

CLiC-it 2023

Ninth Italian Conference
on Computational Linguistics

November 30 - December 2
2023, Venice

CLiC-it is endorsed by



Program at a Glance

Thursday, November 30, 2023

Location: Auditorium Santa Margherita, Dorsoduro 3689

8:30–9:00	Registration		
9:15–10:00	Opening Session		
10:00–11:00	IT	David Bamman School of Information, UC Berkeley	The Promise and Peril of Large Language Models for Cultural Analytics
11:00–11:30	Coffee Break		
11:30–13:00	OS I	Chair: Rachele Sprugnoli	NLP 4 Social Good
13:00–14:30	Lunch Break		
14:30–15:30	OS II	Chair: Tommaso Caselli	Large Language Models: Foundations
15:30–16:30	PS ⁺ I	Chair: Elisabetta Fersini	Digital Humanities and NLP 4 Social Good (and coffee break)
16:30–17:30	RC	Chair: Alessandro Lenci	Research Communications
17:30–18:30	AILC Meeting		
19:30	Social Dinner		

IT: Invited Talk, OS: Oral Session, PS⁺: Poster Session and Coffee Break, RC: Research Communications.

Friday, December 01, 2023

Location: Auditorium Santa Margherita, Dorsoduro 3689

9:00-10:30	OS III	Chair: Gloria Gagliardi	Linguistic Issues in CL and NLP
10:30-11:00	Coffee Break		
11:00-12:00	IT	Vera Demberg Saarland University	Pragmatic Processing in Humans and Language Models
12:00-13:00	PS II	Chair: Giulia Venturi	(Psycho)Linguistics Issues
13:00-14:30	Lunch Break		
14:30-15:30	OS IV	Chair: Pierpaolo Basile	Digital Humanities and Related Topics
15:30-16:30	PS III	Chair: Elisabetta Ježek	Core and Applied NLP Tasks
16:30-18:00	OS V	Chair: Valerio Basile	Large Language Models Applications
18:00-18:30	Closing Session		
19:30	Goodbye Drink		

Saturday, December 02, 2023

Location: Sala Morelli, ground floor of Malcantone Marcorà, Dorsoduro 3484

09:00-12:30	TU	Danilo Croce and Claudiu Daniel Hromei University of Rome, Tor Vergata	Large Language Models and How to Instruction Tune Them (in a Sustainable Way)
-------------	----	--	--

IT: Invited Talk, OS: Oral Session, PS: Poster Session, TU: Tutorial.

Detailed Program

Thursday, November 30, 2023

Location: Auditorium Santa Margherita, Dorsoduro 3689

08:30–09:00 Registration

09:15–10:00 Opening Session

9:15–9:40 Greetings by Gianluca Lebani, Marina Buzzoni, Simonetta Montemagni and Oliviero Stock

9:40–10:00 Rodolfo Del Monte. **Linguistica Computazionale a Venezia: quando tutto è cominciato**

10:00–11:00 David Bamman: **The Promise and Peril of Large Language Models for Cultural Analytics**

11:00–11:30 Coffee Break

11:30–13:00 Oral Presentations I - **NLP 4 Social Good**

Session Chair: **Rachele Sprugnoli**

11:30–11:45 Gianmaria Balducci, Giulia Rizzi, Elisabetta Fersini. **Bias Mitigation in Misogynous Meme Recognition: A Preliminary Study**

11:45–12:00 Marco Madeddu, Simona Frenda, Mirko Lai, Viviana Patti, Valerio Basile. **DisaggregHate It Corpus: A Disaggregated Italian Dataset of Hate Speech**

12:00–12:15 Kamyar Zeinalipour, Tommaso Iaquinta, Asya Zanollo, Giovanni Angelini, Leonardo Rigutini, Marco Maggini, Marco Gori. **Italian Crossword Generator: Enhancing Education through Interactive Word Puzzles**

12:15–12:30 Giovanni Valer, Alan Ramponi, Sara Tonelli. **When You Doubt, Abstain: A Study of Automated Fact-Checking in Italian Under Domain Shift**

12:30–12:45 Jacopo Gili, Lucia Passaro, Tommaso Caselli. **Check-IT!: A Corpus of Expert Fact-checked Claims for Italian**

12:45–13:00 Marta Maffia, Loredana Schettino, Vincenzo Norman Vitale. **Automatic Detection of Parkinson’s Disease with Connected Speech Acoustic Features: towards a Linguistically Interpretable Approach**

