

Carlos Ramisch

February 15, 2016

LIF-TALEP

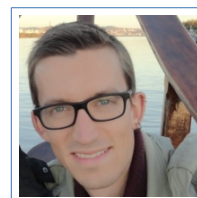
163, avenue de Luminy - Case 901

13288 MARSEILLE CEDEX 9

☎ (+33) 6 71 73 64 74

✉ Carlos.Ramisch [at] lif.univ-mrs.fr

📄 pageperso.lif.univ-mrs.fr/~carlos.ramisch



*Assistant professor (maître de conférences)
Computer science and computational linguistics*

Education

2009–2012 **PhD in computer science.**

University of Grenoble (France) & Federal University of Rio Grande do Sul (Brazil)

A generic and open framework for multiword expressions treatment: from acquisition to applications

Advisers: Aline Villavicencio & Christian Boitet

2007–2009 **Research master in informatics (M2R).**

ENSIMAG - Grenoble Institute of Technology (France)

Multi-word terminology extraction for domain-specific documents

Advisers: Christian Boitet & Aline Villavicencio

2004–2009 **Bachelor in computer science.**

Federal University of Rio Grande do Sul (Brazil)

Professional experiences

2013–now **Assistant professor (maître de conférences), LIF-TALEP & Aix-Marseille University.**

Research: Computational linguistics, multiword expressions, syntax and semantics

Teaching: algorithms in C, Linux (1st year); compilers, natural lang. processing (3rd year)

2012–2013 **Teaching and research assistant (ATER), LIG-GETALP & Joseph Fourier University.**

Research: Computational linguistics, machine translation, lexical resources

Teaching: natural lang. processing (4th year), human-computer interaction (3rd year)

2009–2012 **PhD researcher and lecturer, LIG-GETALP & University of Grenoble, Grenoble.**

Research: Computational linguistics, development/evaluation of framework for multiword expressions

Teaching: algorithms in C (1st year)

2008–2009 **Junior researcher, Xerox Research Centre Europe, CLT team, Grenoble.**

Improvements and optimizations for a dynamic translation memory

Advisers: Marc Dymetman & Lucia Specia

04–08/2008 **Junior researcher, LIG-GETALP, Grenoble.**

Development of a generic interactive multilingual access gateway

Advisers: Christian Boitet & Valérie Bellynck

2006–2007 **Junior researcher, Federal University of Rio Grande do Sul, NLP team, Porto Alegre.**

Autonomous language acquisition: multiword expressions and corpora

Advisers: Aline Villavicencio & Marco Idiart

2004–2006 **Junior researcher, Federal University of Rio Grande do Sul, BDI team, Porto Alegre.**

Artificial intelligence applied to the evaluation of petrol reservoirs

Adviser: Mara Abel

Keywords

Scientific Computational linguistics, artificial intelligence, multiword expressions, distributional semantics

Technical Python, C/C++, Java/J2EE, Linux, PHP, scripts for text processing (bash, awk, sed)

Leisure Traveling, cycling, running, working out, learning foreign languages

Projects

- 2016-2019 Local coordinator, PARSEME-FR (ANR) — <http://parsemefr.lif.univ-mrs.fr>
- 2013-2017 Participant, PARSEME (EU COST Action) — <http://typo.uni-konstanz.de/parseme/>
- 2014-2016 Local coordinator, AIM-WEST (CNRS-FAPERGS-FAPESP) — <http://aim-west.imag.fr/>
- 2011-2014 Co-leader, CAMELEON (CAPES-COFECUB 707-11) — <http://cameleon.imag.fr>

Events

- Tutorials on mwetoolkit at PARSEME 2015 General meetings
- Tutorial on MWE processing at EBRALC 2014 - <http://www.elc-ebralc-2014.com.br/>
- Member of local committee of TALN 2014 in Marseille — <http://www.taln2014.org/>
- Guest editor ACM TSLP special issue: MWEs — <http://multiword.sf.net/tslp2011si>
- Co-chair of MWE 2013, 2011 and 2010 — <http://multiword.sf.net/>
- Area chair for multiword expressions in *SEM 2012 — <http://ixa2.si.ehu.es/starsem>

Reviewing

- Journals Natural lang. engineering, Lang. resources & evaluation, Linguistic issues in lang. technology
- Conferences ACL, COLING, LREC, EMNLP, MWE workshop, PROPOR, *SEM, EAMT ...

