Towards Uniform Meaning Representation for Brazilian Portuguese: Building a First UMR-Annotated Dataset

Adriana S. Pagano[1], **Magali Duran**[2], **Federica Gamba**[3], **Daniel Zeman**[3]

[1] Faculdade de Letras – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – Brazil

[2]Instituto de Ciências Matemáticas e de Computação
Universidade Federal de São Carlos (UFSCar)
São Carlos – Brazil

[3]Institute of Formal and Applied Linguistics – Charles University
Prague – Czech Republic

apagano@ufmg.br, magali.duran@gmail.com,
{gamba,zeman}@ufal.mff.cuni.cz

Abstract. *This paper presents the first Brazilian Portuguese dataset annotated for Uniform Meaning Representation (UMR). It contains 96 sentences from the Brazilian Portuguese portion of the Parallel Universal Dependencies treebank, parallel to existing UMR annotations in English, Czech, and Italian. The sentences were parsed with PortParser, revised in Arborator-Grew, converted from CoNLL-U into preliminary sentence-level UMR graphs, and manually revised in PENMAN format. The results show that CoNLL-U is a useful starting point, but semantic graph construction requires manual interpretation and language-specific lexical resources. The dataset expands multilingual UMR coverage and supports future Portuguese semantic annotation.*

1. Introduction

As natural language processing moves beyond surface-level linguistic analysis toward deeper forms of meaning representation, semantic annotation has become increasingly important for the development of interpretable, cross-linguistic, and semantically robust language technologies. This need is especially relevant for less represented languages, for which semantically annotated resources remain limited. Brazilian Portuguese is one such case: although it is represented in syntactic resources such as Universal Dependencies [Rademaker et al. 2017, Pardo et al. 2021], there are still few resources that encode sentence-level meaning in a structured graph format.

Uniform Meaning Representation (UMR) offers a framework for building such resources, since it supports cross-linguistic semantic annotation while representing phenomena such as predicate-argument structure, aspect, modality, reference, temporal relations, and reported content [Van Gysel et al. 2021]. Recent work has also emphasized the development of UMR as a broader multilingual infrastructure, including resources and tools for annotation across languages [Bonn et al. 2024].

This paper presents the first UMR-annotated dataset for Brazilian Portuguese. The dataset consists of 96 sentences selected from the Brazilian Portuguese portion of the Parallel Universal Dependencies treebank [Zeman et al. 2017]. These sentences were selected because their counterparts had already been annotated in UMR for English, Czech,

and Italian. As a result, the Brazilian Portuguese annotation can be integrated into an existing multilingual UMR resource rather than standing as an isolated dataset.

The paper makes two main contributions. First, it provides a pilot UMR dataset for Brazilian Portuguese, contributing to the development of semantic resources for Portuguese and to the broader multilingual expansion of UMR. Second, it analyzes the revision process itself, showing which aspects of meaning can be partially derived from CoNLL-U and which require manual semantic interpretation.

The remainder of the paper is organized as follows. Section 2 introduces the UMR framework and related work on multilingual semantic annotation. Section 3 describes the source data and the conversion and revision workflow. Section 4 characterizes the resulting Brazilian Portuguese UMR dataset and analyzes the main types of revision. Section 5 discusses the implications of the dataset for Portuguese semantic annotation and for future UD-to-UMR conversion. Section 6 concludes the paper and outlines next steps.

2. Background and Related Work

Uniform Meaning Representation (UMR) is a graph-based semantic annotation framework designed to support cross-linguistic meaning representation [Van Gysel et al. 2021]. It builds on Abstract Meaning Representation (AMR), originally developed primarily for English [Banarescu et al. 2013], while extending its scope to typologically diverse and under-resourced languages. At the sentence level, UMR represents predicate-argument structure, semantic roles, entity reference, modality, aspect, quantification, and named entities. At the document level, it also provides mechanisms for representing coreference, temporal relations, and modal dependencies [Bonn et al. 2024]. These features make UMR suitable for multilingual semantic annotation, since it provides a shared representational layer while allowing language-specific grammatical and lexical distinctions to be captured.