13:00–14:30 Lunch break

14:30–15:30 Oral Presentations II - **Large Language Models: Foundations**

Session Chair: Tommaso Caselli

14:30–14:45 Andrea Santilli, Emanuele Rodolà. **Camoscio: an Italian Instruction-tuned LLaMA**

14:45–15:00 Federico Bianchi, Giuseppe Attanasio, Raphael Pisoni, Silvia Terragni, Gabriele Sarti, Dario Balestri. **Contrastive Language–Image Pre-training for the Italian Language**

15:00–15:15 Debora Nozza, Giuseppe Attanasio. **Is It Really That Simple? Prompting Large Language Models for Automatic Text Simplification in Italian**

15:15–15:30 Vincenzo Norman Vitale, Loredana Schettino, Francesco Cutugno. **On incrementing interpretability of machine learning models from the foundations: a study on syllabic speech units**

15:30–16:30 Coffee Break and Poster Session I - **Digital Humanities and NLP 4 Social Good**

Session Chair: Elisabetta Fersini

Conference Papers

- Alessandra Teresa Cignarella, Simona Frenda, Tom Bourgeade, Cristina Bosco, Francesca D’Errico. **Linking Stance and Stereotypes About Migrants in Italian Fake News**
- Melissa Donati, Ludovica Polidori, Paola Vernillo, Gloria Gagliardi. **Building a corpus on Eating Disorders from TikTok: challenges and opportunities**
- Federica Gamba, Marco Carlo Passarotti, Paolo Ruffolo. **Linking the Dictionary of Medieval Latin in the Czech Lands to the LiLa Knowledge Base**
- Marco Gaido, Dennis Fucci, Matteo Negri, Luisa Bentivogli. **How To Build Competitive Multi-gender Speech Translation Models For Controlling Speaker Gender Translation**
- Pier Felice Balestrucci, Alessandro Mazzei, Luca Anselma, Cristian Bernareggi. **Building a Spoken Dialogue System for Supporting Blind People in Accessing Mathematical Expressions**
- Valeria Irene Boano, Francesco Mambrini, Marco Carlo Passarotti, Riccardo Ginevra. **Modelling and Publishing the Lexicon der indogermanischen Verben as Linked Open Data**
- Irene De Felice, Lucia Tamponi, Federica Iurescia, Marco Carlo Passarotti. **Linking the Corpus CLaSSES to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin**

- Elena Sofia Ruzzetti, Dario Onorati, Leonardo Ranaldi, Davide Venditti, Fabio Massimo Zanzotto. **Investigating Gender Bias in Large Language Models for the Italian Language**
- Oscar Araque, M^a Felipa Ledesma Corniel, Kyriaki Kalimeri. **Towards a Multilingual System for Vaccine Hesitancy using a Data Mixture Approach**
- Arthur Thomas Edward Capozzi Lupi, Alessandra Teresa Cignarella, Simona Frenda, Mirko Lai, Marco Antonio Stranisci, Alessandra Urbinati. **Debunker Assistant: a support for detecting online misinformation**
- Martina Rosola, Simona Frenda, Alessandra Teresa Cignarella, Matteo Pellegrini, Andrea Marra, Mara Floris. **Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian**
- Alessandro Astorino, Giulia Rizzi, Elisabetta Fersini. **Integrated Gradients as Proxy of Disagreement in Hateful Content**
- Paolo Gajo, Silvia Bernardini, Adriano Ferraresi, Alberto Barrón-Cedeño. **Hate Speech Detection in an Italian Incel Forum Using Bilingual Data for Pre-Training and Fine-Tuning**
- Federico Boschetti, Laura Chilla, Maria Konstantinidou, John Pavlopoulos. **Annotating Homeric Emotions by a Domain-Specific Language**
- Andrea Zaninello, Sofia Brenna, Bernardo Magnini. **Textual Entailment with Natural Language Explanations: The Italian e-RTE-3 Dataset**
- Andrea Zugarini, Thomas Røthenbacher, Kai Klede, Marco Ernandes, Bjoern Eskofier, Dario Zanca. **Die Rätselrevolution: Automated German Crossword Solving**
- Olga Uryupina. **Qualitative Analysis of Persuasive Emotion Triggering in Online Content**
- Ivan Spada, Mirko Lai, Viviana Patti. **Inters8: A Corpus to Study Misogyny and Intersectionality on Twitter**

