

Manuela Sanguinetti

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

● WORK EXPERIENCE

31/05/2020 – CURRENT

POST-DOC RESEARCH FELLOW – UNIVERSITY OF CAGLIARI

Title: *User-Friendly Access Algorithms to Open Data*

Duration: 2 years (ongoing)

Cagliari, Italy

10/03/2021 – 26/05/2021

LAB TUTOR – UNIVERSITY OF CAGLIARI

Teaching course: Programmazione 2

Corso di Laurea in Informatica, Facoltà di Scienze (L-31, SSD INF/01 SC 01/B1)

Teaching hours: 32

20/02/2020 – 30/05/2020

RESEARCH INTERN – UNIVERSITY OF TURIN

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

Duration: 3 months

Turin, Italy

31/07/2018 – 31/12/2019

POST-DOC RESEARCH FELLOW – UNIVERSITY OF TURIN

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

Duration: 1 year + 5-month maternity leave (28/11/2018-28/04/2019)

Turin, Italy

30/06/2017 – 30/06/2018

POST-DOC RESEARCH FELLOW – UNIVERSITY OF TURIN

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

Duration: 1 year

Turin, Italy

31/07/2016 – 30/03/2017

RESEARCH INTERN – UNIVERSITY OF PISA

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

Duration: 8 months

Pisa, Italy

● EDUCATION AND TRAINING

31/12/2011 – 18/09/2016 – Turin, Italy

PHD IN COMPUTER SCIENCE (PHD PROGRAM IN COMPUTATIONAL LINGUISTICS) – University of Turin

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

- **10/2015 - 04/2016:** Maternity leave
- **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

31/08/2008 – 13/07/2011 – Turin, Italy

MASTER'S DEGREE IN TRANSLATION – University of Turin

31/08/2002 – 07/07/2008 – Cagliari, Italy

BACHELOR'S DEGREE IN EUROPEAN LANGUAGES AND CULTURES – University of Cagliari

● LANGUAGE SKILLS

Mother tongue(s): **ITALIAN**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C1	B2	C1	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

● JOB-RELATED SKILLS

Publications

- Full list: [DBLP](#), [Google Scholar](#)

- Selected publications:

- 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 UD based overview of guidelines, corpora and unified recommendations. *Language Resources and Evaluation*, 2022
- M.Sanguinetti, G.Comandini, E. Di Nuovo, S.Frenda, C.Bosco, T.Caselli, V.Patti, I.Russo, M.Stranisci. HaSpeeDe 2@EVALITA2020: Overview of the EVALITA 2020 Hate Speech Detection Task. In *Proceedings of EVALITA 2020*
- F.Poletto, V.Basile, M.Sanguinetti, C.Bosco, V.Patti. *Resources and benchmark corpora for hate speech detection: a systematic review*. *Language Resources and Evaluation*, 2020

- M.Sanguinetti, A.Mazzei, V.Patti, M.Scalerandi, D.Mana, R.Simeoni. Annotating Errors and Emotions in Human-Chatbot Interactions in Italian. In Proceedings of LAW@COLING, 2020
- A.T.Cignarella, V.Basile, M.Sanguinetti, C.Bosco, P.Rosso, F.Benamara. Multilingual Irony Detection with Dependency Syntax and Neural Models. In Proceedings of COLING, 2020
- 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*. In Proceedings of LREC, 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*. In Proceedings of LREC, 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, M.Stranisci. *Computational linguistics against hate: Hate speech detection and visualization on social media in the "Contro l'Odio" project*. In Proceedings of CliC-it, 2019
- E.Di Nuovo, C.Bosco, A.Mazzei, M.Sanguinetti. *Towards an Italian learner treebank in Universal Dependencies*. In Proceedings of CliC-it, 2019
- V.Basile, C.Bosco, E.Fersini, D.Noza, V.Patti, F.M.Rangel Pardo, P.Rosso, M.Sanguinetti. *Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter*. In Proceedings of SemEval, 2019
- C.Bosco, F.Dell'Orletta, F.Poletto, M.Sanguinetti, M.Tesconi. *Overview of the EVALITA 2018 Hate Speech Detection task*. In Proceedings of EVALITA, 2018
- M.Sanguinetti, C.Bosco, A.Lavelli, A.Mazzei, O.Antonelli, F.Tamburini. *PostWITA-UD: an Italian Twitter treebank in Universal Dependencies*. In Proceedings of LREC, 2018.
- M.Sanguinetti, F.Poletto, C.Bosco, V.Patti, M.Stranisci. *An Italian Twitter corpus of hate speech against immigrants*. In Proceedings of LREC, 2018
- D.Zeman, et al. *CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies*. In Proceedings of the CoNLL 2017 Shared Task, 2017
- M.Sanguinetti, C.Bosco. ParTUT: The Turin University Parallel Treebank. Studies in Computational Intelligence, 2015

