Preface to the Special Issue *Bridging Theoretical Linguistics and Automated Language Processing: Emerging Synergies and Advances*

Alessandro Lenci*
Università di Pisa

Marco Passarotti**
Università Cattolica del Sacro Cuore

Rachele Sprugnoli†
Università Cattolica del Sacro Cuore

Fabio Tamburini‡
Università di Bologna

In recent years, Natural Language Processing (NLP) has gone through a radical breakthrough due to the success of Large Language Models (LLMs). These models have demonstrated that an unsupervised inductive approach, based on processing large amounts of raw data, can achieve extraordinary accuracy rates in many NLP-related tasks. This progress seems to have reduced the role of theoretical linguistics in computational approaches to linguistic analysis. For instance, metalinguistic annotation, a central element in supervised learning, which traditionally relies on established linguistic theories, has lost its core place in developing trained models for different NLP tasks.

In this context, a closer dialogue between theoretical and computational linguistics becomes essential to integrate theoretical and practical perspectives, bridging the gap between these two approaches to language study. This is particularly important for both the practical implementation of linguistic theories in computational models and the theoretical interpretation of results produced by NLP systems. This special issue of the Italian Journal of Computational Linguistics aims to bridge this gap by exploring the emerging synergies between theoretical linguistics and computational linguistics and the implications these have for the future of the two approaches.

The six articles collected in the issue address various aspects of this interaction, highlighting both the theoretical background and the practical implications of several studies in linguistics. Below, we present a brief summary of the contributions included.

In her paper entitled "Bridging Linguistics and Computational Linguistics: Insights into Synergies and Challenges from a Case Study", Simonetta Montemagni analyzes the evolution of the interactions between theoretical linguistics and computational linguistics, mapping the current state of their synergy and identifying areas for improvement. Through a pilot study, the author demonstrates how linguistic resources and computational modeling can answer long-standing questions about language typology, opening

* Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa - via S. Maria, 36, 56126 Pisa PI. E-mail: alessandro.lenci@unipi.it
** Dipartimento di Scienze Linguistiche e Letterature Straniere, Università Cattolica del Sacro Cuore - Largo Gemelli, 1 - 20123 Milano. E-mail: marco.passarotti@unicatt.it
† Dipartimento di Scienze Linguistiche e Letterature Straniere, Università Cattolica del Sacro Cuore - Largo Gemelli, 1 - 20123 Milano. E-mail: rachele.sprugnoli@unicatt.it
‡ FICLIT, Università di Bologna - via Zamboni, 32. E-mail: fabio.tamburini@unibo.it

up new research perspectives. The importance of fully exploiting the capabilities of computational methods to meet the needs of language research is emphasized.

Arianna Bienati, Mariachiara Pascucci, Jennifer-Carmen Frey and Alessio Palmero Aprosio ("Theoretical implications of automated discourse parsing in student writing") explore the automatic annotation of discourse relations in essays written by Italian students, comparing human annotations with the results of large generative models. The authors discuss theoretical challenges related to connective definition and textual segmentation, showing how computational approaches can illuminate linguistic theories while being guided by them to improve their application.

Gaia Caligiore ("The Development of a Medical Dataset in Italian Sign Language (LIS): Theoretical Considerations and Practical Applications") presents the development of the MultiMedaLIS dataset, a multimodal corpus in the medical domain for Italian Sign Language. The author addresses the theoretical and practical challenges involved in creating processable resources for automatic LIS recognition, highlighting the importance of a multilevel annotation system and multimodal tools to capture the visual-gestural complexity of sign language.

In "Large Language Models Under Evaluation: An Acceptability, Complexity And Coherence Assessment In Italian", Cristiano Chesi, Francesco Vespignani and Roberto Zamparelli evaluate the morphosyntactic and semantic competence of four LLMs in Italian, through tests of acceptability, complexity and coherence. The results show how the more advanced models, while not descriptively or explanatorily adequate, manage to capture some linguistic generalizations that cannot be directly inferred from the primary data, raising questions about stimulus poverty theory.

Martina Di Bratto and Maria Di Maro ("The pragmatic utility of asking the right question in a recommendation scenario") explore the relationship between the semantic-syntactic structure of questions and their pragmatic features, analyzing how these influence the quality of responses in recommendation scenarios. The authors suggest that polarized questions tend to generate more informative responses, while content questions show more variability. The study opens new perspectives for the design of argumentation-based dialogue systems.

Finally, Alice Suozzi, Simone Mazzoli and Gianluca E. Lebani ("Italian-based Large Language Models at the Syntax-Semantics Interface: the Case of Instrumental Role") investigate the capabilities of Italian language models in determining the instrumental role through psycholinguistic experiments. The results show that the performance of the models depends on the semantic selectivity of verbs and the presence of syntactic context, highlighting gaps and potentials of LLMs in dealing with the syntax-semantics interface.

Overall, this special issue highlights how computational linguistics offers theoretical linguistics valuable tools and resources to support theoretical insights with empirical evidence and (replicable) experiments. At the same time, theoretical linguistics remains crucial to computational linguistics, providing explanations of how language works and abstract models that represent linguistic phenomena. This joint work between theory and practice represents a challenge and an opportunity for both approaches, paving the way for new synergies and emerging advances.