Research Communications

- Cristina Bosco, Viviana Patti, Simona Frenda, Alessandra Teresa Cignarella, Marinella Paciello, Francesca D'Errico. **Detecting racial stereotypes: An Italian social media corpus where psychology meets NLP**
- Gosse Minnema, Sara Gemelli, Chiara Zanchi, Malvina Nissim, Tommaso Caselli. **Dead or Murdered? Predicting Responsibility Perception in Femicide News Reports**
- Daniel Russo, Serra Sinem Tekiroglu, Marco Guerini. **Benchmarking the Generation of Fact Checking Explanations**
- Helena Bonaldi, Sara Dellantonio, Serra Sinem Tekiroglu, Marco Guerini. **Human-Machine Collaboration Approaches to Build a Dialogue Dataset for Hate Speech Countering**

16:30–17:30 Research Communications

Session Chair: **Alessandro Lenci**

16:30–16:45 Roberto Dessì, Michele Bevilacqua, Eleonora Gualdoni, Nathanaël Carraz Rakotonirina, Francesca Franzon, Marco Baroni. **Cross-Domain Image Captioning with Discriminative Finetuning**

16:45–17:00 Marco Antonio Stranisci, Rossana Damiano, Enrico Mensa, Viviana Patti, Daniele Paolo Radicioni, Tommaso Caselli. **WikiBio: a Semantic Resource for the Intersectional Analysis of Biographical Events**

17:00–17:15 Andrea Santilli, Silvio Severino, Emilian Postolache, Valentino Maiorca, Michele Mancusi, Riccardo Marin, Emanuele Rodolà. **Accelerating Transformer Inference for Translation via Parallel Decoding**

17:15–17:30 Simone Tedeschi, Johan Bos, Thierry Declerck, Jan Hajič, Daniel Hershcovich, Eduard H. Hovy, Alexander Koller, Simon Krek, Steven Schockaert, Rico Sennrich, Ekaterina Shutova, Roberto Navigli. **What's the Meaning of Superhuman Performance in Today's NLU?**

17:30–18:30 AILC Meeting

19:30 Social Dinner

Friday, December 1, 2023

Location: Auditorium Santa Margherita, Dorsoduro 3689

09:00–10:30 Oral Presentations III - Linguistic Issues in CL and NLP

Session Chair: **Gloria Gagliardi**

9:00–9:15 Paolo Brasolin, Greta Hayley Franzini, Stefania Spina. **“Ti blocco perché sei un trollazzo”**. Lexical innovation in contemporary Italian in a large Twitter corpus

09:15–09:30 Matteo Melis, Anastasiia Salova, Roberto Zamparelli. **Is Change the Only Constant? An Inquiry Into Diachronic Semantic Shifts in Italian and Spanish**

09:30–09:45 Paola Merlo, Giuseppe Samo, Vivi Nastase, Chunyang Jiang. **Building structured synthetic datasets: The case of Blackbird Language Matrices (BLMs)**

09:45–10:00 Maria Cassese, Alessandro Bondielli, Alessandro Lenci. **Assessing Language and Vision-Language Models on Event Plausibility**

10:00–10:15 Federica Papa, Luca Dini, Dominique Brunato, Felice Dell'Orletta. **Unraveling Text Coherence from the Human Perspective: a Novel Dataset for Italian**

10:15–10:30 Claudia Collacciani, Giulia Rambelli. **Interpretation of Generalization in Masked Language Models: An Investigation Straddling Quantifiers and Generics**

10:30–11:00 Coffee Break

11:00–12:00 Vera Demberg: **Pragmatic Processing in Humans and Language Models**

12:00–13:00 Poster Session II - (Psycho)Linguistics Issues

Session Chair: **Giulia Venturi**

Conference Papers

- Melissa Donati, Carlo Strapparava. **A Cognitive Linguistics analysis of Phrasal Verbs representation in Distributional Semantics**
- Muhadj Adnan, Nicolo' Brandizzi. **CorpusCompass: A Tool for Data Extraction and Dataset Generation in Corpus Linguistics**
- Matteo Gay, Cristiano Chesi. **Extracting an expectation-based lexicon for UD treebanks**
- Evan De Francesco, Carlo Strapparava. **The PBSDS: A Dataset for the Detection of Pseudo-profound Bullshit**
- Elena Spaziani, Eliana Di Palma. **Let's keep an eye on Russian: testing sensitivity to the change in the grammatical number in somatic idioms with ruBERT**

