

Alessandra Teresa Cignarella

E-mail: cigna@di.unito.it

28 December 2020

Education:

- **Languages & Literature Department, Università degli Studi di Torino** **Torino**
Bachelor Student *October 2011 - November 2014*
- **Linguistics Department, Università degli Studi di Torino** **Torino, Italy**
Master Student (LM-39) *October 2014 - April 2017*
- **Computerlinguistik, Universität des Saarlandes** **Saarbrücken, Germany**
Erasmus+ Student *March 2016 - September 2016*
- **PRHLT Research Center, Universitat Politècnica de València** **Valencia, Spain**
Erasmus+ Trainee *April 2017 - August 2017*
- **Computer Science Department, Università degli Studi di Torino** **Torino, Italy**
Co-tutelle Ph.D student *October 2017 - November 2020*
- **PRHLT Research Center, Universitat Politècnica de València** **Valencia, Spain**
Co-tutelle Ph.D student *October 2018 - today*

Technical and personal skills:

- **Computer skills:**
 - excellent command of programming with Python (2.7+)
 - good command of Python libraries and APIs for NLP (scikit-learn, NLTK, spaCy, gensim)
 - good command of machine learning platforms and APIs (TensorFlow, Keras)
 - good command of database management (MongoDB, WorkBench) and knowledge representation (OWL)
 - good knowledge of the Artificial Intelligence Markup Language (AIML)
 - excellent command of spreadsheets, LaTeX, e-mails and social media
 - +10 years of experience with both Linux and Windows operating systems
- **Softwares:**
 - Linguistic: Praat, ELAN
 - Ontology editor: Protégé
 - Audio/Video: Audacity, Kazam
 - Graphics: Photoshop, GIMP
- **Foreign languages:**
 - Italian (native speaker)
 - English (C1)
 - Spanish (C1)
 - German (B2)
 - French (A1)
 - Esperanto (basics)

- **Other**
 - Good presentations skills
 - Works well in a team
 - Can write well organized and structured reports
 - Creative and efficient in problem-solving
 - Adapts fast in new environments and situations

Complementary courses:

- **Coursera [September 2017 - October 2018]**
 - Programming for Everybody (Getting Started with Python)
 - Python Data Structures
 - Using Databases with Python
 - Using Python to Access Web Data
 - Capstone: Retrieving, Processing, and Visualizing Data with Python
 - Advanced Machine Learning Specialization
 - Applied Data Science with Python Specialization
- **Spring/summer schools and courses [March 2018 - May 2018]**
 - Bertinoro International Spring School 2018 (30 hours)
 - Lectures on Computational Linguistics 2018 (16 hours)
 - European Summer School in Logic, Language and Information 2018 (2 weeks)
- **Universitat Politècnica de València [January 2019 - March 2019]**
 - Scientific communication: High standards for scientific production and communication (20 hours)
 - Tools for research: Composition of high quality documents and presentations with \LaTeX (20 hours)
 - Rewriting science. How to avoid errors in written and spoken scientific English (20 hours)
- **Università degli Studi di Torino [May 2020]**
 - English for Scientific Academic Purposes (30 hours)
- **University of the Basque Country [July 2020]**
 - Deep Learning for Natural Language Processing (30 hours)

Research activities:

- **Tirocinio “Annoluce” markup corpus *Jus Jurium sentenze corte cassazione*.**
- **Holder of a short-term fellowship in Linguistics at the Università degli Studi di Torino:** within the research project *“Ricerca Locale 2015 quota A – Mutamento e variazione nell’italiano contemporaneo: fenomeni di contatto e spinte evolutive interne” (cod. ANDC_RIC_LOC_15_01)*.
I developed a template for transcription and annotation of data, based on the study of relevant bibliography (on speech transcription) and congruent with the specifications required by the ELAN software. I transcribed a corpus of videotaped data of classroom speech of about 15 hours of total duration. The data had to be converted to an appropriate audio-video standard, transcribed into orthographic form and aligned with the audio-video signal. The corpus thus constituted has been added to a *SayMore* archive. I also provided assistance to students who carried out part of the transcription work.
- **Co-organizer of a shared task: IronITA 2018: Irony Detection in Italian Tweets.**
The IronITA shared task was organized within EVALITA 2018, the 6th evaluation campaign of Natural

Language Processing and Speech tools for Italian, which was held in Turin, Italy, on December 12th and 13th, 2018. Evalita 2018 was co-located with CLiC-it 2018.

