

Tommaso Pasini

✉ p.tommaso@gmail.com • 🌐 pasinit.github.io

Education

La Sapienza

PhD Degree

PhD in Natural Language Processing

Rome

01/10/2015–16/09/2019

La Sapienza

Master's Degree, 110/110 cum laude

Master's degree in computer science

Rome

01/09/2013–27/10/2015

La Sapienza

Bachelor's Degree, 108/110

Bachelor's degree in computer science

Rome

01/09/2010–19/12/2013

Professional Experience

Research Fellow

La Sapienza, CS Department, Rome

01/09/2019–Present

Automatic generation of training sets for Word Sense Disambiguation in multiple languages. I am also interested in learning representations for word meanings and their distribution within a given corpus.

Bloomberg Intern

Machine Learning Team, Bloomberg LP.

01/06/2018–31/10/2018

Research focused internship in NLP. I studied the benefits and the drawbacks of using a Knowledge Base as background knowledge in NLP tasks and neural networks. In the second part of my internship, I researched methods to interpret neural networks outcomes on NLP tasks such as POS tagging and Entity Recognition.

Tutor Unitelma Object-Oriented Programming

Unitelma Sapienza

01/09/2016 – 31/03/2018

I tutored "Metodologie di Programmazione" online course of the bachelor degree in CS. The course is intended to teach students Object Oriented Programming via the Java programming language.

Natural Language Processing Tutor

La Sapienza, CS Department, Rome

01/03/2015 – 30/09/2018

I have tutored the master course of Natural Language Processing of prof. Roberto Navigli. Inter alia, I followed and evaluated the students on the three homework we delivered (on morpheme splitting, pos tagging with neural architectures, relation extraction) during the course and on the final projects.

PhD in Natural Language Processing

La Sapienza, CS Department, Rome

01/11/2015–30/09/2019

PhD in Natural Language Processing supervised by prof. Roberto Navigli. My thesis was about the automatic generation of semantically-annotated training data in different languages and potentially specific to a target domain of interest.

Research Scholarship

La Sapienza, CS Department, Rome

01/01/2014–30/06/2015

WiBi Taxonomy : automatic taxonomy creation of English Wikipedia page and categories with Prof. Roberto Navigli, funded by ERC Starting Grant, EU FP7 project number 259234.

References

1. **Tommaso Pasini**, Federico Scozzafava, and Bianca Scarlini. CluBERT: A Cluster-Based Approach for Learning Sense Distributions in Multiple Languages. In *Proceedings of ACL*, 2020; Citations: 1.
2. **Tommaso Pasini**. The Knowledge Acquisition Bottleneck Problem in Multilingual Word Sense Disambiguation. In *Proceedings of IJCAI, Survey Track*, 2020; Citations: 2.
3. Edoardo Barba, Procopio Luigi, Niccolò Campolungo, **Tommaso Pasini**, and Roberto Navigli. MuLaN: Multilingual Label propagation for Word Sense Disambiguation. In *Proceedings of IJCAI*, 2020; Citations: 0.

4. **Tommaso Pasini** and José Camacho-Collados. A short survey on sense-annotated corpora. In *Proceedings of LREC 2020*, Marseille, France, 2020; Citations: 2.
5. Bianca Scarlini, **Tommaso Pasini**, and Roberto Navigli. Sense-annotated corpora for word sense disambiguation in multiple languages and domains. In *Proceedings of LREC 2020*, Marseille, France, 2020; Citations: 0.
6. Bianca Scarlini, **Tommaso Pasini**, and Roberto Navigli. SensEmBERT: Context-Enhanced Sense Embeddings for Multilingual Word Sense Disambiguation. In *Proceedings of the Thirty-Fourth Conference on Artificial Intelligence*. Association for the Advancement of Artificial Intelligence, 2020; Citations: 7.
7. Caterina Lacerra, Michele Bevilacqua, **Tommaso Pasini**, and Roberto Navigli. CSI: A coarse sense inventory for 85% word sense disambiguation. In *Proceedings of the Thirty-Fourth Conference on Artificial Intelligence*. Association for the Advancement of Artificial Intelligence, 2020; Citations: 0.
8. **Tommaso Pasini** and Roberto Navigli. Train-O-Matic: Supervised Word Sense Disambiguation with No (Manual) Effort. *Artificial Intelligence Journal*, 2020; Citations: 1.
9. Bianca Scarlini, **Tommaso Pasini**, and Roberto Navigli. Just “OneSeC” for producing multilingual sense-annotated data. In *Proceedings of ACL*, Florence, Italy, 2019; Citations: 2.
10. **Tommaso Pasini** and Roberto Navigli. Two knowledge-based methods for high-performance sense distribution learning. In *Proceedings of AAAI*, pages 5374–5381, New Orleans, Louisiana, USA, 2018; Citations: 8.
11. **Tommaso Pasini**, Francesco Elia, and Roberto Navigli. Huge Automatically Extracted Training-Sets for Multilingual Word Sense Disambiguation. In *Proceedings LREC 2018*, Miyazaki, Japan, 2018; Citations: 0.
12. Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa Anke, Sergio Oramas, **Tommaso Pasini**, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. Semeval-2018 task 9: Hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 712–724, 2018; Citations: 45.
13. **Tommaso Pasini** and Roberto Navigli. Train-O-Matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In *EMNLP*, 2017; Citations: 26.
14. Tiziano Flati, Daniele Vannella, **Tommaso Pasini**, and Roberto Navigli. Multiwibi: The multilingual wikipedia bitaxonomy project. *Artificial Intelligence*, 241:66–102, 2016; Citations: 23.
15. Tiziano Flati, Daniele Vannella, **Tommaso Pasini**, and Roberto Navigli. Two is bigger (and better) than one: the wikipedia bitaxonomy project. In *ACL (1)*, pages 945–955, 2014; Citations: 95.