- Agnese Daffara, Elisabetta Ježek. **Towards an Italian Corpus for Implicit Object Completion**
- Dario Del Fante, Federico Manzella, Guido Sciacvico, Eduard Ionel Stan. **A Post-Modern Approach to Automatic Metaphor Identification**
- Olga Loginova, Raffaella Bernardi. **The Inherence of Telicity: Unveiling Temporal Reasoning in Video Question Answering**
- Flavia Sciolette, Simone Marchi, Emiliano Giovannetti. **Towards a New Computational Lexicon for Italian: building the morphological layer by harmonizing and merging existing resources**
- Maria De Martino. **Processing effort during reading texts in young adults: text simplification, readability assessment and preliminary eye-tracking data**
- Manuela Marchisio, Alessandro Mazzei, Dario Sammaruga. **Introducing Deep Learning with Data Augmentation and Corpus Construction for LIS**
- Davide Buscaldi, Paolo Rosso. **How good is NLLB for low-resource languages? A study on the Genoese Language**
- Ilaria Manfredi, Lorenzo Gregori. **Identification of Multiword Expressions: comparing the performance of a Conditional Random Fields model on corpora of written and spoken Italian**
- Francesca Chiusaroli, Tiberio Uricchio, Johanna Monti, Maria Laura Pierucci, Federico Sangati. **GPT-based Language Models meet Emojitaliano: A Preliminary Assessment Test between Automation and Creativity**
- Valentino Frasnelli, Alessio Palmero Aprosio. **A preliminary release of the Italian Parliamentary Corpus**

Research Communications

- Pierluigi Cassotti, Lucia Siciliani, Marco de Gemmis, Giovanni Semeraro, Pierpaolo Basile. **XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic change**
- Simona Frenda, Alessandro Pedrani, Valerio Basile, Soda Marem Lo, Alessandra Teresa Cignarella, Raffaella Panizzon, Cristina Marco, Bianca Scarlini, Viviana Patti, Cristina Bosco, Davide Bernardi. **EPIC: Multi-Perspective Annotation of a Corpus of Irony**
- Alkis Koudounas, Moreno La Quatra, Lorenzo Vaiani, Luca Colomba, Giuseppe Attanasio, Eliana Pastor, Luca Cagliero, Elena Baralis. **ITALIC: An Italian Intent Classification Dataset**
- Alessandro Maisto. **Extract Similarities from Syntactic Contexts: a Distributional Semantic Model Based on Syntactic Distance**
- Giulia Rambelli, Marianna M. Bolognesi. **Contextual Variability Depends on Categorical Specificity rather than Conceptual Concreteness: A Distributional Investigation on Italian data**
- Giulia Rambelli, Emmanuele Chersoni, Philippe Blache, Alessandro Lenci. **Compositionality as an Analogical Process: Introducing ANNE**

13:00–14:30 Lunch Break

14:30–15:30 Oral Presentations IV - Digital Humanities and Related Topics

Session Chair: **Pierpaolo Basile**

14:30–14:45 Rachele Sprugnoli, Marco Sartor. **“That branch of the Lake of Como...”: Developing a New Resource for the Analysis of I Promessi Sposi and its Historical Translations**

14:45–15:00 Claudia Corbetta, Marco Carlo Passarotti, Flavio Massimiliano Cecchini, Giovanni Moretti. **Highway to Hell. Towards a Universal Dependencies Treebank for Dante Alighieri’s Comedy**

15:00–15:15 Cristina Bosco, Muhammad Okky Ibrohim, Valerio Basile, Indra Budi. **How green is Sentiment Analysis? Environmental Topics in Corpora at the University of Turin**

15:15–15:30 Chiara Alzetta, Felice Dell’Orletta, Chiara Fazzone, Alessio Miaschi, Giulia Venturi. **Unmasking the Wordsmith: Revealing Author Identity through Reader Reviews**

