

## PERSONAL INFORMATION

## Manuela Sanguinetti

 manuela.sanguinetti@unica.it <https://msang.github.io/>

## WORK EXPERIENCE

01/06/2020–Present

## Post-Doc Research Fellow

University of Cagliari, Cagliari (Italy)

Title: *Simplified Algorithms to Access Open Data*

21/02/2020–31/05/2020

## Research Intern

University of Turin, Turin (Italy)

Title: *Natural Language Technologies for Intelligent Conversational Interfaces: Explanation and Emotions in Human-Chatbot Conversations*

01/08/2018–01/01/2020

## Post-Doc Research Fellow

University of Turin, Turin (Italy)

Title: *Data Annotation and Analysis for Developing Hate Speech Detection Systems and Methodologies*

\*28/11/2018-28/04/2019: maternity leave

01/07/2017–01/07/2018

## Post-Doc Research Fellow

University of Turin, Turin (Italy)

Title: *Developing Linguistic Resources for Sentiment Analysis Oriented to Hate Speech Detection*

01/08/2016–31/03/2017

## Research intern

University of Pisa, Pisa (Italy)

Title: *Development of a Parallel Multilingual Dependency Treebank in the Universal Dependency Format*

## EDUCATION AND TRAINING

01/01/2012–19/09/2016

## PhD in Computer Science (PhD program in Computational Linguistics)

University of Turin, Turin (Italy)

Thesis Title: *Combining Approaches for a Parallel Treebank Development and Alignment*

- **02/2014 - 05/2014:** Research stay at the Institute of Formal and Applied Linguistics (Charles University, Prague), funded by the **Erasmus Placement** program
- **08/2013:** European Summer School in Logic, Language and Information (ESSLLI), Dusseldorf, Germany
- **03/2012:** Bertinoro International Spring School (BISS), Bertinoro (FC), Italy

\*27/10/2015-27/03/2016: maternity leave

01/09/2008–14/07/2011 **Master's Degree in Translation**  
University of Turin, Turin (Italy)

01/09/2002–08/07/2008 **Bachelor's Degree in European Languages and Cultures**  
University of Cagliari, Cagliari (Italy)

## PERSONAL SKILLS

Mother tongue(s) Italian

### Foreign language(s)

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	C1	C1	C1	B2	C1
French	C1	C1	B2	B2	B1
Russian	B1	B1	A2	A2	A2

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user  
Common European Framework of Reference for Languages - Self-assessment grid

### Organisational / managerial skills

I participated in the organization of the following workshops and academic evaluation campaigns:

- 2nd Hate Speech Detection ([HaSpeeDe2](#)) task at EVALITA 2020
- Multilingual detection of hate speech against immigrants and women in Twitter ([HatEval](#)) at the International Workshop on Semantic Evaluation (SemEval) 2019
- Hate Speech Detection ([HaSpeeDe](#)) task at EVALITA 2018
- [Third International Workshop on Artificial Intelligence and Cognition](#) (AIC 2015), Turin
- [Doctoral Consortium on Artificial Intelligence](#), co-located with the XIII International Symposium on Advances in Artificial Intelligence (AI\*IA 2014), Pisa
- Shared task on [Dependency Parsing](#) at EVALITA 2014

### Job-related skills

#### **Research Activities:**

##### **Research Areas**

**Main areas:** Computational linguistics, Language resources, Dependency syntax and dependency-based annotation, with particular interest in Universal Dependencies, Hate speech in social media

**Secondary areas:** Irony detection and annotation, learner corpora, dialogue act tagging

#### **Supervision activities:**

I am currently co-advising a bachelors's student at the Department of Computer Science of Turin

#### **Projects:**

I participated in the following research projects:

- [IHatePrejudice. Immigrants, Hate and Prejudice in Social Media](#). Funded by Compagnia San Paolo Foundation, Turin
- [Hate Speech and Social Media. Analysis of Hate Speech Contents in Piedmont Region](#). Funded by CRT Foundation, Turin

