

Alessio Miaschi | Curriculum Vitae

Piazza L. Gereschi 71, San Giuliano Terme, Italy, 56017
CNR-ILC, Via G. Moruzzi 1, Pisa, Italy, 56124

✉ alessio.miaschi@ilc.cnr.it • 🌐 <https://alemmaschi.github.io/>

I am a PostDoc at the ItaliaNLP Lab (ILC-CNR), Institute for Computational Linguistics (ILC-CNR). In 2022, I received my PhD in Computer Science at the University of Pisa.

My research interests lie primarily in the context of Natural Language Processing (NLP) and in the study of Language Models (LM). I am particularly interested in the interpretability of large-scale LMs and in the evaluation of their internal representations, with a specific emphasis on understanding their inner linguistic abilities. Furthermore, I work in the development of NLP tools tailored for building educational applications.

Experience

- **Postdoctoral Research Fellowship** **Pisa, Italy**
Institute for Computational Linguistics "A. Zampolli", National Research Council June 2022 – present
- **Research Fellowship** **Pisa, Italy**
Institute for Computational Linguistics "A. Zampolli", National Research Council Febr. 2022 – May 2022
- **PhD Fellowship** **Pisa, Italy**
Institute for Computational Linguistics "A. Zampolli", National Research Council Nov. 2018 – Oct. 2021
 - Research interests on NLP for education and on the analysis and interpretation of Neural Language Models.
- **Junior Developer** **Modena, Italy**
ADmantX (Expert System) Oct. 2017 – Sept. 2018
 - Software development and brand planning/targeting through advanced Semantic Cognitive Technology and AI propensity profiling solution.
 - Improvement and development of client/server web based applications.
- **Internship** **Pisa, Italy**
Institute for Computational Linguistics "A. Zampolli", National Research Council Oct. 2016 – Jan. 2017
 - Definition of a NLP system for tracking and evaluating the writing skill's evolution of junior high school's students.
- **Internship** **Pisa, Italy**
University of Pisa April 2014 – Sept. 2014
 - Definition of a TEI scheme, manual annotation and web development for the realization of an ancient medieval manuscript's digital edition.

Education

- **University of Pisa** **Pisa, Italy**
PhD in Computer Science Nov. 2018 – May 2022
 - Thesis: *Tracking Linguistic Abilities in Neural Language Models*
 - Abstract: In the last few years, the analysis of the inner workings of state-of-the-art Neural Language Models (NLMs) has become one of the most addressed line of research in Natural Language Processing (NLP). Several techniques have been devised to obtain meaningful explanations and to understand

how these models are able to capture semantic and linguistic knowledge. The goal of this thesis is to investigate whether exploiting NLP methods for studying human linguistic competence and, specifically, the process of written language evolution is it possible to understand the behaviour of state-of-the-art Neural Language Models (NLMs). First, we present an NLP-based stylometric approach for tracking the evolution of written language competence in L1 and L2 learners using a wide set of linguistically motivated features capturing stylistic aspects of a text. Then, relying on the same set of linguistic features, we propose different approaches aimed at investigating the linguistic knowledge implicitly learned by NLMs. Finally, we propose a study in order to investigate the robustness of one of the most prominent NLM, i.e. BERT, when dealing with different types of errors extracted from authentic texts written by L1 Italian learners.

- Supervisors: Felice Dell'Orletta, Anna Monreale

University of Pisa

Pisa, Italy

- *Master's Degree in Language Technology (Digital Humanities)* Feb. 2015 – May 2017
 - Graduated Summa Cum Laude.
 - Thesis: *Definizione di modelli computazionali per lo studio dell'evoluzione delle abilità di scrittura a partire da un corpus di produzioni scritte di apprendenti della scuola secondaria di primo grado* (Definition of NLP-approaches for studying the evolution of writing competence starting from a corpus of junior high schools essays, AILC Master Thesis Prize 2017)
 - Coursework in Programming, Natural Language Processing (NLP), Data Mining and Linguistics.
- *Bachelor's Degree in Digital Humanities* Sept. 2011 – Feb. 2015
 - Coursework in Programming, Text encoding and Web design.

Technical and Personal skills

- **Programming Languages:** Proficient in: Python, Javascript
Also basic ability with: Java, R, C#.
- **Frameworks:** Tensorflow, Pytorch, Keras, Scikit-learn, Pandas, Numpy, NLTK
- **Web Skills:** HTML (CSS), Angular JS.
- **Languages:** Italian, English.

Selected Publications

For the complete list reference to the following link: <https://scholar.google.it/citations?user=BaL8eJOAAAAJ&hl=it>

- Miaschi A., Alzetta C., Brunato D., Dell'Orletta F. and Venturi G. (2023). "Testing the Effectiveness of the Diagnostic Probing Paradigm on Italian Treebanks". In *Information*, vol. 14, n. 3, 2023, doi: 10.3390/info14030144.
- Miaschi A., Brunato D., Dell'Orletta F. and Venturi G. (2022). "On Robustness and Sensitivity of a Neural Language Model: A Case Study on Italian L1 Learner Errors". In *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 426-438, 2023, doi: 10.1109/TASLP.2022.3226333.
- Miaschi A., Sarti G., Brunato D., Dell'Orletta F. and Venturi G. (2022). "Probing Linguistic Knowledge in Italian Neural Language Models across Language Varieties". In *Italian Journal of Computational Linguistics (IJCoL)*, Vol. 8, N. 1, pages 25-44.
- Miaschi A., Brunato D. and Dell'Orletta F. (2021). "A NLP-based stylometric approach for tracking the evolution of L1 written language competence". In *Journal of Writing Research (JoWR)*, Vol. 13, Issue 1, pages 71-105.