15:30–16:30 PM Coffee Break and Poster Session III - Core and Applied NLP tasks

Session Chair: **Elisabetta Ježek**

Conference Papers

- Elisa di Nuovo. **An experiment in error analysis of real-time speech machine translation using the example of the European Parliament’s Innovation Partnership**
- Francesco Russo, Michele Filannino. **Blaze-IT: a lightweight BERT model for the Italian language**
- Lucia Siciliani, Emanuele Tanzi, Pierpaolo Basile, Pasquale Lops. **Automatic Generation of Common Procurement Vocabulary Codes**
- Romane Werner, Sonja Bitzer, Thomas Francois. **Drug name recognition in the cryptomarket forum of Silk Road 2**
- Leonardo Ranaldi, Giulia Pucci, Elena Sofia Ruzzetti, Fabio Massimo Zanzotto, Andre Freitas. **Teasing It-LLMs adapted to Italian**
- Pierpaolo Basile, Pierluigi Cassotti, Marco Polignano, Lucia Siciliani, Giovanni Semeraro. **On the impact of Language Adaptation for Large Language Models: A case study for the Italian language using only open resources**
- Manfred Klenner, Dylan Massey, Susanna Tron. **An Italian Verb Lexicon for Sentiment Inference**
- Flavia Achena, David Preti, Davide Venditti, Leonardo Ranaldi, Cristina Giannone, Fabio Massimo Zanzotto, Andrea Favalli, Raniero Romagnoli. **Legal Summarization: to each Court its own model**
- Federico Ranaldi, Elena Sofia Ruzzetti, Leonardo Ranaldi, Davide Venditti, Cristina Giannone, Andrea Favalli, Raniero Romagnoli, Fabio Massimo Zanzotto. **Prompting LLMs in Italian language for Text-to-SQL translation**

- Robert J. Alexander, Matteo Bartocci, Oriana Persico, Guido Vetere. **Harnessing 'Il Manifesto' Newspaper Archive for Knowledge Base Creation: Techniques and Findings in the 'MeMa' Project**
- Silvia Casola, Tiziano Labruna, Alberto Lavelli, Bernardo Magnini. **Testing ChatGPT for Stability and Reasoning: A Case Study Using Italian Medical Specialization Tests**
- Massimo Aria, Corrado Cuccurullo, Luca D'Aniello, Michelangelo Misuraca, Maria Spano. **TALL: a new Shiny app of Text Analysis for All**
- Nicola Cirillo, Daniela Vellutino. **Towards a Multi-Level Annotation Format for the Interoperability of Automatic Term Extraction Corpora**
- Salar Mohtaj, Vera Schmitt, Razieh Khamsehashari, Sebastian Möller. **Multi-task Learning for German Text Readability Assessment**
- Marco Grazioso, Alessandro Suglia. **An Analysis of Visually Grounded Instructions in Embodied AI Tasks**
- Murad Hossain, Luca Anselma, Alessandro Mazzei. **Exploring sentiments in summarization: SentiTextRank, an Emotional Variant of TextRank**

Research Communications

- Edoardo Barba, Luigi Procopio, Roberto Navigli. **ExtEnD: Extractive Entity Disambiguation**
- Pere-Lluís Huguet Cabot, Simone Tedeschi, Axel-Cyrille Ngonga Ngomo, Roberto Navigli. **RED^{FM}: a Filtered and Multilingual Relation Extraction Dataset**
- Roberto Dessi, Eleonora Gualdoni, Francesca Franzon, Gemma Boleda, Marco Baroni. **Communication breakdown: On the low mutual intelligibility between human and neural captioning**
- Moreno La Quatra, Luca Cagliero. **BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization**
- Marco Polignano, Valerio Basile, Pierpaolo Basile, Giuliano Gabrieli, Marco Vassallo, Cristina Bosco. **A hybrid lexicon-based and neural approach for explainable polarity detection**
- Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, Marco Turchi. **Under the Morphosyntactic Lens: A Multifaceted Evaluation of Gender Bias in Speech Translation**

16:30-18:00 Oral Presentations V - Large Language Models: Applications

Session Chair: **Valerio Basile**

16:30–16:45 Giulia Pucci, Leonardo Ranaldi, Fabio Massimo Zanzotto. **Are All Languages Equal? Curriculum Learning over Different Languages**

16:45–17:00 Michele Papucci, Alessio Miaschi, Felice Dell'Orletta. **Lost in Labels: An Ongoing Quest to Optimize Text-to-Text Label Selection for Classification**

17:00–17:15 Federico Martelli, Andrei Stefan Bejgu, Cesare Campagnano, Jaka Čibej, Rute Costa, Apolonija Gantar, Jelena Kallas, Svetla Peneva Koeva, Kristina Koppel, Simon Krek, Margit Langemets, Veronika Lipp, Sanni Nimb, Sussi Olsen, Bolette S. Pedersen, Valeria

Quochi, Ana Salgado, László Simon, Carole Tiberius, Rafael Ureña-Ruiz, Roberto Navigli. **XL-WA: a Gold Evaluation Benchmark for Word Alignment in 14 Language Pairs**

17:15–17:30 Serena Auriemma, Mauro Madeddu, Martina Miliani, Alessandro Bondielli, Lucia Passaro, Alessandro Lenci. **Challenging specialized transformers on zero-shot classification**

