Empirical Study Towards RBAMR-News: a Rule-Based AMR Parser for News in Portuguese

Maria Julia Bernardo Comarim[1,2], **Ariani Di-Felippo**[1,2,3]

[1]Interinstitutional Center for Computational Linguistics (NILC/USP)

[2]Graduate Program in Linguistics - Federal University of São Carlos (PPGL/UFSCar)

[3]Language and Literature Department - (DL/UFSCar)

mjbcomarim@estudante.ufscar.br, ariani@ufscar.br

Abstract. *This work evaluates AMR graph construction rules, originally developed for literary texts, in Portuguese news data to support building a symbolic AMR parser for this genre. Using the AMRNews-PT gold-standard corpus, we conduct an ablation analysis to assess rule contributions to AMR representation quality. The results reveal a hierarchical organization in which a small set of core structural rules drives most parsing performance, while additional rules provide finer semantic refinements. This rule hierarchy serves as the basis for building the parser and enabling the annotation of a larger news corpus.*

1. Introduction

Semantic representation refers to structured models of natural language meaning. This capability is essential to enhance machine understanding of human language [Jurafsky and Martin, 2026]. Among deep semantic formalisms, Abstract Meaning Representation (AMR) [Banarescu et al., 2013] emerges as a versatile framework, proving effective in diverse applications such as machine translation, summarization, and information extraction [Wein and Opitz, 2024]. The success of AMR in Natural Language Processing (NLP) can be attributed to three main factors: (i) its abstract, graph-based semantic representation that abstracts away from surface syntax; (ii) its standardized annotation scheme with clear guidelines[1] ensuring consistency and enabling gold-standard corpora; and (iii) its grounding in PropBank frame semantics [Palmer et al., 2005], providing a stable lexical-semantic foundation [Sadeddine et al., 2024; Mansouri, 2025].

The most important corpus is the AMR Annotation Release [Banarescu et al., 2013]. It was constructed fully manually by trained annotators, and contains about 60,000 sentences in its latest (3.0) version. This resource has enabled the development of monolingual[2] AMR parsing across multiple modeling paradigms, fostering state-of-the-art (SOTA) transformer-based models [Bevilacqua et al., 2021; Zhou et al., 2021; Bai et al., 2022; Vasylenko et al., 2023] and competitive approaches based on parameter-efficient strategies [Ho, 2025]. However, these approaches remain highly data-dependent, relying heavily on large annotated corpora such as AMR 3.0 for training and performance gains. This limits their applicability to low-resource settings such as Brazilian Portuguese (BP), whose largest available corpus, AMR-LittlePrince [Anchiêta and Pardo, 2018], comprises around 1.5k sentences and is restricted to a specific literary genre (allegorical narrative).

[1]https://github.com/amrisi/amr-guidelines/blob/master/amr.md

[2]This work focuses exclusively on monolingual AMR parsing, enabling direct modeling of language-specific phenomena without reliance on cross-lingual transfer from English resources.

To address this gap, we focus on approaches that provide stronger structural control while reducing dependence on large annotated corpora, aiming to support reliable AMR parsing in under-resourced languages. Although large language models (LLMs) with in-context learning (ICL) strategies have recently emerged as a promising alternative for low-resource AMR parsing, results suggest that they are not yet reliable parsers [Ettinger et al., 2023; Li and Fowlie, 2025]. Consequently, symbolic approaches remain relevant due to their interpretability and explicit structural modeling, particularly in settings such as Portuguese news AMR parsing.

Thus, we aim to develop RBAMR-News, a symbolic AMR parsing method for the Portuguese news genre. To this end, we take RBAMR (which stands for "Rule-Based AMR Parsing") [Anchiêta and Pardo, 2018] as a starting point. It is a pipeline-based architecture composed of syntactic parsing, semantic role labeling (SRL), and rule-based AMR generation modules. This symbolic method was originally built on the AMR-LittlePrince corpus and has not previously been evaluated on the news genre. As the original parser is no longer executable, we reconstruct its pipeline scheme using recently available tools employed independently, without pipeline-level integration. This reconstruction enables a broader four-stage framework for developing RBAMR-News: (i) evaluation of the original RBAMR rules on news text, (ii) coverage expansion through the investigation of genre-specific phenomena and the proposal of new rules from a larger corpus annotated with gold-standard UD and PropBank semantic roles, (iii) iterative diagnostic projection and rule refinement, and (iv) final validation of the resulting rule hierarchy.

