

From Syntax to Semantics: Introducing UMR for NLP Annotation

Adriana S. Pagano¹, Magali Sanches Duran², and Federica Gamba³

¹ Universidade Federal de Minas Gerais, Brazil

² Universidade de São Paulo, Brazil

³ Charles University, Czechia

apagano@ufmg.br magali.duran@gmail.com gamba@ufal.mff.cuni.cz

Abstract

Uniform Meaning Representation (UMR) is a cross-linguistic semantic representation framework designed to encode sentence meaning in a structured and interpretable way. Building on the foundations of Abstract Meaning Representation (AMR), UMR extends semantic coverage to events, participants, semantic roles, temporal/aspectual information, modality, and discourse links. It is language-agnostic and therefore suitable for multilingual exploration.

This tutorial provides a beginner's introduction to UMR aimed at an audience with no prior experience with AMR, UMR, or meaning representations. The tutorial begins with a simple introduction to the essentials of Universal Dependencies (UD) needed to understand how UMR graphs can be constructed from syntactic information. Using simple Portuguese examples, the tutorial illustrates how basic UD structures guide the creation of UMR graphs. Participants will leave with a foundational understanding of what UMR is; how it relates to syntax and semantic roles; how to create minimal UMR graphs, and how Portuguese UD treebanks can support UMR annotation.

Adriana S. Pagano

Universidade Federal de Minas Gerais

Adriana S. Pagano is a Full Professor of Applied Linguistics and Translation Studies at Universidade Federal de Minas Gerais (UFMG). She holds a BA in Translation (UNLP), an MA in English Language and Literature (UFSC), and a PhD in Linguistic and Literary Studies (UFMG). She has led and collaborated on several NLP projects involving translation and post-editing, natural language understanding, and natural language generation. She currently coordinates dependency syntax annotation projects in the healthcare domain and collaborates with annotation initiatives at the Center for Ar-

tificial Intelligence (C4AI). Her research interests include systemic-functional grammar, NLU/NLG, and linguistically informed approaches to annotation.

Magali Sanches Duran

Universidade de São Paulo

Magali Sanches Duran holds a degree in Translation Studies from UNESP (1985), an MBA from FGV-São Paulo (1992), and a Master's (2004) and PhD (2008) in Linguistics from UNESP. She completed multiple postdoctoral research projects at the Núcleo Interinstitucional de Linguística Computacional (NILC/USP-São Carlos), working between 2009 and 2025 on initiatives funded by Microsoft, Samsung, and IBM at the Centro de Inteligência Artificial (C4AI). Her expertise includes syntactic and semantic corpus annotation, sentiment analysis, text complexity metrics, and the development of lexical resources for NLP.

Federica Gamba

Charles University

Federica Gamba is a PhD candidate at the Institute of Formal and Applied Linguistics (UFAL), Charles University, specializing in semantic and syntactic annotation with a focus on Uniform Meaning Representation (UMR) and Universal Dependencies (UD). She has worked on multilingual resource development as a Visiting Researcher at the University of Colorado Boulder and previously as a Research Fellow at the Institute of Computational Linguistics (CNR-ILC) in Pisa. Her background includes work on lexical and textual resources in low-resource and historical languages, supported by advanced training at the University of Pavia, IUSS Pavia, and the Université Paris-Sorbonne.