- Miaschi A., Sarti G., Brunato D., Dell'Orletta F. and Venturi G. (2020). "Italian Transformers Under the Linguistic Lens". In *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020)*
- Alzetta C., Miaschi A., Dell'Orletta F., Koceva F. and Torre I. (2020). "PRELEARN @ EVALITA 2020: Overview of the Prerequisite Relation Learning Task for Italian". In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org
- Miaschi A., Brunato D., Dell'Orletta F. and Venturi G. (2020). "Linguistic Profiling of a Neural Language Model". In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, December 2020, Barcelona (Online).
- Miaschi A., and Dell'Orletta F. (2020). "Contextual and Non-Contextual Word Embeddings: an in-depth Linguistic Investigation". In *Proceedings of the 5th Workshop on Representation Learning for NLP (ACL 2020)*, July 2020, Online.
- Miaschi A., Davidson S., Brunato D., Dell'Orletta F., Sagae K., Sanchez-Gutierrez C.H., and Venturi G. (2020). "Tracking the Evolution of Written Language Competence in L2 Spanish Learners". In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications (ACL 2020)*, July 2020, Online.
- Miaschi A., Alzetta C., Cardillo F. A., and Dell'Orletta F. (2019). "Linguistically-Driven Strategy for Concept Prerequisites Learning on Italian". In *Proceedings of 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2019)*, 2 August 2019, Florence, Italy.

Teaching

- Linguistica Computazionale (*Computational Linguistics*), MSc Linguistics, University of Padova, ongoing.
- Linguistica Computazionale II (*Computational Linguistics II*), MSc Digital Humanities, University of Pisa, ongoing.
- Python per umanisti principianti (*Python for beginners*), Hands on: Strumenti digitali per le DH, May 10 – June 3 2021, Laboratorio di Cultura Digitale, University of Pisa (Online).
- Linguistica Computazionale (*Computational Linguistics*), BSc Digital Humanities, University of Pisa, a.y. 2020/21, 2021/22 (Teaching assistant).
- Progettazione e programmazione web (*Web Design and Programming*), BSc in Digital Humanities, University of Pisa, a.y. 2018/19, 2019/20 (Teaching assistant).

Activities

- Reviewing for:
 - Conferences: ACL, EMNLP, COLING, LREC, CLiC-it, EVALITA, ACM UMAP, BEA Workshop (ACL)
 - Journals: Soft Computing
- Co-organization of the LangLearn Shared Task at EVALITA 2023.
- Co-organization of PRELEARN and ATE_ABSITA Shared Tasks at EVALITA 2020 (co-located event of CLiC-it 2020, Bologna).
- Co-organization of dissemination activities:
 - È Dante o non è Dante? Questo è il problema, October 27-29 2021, Festival della Scienza, Genova (Online), http://www.festivalcienza.it/contents/instance2/static/programma_21_def.pdf
 - Ehi Siri, che cos'è la Linguistica Computazionale?, April 17 2021, Science Web Festival (Online), <https://sciencewebfestival.it/index.php/2021/04/09/ehi-siri-che-cose-la-linguistica-computazionale/>

Invited Talks

- 09/06/2023: Invited Talk at the DCP23 Workshop, University of Pisa - Opening Large Language Models
- 31/05/2023: Lab at the Lectures on Computational Linguistics 2023 - Explaining Neural Language Models from Internal Representations to Model Predictions
- 08/03/2023: Invited talk at "Seminario di Cultura Digitale" (Informatica Umanistica - Università di Pisa) - Le risorse linguistiche al tempo delle reti neurali
- 13/12/2022: Invited Talk at "Ab urbe condita" seminar series organized by KDD Lab - Profiling Neural Language Models
- 27/10/2022: Tech Talk at the School of AI (Pi School) 2022 - Interpreting Neural Language Models
- 21/09/2022: Invited Talk at the International Summer School on "Advances in AI" 2022 - Profiling Neural Language Models

Honors and Awards

- Research Prize "Dino Buzzetti" 2023. Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD).
- Special mention at CLiC-it 2020 (*Italian Transformers Under the Linguistic Lens*). Seventh Italian Conference on Computational Linguistics (CLiC-it 2020).
- Outstanding Paper for COLING 2020 (*Linguistic Profiling of a Neural Language Model*). 28th International Conference on Computational Linguistics (COLING 2020), December 2020, Barcelona (Online).
- AILC Master Thesis Prize 2017. Fourth Italian Conference on Computational Linguistics, Rome, December 11-13 2017.