A central challenge for UMR is the effort demanded by annotation. As with other forms of deep semantic annotation, producing UMR graphs requires linguistic expertise and careful manual work. For this reason, recent work has explored strategies for bootstrapping UMR annotation from already available linguistic resources. One approach is to convert AMR corpora into UMR [Bonn et al. 2023]. This relies on the fact that UMR inherits many representational principles from AMR. However, relying only on AMR-to-UMR conversion limits the development of UMR resources to languages for which AMR annotations already exist. This is problematic for multilingual expansion, especially in the case of less represented languages.

Other initiatives have explored conversion from different types of linguistic resources. [Buchholz et al. 2024] propose bootstrapping UMR from interlinear glossed text, generating preliminary UMR structures for Arapaho from language documentation resources. Another line of work converts semantically richer dependency resources, such as the Prague Dependency Treebank, into UMR [Lopatková et al. 2025]. The tectogrammatical layer of the Prague Dependency Treebank includes deep syntactic-semantic information, which provides richer input for conversion than basic dependency trees. However, this type of resource is not available for most languages.

The UD-to-UMR approach proposed by [Gamba et al. 2025] is relevant to our

work because it aims at a language-independent conversion pipeline. Universal Dependencies (UD) provides broad multilingual coverage and a relatively consistent morphosyntactic annotation scheme across languages [de Marneffe et al. 2021]. Although UD is not a semantic representation, its dependency structures encode information about predicates, dependents, grammatical relations, and morphological features that can be partially mapped to UMR structures. In the UD-to-UMR converter, concept nodes are generally derived from lemmas, while participant roles are assigned through linguistically informed rules. For example, UD relations such as `nsubj`, `csubj`, and `obl:agent` may be mapped to UMR `:actor`, while `obj`, `nsubj:pass`, and `csubj:pass` may be mapped to `:undergoer`. Morphological information can also contribute to semantic interpretation; for instance, dative obliques may be interpreted as recipients.

The resulting graph is not intended to be a complete UMR representation, but rather a structure that annotators can revise and complete. This is important because many UMR distinctions cannot be reliably inferred from syntax alone. Named entity types, abstract predicates, reentrancies, implicit arguments, fine-grained aspectual distinctions, and discourse-level relations require lexical-semantic knowledge or manual interpretation. [Gamba et al. 2025] therefore treat conversion as a way to facilitate annotation rather than as a fully automatic semantic analysis.

3. Methodology

This section describes the construction of the Brazilian Portuguese UMR dataset. The goal was to produce a manually revised semantic dataset while assessing which aspects of UMR annotation can be supported by syntactic input.

The source data consist of 96 sentences selected from the Brazilian Portuguese portion of the Parallel Universal Dependencies treebank (PUD) [Zeman et al. 2017]. These sentences were selected because their corresponding English, Czech, and Italian sentences had already been annotated in UMR. This design makes the Portuguese dataset compatible with an existing multilingual UMR resource and allows the annotation process to benefit from parallel semantic analyses.

The Brazilian Portuguese sentences were first automatically parsed with PortParser ([Lopes and Pardo 2024] and subsequently revised in Arborator-Grew ([Guibon et al. 2020]. This process produced CoNLL-U files containing tokenization, lemmas, part-of-speech tags, morphological features, dependency relations, and syntactic heads. These revised CoNLL-U files served as the morphosyntactic input for the initial conversion into UMR.

The first stage consisted in automatically converting the CoNLL-U sentences into preliminary UMR graphs. The conversion followed the general UD-to-UMR strategy proposed by [Gamba et al. 2025]. In this approach, the converter creates preliminary UMR concept nodes from the lemmas in the CoNLL-U file. Thus, inflected word forms are initially represented through their base forms. Dependency relations and morphological features provide cues for building the initial graph structure -dependency relations provide cues for participant structure, and morphological features contribute to the assignment of semantic attributes where possible.