17:30–17:45 Claudiu Daniel Hromei, Danilo Croce, Roberto Basili. **End-to-end Dependency Parsing via Auto-regressive Large Language Model**

17:45–18:00 Martina Miliani, Fernando Alva-Manchego, Alessandro Lenci. **Simplifying Administrative Texts for Italian L2 Readers with Controllable Transformers Models: A Data-driven Approach**

18:00–18:30 Closing Session

Emanuele Pianta Award for the Best Master Thesis, Best Student Paper

19:30 Goodbye Drink

Saturday, December 2, 2023

Location: Sala Morelli, ground floor of Malcanton Marcorà, Dorsoduro 3484

9:00–10:30 Danilo Croce and Claudiu Daniel Hromei: **Large Language Models and How to Instruction Tune Them (in a Sustainable Way)** - PART 1

10:30–11:00 Coffee Break

11:00–12:30 Danilo Croce and Claudiu Daniel Hromei: **Large Language Models and How to Instruction Tune Them (in a Sustainable Way)** - PART 2

Invited Speakers

Thursday, November 30, 2023 - 10:00-11:00

David Bamman



David Bamman is an associate professor in the School of Information at UC Berkeley, where he works in the areas of natural language processing and cultural analytics, applying NLP and machine learning to empirical questions in the humanities and social sciences. His research focuses on improving the performance of NLP for underserved domains like literature (including LitBank and BookNLP) and exploring the affordances of empirical methods for the study of literature and culture. Before Berkeley, he received his PhD in the School of Computer Science at Carnegie Mellon University and was a senior researcher at the Perseus Project of Tufts University. Bamman's work is supported by the National Endowment for the Humanities, National Science Foundation, an Amazon Research Award, and an NSF CAREER award.

The Promise and Peril of Large Language Models for Cultural Analytics

Much work at the intersection of NLP and cultural analytics/computational social science is focused on creating new algorithmic measuring devices for constructs we see encoded in text (including agency, respect, and power, to name a few). How does the paradigm shift of large language models change this? In this talk, I'll discuss the role of LLMs (such as ChatGPT, GPT-4 and open alternatives) for research in cultural analytics, both raising issues about the use of closed models for scholarly inquiry and charting the opportunity that such models present. The rise of large pre-trained language models has the potential to radically transform the space of cultural analytics by both reducing the need for large-scale training data for new tasks and lowering the technical barrier to entry, but need care in establishing the reliability of results.

Friday, December 1, 2023 - 11:00–12:00

Vera Demberg



Vera Demberg is a Professor of Computer Science and Computational Linguistics at Saarland University and a Fellow of ELLIS. She received her PhD from the School of Informatics at the University of Edinburgh in 2010 and subsequently held a position as an independent research group leader at the Cluster of Excellence for Multimodal Computing and Interaction at Saarland University. Vera Demberg received an ERC Starting Grant “Individualized Interaction in Discourse” in 2020. Her current research spans experimental and computational research on individual differences in discourse and pragmatic processing as well as computational models for natural language generation and discourse understanding.

Pragmatic Processing in Humans and Language Models

Pragmatic processing concerns inferences that go beyond the literal meaning of a text or utterance. In my talk, I will go through different types of pragmatic inferences, including scalar implicatures, atypicality inferences and other tasks requiring theory of mind reasoning; for each of these, I will discuss recent work from our lab and others, regarding the ability of large language models, and of humans, to handle these phenomena.

Saturday, December 2, 2023 - 09:00-12:30

Danilo Croce and Claudiu Daniel Hromei



Danilo Croce is a Tenure track Assistant Professor in Computer Science at the University of Rome, Tor Vergata, Faculty of Engineering. He holds a Ph.D. in Informatics Engineering from the same university. His expertise spans both theoretical and applied Machine Learning. He primarily delves into Natural Language Processing, Information Retrieval, and Data Mining domains. His particular interest lies in innovative Kernels within Support Vectors, Deep Neural Machines, Transformer-based Neural Architectures, and Large Language Models. His focus centers on advanced syntactic and semantic processing within the realms of Natural Language Processing and Computer Vision. Throughout his academic career, he has

actively participated in various roles across numerous national and European projects. He has authored over 150 publications in international journals, conference proceedings, and workshops (H-index: 24), with 10 of these being recognized as “Best Paper” and more than 15 top ranks in national and international benchmarking campaigns. Danilo has been a member of the Program Committee for more than 25 International Conferences and Workshops. He is also a regular reviewer for esteemed International Conferences including ACL, EMNLP, COLING, and AAAI.