In the present paper, we focus on the reconstruction of RBAMR pipeline scheme and on the first stage of the proposed framework, assessing the transferability of its original rule inventory to the news domain through an ablation-based analysis on AMRNews-PT [Sobrevilla Cabezudo and Pardo, 2019], the only news corpus with manual AMR annotation, containing 870 sentences. The remainder of this article is structured as follows. Section 2 briefly introduces AMR fundamentals, while Section 3 provides the main related work. Section 4 presents the reconstruction of the RBAMR parser, detailing the adaptation of each component, especially the original rules, to the current setup. Section 5 reports the experimental design, including the dataset and evaluation metrics. Section 6 summarizes the main findings and outlines directions for future work.

2. Abstract Meaning Representation

AMR is a deep semantic formalism characterized by graph-based representation, abstractness, and uniformity. It represents sentence meaning as rooted, directed, acyclic graphs. (Figure 1(c))[3]. Nodes of an AMR graph are concepts, and edges are relationships between those concepts. AMR concepts are either English lexical items (e.g., "boy"), PropBank frames [Palmer et al., 2005] (e.g., *desire-01*), or specialized keywords, which include special named entity types, quantities, and logical operators. Leaf nodes are labeled with semantic concepts, such that *(b / boy)* denotes an instance of the concept *boy*, where the variable *b* serves as the identifier of the node in the graph. For example, in *(d / desire-01 :ARG0 (b / boy))*, the variable *d* introduces an event node instantiating the PropBank frame *desire-01*, which is linked via the *:ARG0* edge to the variable *b*, corresponding to the entity *boy*. This structure encodes that *b* fills the *:ARG0* role (wanter) of the *desire-*

[3]The PENMAN notation (Figure 1(b)) is used as a linearization format for human reading and writing.

01 event, establishing a semantic relation between event and participant. The roleset is derived from the PropBank frame repository [Palmer et al., 2005]. In the case, the frame *desire-01* has a roleset with two pre-defined arguments (:arg0=wanter and :arg1=wanted), which are filled by *boy* and *believe-01* in the sentences shown in Figure 1.

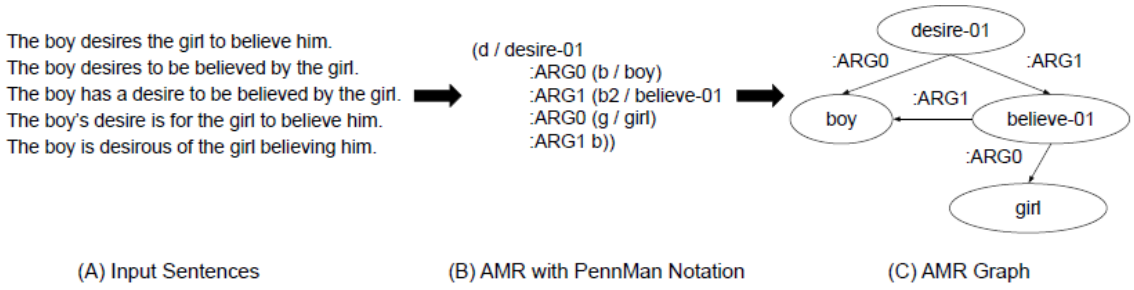

Figure 1. Example of AMR formalism [Mansouri 2025].

Edges show the relations between nodes in the graph, supporting compositional meaning construction. When a single participant fulfills multiple semantic roles, AMR encodes this via reentrancy (nodes with multiple parents) or variable re-use in PENMAN notation. The AMR relation inventory comprises approximately one hundred semantic relations, grouped into functional classes: (i) frame arguments (PropBank roles), (ii) general relations, (iii) quantities, (iv) date-entities, and (v) relations for lists. AMR also defines inverse relations (e.g., *:location-of*), which allow reverse role interpretation within the graph structure. Additionally, each relation admits a reified counterpart, enabling the relation itself to be treated as a first-class semantic entity; for instance, *:location* can be reified as *be-located-at-91* [Banarescu et al., 2013].