In the initial graphs, verbal and nominal predicates were represented as concept nodes whenever they could be identified from the syntactic structure. Core syn-

tactic dependents were mapped to preliminary semantic roles. For example, subjects and agents were generally mapped to actor-like roles, while objects and passive subjects were mapped to undergoer-like roles. Other dependents, such as obliques and modifiers, were often assigned broader placeholder relations when a more specific UMR relation could not be reliably inferred from syntax alone.

The conversion also introduced placeholders for categories that require manual interpretation. For instance, named entities could be detected from proper-name structures, but their semantic type, such as person, organization, or place, often required manual correction. Similarly, aspectual values could not always be inferred from UD morphology and were therefore revised manually. This is consistent with the purpose of UD-to-UMR conversion: the converted graph provides a useful starting point, but several semantic distinctions require annotator intervention.

After conversion, the preliminary UMR graphs were manually revised through direct text editing in PENMAN format. This allowed the annotator to inspect and modify the Penman-style graph structure, correct concept labels, normalize semantic roles, resolve placeholder values, and add missing semantic information.

Manual revision focused on five main areas. First, predicate-argument structures were revised: subjects, objects, passive constructions, oblique arguments, control structures, and reported clauses were checked in relation to the predicate involved. We also examined eventive nouns and verbal periphrases, since these constructions often require decisions that go beyond syntactic dependency structure. Eventive nouns were evaluated to determine whether they introduced event concepts in the UMR graph, while verbal periphrases were analyzed to decide which verbs should be represented as semantic predicate nodes and which should instead be captured through aspectual annotation.

Second, named entities were typed and normalized, including the use of `:name` structures and appropriate entity labels. Third, aspect and modality values were revised, since these categories often depend on lexical and constructional interpretation. Fourth, temporal, spatial, and discourse-related relations were checked, especially when the initial graph contained generic placeholder relations. Fifth, parallel UMR annotations in Italian were consulted when useful, while preserving language-specific semantic decisions for Brazilian Portuguese.

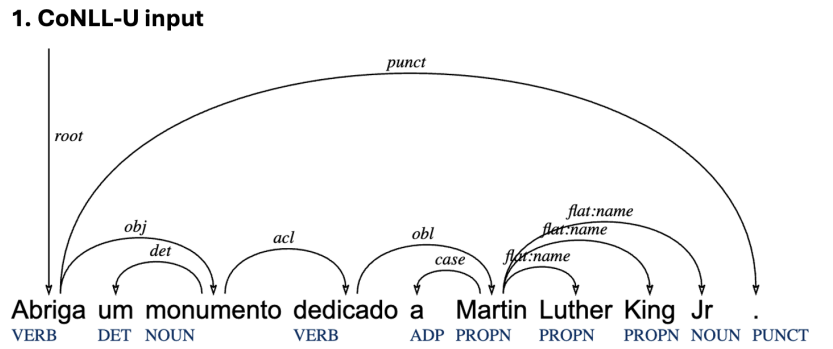
The revision process also included the treatment of implicit or underspecified participants. In some sentences, Portuguese allows arguments to remain unexpressed when they are recoverable from context or from the construction. In these cases, the graph was revised to include the semantic participant when required by the UMR analysis.

A central part of the revision process involved the normalization of predicate-specific argument roles. While the initial conversion produced general participant roles such as `:actor`, `:undergoer`, and related semantic labels, the manually revised graphs were normalized into frame-based `:ARG` roles whenever possible. For this purpose, we used Verbo Brasil [Duran et al. 2013, Duran and Aluísio 2015] as a lexical-semantic resource for Brazilian Portuguese.

For each verbal predicate, we checked whether a corresponding Verbo Brasil frame was available. When a frame existed, its argument structure was used to guide the assignment of UMR roles. When no Verbo Brasil frame was available, or when the avail-

able frame did not adequately cover the sentence, the annotation was completed through manual analysis. In such cases, we considered the predicate meaning, the syntactic structure, and the parallel UMR annotations in Italian. This procedure ensured that the final annotation remained semantically motivated even when lexical resources were incomplete.