#### **Publications:**

- Scopus h-index: 5

#### **Conference/Workshop papers:**

- M. Sanguinetti, C. Bosco, L. Cassidy, Ö. Çetinoğlu, A. T. Cignarella, T. Lynn, I. Rehbein, J. Ruppenhofer, D. Seddah, A. Zeldes. *Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies*. Accepted paper at the 12th Conference on Language Resources and Evaluation (LREC 2020), 2020.
- A. T. Cignarella, M. Sanguinetti, C. Bosco, P. Rosso. *Marking Irony Activators in a Universal Dependencies Treebank: The Case of an Italian Twitter Corpus*. Accepted paper at the 12th Conference on Language Resources and Evaluation (LREC 2020), 2020.
- A. T. E. Capozzi, M. Lai, V. Basile, C. Musto, M. Polignano, F. Poletto, M. Sanguinetti, C. Bosco, V. Patti, G. Ruffo, G. Semeraro, and M. Stranisci. *Computational linguistics against hate: Hate speech detection and visualization on social media in the "Contro l'Odio" project*. In Proceedings of the Sixth Italian Conference on Computational Linguistics (CliC-it 2019), 2019
- A. T. Cignarella, M. Sanguinetti, C. Bosco, and P. Rosso. *Is this an effective way to annotate irony activators?* In Proceedings of the Sixth Italian Conference on Computational Linguistics (CliC-it 2019), 2019
- C. Francesconi, C. Bosco, F. Poletto, and M. Sanguinetti. *Error analysis in a hate speech detection task: The case of HaSpeeDe-TW at EVALITA 2018*. In Proceedings of the Sixth Italian Conference on Computational Linguistics (CliC-it 2019), 2019
- E. Di Nuovo, C. Bosco, A. Mazzei, and M. Sanguinetti. *Towards an Italian learner treebank in Universal Dependencies*. In Proceedings of the Sixth Italian Conference on Computational Linguistics (CliC-it 2019), 2019
- V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, and M. Sanguinetti. *Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter*. In Proceedings of the 13th International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2019
- V. Basile, M. Lai, and M. Sanguinetti. *Long-term social media data collection at the University of Turin*. In Proceedings of the Fifth Italian Conference on Computational Linguistics (CliC-it 2018), 2018
- C. Bosco, F. Dell'Orletta, F. Poletto, M. Sanguinetti, and M. Tesconi. *Overview of the EVALITA 2018 Hate Speech Detection task*. In Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018), 2018
- M. Sanguinetti, C. Bosco, A. Lavelli, A. Mazzei, Oronzo Antonelli, and F. Tamburini. *PoSTWITA-UD: an Italian Twitter treebank in Universal Dependencies*. In Proceedings of LREC 2018, Miyazaki, Japan. European Language Resources Association (ELRA), 2018.
- M. Sanguinetti, F. Poletto, C. Bosco, V. Patti, and M. Stranisci. *An Italian Twitter corpus of hate speech against immigrants*. In Proceedings of LREC 2018, Miyazaki, Japan. European Language Resources Association (ELRA), 2018
- F. Poletto, M. Stranisci, M. Sanguinetti, V. Patti, and C. Bosco. *Hate speech annotation: Analysis of an Italian Twitter corpus*. In Proceedings of the Fourth Italian Conference on Computational Linguistics (CliC-it 2017), Rome, Italy, 2017
- Daniel Zeman, et al. *Conll 2017 shared task: Multilingual parsing from raw text to Universal Dependencies*. In Proceedings of the CoNLL 2017 Shared Task, Vancouver, Canada. Association for Computational Linguistics, 2017
- M. Sanguinetti, C. Bosco, A. Mazzei, A. Lavelli, and F. Tamburini. *Annotating Italian social media texts in Universal Dependencies*. In Proceedings of the 4th Conference on Dependency Linguistics (DepLing2017), Pisa, Italy. Linköping University Electronic Press, 2017
- E. Sulis, M. Lai, M. Vinai, and M. Sanguinetti. *Exploring sentiment in social media and official statistics: a general framework*. In Proceedings of the 2nd Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM 2015), Istanbul, Turkey, 2015
- M. Sanguinetti. *Experimenting the use of catenae in Phrase-Based SMT*. In Proceedings of the Second Italian Conference of Computational Linguistics (CliC-it 2015), Trento, Italy, 2015
- C. Bosco, L. Allisio, V. Mussa, V. Patti, G. Ruffo, M. Sanguinetti, and E. Sulis. *Detecting happiness in Italian tweets: Towards an evaluation dataset for sentiment analysis in Felicità*. In Proceedings of the International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data (ES3LOD 2014), Reykjavik, Iceland. European Language Resources Association (ELRA), 2014
- C. Bosco and M. Sanguinetti. *Converting the parallel treebank ParTUT in Universal Stanford Dependencies*. In Proceedings of the First Italian Conference of Computational Linguistics (CliC-it), Pisa, Italy, 2014
- C. Bosco and M. Sanguinetti. *Towards a Universal Stanford Dependencies parallel treebank*. In Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13),