Another principle is to abstract away from syntactic structure, meaning that sentences with different word orders or syntactic constructions conveying the same meaning will have the same AMR (Figure 1(a)). This abstraction facilitates better generalization for downstream tasks. Note that functional words like “as” or “it” are not annotated, as they are considered not to contribute additional semantic content. AMR also aims to provide a uniform, consistent semantic representation for different linguistic constructions. For example, both active and passive voice sentences have the same AMR as long as they convey the same meaning, regardless of differences in syntactic structure.

AMR annotation follows a set of guidelines originally introduced by Banarescu et al. [2013] and subsequently refined to incorporate a broader range of linguistic phenomena, and increase their application for downstream tasks. Efforts to develop annotation guidelines for AMR have extended to several non-English languages. For Portuguese, these efforts adapt the model to language-specific phenomena such as null subjects, diminutives, possessive pronoun ambiguity, light verbs, and idiomatic expressions [Inácio et al., 2023], as well as to the specificity of stock market tweets [Ceregatto, 2025]. In the following section, we present the reengineering of RBAMR, focusing mainly on the adaptation of its components to evaluate their suitability for the news genre.

3. Related Work

The foundational AMR work introduced in 2013 [Banarescu et al., 2013] established the annotation guidelines and graph-based semantic representation framework, leading to manually annotated corpora (including *The Little Prince*) and the subsequent release of

large-scale English AMR datasets by the Linguistic Data Consortium[4] (LDC). The LDC has released three successive versions of AMR annotations and datasets, with each release incorporating updates and refinements over previous versions. These datasets are AMR 1.0, 2.0 e 3.0, with 13,051, 39,260 and 59,255 sentences, respectively. AMR 3.0 is the most comprehensive, with the largest number of sentences and the most refined and consistent annotations, including natural language sentences from broadcast conversations, newswires, weblogs, web discussion forums, fiction, and web text.

These resources have enabled the development of monolingual parsers across a variety of NLP paradigms, with Transformer-based approaches driving the most significant advances [Mansouri, 2025]. Recent systems rely on full fine-tuning of seq-to-seq models such as BART, formulating AMR parsing as text-to-graph generation via graph linearization, and have substantially improved Smatch performance [Bevilacqua et al., 2021; Zhou et al., 2021; Bai et al., 2022; Vasylenko et al., 2023]. For example, LeakDistill [Vasylenko et al., 2023] achieves Smatch scores of 0.861 on AMR 2.0 and 0.856 on AMR 3.0, positioning it as a current SOTA AMR parser. More recently, parameter-efficient fine-tuning (PEFT) methods such as LoRA have been applied to large language models (such as Llama 3.2) for AMR parsing, reducing adaptation costs while maintaining competitive performance (0.80) [Ho, 2025]. In parallel, researchers have begun investigating the ability of LLM to handle AMR parsing through ICL [Ettinger et al., 2023; Li and Fowlie, 2025]. The results indicate that, despite generating well-formed and partially correct AMR structures, LLMs are still not reliable out-of-the-box AMR parsers, with maximum Smatch scores reaching only around 0.60 with GPT-4o in a 5-shot setting.

Given this, top-performance approaches rely on large annotated corpora and high computational cost, and they inherit limitations from linearization-based strategies, particularly in representing reentrancies and long-distance dependencies. These constraints are even more critical in low-resource languages, where the scarcity of annotated data further limits the applicability of large-scale neural models. AMR resources[5] for Portuguese remain limited in size and domain coverage (Table 1), with the largest corpus, AMR-LittlePrince [Anchiêta and Pardo, 2022] (1,527 sentences), restricted to literary text. All the corpora were manually or semi-automatically (with human revision) annotated based on Verbo-Brasil[6] [Duran and Aluísio, 2012; Sanches Duran and Aluísio, 2015] - a repository of verbs and frames inspired by the English PropBank.

Table 1. Corpora with AMR annotation in Portuguese.