Software and resources

- Main developer and maintainer of mwetoolkit — <http://mwetoolkit.sf.net>
- MWE-SIGLEX Section website — <http://multiword.sf.net>

Lectures and courses (computer science)

- | | | | |
|-------|-----------------------------------------------------------|---------|---------------------------------------------|
| 2014– | Natural lang. processing (3rd/4th year, AMU) | 2013 | Natural lang. processing (4th year, UJF) |
| 2014– | Compilers (3rd year, AMU) | 2013 | Interactive design in C (1st year, ENSIMAG) |
| 2013 | Algorithms in C (1st year, AMU) | 2012 | Human-computer interaction (3rd year, UJF) |
| 2013– | Linux and L ^A T _E X (1st year, AMU) | 2009–12 | Algorithms in C (1st year, UJF) |

Adviser and co-adviser

- | | | | |
|------------|----------------------------------|------------|---------------------------------|
| 2014–2017 | Silvio Cordeiro (PhD) | 02–06/2013 | Alexander Kobzar (master 2) |
| 04–08/2016 | Manon Scholivet (master 1) | 02–06/2013 | Samy Sassi (master 1) |
| 08–12/2014 | Alexandre Rondon (undergraduate) | 01–09/2012 | Sandra Castellanos (master 1) |
| 01–09/2014 | Sandra Castellanos (master 2) | 2011–2012 | Vitor de Araujo (undergraduate) |

Awards, fellowships and grants

- 2014 Merit Scholarship from FRQNT for scientific visit to UQAM (Canada)
- 2013 Qualification French CNU 27 (informatics) and 7 (linguistics)
- 2012 PhD double degree with honors from University of Grenoble & UFRGS
- 2009–2012 PhD fellowship of the French Ministry of Higher Education and Research (*allocation*)
- 2007–2008 CAPES Brafitec scholarship for exchange student
- 2007 Bachelor degree with honors from UFRGS (*l^áurea acad^êmica*)

Languages

Portuguese	mother tongue	German	fluent
English	fluent (<i>TOEFL 117/120</i>)	Spanish	intermediate
French	fluent (<i>DALF C1 98/100</i>)	Italian	beginner
		Arabic	beginner

Publications

2016

Silvio Cordeiro, Carlos **Ramisch**, and Aline Villavicencio. mwetoolkit+sem: Integrating word embeddings in the mwetoolkit for semantic MWE processing. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, May 2016 (to appear). European Language Resources Association.

Carlos **Ramisch**, Alexis Nasr, André Valli, and José Deulofeu. DeQue: A lexicon of complex prepositions and conjunctions in French. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, May 2016 (to appear). European Language Resources Association.

2015

Alexis Nasr, Carlos **Ramisch**, José Deulofeu, and André Valli. Joint dependency parsing and multiword expression tokenization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1116–1126, Beijing, China, July 2015. Association for Computational Linguistics.

Alexandre Rondon, Helena Caseli, and Carlos **Ramisch**. Never-ending multiword expressions learning. In *Proceedings of the 11th Workshop on Multiword Expressions*, pages 45–53, Denver, Colorado, June 2015. Association for Computational Linguistics.

Carlos **Ramisch**. *Multiword Expressions Acquisition: A Generic and Open Framework*, volume XIV of *Theory and Applications of Natural Language Processing*. Springer, 2015. <http://link.springer.com/book/10.1007%2F978-3-319-09207-2>.

2014

Bruno Laranjeira, Viviane Moreira, Aline Villavicencio, Carlos **Ramisch**, and Maria José Bocorny Finatto. Comparing the quality of focused crawlers and of the linguistic resources obtained from them. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, May 2014. European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2014/pdf/1095_Paper.pdf.

Muntsa Padró, Marco Idiart, Aline Villavicencio, and Carlos **Ramisch**. Comparing similarity measures for distributional thesauri. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, May 2014. European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2014/pdf/619_Paper.pdf.

Muntsa Padró, Marco Idiart, Aline Villavicencio, and Carlos **Ramisch**. Nothing like good old frequency: Studying context filters for distributional thesauri. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014) - short papers*, Doha, Qatar, October 2014. <http://www.aclweb.org/anthology/D14-1047>.

2013

Valia Kordoni, Carlos **Ramisch**, and Aline Villavicencio, editors. *Proceedings of the of the 9th Workshop on Multiword Expressions (MWE 2013)*. Association for Computational Linguistics, Atlanta, GA, USA, June 2013. <http://www.aclweb.org/anthology/W13-10>.

Magali Sanches Duran, Carolina Evaristo Scarton, Sandra Maria Aluísio, and Carlos **Ramisch**. Identifying pronominal verbs: Towards automatic disambiguation of the clitic 'se' in Portuguese. In *Proceedings of the 9th Workshop on Multiword Expressions*, pages 93–100, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics. <http://www.aclweb.org/anthology/W13-1014>.

