (France), UFRGS and UFSCar (Brazil)
- 10k€ / year on French side - cannot be transferred to year N+1
- Missions in both directions + joint workshop
- Funded by CNRS, FAPERGS and FAPESP
- Website, mailing list
- Paper acknowledgements

Some results

- Annotated corpora and MWE resources for French and Portuguese
- Statistical machine translation systems developed for French-English, Portuguese-English and French-Portuguese language pairs
- Protocols and frameworks for MWE translation evaluation
- 27 publications (11 with Fr-Br authors)
- Tutorials of different topics: distributional semantics, automatic speech recognition, statistical machine translation
- 3 workshop (San Carlos, Porto-Alegre, Grenoble)
- Students Internships: Fr=>Br (2) - Br=>Fr (4)
- More on AIM-WEST web site ¹

¹<https://aim-west.imag.fr>

Sample publications

- Alexandre Rondon, Helena Caseli, Carlos Ramisch. **Never-Ending Multiword Expressions Learning**, Proceedings of the 11th Workshop on Multiword Expressions (MWE 2015), Denver, USA, p. 45-53, June, 2015.
- Zied elloumi, Olivier Kraif, Laurent Besacier. **Integrating Multi Word Expressions in Statistical Machine Translation**, Proceedings of Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT2015) – part of EUROPHRAS2015 – Malaga, Spain, 1-2 July 2015.
- Muntsa Padró, Marco Idiart, Aline Villavicencio, Carlos Ramisch, **Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri**, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014) – short papers, Doha, Qatar, October, 2014.