Tubingen, Germany, 2014

- C. Bosco, F. Dell'Orletta, S. Montemagni, M. Sanguinetti, and M. Simi. *The Evalita 2014 dependency parsing task*. In Proceedings of Evalita 2014. Evaluation of Natural Language and Speech Tools for Italian. Pisa University Press, 2014
- M. Sanguinetti, E. Sulis, V. Patti, G. Ruffo, L. Allisio, V. Mussa, and C. Bosco. *Developing corpora and tools for sentiment analysis: the experience of the University of Turin group*. In Proceedings of the First Italian Conference of Computational Linguistics (CLiC-it), Pisa, Italy, 2014.
- M. Sanguinetti, C. Bosco, and L. Cupi. *Exploiting catenae in a parallel treebank alignment*. In Proceedings of the 9th Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland. European Language Resources Association (ELRA), 2014
- M. Sanguinetti, C. Bosco, and L. Lesmo. *Dependency and constituency in translation shift analysis*. In Proceedings of the 2nd Conference on Dependency Linguistics (DepLing2013), Prague, Czech Republic. MATFYZPRESS/ACL, 2013
- M. Sanguinetti and C. Bosco. *Translational divergences and their alignment in a parallel multilingual treebank*. In Proceedings of the 11th Workshop on Treebanks and Linguistic Theories (TLT11), Lisbon, Portugal, 2012
- C. Bosco, M. Sanguinetti, and L. Lesmo. *The parallel-TUT: a multilingual and multifformat treebank*. In Proceedings of the 8th Conference on Language Resources and Evaluation (LREC2012), Istanbul, Turkey. European Language Resources Association (ELRA), 2012
- M. Sanguinetti, C. Bosco, *Building the Multilingual TUT Parallel Treebank*. In Proceedings of the 2nd Workshop on Annotation and Exploitation of Parallel Corpora (AEPC2), Hissar, Bulgaria, 2011

#### **Parts in Books/Collections:**

- M. Sanguinetti and C. Bosco. *ParTUT: The Turin University Parallel Treebank*. In Italian Natural Language Processing within the PARLI Project, volume 589 of Studies in Computational Intelligence, pages 51–69. Springer, 2015

#### **Editorship:**

- A. Lieto, C. Battaglini, D. P. Radicioni, and M. Sanguinetti, editors. *Proceedings of the 3rd International Workshop on Artificial Intelligence and Cognition*, Turin, Italy, September 28-29, 2015, volume 1510 of CEUR Workshop Proceedings. CEUR-WS.org, 2015
- L. Di Caro, C. Dodaro, A. Loreggia, R. Navigli, A. Perotti, and M. Sanguinetti, editors. *Proceedings of the Second Doctoral Workshop in Artificial Intelligence (DWAi 2014)*. An official workshop of the 13th Symposium of the Italian Association for Artificial Intelligence (AI\*IA 2014), Pisa, Italy, December 11, 2014, volume 1334 of CEUR Workshop Proceedings. CEUR-WS.org, 2014

For an up-to-date list, see also my [Google Scholar](#) profile.

#### **Committees:**

I served as a member of the following scientific/program committees:

- Scientific committee of the 12th Conference on Language Resources and Evaluation (LREC 2020)
- Program committee of the Workshop on Resources and Techniques for User and Author Profiling in Abusive Language (ResT UP)
- Program committee for the *Language Resources* track of the Italian Conference on Computational Linguistics (CLiC-it, 2017, 2018 and 2019 edition)
- Program committee for the IEEE DSAA Special Session on Sentiment, Emotion, and Credibility of Information in Social Data (SeCredISData) @DSAA 2018: International Conference on Data Science and Advanced Analytics

#### **Invited Talks:**

"Introducing the ParTUT parallel treebank": invited talk at the [ACor4French Workshop at TALN 2017](#), Orleans, France

#### **Seminars:**

- 1 June 2020 - "Hate Speech Detection and Corpora: An Overview of the Italian Scenario": seminar held for the teaching course on *Artificial Intelligence and Natural Language Processing*, Department of Mathematics and Computer Science, University of Cagliari (Instructor: Prof.

Maurizio Atzori)

- 17-18 March 2015 - 2-day series of seminars on MT and alignment issues, with a practical session on manual word alignment, held for the teaching course on "*Informatica Applicata alla Comunicazione Multimediale*", Department of Foreign Languages, University of Torino (Instructor: Cristina Bosco, PhD)
- 18 March 2013 - "Building a parallel multilingual treebank: a case study" : seminar held for the teaching course on "*Informatica Applicata alla Comunicazione Multimediale*", Department of Foreign Languages, University of Torino (Instructor: Cristina Bosco, PhD)
- May 2012 - "Annotazione sintattica di corpora multilingue paralleli" : seminar held for the teaching course on "*Informatica Applicata alla Comunicazione Multimediale*", Department of Foreign Languages, University of Torino (Instructor: Cristina Bosco, PhD)

**Reviewing activities:**

Invited reviewer for IEEE Access and Future Generation and Computer Systems journals.