Carlos **Ramisch**, Laurent Besacier, and Alexander Kobzar. How hard is it to automatically translate phrasal verbs from English to French? In *MT Summit 2013 Workshop on Multi-word Units in Machine Translation and Translation Technology*, Nice, France, September 2013.

Carlos **Ramisch**, Aline Villavicencio, and Valia Kordoni. Introduction to the special issue on multiword expressions: From theory to practice and use. *ACM Transactions on Speech and Language Processing*, 10(2):3:1–3:10, June 2013.

Leonardo Zilio, Carlos **Ramisch**, and Maria José Bocorny Finatto. Desenvolvimento de um recurso léxico com papéis semânticos para o português. *Linguamática*, 5(2):23–41, December 2013. <http://www.linguamatica.com/index.php/linguamatica/article/view/167>.

2012

Roger Granada, Lucelene Lopes, Carlos **Ramisch**, Cassia Trojahn Dos Santos, Renata Vieira, and Aline Villavicencio. A Comparable Corpus Based on Aligned Multilingual Ontologies. In *Proceedings of the ACL 2012 First Workshop on Multilingual Modeling (MM 2012)*, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. <http://aclweb.org/anthology-new/W/W12/W12-3904.pdf>.

Mathieu Mangeot and Carlos **Ramisch**. A Serious Lexical Game for Building a Portuguese Lexical-Semantic Network. In *Proceedings of the ACL 2012 3rd Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP*, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. <http://aclweb.org/anthology-new/W/W12/W12-4002.pdf>.

Carlos **Ramisch**. *A generic and open framework for multiword expressions treatment: from acquisition to applications*. PhD thesis, University of Grenoble (France) and Federal University of Rio Grande do Sul (Brazil), Grenoble, France, September 2012. http://www.inf.ufrgs.br/~ceramisch/download_files/thesis-getalp.pdf.

Carlos **Ramisch**. A Generic Framework for Multiword Expressions Treatment: from Acquisition to Applications. In *Proceedings of the ACL 2012 Student Research Workshop*, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. <http://aclweb.org/anthology-new/W/W12/W12-3311.pdf>.

Carlos **Ramisch**. Une plate-forme générique et ouverte pour le traitement des expressions polylexicales. In Jorge Mauricio Molina Mejia and Didier Schwab, editors, *Actes de 14e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2012)*, Grenoble, France, June 2012. <http://aclweb.org/anthology-new/F/F12/F12-3011.pdf>.

Carlos **Ramisch**, Vitor De Araujo, and Aline Villavicencio. A Broad Evaluation of Techniques for Automatic Acquisition of Multiword Expressions. In *Proceedings of the ACL 2012 Student*

Research Workshop, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. <http://aclweb.org/anthology-new/W/W12/W12-3301.pdf>.

Aline Villavicencio, Marco Idiart, Carlos **Ramisch**, Vitor Araujo, Beracah Yankama, and Robert Berwick. Get out but don't fall down: verb-particle constructions in child language. In *Proceedings of the Workshop on Computational Models of Language Acquisition and Loss*, pages 43–50, Avignon, France, April 2012. Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W12/W12-0911.pdf>.

2011

Vitor De Araujo, Carlos **Ramisch**, and Aline Villavicencio. Fast and Flexible MWE Candidate Generation with the mwetoolkit. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pages 134–136, Portland, OR, USA, June 2011. Association for Computational Linguistics. <http://www.aclweb.org/anthology/W11-0822.pdf>.

Magali Sanches Duran and Carlos **Ramisch**. How do you feel? Investigating lexical-syntactic patterns in sentiment expression. In *Proceedings of Corpus Linguistics 2011: Discourse and Corpus Linguistics Conference*, Birmingham, UK, July 2011. http://www.inf.ufrgs.br/~ceramisch/download_files/publications/2011/p07.pdf.

Magali Sanches Duran, Carlos **Ramisch**, Sandra Maria Aluísio, and Aline Villavicencio. Identifying and Analyzing Brazilian Portuguese Complex Predicates. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pages 74–82, Portland, OR, USA, June 2011. Association for Computational Linguistics. <http://www.aclweb.org/anthology/W11-0812.pdf>.

Valia Kordoni, Carlos **Ramisch**, and Aline Villavicencio, editors. *Proceedings of the ACL Workshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011)*. Association for Computational Linguistics, Portland, OR, USA, June 2011. <http://www.aclweb.org/anthology/W/W11/W11-08.pdf>.

