

Advancing Medical Communication: Multilingual, Multicultural, and Multimodal Processing for Translation and Simplification

Maria Pia di Buono

University of Naples "L'Orientale"
Via Duomo, 219 - Napoli (Italy)
mpdibuono@unior.it

Abstract

This paper presents the Multilingual, Multicultural, and Multimodal Medical Language Processing (4MLP) Project, funded through a competitive call of the University of Naples "L'Orientale" (Italy).

1 Introduction

Multilingual, Multicultural, and Multimodal Medical Language Processing¹ (4MLP) is a project started in December 2025 and funded through a competitive call of the University of Naples "L'Orientale" (Italy) under the University Research Projects funding scheme (PRA 2025). In addition to the principal investigator (PI), the team consists of one full professor and three doctoral students.

Although the project does not involve a formal consortium, external partners will be engaged through framework agreements. Among those, a recent agreement has been established with Senaso srl, a local social enterprise that provides support to immigrants, including medical escort services. 4MLP advances linguistic research and language technologies in healthcare by developing multilingual, multicultural, and multimodal resources, alongside specialized medical language models. It aims to improve healthcare communication by enhancing patients' access to, understanding of, and engagement with medical information and treatments, ultimately supporting more efficient and patient-centered care.

Effective communication is crucial for successful treatment outcomes, yet misunderstandings, medical jargon, and fragmented information often create barriers. Moreover, as cultural and linguistic factors critically shape doctor–patient communication (Schouten and Meeuwesen, 2006), including differences in explanatory models of health and illness, cultural values, patient preferences regarding doctor–patient relationships, racism and perceptual biases, and linguistic barriers, medical language processing requires approaches that account for multilingual, multicultural, and multimodal dimensions.

Several studies have explored the use of Natural Language Processing (NLP) and large language models (LLMs) to improve healthcare communication, medical information retrieval, and patient comprehension. Reddy (2023) proposed a framework for assessing the translational value of LLMs in healthcare, while Sakakini *et al.* (2020) investigated context-aware simplification of health materials in low-resource settings. Other works focused on specialized medical resources and models, such as LeMe-PT (Simões and Gamallo, 2021), ChatDoctor (Li *et al.*, 2023), and PharmMT (Li *et al.*, 2020), aimed respectively at improving medical language understanding, medical advice generation, and the simplification of prescription instructions.

2 Objectives and Methodology

Building on previous research by the PI and co-authors, including studies on the use of LLMs as virtual assistants for patients in the context of drug administration (Giordano and di Buono, 2024), 4MLP develops responsible NLP-based approaches which combine symbolic knowledge and neural methods to provide clear, reliable, and ac-

cessible healthcare communication.

In line with current research in the field, 4MLP leverages language models for medical information retrieval and text simplification, and is structured around a four-step workflow: (i) development of domain-specific enriched and linked resources, integrating cultural and multilingual dimensions to ensure interpretability and cultural grounding; (ii) subsymbolic modeling including adaptation and fine-tuning of language models and integration of multimodal data for robust representation and flexible processing; (iii) neurosymbolic integration to combine symbolic transparency with subsymbolic adaptability; (iv) evaluation and application to test developed tools in different healthcare communication scenarios, including iterative refinement of models and resources.

3 Preliminary Results

The project is currently in its initial phase, and the results presented here represent a first step in its development. The research initially focuses on Italian and will progressively extend to multilingual and multimodal settings involving languages spoken by migrant communities. Particular attention is devoted to contexts in which institutional languages (e.g., French or Portuguese) co-exist with widely used vehicular languages such as Wolof (Minerba, 2021) or Portuguese Creole.

A first outcome of the project is the Italian Medical Term Simplification (I-MTS) resource, comprising 1,356 Italian term-simplification pairs enriched with semantic and lexical information (di Buono, 2026). Future work will expand both the linguistic coverage and the range of medical documents considered, including heterogeneous healthcare materials and multilingual resources, with the aim of supporting AI-assisted translation into lower-resourced languages and accessible multilingual healthcare applications.

Ethics Statement

The project addresses key ethical considerations, with particular focus on fairness and the mitigation of socio-demographic, religious, and ethnic biases through rigorous evaluation, careful data curation, and synthetic data generation. Human oversight by domain experts supports the prevention of inappropriate outputs and unintended outcomes.

Acknowledgements

This research is funded by the University of Naples "L'Orientale" under the University Research Projects funding scheme (PRA 2025) - CUP C66125002860005.

References

- di Buono, Maria Pia. 2026. Italian medical term simplification: From patient information leaflets to simplified language resources. In *Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC 2026*.
- Giordano, Luca and Maria Pia di Buono. 2024. Large language models as drug information providers for patients. In *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC-COLING 2024*, pages 54–63.
- Li, Jiazhao, Corey Lester, Xinyan Zhao, Yuting Ding, Yun Jiang, and VG Vinod Vydiswaran. 2020. Pharmmt: a neural machine translation approach to simplify prescription directions. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2785–2796.
- Li, Yunxiang, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. 2023. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. *Cureus*, 15(6).
- Minerba, Emiliano. 2021. Wolof language and literature: an introduction. In *Language and Identity Theories and experiences in lexicography and linguistic policies in a global world*, pages 202–216. EUT Edizioni Università di Trieste.
- Reddy, Sandeep. 2023. Evaluating large language models for use in healthcare: A framework for translational value assessment. *Informatics in Medicine Unlocked*, 41:101304.
- Sakakini, Tarek, Jong Yoon Lee, Aditya Duri, Renato FL Azevedo, Victor Sadauskas, Kuangxiao Gu, Suma Bhat, Dan Morrow, James Graumlich, Saqib Walayat, et al. 2020. Context-aware automatic text simplification of health materials in low-resource domains. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 115–126.
- Schouten, Barbara C and Ludwien Meeuwesen. 2006. Cultural differences in medical communication: a review of the literature. *Patient education and counseling*, 64(1-3):21–34.
- Simões, Alberto and Pablo Gamallo. 2021. Lemept: A medical package leaflet corpus for portuguese. In *10th Symposium on Languages, Applications and Technologies (SLATE 2021)*, pages 10–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.