Corpus	Genre	Size	Reference
AMR-LittlePrince	Literature	1,527	Anchieta and Pardo (2018a)
DANTESTocks-AMR	Tweet	1,128	Ceregatto and Di-Felippo (2025)
AMRNews-PT	News	870	Sobrevilla Cabezudo and Pardo (2019)
OpiSums-PT-AMR	Online review	404	Inácio et al. (2023)
AMRScien-Br-Corpus	Popular science	200	Seno et al. (2022)

This scarcity of annotated resources is also reflected in the limited availability of AMR parsers. For Portuguese, three main parsers have been evaluated on the manually anno-

[4]https://catalog.ldc.upenn.edu/

[5]https://github.com/nilc-nlp/AMR-BP

[6]http://143.107.183.175:21380/verbobrasil/

tated AMR-LittlePrince corpus. RBAMR is a rule-based parser that leverages syntactic and semantic information (SRL) to construct AMR graphs. Its rules were derived from AMR-LittlePrince, and when tested on the same dataset, it achieved the best results, with an average Smatch F-score of 0.535. Later, with the addition of new rules and a pruning method, its average Smatch increased to 0.57. CAMR [Wang et al., 2015] and AMREager [Damonte et al., 2017] are transition-based parsers trained and tested on the same corpus, adapted to Portuguese with pre-processing tools such as Stanza, LX-Parser, and spaCy, as well as translated lexical resources and pretrained Portuguese embeddings [Hartmann et al., 2017]. CAMR converts dependency trees into AMR graphs through aligned transitions and reached an average Smatch F-score of 0.455, while AMREager processes sentences incrementally from left to right and obtained 0.445.

Thus, advancing AMR parsing for Portuguese requires not only improvements in parsing methods, but also a substantial expansion of annotated corpora. In its low-resource setting, rule-based approaches remain relevant for the semi-automatic construction of gold-standard AMR corpora, as they provide controlled and transparent initial representations that facilitate manual revision and improve annotation consistency. Hence, we are currently developing the RBAMR-News parser to the news genre, which is underrepresented in the AMR resources (Table 1). Here, we investigate the portability of the pioneering RBAMR parsing rules to news texts, aiming to assess its adequacy for supporting semi-automatic AMR corpus construction in this genre. In the next section we introduce AMR and explain its fundamental concepts and structures.

4. RBAMR Reengineering

The RBAMR parser was implemented as a pipeline architecture composed of three modules: (i) syntactic parsing (Palavras [Bick, 2000]), (ii) SRL (Brazilis [Hartmann et al., 2016] based on [Alva-Manchego and Rosa, 2012]), and (iii) rule-based generation. Since the parser relies on outdated libraries and legacy dependencies, the system had to be reengineered to support this experiment. To preserve the original architecture as closely as possible, we replaced each module with contemporary tools providing analogous functionality, although without fully restoring the original end-to-end integration (Figure 2).

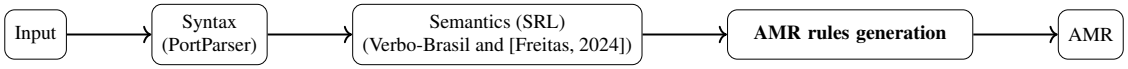

Figure 2. Pipeline-inspired architecture for rule evaluation.

For the syntactic module, we used PortParser v.2 [Lopes and Pardo, 2024]. Built upon the LatinPipe architecture [Straka et al., 2024], the parser achieves the best performance for Portuguese under the Universal Dependencies (UD) framework [de Marneffe et al., 2021], reaching 97.17% UAS (Unlabeled Attachment Score) and 96.19% LAS (Labeled Attachment Score). It was trained on the Porttinari-base v2 corpus[7] using BERTimbau-*large* [Souza et al., 2020] (*bert-base-portuguese-cased*) embeddings. The output of the parser consists of sentences in CoNLL-U format, containing token-level morphological and syntactic dependency annotations compliant with the UD formalism.

[7]https://sites.google.com/icmc.usp.br/poetisa/porttinari-2-1

For the SRL component, we consider BERTimbau (large) trained on PropBank.Br[8] (first version - v1.1)[9], since it is among the best-performing SRL models for Portuguese (see Oliveira et al. [2021]). However, the model requires predicate positions to be explicitly provided as input and produces argument annotations as flat BIO label sequences. Both characteristics pose practical limitations for the experimental setting and would require substantial architectural reformulation for integration into our approach.