- **Co-organizer of a shared task:** *SardiStance 2020: Stance Detection in Italian Tweets*.
The SardiStance shared task has been organized within EVALITA 2020, the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian, which will be held online (due to the state of emergency regarding the covid-19 pandemic), on December 16th and 17th, 2020.
- **Contributor to the *Universal Dependencies* project:** *TWITTIRÒ-UD*.
TWITTIRÒ-UD is a collection of ironic Italian tweets annotated in *Universal Dependencies*. The treebank can be exploited for the training of NLP systems to enhance their performance on social media texts, and in particular, for irony detection purposes.
- **Co-creator of an inflected lexicon for Italian:** *LeFFI*.
LeFFI (Lessico delle Forme Flesse dell'Italiano) is an inflected lexicon of Italian listing all the available wordforms of 2,053 verbs. It can be used to perform an entropy-based analysis of the mutual predictability of wordforms within Italian verb paradigms.

Other research activities:

- **Reviewer for scientific journals**
 - *Information Fusion*
 - *Sādhanā*
 - *Information Processing and Management*
 - *Industrial Marketing Management*
 - *Language Resources and Evaluation*
 - *Procesamiento del Lenguaje Natural*
 - *Journal of Intelligent and Fuzzy Systems*
- **Program Committee member for international conferences and workshops**
 - 18th EPIA Conference on Artificial Intelligence (EPIA 2017)
 - 19th International Conference on Speech and Computer (SPECOM 2017)
 - 23rd International Conference on Applications of Natural Language to Information Systems (NLDB2018)
 - 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)
 - 6th International Conference on Mining Intelligence and Knowledge Exploration (MIKE 2018)
 - 5th Italian Conference on Computational Linguistics (CLiC-it 2018)
 - Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2018)
 - 24th International Conference on Applications of Natural Language to Information Systems (NLDB2019)
 - 41st European Conference on Information Retrieval (ECIR 2019)
 - International Conference Recent Advances in Natural Language Processing (RANLP 2019)
 - 5th SwissText & 16th KONVENS Joint Conference 2020 (SwissText at KONVENS 2020)
 - 7th Italian Conference on Computational Linguistics (CLiC-it 2020)

○ **Organization of international conferences, workshops and shared tasks**

- Organization of the IronITA shared task (Irony Detection in Italian Tweets) within EVALITA 2018, the 6th Evaluation Campaign of Natural Language Processing and Speech tools for Italian.
- Member of the local organizing committee of the 5th Italian Conference on Computational Linguistics (CLiC-it 2018).
- Organization of the SardiStance shared task (Stance Detection in Italian Tweets) within EVALITA 2020, the 7th Evaluation Campaign of Natural Language Processing and Speech tools for Italian.

○ **Additional organizational activity**

- Organizer of a bi-weekly meeting for a reading activity with other PhD students of the 33rd cycle at the Computer Science department at the Università degli Studi di Torino. [January 2018 - May 2018].
- Co-manager of the mailing list of the Content Centered Computing group at the Università degli Studi di Torino (CCC-group).

Talks:

- Oral presentation at the 4th Italian Conference on Computational Linguistics (CLiC-it 2017)
- Oral presentation at the 6th Evaluation Campaign on NLP and Speech Tools for Italian (EVALITA 2018)
- Oral presentation at the 3rd Conference on Recent Advances in Artificial Intelligence (RAAI 2019)
- Oral presentation at the 5th International Conference on Dependency Linguistics (DepLing 2019)
- Oral presentation at the at the Doctoral Symposium inside SEPLN 2019
- Oral presentation at the 6th Italian Conference on Computational Linguistics (CLiC-it 2019)
- Video presentation at the 28th International Conference on Computational Linguistics (COLING 2020)
- Video presentation at the 7th Evaluation Campaign on NLP and Speech Tools for Italian (EVALITA 2020)

Seminars/Lessons:

- 17th December 2020 (presso Prin LAILA - Unità di ricerca pavese)
Nozioni di base di linguistica computazionale e Natural Language Processing

Awards:

- Special Mention for the "best discussion, critical insights, and error analysis" at 6th Evaluation Campaign of Natural Language Processing and Speech tools for Italian, for the paper "Automatic Identification of Misogyny in English and Italian Tweets at EVALITA 2018 with a Multilingual Hate Lexicon".

Publications in journals:

- **Cignarella, A. T.**, Bosco, C., Patti, V., and Lai, M. (2019). TWITTIRO: an Italian Twitter Corpus with a Multi-layered Annotation for Irony. *Italian Journal of Computational Linguistics (IJCoL)*.