Figure 1 illustrates our workflow with the sentence: "Abriga um monumento dedicado a Martin Luther King Jr." The CoNLL-U analysis identifies *Abriga* as the root predicate, *monumento* as its object, and *dedicado a Martin Luther King Jr.* as an adjectival clausal modifier of *monumento*. The automatic conversion yields a preliminary semantic graph, using broad roles such as `:actor`, `:undergoer`, and `:OBLIQUE`. In the manually revised graph, these relations are normalized into predicate-specific roles: *monumento* is analyzed as `:ARG1` of `abrigar`, and, following the Verbo Brasil roleset for *dedicar*, *monumento* is also `:ARG1` of `dedicar`, while *Martin Luther King Jr.* is represented as `:ARG2`. The revision also replaces placeholder values for named entity type, aspect, and modal strength with `person`, `State`, and `FullAff`.



2. Automatically converted UMR graph

```
(s20a / abrigar
  :actor (s20f / FILL
    :refer-number singular)
  :undergoer (s20m / monumento
    :refer-number singular
    :undergoer-of (s20d / dedicar
      :OBLIQUE (s20n / NE-TYPE
        :name (s20n2 / name
          :op1 "Martin"
          :op2 "Luther"
          :op3 "King"
          :op4 "Jr"))
      :modal-strength MS-affirmative
      :aspect ASP))
  :modal-strength full-affirmative
  :aspect ASP)
```

3. Manually revised UMR graph

```
(s20a / abrigar
  :ARG1 (s20m / monumento
    :refer-number singular
    :ARG1-of (s20d / dedicar
      :ARG2 (s20p / person
        :name (s20n / name
          :op1 "Martin"
          :op2 "Luther"
          :op3 "King"
          :op4 "Jr"))
      :modal-strength FullAff
      :aspect State))
  :modal-strength FullAff
  :aspect State)
```

Figure 1. Annotation workflow from CoNLL-U input to automatically converted and manually revised UMR graphs.

To analyze the contribution of manual revision, we compared the initial converted graphs with the final revised graphs. The comparison was qualitative and focused on the main types of changes introduced during revision.

The comparison focused on recurring revision patterns, including: replacement of

general participant roles with predicate-specific :ARG roles; correction of named entity types; resolution of placeholder relations such as generic obliques; addition or removal of participants; normalization of aspect and modality values; restructuring of reported speech and embedded clauses; and treatment of constructions where the syntax-semantics mapping was non-isomorphic. These categories were used to characterize which aspects of the final UMR graphs could be supported by CoNLL-U and which required manual semantic interpretation.

4. Results

This section characterizes the final manually revised Brazilian Portuguese UMR dataset. It reports the main annotation layers, including predicate-specific argument roles, aspect, modality, reported speech, special UMR predicates, and named entities, and briefly comments on the revision patterns that shaped the final graphs.

The first version of the dataset was produced through automatic conversion from CoNLL-U into preliminary UMR graphs. This conversion provided a useful starting point for annotation, since it identified candidate predicates, introduced initial concept nodes, and mapped some syntactic dependents to general semantic roles. In particular, syntactic relations such as subjects, objects, passive subjects, and obliques were converted into broad participant or modifier relations.

The automatically converted graphs thus contained general participant labels, such as :actor, :undergoer, and :theme, as well as placeholder values for categories that require semantic interpretation. These included generic aspect labels, modal-strength values, named entity types, and broad oblique relations.

Table 1 provides a general overview of the final Brazilian Portuguese sentence-level UMR dataset. The corpus contains 96 sentence-level UMR graphs, corresponding to 1,970 word tokens when punctuation is excluded and 2,216 tokens when punctuation is included. The table also summarizes key semantic annotation layers in the dataset, including 559 predicate-specific ARG relations, 11 reported-speech relations encoded with :quot, and 56 occurrences of special UMR predicates. In addition, the corpus contains 123 named entities, 71 of which are linked to external identifiers through :wiki values.