Therefore, a heuristic-based symbolic SRL module was developed to operate directly on UD annotations by first identifying verbal predicates and their syntactic arguments from dependency relations such as *nsubj*, *obj*, and *obl*. Next, the module consults Verbo-Brasil to select the most compatible roleset according to the extracted verbal valency. Semantic roles are assigned by mapping UD dependency relations to semantic arguments, with *nsubj* and *obj* generally mapped to *Arg0* and *Arg1*, respectively, accounting for passive constructions Freitas [2024]. The module also handles implicit subjects in subordinate and coordinated clauses through recursive dependency-tree traversal. When a predicate is not found in Verbo-Brasil, a deterministic fallback strategy generates a generic roleset (*lemma.01*), preserving the continuity of the semantic representation. In the end, the SRL returns the predicates with their respective arguments. The SRL module was evaluated on the Porttinari-Base PropBank (PBP) corpus[10], comprising 13,395 verbal predicates annotated with semantic roles. Our module achieved an F1-score of 0.413. By comparison, the supervised SRL system originally employed by RBAMR obtained an F1-score of 0.673 on PropBank [Alva-Manchego and Rosa, 2012]. Although this provides a useful point of reference, the two scores are not directly comparable because they were obtained on different corpora and under different evaluation protocols.

Particularly relevant to this work are the handcrafted rules derived from the manual annotation of the AMR-LittlePrince corpus, which are responsible for constructing the final AMR graph from the intermediate representations. The adoption of the PortParser required the reformulation of the rules, adapting surface-based patterns into syntactic dependency structures compatible with its output. Both the original rules and their adapted versions used in the experiment are presented in Table 2. We exemplify the adaptation with the direct object rule. In the original parser, direct objects marked with the ACC (accusative) tag by PALAVRAS were mapped to the AMR relation *:ARG1*. In our experiment, we adapted this rule so that the UD dependency relation *obj* is mapped to *:ARG1*.

The rules can be broadly organized into four main categories according to the linguistic level they operate on. Lexical and morphosyntactic rules operate over lexical units, POS tags, and morphological features to support AMR concept generation, auxiliary detection, implicit subject recovery, and functional word filtering. Named-entity rule generates standardized AMR subgraphs for entities through canonical *:name* structures and their corresponding *:opx*[11] children. Argument-structure rules are driven by the output of SRL, mapping Verbo-Brasil semantic roles to core AMR relations (e.g., *:ARG0*, *:ARG1*)

[8]PropBank.Br is the (Brazilian) Portuguese version of PropBank and it follows the same annotation style as the original, having a similar role set [Duran and Aluísio, 2012]

[9]It contains 5,931 manually annotated instances covering 3,348 sentences from the Bosque Treebank [Afonso et al., 2002], which is derived from the 1994 *Folha de São Paulo* newspaper.

[10]https://sites.google.com/icmc.usp.br/poetisa/porttinari-base-propbank

[11]The named entity "Nobel Prize", for example, is represented through a `:name` structure whose ordered tokens are encoded by the relations *:op1* "Nobel" and *:op2* "Prize".

while also resolving reentrancies and shared arguments. Finally, syntactic-semantic rules transform dependency relations into AMR edges, handling phenomena such as modification, negation, quantification, and copular inversions. For adverbial modification, we employ a lexical heuristic to refine the generic UD relation *advmod* produced by PortParser into more semantically specific AMR relations, such as *:time*, *:degree*, and *:manner*.

Table 2. Original rules and adaptations.

Phenomenon	**Original Rules**	**Adaptations**
Lexical and Dictionary-Based Rules		
Auxiliary detection	PALAVRAS (AUX) labels	UD AUX + Hartmann's rules
Subject insertion	Morphology (Person/Number)	Person/Number→virtual pronoun
Ordinals	Lexical ordinal rules	ADJ+°/ª→ordinal-entity
Stopword filtering	Functional word list	Stopword list filtering
Named Entity Rule	PROPN+Gazetteers (Gaz.)	PROPN+Gaz.→named AMR node
Argument Structure Rules		
Subject	SUBJ → :ARG0	nsubj → :ARG0
Passive subject	v-pcp + SUBJ → :ARG1	nsubj:pass → :ARG1
Direct object	ACC → :ARG1	obj → :ARG1
Indirect object	DAT → :ARG2	iobj → :ARG2
Open clausal complement	Infinitival (IV) rules	xcomp → :ARG2
Nominal modifier	Nominal dependency rules	nmod → :ARG1
Syntactic-Semantic Rules		
Adjectival modifier	ADJ attachment rules	amod → :mod
Adverbial modifier	ADV rules	advmod → :time, :degree, :manner
Coordination	Coordination (KC) → :op	cc/conj → :op
Quantification	NUM attachment rules	nummod → :quant
Possession	Pronoun/Genitive → :poss	nmod:poss → :poss
Negation	ADV(*não/jamais*)→:polarity -	*não/nunca*→ :polarity -
Obligation modality	Aux (*dever*)→ obligate-01	*ter/dever/haver*→ :obligate
Possibility modality	Aux (*poder*)→possible-01	*poder*→ :possible
Passive voice	Pattern *ser*+v-pcp	aux:pass
Copular constructions	*ser/estar*→be-01	*ser/estar*→be-03
Contrast	Discourse (*mas/porém*)	*mas/porém/todavia*→:contrast
Questions	Punctuation (?)+interrogatives	Interrogative sent. heuristics