Claudiu Daniel Hromei is a 3rd year Ph.D. student at the University of Rome Tor Vergata, with a keen interest in the dynamic field of Human-Robot Interaction. His research primarily revolves around Natural Language Processing (NLP) and Generative AI. In 2020, Claudiu successfully earned his Master’s Degree in Computer Science from Tor Vergata, in the context of Human-Robot Interaction, with a specific focus on medical scenarios. Over time, his research focused on exploring Kernel-based Support Vector Machines and the world of Large Language Models (LLMs), especially in light of the Era of ChatGPT. His approach extends to the development and maintenance of LLMs while ensuring their sustainability through innovative approaches such as Low-Rank Adaptation (LoRA) and model quantization with lower floating-point precision. Additionally, he is focused on architectures that integrate multi-modal inputs with LLMs.



Large Language Models and How to Instruction Tune Them (in a Sustainable Way)

In recent years, the evolution of Large Language Models (LLMs) has marked a profound transformation in computational linguistics and computer science. This tutorial aims to provide participants with a comprehensive understanding of the state-of-the-art LLMs introduced in the literature. We

will particularly focus on the progression that led models like GPT or LLaMA to be adapted for language inference tasks and their capability to be instruction fine-tuned, eventually leading to the development of models such as ChatGPT, Alpaca and Vicuna.

A notable challenge with LLMs has been their computational complexity, making them appear unfeasible for common usage. To (partially) solve such issues, this tutorial will delve into techniques like quantization and, more prominently, low-rank adaptation (LoRA). These techniques make it possible to fine-tune an LLM with 7 or 13 billion parameters on a standard 16 GB GPU. The application of these methods has enabled the adaptation of foundational LLMs pre-trained on large-scale text collections to a myriad of tasks, leading to a remarkable proliferation of models being released on the web daily.

Building on this, we will explore the development of a unified architecture that participated under the name of ExtremITA in all tasks of EVALITA 2023. By effectively combining prompting engineering and sustainable learning techniques, this monolithic architecture based on LLaMA tackled twenty-two complex semantic processing tasks in the Italian language, across varied semantic dimensions, including Affect Detection, Authorship Analysis, Computational Ethics, Named Entity Recognition, Information Extraction, and Discourse Coherence.

Lastly, attendees will gain insights into training a state-of-the-art foundational model (LLaMA2) using data from the competition, replicating the approach in ExtremITA. This will empower them with the knowledge to develop a monolithic model capable of addressing all tasks and extending beyond.

Social Dinner

November 30, 19:30



OKE Zattere (Red tickets)

Sestiere Dorsoduro, 1414 - 30124 Venice

10 minutes walking distance from the conference venue

OKE Santa Lucia (Blue tickets)

Sestiere Cannaregio, 158 - 30100 Venice

15 minutes walking distance from the conference venue

Goodbye Drink

December 1, 19:30



HUM.US - Appetite for Growth

Calle Larga S. Marta, 2137 - 30123 Venice

15 minutes walking distance from the conference venue

Restaurants

Bakarò (✓): Sestiere Dorsoduro, 3662 (Campo Santa Margherita). **1 minute** walking distance from the conference venue.

Tel.: +39 041 639 9946

Orient Experience II (✓ , 🍷): Sestiere Dorsoduro, 2920 (Campo Santa Margherita). **2 minutes** walking distance from the conference venue.

Tel.: +39 041 241 2669

Osteria Alla Bifora (✓): Campo Santa Margherita, 2930. **2 minutes** walking distance from the conference venue.

Tel.: +39 041 523 6119

Hostaria Osottoosopra (✓ , 🍷): Calle S. Pantalon, 5754. **2 minutes** walking distance from the conference venue.

Tel.: +39 333 802 7830

L'impronta (✓): Sestiere Dorsoduro, 3815. **3 minutes** walking distance from the conference venue.

Tel.: +39 041 275 0386

Veniceat (✓): Sestiere Dorsoduro, 3246. **4 minutes** walking distance from the conference venue.

Tel.: +39 334 788 1517

Bacaro vintidò 22 (✓ , 🍷): Calle Dona Onesta, 3928/29. **4 minutes** walking distance from the conference venue.

Tel.: +39 351 368 3808

Oniga : Sestiere Dorsoduro, 2852. **5 minutes** walking distance from the conference venue.

Tel.: +39 041 522 4410

Al Profeta (✓ , 🍷): C. Lunga S. Barnaba, 2671. **5 minutes** walking distance from the venue.