Dataset feature	Count
Sentence-level UMR graphs	96
Word tokens, excluding punctuation	1,970
Word tokens, including punctuation	2,216
Predicate-specific ARG relations	559
Reported-speech relations encoded with :quot	11
Special UMR predicates	56
Named entities	123
Named entities linked through :wiki values	71

Table 1. Overview of the Brazilian Portuguese sentence-level UMR dataset.

To characterize the predicate-argument layer of the corpus, Table 2 reports the distribution of predicate-specific ARG roles after manual revision. Counts combine equiva-

lent direct and inverse realizations of the same role, since both encode the same participant function in the graph.

ARG role	Count	% of ARG relations
ARG0	129	23.1
ARG1	289	51.7
ARG2	112	20.0
ARG3	24	4.3
ARG4	5	0.9
Total	**559**	**100.0**

Table 2. Distribution of predicate-specific argument roles in the Brazilian Portuguese sentence-level UMR dataset. Counts combine direct and inverse realizations of the same ARG role.

The distribution shows that ARG1 is the most frequent role, accounting for more than half of all predicate-specific argument relations. This reflects the high number of undergoer, theme, or affected-participant roles in the corpus. ARG0 and ARG2 are also well represented, while ARG3 and ARG4 are comparatively rare and occur mainly in predicates with more specialized argument structures.

Aspect and modality are central components of UMR annotation and were therefore checked during manual revision. Table 3 presents the distribution of aspectual values in the manually revised dataset. The most frequent values are State and Performance, with 115 occurrences each, followed by Activity with 48 and Process with 18. The remaining aspectual values occur less frequently.

Aspect value	Frequency
State	115
Performance	115
Activity	48
Process	18
Endeavor	2
Generic	2
Habitual	1

Table 3. Aspect values in the manually revised UMR dataset.

The distribution indicates a balance between stative and eventive predications. The high frequency of State reflects descriptions, classifications, and property attributions, while Performance and Activity account for dynamic situations. These distinctions are not directly available in the original CoNLL-U annotation and therefore required manual revision.

Table 4 shows the distribution of modal-strength values. FullAff is by far the most frequent value, with 256 occurrences. Negative, neutral, and partial modal values are less frequent, but their presence shows that the annotation captures variation in polarity and modal commitment.

The concept inventory contains 625 unique concept labels. Table 5 lists the most frequent concepts in the manually revised dataset. The most frequent concept is name,

Modal strength	Frequency
FullAff	256
FullNeg	17
NeutAff	16
PrtAff	7
PrtNeg	3
NeutNeg	2

Table 4. Modal-strength values in the manually revised UMR dataset.

with 124 occurrences, followed by person, with 81 occurrences. This reflects the high number of named entities and named individuals in the PUD sentences.

Concept	Frequency
name	124
person	81
and	43
date-entity	28
identity-91	20
country	17
have-mod-91	10
thing	10
human-settlement	10
have-role-91	9
after	7
outro	7
event	7
grande	6
muito	6
esse	6
have-rel-role-92	6
include-91	5
ethnic-group	5
company	5

Table 5. Twenty most frequent concepts in the manually revised UMR dataset.

The frequent occurrence of abstract predicates such as identity-91, have-mod-91, have-role-91, have-rel-role-92, and include-91 shows that the revised annotation introduces abstract semantic structures required by UMR.

The main effect of manual revision was the transformation of a general UD-derived graph backbone into a predicate-specific semantic representation. General participant labels introduced during conversion were replaced with :ARG roles whenever possible, guided by Verbo Brasil and by manual analysis of predicate meaning. This is reflected in the high frequency of :ARG0, :ARG1, :ARG2, and :ARG3 in the final dataset.

Manual revision also resolved placeholder values introduced during conversion. Generic aspect labels were replaced by specific aspectual values such as State,

Performance, Activity, and Process. Generic modal labels were replaced by values such as FullAff, FullNeg, NeutAff, and PrtAff. Named entities were typed, :name structures were checked, and :wiki links were added when appropriate.