5. RBAMR Rules Evaluation on News Texts

The RBAMR rules were evaluated using the AMRNews-PT corpus [Sobrevilla Cabezudo and Pardo, 2019] (see Table 1), whose texts were extracted from different sections of the Folha de São Paulo news agency, and also from PropBank.Br. The corpus contains 870 manually annotated sentences, with length limited to up to 23 tokens[12]. The sentences were annotated based on the original English AMR guidelines with adaptations for Portuguese and using Verbo-Brasil. The annotation was carried out by trained annotators from Computer Science and Linguistics. The process involved four steps: (i) identification of sentence type to handle complex structures (e.g., coordination and subordination); (ii) concept identification, (iii) semantic role assignment, and (iv) AMR graph construction using the AMR Editor tool [Hermjakob, 2013][13]. Using a set of 50 sentences, the overall agreement for AMRNews-PT achieved a Smatch [Cai and Knight, 2013] value of 0.73, which is a good value, considering that the IAA in the original AMR project ranged between 0.70 and 0.80.

[12]Due to the complexity of annotation, the corpus includes only sentences shorter than the average length.

[13]AMR Editor is no longer available and has since been replaced by metAMoRphosED [Heinecke, 2023]

The evaluation is conducted at two complementary levels. At the rule level, we first assess the relevance of the symbolic AMR rules in the target journalistic corpus using support, defined as the mean number of rule applications per sentence. Unlike the classical definition of support in association rule mining, which represents a relative frequency bounded by 1 [Agrawal and Srikant, 1994], this metric measures rule application frequency and may therefore exceed 1 when a rule is applied multiple times within the same sentence.

This is particularly important for evaluating the transferability of rules originally designed for a different genre. We then assess rule performance using precision, recall, and F1-score, following standard definitions in information extraction [Jurafsky and Martin, 2026]. Precision is defined as the proportion of correct rule applications among all triggered instances (TP/(TP+FP)), while recall measures the proportion of correctly identified instances among all relevant cases in the corpus (TP/(TP+FN)). The F1-score corresponds to the harmonic mean of precision and recall, providing a balanced measure between correctness and coverage. This stage enables a fine-grained analysis of both the accuracy and coverage of each rule in capturing relevant linguistic patterns. At the (pipeline-inspired) method level, we evaluate the quality of the resulting AMR structures using Smatch, the standard metric for AMR graph comparison [Cai and Knight, 2013], which computes the F-score over concept and relation overlaps between generated and gold-standard graphs. To further assess the contribution of individual rules to the overall system performance, we introduce an ablation-based analysis, which is described in detail in the following subsection.

We apply an ablation-based evaluation [Meyes et al., 2019] to quantify the impact of each symbolic AMR rule on the overall pipeline-inspired parsing, identifying the rules most relevant for semantic representation and cross-genre generalization. The procedure consists of iteratively removing one rule at a time from the system while keeping all other components unchanged. After each removal, the full AMR parsing "pipeline" is re-executed on the evaluation corpus. Performance is then measured using Smatch at the system level, as well as precision and recall where applicable. To quantify the contribution of each rule, We compute the performance delta for each ablation as:

$$\Delta S = S_{\text{ablated}} - S_{\text{full}}, \quad \Delta P = P_{\text{ablated}} - P_{\text{full}}, \quad \Delta R = R_{\text{ablated}} - R_{\text{full}}$$

The equations represent the performance variation induced by each ablation setting. Specifically, ΔS, ΔP, and ΔR correspond to the differences in Smatch, precision, and recall, respectively, between the ablated version (ablated) and the complete system (full). Positive values indicate that removing a module improves performance, suggesting it is detrimental to the system, while negative values indicate that removing it degrades performance, suggesting it is beneficial.