Projects

- [Universality, Diversity and Idiosyncrasy in Language Technology \(UniDive\)](#), COST Action funded by European Cooperation in Science and Technology. Duration: 2022-2026
- [HOPE \(High quality Open data Publishing and Enrichment\)](#), funded by the Ministry of Education, University and Research under the PRIN (Progetti di Ricerca di Interesse Nazionale) program. Duration: 2020-2023
- [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

Scientific Committees and Reviewing Activities

- Program committee of the Workshop on Resources and Techniques for User and Author Profiling in Abusive Language (ResT UP, ResT-UP2)
- Program committee of the Workshop on Perspectivist Approaches to Disagreement in NLP (NLPerspectives 2022)
- Track chair of the track on Machine Translation and Multilingualism at the 8th Italian Conference on Computational Linguistics (CliC-it 2021)
- Since 2020: Program committee of the Conference on Language Resources and Evaluation (LREC)
- Since 2017: Program committee for the Language Resources track of the Italian Conference on Computational Linguistics (CliC-it)
- 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 reviewer for international journals: Expert Systems with Applications, Language Resources and Evaluation, Quality and Quantity, IEEE Transactions on Affective Computing, Journal of Experimental & Theoretical Artificial Intelligence, PeerJ, IEEE Access, Future Generation and Computer Systems, Natural Language Engineering

Thesis Supervision

Thesis co-advisor of the following undergraduate students:

- Michela Lecca. Corso di Laurea Triennale in Informatica, Università degli Studi di Cagliari. Thesis title (to be discussed): "*Set Expansion and Labeling Using Masked Language Models*". Academic year 2021/22
- Matteo Gioele Collu and Federico Sanna (joint thesis). Corso di Laurea Triennale in Informatica, Università degli Studi di Cagliari. Thesis title: "*A Data-to-Text Conversational Agent for Dieting Support*". Academic year 2020/21
- Marco Scalerandi. Corso di Laurea Triennale in Informatica, Università degli Studi di Torino. Thesis title: "*Strumenti per l'analisi e il miglioramento di conversazioni utente-chatbot*". Academic year 2019/20

Seminars and Talks

- "Introducing the ParTUT parallel treebank": invited talk at the ACor4French Workshop at TALN 2017, Orleans, France
- 19-23 May 2022 - 2-day series of seminars on NLP benchmark datasets and Knowledge-based Question Answering, with hands-on sessions on text classification and Question Answering, 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)
- May 2021 - "Datasets for NLP": 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)
- 29 October 2021 - "NLP & Corpora": seminar held for the teaching course of Applied Linguistics, Dipartimento di Lettere, Lingue e Beni Culturali, University of Cagliari (Instructor: Prof. Nicoletta Puddu)
- 4 May 2021 - "Elaborazione Automatica del Linguaggio Naturale e Applicazioni nei Social Media": online talk at the UniCa OpenDay
- 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)

Organizing committees

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

07/2022 – CURRENT

Management committees

- Member of the [Management Committee](#) of the COST Action UniDive