Tel.: +39 041 523 7466

Ai Artisti : Fondamenta Toletta, 1169/A. **6 minutes** walking distance from the conference venue.

Tel.: +39 041 523 8944

Frary's ( , ): Fondamenta Frari, 2558. **7 minutes** walking distance from the conference venue.

Tel.: +39 041 720050

Sudest 1401 : Dorsoduro, 1401. **9 minutes** walking distance from the conference venue.

Tel.: +39 041 099 6843

Osteria Ca' del Vento ( , ): Fondamenta de S. Basegio, 1518. **9 minutes** walking distance from the conference venue.

Tel.: +39 041 850 1909

Taverna di San Trovaso (): Calle Contarini Corfù, 1016. **9 minutes** walking distance from the conference venue.

Tel.: +39 041 520 3703

Ae Saracche (): Fondamenta Rio Marin, 847/D. **10 minutes** walking distance from the conference venue.

Tel.: +39 041 309 9283

Birraria La Corte (): Campo San Polo, 2168. **12 minutes** walking distance from the conference venue.

Tel.: +39 041 275 0570

La Tecia Vegana (): Calle dei Secchi, 2104. **13 minutes** walking distance from the conference venue.

Tel.: +39 041 524 6244

 = Vegetarian options available

 = Vegan options available

Organization

Conference Chairs

Federico Boschetti (Ca' Foscari University of Venice/Institute for Computational Linguistics "Antonio Zampolli", CNR)

Gianluca Lebani (Ca' Foscari University of Venice)

Bernardo Magnini (Fondazione Bruno Kessler)

Nicole Novielli (Aldo Moro University of Bari)

Local Organizing Committee

Franz Fischer (Ca' Foscari University of Venice)

Cristina Procentese (Ca' Foscari University of Venice)

Matteo Radaelli (Norwegian University of Science and Technology)

Proceedings Chairs

Luca Molinari (Ca' Foscari University of Venice)

Alice Suozzi (Ca' Foscari University of Venice)

Andrea Zaninello (Fondazione Bruno Kessler)

Publicity Chair

Sofia Brenna (Fondazione Bruno Kessler)

Webmasters

Eleonora Ghizzota (Aldo Moro University of Bari)

Tiziano Labruna (Fondazione Bruno Kessler)

Program Committee

Pierpaolo Basile (Università degli Studi di Bari Aldo Moro)

Valerio Basile (Università di Torino)

Tommaso Caselli (Rijksuniversiteit Groningen)

Danilo Croce (University of Rome Tor Vergata)

Francesco Cutugno (University of Naples Federico II)

Elisabetta Fersini (University of Milano-Bicocca)

Gloria Gagliardi (University of Bologna)

Elisabetta Ježek (University of Pavia)

Alessandro Lenci (University of Pisa)

Rachele Sprugnoli (University of Parma)

Sara Tonelli (Fondazione Bruno Kessler)

Giulia Venturi (Italian National Research Council)



WE MAKE LAW ACCESSIBLE THROUGH AI.



WANT TO JOIN OUR REVOLUTION?

 **APTUS.AI**
shaping knowledge

almawave.com



ALMAWAVE

300+	400+	35	6	20+	70+
CLIENTI	RISORSE	PARTNER	LABORATORI TECNOLOGICI	SOLUZIONI PROPRIETARIE	IT REPORT

#AI4you | La naturale interazione
con la tecnologia



Talia is a **spin-off company of CNR ILC ('A. Zampolli' Institute of Computational Linguistics) and GruppoMeta; among the first computational linguistic spin off** to work **directly** with commercial clients. It combines the experience of CNR- ILC, the international standard in the development of technologies, products and services for automatic language processing, with the know-how of GruppoMeta, a leading software house.

Talia.cloud is the ecosystem of applications developed by Talia that uses **artificial intelligence** and **automatic language processing** to transform texts into data and knowledge. It provides developers, system integrators, companies and public administration with tools and services for integration with any application: text editors, e-learning platforms, CRM, knowledge discoveries, cataloguing software and document management systems. It enables **automatic analysis** of vast amounts of text -documents, texts, archives- during the management, interpretation, authoring and writing phases.

Talia solutions are delivered with a **SaaS (software as a service)** approach to maximise efficiency, remote access to services, scalability and security, and optimise investments.

Info:

info@talía.cloud <https://talía.cloud/>



Ca' Foscari
University
of Venice

Department of Linguistics
and Comparative Cultural Studies