For the evaluation of the rules, support was calculated over the entire corpus, while the ablation experiments were conducted on a randomly selected subset of 435 sentences. This decision was motivated by the empirical stability observed in preliminary experiments, suggesting that the subset provided a representative approximation of the parser's overall behavior, while also reducing processing time and facilitating iterative inspection of the linguistic heuristics. Additionally, some ablation settings produced structurally inconsistent graphs, making large-scale one-to-one comparisons for precision, recall, and Smatch calculations less reliable.

6. Results and Discussion

Table 3 reports the Smatch scores obtained by comparing the output of our method against the gold-standard AMR annotations of the AMRNews-PT corpus. These results serve as the baseline for our experiments, as they correspond to the full configuration in which all rules are activated, providing a reference point for the subsequent ablation analysis. The overall performance is shown in Table 3a, where the system achieves a Smatch F1-score of 0.51, with precision of 0.43 and recall of 0.64. The fine-grained analysis in Table 3b indicates that concept identification (0.57 F1) is substantially more reliable than relation prediction (0.39 F1), reflecting the current dependency of the model on UD-based mappings for semantic role assignment. Compared to the Smatch F1-score of 0.535 obtained by the original RBAMR on the AMR-LittlePrince corpus, our pipeline-inspired method achieves similar performance. Table 3 summarizes the overall behavior of the full system without isolating the contribution of each symbolic rule. In order to address this aspect, we introduce an ablation study. For comparison, we also computed Smatch scores on a 435-sentence subset of the corpus used in the ablation study, as indicated in Table 4.

The ablation results reveal a stratification of rule importance in the news corpus. Core semantic backbone rules—subject (*nsubj*→:ARG0), direct object (*obj*→:ARG1), and named entities (highlighted in blue in Table 4)—consistently yield the largest drops in Smatch (ΔS) when removed, indicating their central role in preserving predicate–argument structure and ensuring faithful entity grounding in the resulting AMR graphs. In contrast, a set of grammatical phenomena (highlighted in orange), including copular constructions, possession, ordinals, and indirect objects, shows negligible variation across Smatch, precision, and recall (ΔS, ΔP, $\Delta R \approx 0$), in some cases with zero corpus support, suggesting that their contribution to global semantic reconstruction is minimal. An intermediate group (highlighted in purple), comprising adverbial modification, nominal modification, and coordination, exhibits small but consistent effects ($\Delta S \sim 0$ to -0.02), reflecting a secondary role in refining event semantics rather than defining core structure. Finally, auxiliary-related phenomena cannot be evaluated in isolation due to interaction effects with other rules, indicating non-modular dependencies within the system.

Table 3. Pipeline-inspired method evaluation on AMRNews-PT.

(a) Overall performance

Metric	Value
Precision	0.43
Recall	0.64
F1-score	**0.51**

(b) Fine-grained performance

Category	F1-score
Concepts	0.56
Relations	0.39

7. Final Remarks and Future Work

The ablation results provide a hierarchy in which core structural rules are prioritized, intermediate rules support semantic refinement, and low-impact rules are selectively included or pruned according to performance–cost trade-offs. Future work will focus on expanding the rule hierarchy through the investigation of genre-specific phenomena that may require refinements to existing AMR treatments or even the proposal of new annotation guidelines, which may in turn motivate the formulation of additional rules. This process will be conducted based on the Porttinari-Base PropBank (PBP) corpus [Freitas

Table 4. Ablation experiment for each rule.