Accademic Activities

- **SemEval Organizer:** SemEval-2018 Task 9.
- **Area Chair:** IJCAI 2021.
- **Senior Program Committee Member:** ECAI 2019.
- **Program Committee Member:** ACL 2017, 2018, 2019, 2020; IJCAI 2020, EMNLP 2018, 2019, 2020; AAAI 2018, 2020, 2021; ISWC 2017, 2018; Computational Linguistics 2019, Artificial Intelligence Journal 2019, 2020.

Invited Presentations

Research Seminar

Language Technology Group, University of Helsinki

17/10/2019

Different Approaches for Enabling Multilinguality in Word Sense Disambiguation.

Computer skills

Programming Languages: Python, Java, Bash, C

Frameworks: Pytorch, TensorFlow < 1.0, Keras, Maven, CoreNLP, Git, L^AT_EX

Supervising and Mentoring Activities

I am co-supervising or have co-supervised the following PhD, Master and Bachelor students.

PhD Students.....	
◦ Bianca Scarlini	[2nd year]
Automatic creation for multilingual data for Word Sense Disambiguation and Knowledge Representation.	
◦ Caterina Lacerra	[2nd year]
Coarse-grained Word Sense Disambiguation and Lexical Substitution.	
◦ Edoardo Barba	[1st year]
Automatic creation for multilingual data for Word Sense Disambiguation; data efficient multilingual contextualised embeddings of words and sentences.	
◦ Niccolò Campolungo	[1st year]
Span-based contextualised word embeddings.	
Master Students.....	
◦ Riccardo Orlando	[On Going]
Automatic creation of multilingual data for Semantic Role Labelling.	
◦ Livia Biggi	[On Going]
Enhancing Neural Machine Translation with Word Senses.	
◦ Bianca Scarlini	[July 2018]
Leveraging Wikipedia Categories for Wide Coverage Word Sense Disambiguation.	
◦ Caterina Lacerra	[May 2018]
A Supervised Method for Word Sense Disambiguation with Coarse-Grained Labels.	
◦ Rexhina Blloshmi	[July 2018]
Graph Convolutional Neural Networks for Language Agnostic Document Representation.	
Bachelor Students.....	
◦ Simone Primarosa	[July 2019]
Evaluation of translation-based knowledge graph embedding techniques (TransE, TransR, TransD) on large knowledge graphs (30M edges).	
◦ Kristian Iliev	[June 2017]
An algorithm to merge lexical-semantic resources.	

Languages

Italian: Native

English: Fluent

PhD Thesis

Title: *Knowledge-based approaches to producing large-scale training data from scratch for Word Sense Disambiguation and Sense Distribution Learning*

Supervisors: Prof. Roberto Navigli

Description: The thesis focuses on relieving the burden of human annotations in Word Sense Disambiguation thus enabling the automatic construction of high-quality sense-annotated dataset not only for English, but especially for other languages where sense-annotated data are not available at all. Furthermore, in the thesis I also illustrate the problem of automatically inducing word-sense distributions and, to this end, introduce two knowledge-based and language-independent approaches.