A further effect of revision was the replacement of broad syntactic-semantic relations with more specific UMR relations. Generic oblique relations were revised as :temporal, :place, :manner, :purpose, :reason, :condition, :source, or :goal, depending on the meaning of the construction. Reported content was also manually revised through :quot structures where required.

Overall, these results show that CoNLL-U provides a useful starting point for UMR annotation, but that the final Brazilian Portuguese dataset required manual semantic analysis and language-specific lexical resources. The final graphs are therefore not merely converted syntactic structures, but semantically enriched UMR representations.

5. Discussion

The Brazilian Portuguese UMR epresents a first step toward the development of deeper semantic graph resources for Portuguese. The manually revised graphs encode predicate-argument structure, reference, aspect, modal strength, named entities, temporal and spatial relations, reported content, and abstract predicates, confirming the usefulness of UMR for representing semantic phenomena that go beyond the morphosyntactic information available in CoNLL-U.

The results show that automatic UD-to-UMR conversion provides a useful annotation starting point, but the final graphs required substantial revision because the mapping from syntax to semantics is often non-isomorphic: syntactic subjects, objects, and obliques do not always correspond directly to predicate-specific semantic roles.

A central revision step was the normalization of predicate-argument structure. General roles such as :actor, :undergoer, and :theme were replaced with predicate-specific :ARG roles whenever possible, supported by Verbo Brasil. Other layers, including aspect, modality, named entity typing, :wiki links, abstract predicates, and relations such as :temporal, :place, :manner, and :purpose, also required manual interpretation beyond dependency labels.

Overall, the workflow combines CoNLL-U input, automatic conversion, manual revision, and language-specific lexical-semantic resources. It offers a practical model for developing UMR datasets for languages with syntactic resources but limited semantic graph annotation.

6. Conclusion

This paper presented the first UMR-annotated dataset for Brazilian Portuguese: 96 sentence-level graphs from the Brazilian Portuguese portion of the Parallel Universal Dependencies treebank, manually revised after automatic UD-to-UMR conversion. Because the sentences are parallel to existing UMR annotations in English, Czech, and Italian, the dataset contributes to a broader multilingual UMR resource.

Future work will expand the dataset, refine the use of Brazilian Portuguese lexical resources, compare the Portuguese graphs more systematically with their English, Czech,

and Italian counterparts, improve conversion rules for Portuguese-specific constructions, and release the dataset as part of a multilingual UMR infrastructure.

6.1. Limitations

This study is limited by the small size of the dataset, which contains 96 sentence-level graphs. The annotations were carried out collaboratively by the authors, rather than through a blind procedure, due to the incipient stage of UMR annotation for Brazilian Portuguese and the lack of language-specific guidelines. The dataset is therefore best understood as a pilot resource and a first step toward larger-scale Brazilian Portuguese UMR annotation.

7. Acknowledgments

The authors express their gratitude to the anonymous reviewers for their constructive feedback. Adriana Pagano's research is funded by FAPEMIG and the National Council for Scientific and Technological Development (CNPq, grants 404722/2024-5; 313103/2021-6). Magali Sanches Duran is a member of the Artificial Intelligence Center of the University of São Paulo, funded by the São Paulo Research Foundation (FAPESP grant no. 2019/07665-4) and IBM. She is also supported by the Ministry of Science, Technology and Innovations, with resources from Law No. 8,248 of October 23, 1991, within the scope of PPI-Softex, coordinated by Softex and published as ICT Residency 13, DOU 01245.010222/2022-44. This work received support from the CA21167 COST action UniDive, funded by COST (European Cooperation in Science and Technology) and is conducted under the auspices of the National Institute of Science and Technology in Responsible Artificial Intelligence for Computational Linguistics, Information Treatment, and Dissemination (INCT-TILDIAR), funded by CNPq (grant 408490/2024-1).

8. Ethics disclosure

This work uses sentences from the publicly available Parallel Universal Dependencies treebank. The dataset does not contain private, sensitive, or newly collected personal data. The Brazilian Portuguese UMR graphs were produced for research purposes through automatic conversion and manual revision. The resulting resource is intended to support linguistic research and the development of multilingual semantic annotation, and will be released with appropriate attribution to the source data and in accordance with the relevant dataset licenses.