2010

Helena de Medeiros Caseli, Carlos **Ramisch**, Maria das Graças Volpe Nunes, and Aline Villavicencio. Alignment-based extraction of multiword expressions. *Language Resources and Evaluation Special Issue on Multiword expression: hard going or plain sailing*, 44(1-2):59–77, April 2010. <http://www.springerlink.com/content/H7313427H78865MG>.

Éric Laporte, Preslav Nakov, Carlos **Ramisch**, and Aline Villavicencio, editors. *Proceedings of the COLING Workshop on Multiword Expressions: from Theory to Applications (MWE 2010)*. The Coling 2010 Organizing Committee, Beijing, China, August 2010. <http://aclweb.org/anthology-new/W/W10/W10-37.pdf>.

Evita Linardaki, Carlos **Ramisch**, Aline Villavicencio, and Aggeliki Fotopoulou. Towards the Construction of Language Resources for Greek Multiword Expressions: Extraction and Evaluation. In Stelios Piperidis, Milena Slavcheva, and Cristina Vertan, editors, *Proceedings of the LREC Workshop on Exploitation of multilingual resources and tools for Central and (South) Eastern European Languages*, pages 31–40, Valetta, Malta, 2010. may. <http://www.lrec-conf.org/proceedings/lrec2010/workshops/W22.pdf>.

Carlos **Ramisch**, Helena de Medeiros Caseli, Aline Villavicencio, André Machado, and Maria José Finatto. A Hybrid Approach for Multiword Expression Identification. In Thiago A. S. Pardo, António Branco, Aldebaro Klautau, Renata Vieira, and Vera Lúcia Strube de Lima, editors, *Proceedings of the 9th International Conference on Computational Processing of Portuguese Language (PROPOR 2010)*, volume 6001 of *Lecture Notes in Computer*

Science (Lecture Notes in Artificial Intelligence), pages 65–74, Porto Alegre, RS, Brazil, April 2010. Springer. <http://www.springerlink.com/content/978-3-642-12319-1>.

Carlos **Ramisch**, Aline Villavicencio, and Christian Boitet. Multiword Expressions in the wild? The mwetoolkit comes in handy. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010) — Demonstrations*, pages 57–60, Beijing, China, August 2010. The Coling 2010 Organizing Committee. <http://aclweb.org/anthology-new/C/C10/C10-3015.pdf>.

Carlos **Ramisch**, Aline Villavicencio, and Christian Boitet. mwetoolkit: a Framework for Multiword Expression Identification. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Valetta, Malta, May 2010. European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2010/pdf/803_Paper.pdf.

Carlos **Ramisch**, Aline Villavicencio, and Christian Boitet. Web-based and combined language models: a case study on noun compound identification. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 1041–1049, Beijing, China, August 2010. The Coling 2010 Organizing Committee. <http://aclweb.org/anthology-new/C/C10/C10-2120.pdf>.

Aline Villavicencio, Carlos **Ramisch**, André Machado, Helena de Medeiros Caseli, and Maria José Finatto. Identificação de Expressões Multipalavra em Domínios Específicos. *Linguamática*, 2(1):15–33, April 2010. <http://linguamatica.com/index.php/linguamatica/article/download/43/57>.

2009

Carlos **Ramisch**. Multi-word terminology extraction for domain-specific documents. Master's thesis, École Nationale Supérieure d'Informatique et de Mathématiques Appliquées, Grenoble, France, June 2009. http://www.inf.ufrgs.br/~ceramisch/download_files/publications/2009/p01.pdf.

2008

Carlos **Ramisch**, Paulo Schreiner, Marco Idiart, and Aline Villavicencio. An Evaluation of Methods for the Extraction of Multiword Expressions. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 50–53, Marrakech, Morocco, June 2008. http://www.inf.ufrgs.br/~ceramisch/download_files/publications/2008/p01.pdf.

Carlos **Ramisch**, Aline Villavicencio, Leonardo Moura, and Marco Idiart. Picking them up and Figuring them out: Verb-Particle Constructions, Noise and Idiomaticity. In Alex Clark and Kristina Toutanova, editors, *Proceedings of the Twelfth Conference on Natural Language Learning (CoNLL 2008)*, pages 49–56, Manchester, UK, August 2008. Association for Computational Linguistics. <http://aclweb.org/anthology-new/W/W08/W08-2107.pdf>.

2007

A. Villavicencio, V. Kordoni, Y. Zhang, M. Idiart, and C. **Ramisch**. Validation and Evaluation of Automatically Acquired Multiword Expressions for Grammar Engineering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1034–1043, Prague, Czech Republic, June 2007. Association for Computational Linguistics. <http://www.aclweb.org/anthology/D/D07/D07-1110.pdf>.