- Lai, M., **Cignarella, A. T.**, Farías, D. I. H., Bosco, C., Patti, V., and Rosso, P. (2020). Multilingual stance detection in social media political debates. *Computer Speech and Language*, 101075.
- Sanguinetti, M., Bosco, C., Cassidy, L., Çetinoğlu, Ö., **Cignarella, A. T.**, Lynn, T., Rehbein, I., Ruppenhofer, J., Seddah, D., Zeldes, A. (2020). Treebanking User-Generated Content: a UD Based Overview of Guidelines, Corpora and Unified Recommendations. *Submitted to Language Resources and Evaluation Journal, Special Issue "Annotation of non-standard corpora"*.

Publications in conference/workshop proceedings:

[irony detection]

- **Cignarella, A. T.**, Bosco, C., and Patti, V. (2017). TWITTIRÒ: A social media corpus with a multi-layered annotation for irony. In *Proceedings of the 4th Italian Conference on Computational Linguistics (CLiC-it 2017)*. CEUR-WS.org.
- **Cignarella, A. T.**, Frenda, S., Basile, V., Bosco, C., Patti, V., and Rosso, P. (2018). Overview of the EVALITA 2018 task on irony detection in italian tweets (IronITA). In *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*. CEUR-WS.org.
- **Cignarella, A. T.**, Bosco, C., Patti, V., and Lai, M. (2018). *Application and Analysis of a Multi-layered Scheme for Irony on the Italian Twitter Corpus TWITTIRÒ*. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. ELRA.
- **Cignarella, A. T.**, Sanguinetti, M., Bosco, C., and Rosso, P. (2019). Is This an Effective Way to Annotate Irony Activators? In *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-it 2019)*. CEUR-WS.org.
- **Cignarella, A. T.**, and Bosco, C. (2019). ATC at IroSva 2019: Shallow syntactic dependency-based features for irony detection in Spanish variants. In *35th Conference of the Spanish Society for Natural Language Processing (SEPLN 2019)*. CEUR-WS.org.
- **Cignarella, A. T.**, Sanguinetti, M., Bosco, C., and Paolo, R. (2020). Marking Irony Activators in a Universal Dependencies Treebank: The Case of an Italian Twitter Corpus. In *12th Conference on Language Resources and Evaluation (LREC 2020)*. ELRA.
- **Cignarella, A. T.**, Basile, V., Sanguinetti, M., Bosco, C., Rosso, P., and Benamara F. (2020). Multilingual Irony Detection with Dependency Syntax and Neural Models. *Accepted at COLING 2020, proceedings in press*.

[stance detection]

- Lai, M., **Cignarella, A. T.**, Hernández Farías, D. H., (2017). ITACOS at IberEval2017: Detecting Stance in Catalan and Spanish tweets. In *Proceedings of the 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017)*. CEUR-WS.org.
- Ghanem, B., **Cignarella, A. T.**, Bosco, C., Rosso, P., and Pardo, F. M. R. (2019). UPV-28-UNITO at SemEval-2019 Task 7: Exploiting Post's Nesting and Syntax Information for Rumor Stance Classification. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval 2019)* ACL.
- **Cignarella, A. T.**, Lai, M., Bosco, C., Patti, V., and Rosso, P. (2020). SardiStance@EVALITA2020: Overview of the Task on Stance Detection in Italian Tweets. In *Proceedings of the 6th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. CEUR-WS.org.

[others]

- Pamungkas, E. W., **Cignarella, A. T.**, Basile, V., and Patti, V. (2018). Automatic Identification of Misogyny in English and Italian Tweets at EVALITA 2018 with a Multilingual Hate Lexicon. In *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*. CEUR-WS.org.
- Pamungkas, E. W., **Cignarella, A. T.**, Basile, V., and Patti, V. (2018). 14-ExLab@UniTo for AMI at IberEval2018: Exploiting Lexical Knowledge for Detecting Misogyny in English and Spanish Tweets. In *Proceedings of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018)*. CEUR-WS.org.
- **Cignarella, A. T.**, Bosco, C., and Rosso, P. (2019). Presenting TWITTIRO-UD: An Italian Twitter Treebank in Universal Dependencies. In *Proceedings of the 5th International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*. ACL.
- Sanguinetti, M., Bosco, C., Cassidy, L., Çetinoğlu, Ö., **Cignarella, A. T.**, Lynn, T., Rehbein, I., Ruppenhofer, J., Seddah, D., Zeldes, A. (2020). Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*. ELRA.
- Pellegrini, M. and **Cignarella, A. T.** (2020). (Stem and Word) Predictability in Italian verb paradigms: An Entropy-Based Study Exploiting the New Resource *LeFFI*. *Accepted at CLiC-it 2020, proceedings in press*.

The complete list of my publications is available on my Google Scholar profile <https://scholar.google.it/citations?user=SPV3SSUAAAAJ&hl=it&authuser=1>.