9. Generative AI use

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript.

References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract Meaning Representation for sembanking. In Pareja-Lora, A., Liakata, M., and Dipper, S., editors, *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Bonn, J., Buchholz, M. J., Chun, J., Cowell, A., Croft, W., Denk, L., Ge, S., Hajic, J., Lai, K., Martin, J. H., et al. (2024). Building a broad infrastructure for uniform meaning representations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2537–2547.

Bonn, J., Myers, S., E. L. Van Gysel, J., Denk, L., Vigus, M., Zhao, J., Cowell, A., Croft, W., Hajič, J., H. Martin, J., Palmer, A., Palmer, M., Pustejovsky, J., Urešová, Z., Vallejos, R., and Xue, N. (2023). Mapping AMR to UMR: Resources for adapting existing corpora for cross-lingual compatibility. In Dakota, D., Evang, K., Kübler, S., and Levin, L., editors, *Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023)*, pages 74–95, Washington, D.C. Association for Computational Linguistics.

Buchholz, M. J., Bonn, J., Post, C. B., Cowell, A., and Palmer, A. (2024). Bootstrapping UMR annotations for Arapaho from language documentation resources. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2447–2457, Torino, Italia. ELRA and ICCL.

de Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Duran, M. and Aluísio, S. (2015). Automatic generation of a lexical resource to support semantic role labeling in Portuguese. In Palmer, M., Boleda, G., and Rosso, P., editors, *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 216–221, Denver, Colorado, USA. Association for Computational Linguistics.

Duran, M. S., Martins, J. P., and Aluísio, S. M. (2013). Um repositório de verbos para a anotação de papéis semânticos disponível na web (a verb repository for semantic role labeling bavailable in the web) [in Portuguese]. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.

Gamba, F., Palmer, A., and Zeman, D. (2025). Bootstrapping UMRs from Universal Dependencies for scalable multilingual annotation. In Peng, S. and Rehbein, I., editors, *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, pages 126–136, Vienna, Austria. Association for Computational Linguistics.

Guibon, G., Courtin, M., Gerdes, K., and Guillaume, B. (2020). When collaborative treebank curation meets graph grammars. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5291–5300.

Lopatková, M., Hledíková, H., Štěpánek, J., and Zeman, D. (2025). From the prague dependency treebank to the uniform meaning representation: Gold-standard czech umr data and partial automatic conversion.

Lopes, L. and Pardo, T. (2024). Towards portparser - a highly accurate parsing system for Brazilian Portuguese following the Universal Dependencies framework. In Gamallo, P., Claro, D., Teixeira, A., Real, L., Garcia, M., Oliveira, H. G., and Amaro, R., editors, *Proceedings of the 16th International Conference on Computational Processing*

of Portuguese - Vol. 1, pages 401–410, Santiago de Compostela, Galicia/Spain. Association for Computational Lingustics.

Pardo, T. A. S., Duran, M. S., Lopes, L., Di Felippo, A., Roman, N. T., and Nunes, M. d. G. V. (2021). Porttinari: a large multi-genre treebank for brazilian portuguese. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, Porto Alegre, RS, Brasil. SBC.

Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., and de Paiva, V. (2017). Universal dependencies for portuguese. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 197–206.

Van Gysel, J. E., Vigus, M., Chun, J., Lai, K., Moeller, S., Yao, J., O'Gorman, T., Cowell, A., Croft, W., Huang, C.-R., et al. (2021). Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3):343–360.

Zeman, D., Popel, M., Straka, M., Hajič, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinková, S., Hajič jr., J., Hlaváčová, J., Kettnerová, V., Urešová, Z., Kanerva, J., Ojala, S., Missilä, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., de Paiva, V., Droganova, K., Martínez Alonso, H., Çöltekin, Ç., Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadová, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonça, G., Lando, T., Nitisaroj, R., and Li, J. (2017). CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In Hajič, J. and Zeman, D., editors, *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.