Phenomenon	Supp.	Smatch	ΔS	Prec.	ΔP	Rec.	ΔR
Baseline (435 sents)	—	**0.52**	—	**0.43**	—	**0.64**	—
Stopword filtering	1.4023	0.52	0.00	0.44	+0.01	0.63	-0.01
Named entities	0.5080	0.47	-0.05	0.38	-0.05	0.61	-0.03
Subject (nsubj → :ARG0)	0.4494	0.46	-0.06	0.36	-0.07	0.65	+0.01
Auxiliary detection	0.3391	—	—	—	—	—	—
Direct object (obj → :ARG1)	0.3218	0.47	-0.05	0.37	-0.06	0.65	+0.01
Nominal modification (nmod → :ARG1)	0.3000	0.50	-0.02	0.40	-0.03	0.67	+0.03
Subject insertion	0.2506	0.50	-0.02	0.40	-0.03	0.65	+0.01
Adjectival mod. (amod → :mod)	0.2241	0.51	-0.01	0.41	-0.02	0.65	+0.01
Manner (advmod → :manner)	0.1368	0.52	0.00	0.43	0.00	0.65	+0.01
Time (advmod → :time)	0.0414	0.51	-0.01	0.43	0.00	0.64	0.00
Degree (advmod → :degree)	0.0345	0.52	0.00	0.43	0.00	0.64	0.00
Coordination (cc/conj → :op)	0.1621	0.52	0.00	0.43	0.00	0.66	+0.02
Questions	0.0506	0.52	0.00	0.43	0.00	0.64	0.00
Negation (não/nunca)	0.1379	0.51	-0.01	0.42	-0.01	0.64	0.00
Open clausal comp. (xcomp → :ARG2)	0.1023	0.52	0.00	0.44	+0.01	0.63	-0.01
Quantification (nummod → :quant)	0.0805	0.52	0.00	0.43	0.00	0.65	+0.01
Passive subject (nsubj:pass → :ARG1)	0.0368	0.51	-0.01	0.42	-0.01	0.64	0.00
Passive voice (aux:pass)	0.0000	0.52	0.00	0.43	0.00	0.64	0.00
Contrast (:contrast)	0.0287	0.52	0.00	0.44	+0.01	0.63	-0.01
Obligation (ter/dever)	0.0195	0.52	0.00	0.43	0.00	0.64	0.00
Possibility (poder)	0.0161	0.52	0.00	0.43	0.00	0.64	0.00
Copular constructions (ser/estar)	0.0011	0.52	0.00	0.43	0.00	0.64	0.00
Ordinals (ordinal-entity)	0.0011	0.52	0.00	0.43	0.00	0.64	0.00
Indirect object (iobj → :ARG2)	0.0000	0.52	0.00	0.43	0.00	0.64	0.00
Possession (nmod:poss → :poss)	0.0000	0.52	0.00	0.43	0.00	0.64	0.00

and Pardo, 2024, 2025], which comprises 8,418 sentences annotated with gold-standard UD and PropBank structures. The resulting rule hierarchy, expanded with genre-specific rules, will then be applied to this corpus to automatically generate draft AMR graphs for manual inspection and correction, simultaneously supporting iterative parsing refinement and the semi-automatic construction of additional AMR-annotated data. Finally, the expanded rule hierarchy will be reevaluated on AMRNews-PT through controlled ablation experiments to assess the effectiveness of the proposed adaptation strategy, ultimately resulting in a fully developed RBAMR-News parser. In the broader context of Portuguese, where annotated AMR resources remain scarce and domain coverage is still limited, this strategy may contribute to the incremental construction of larger AMR corpora (PBP), while providing empirically grounded insights into the adaptation of symbolic parsing frameworks across textual genres[14].

Acknowledgements. This work was carried out at the Center for Artificial Intelligence of the University of São Paulo (C4AI - http://c4ai.inova.usp.br/), with support by the São Paulo Research Foundation (FAPESP #2019/07665-4) and by the IBM Corporation. The project was also supported by the Ministry of Science, Technology and Innovation, with resources of Law N.8.248, of Oct 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published as Residence in TIC 13, DOU 01245.010222/2022-44.

Limitations

A current limitation of our experimental setup is the reliance on a rule-based SRL component and the Verbo-Brasil lexicon within the RBAMR pipeline, which constrains the

[14]Experiments, resources and other information related to this paper are available at: https://github.com/mjbcomarim/RBAMR-News

quality of the intermediate semantic representation used for AMR construction. Since the AMR rules operate over SRL outputs, variations in SRL quality can directly affect argument structure identification and, consequently, the downstream semantic graph generation. In particular, this dependency may influence the stability and relative importance of rules when applied to more accurate SRL models. To address this limitation, future work will replace the current SRL module with a higher-quality neural SRL system (e.g., BERT-based models) and re-evaluate the ablation-based rule analysis under this updated intermediate representation. This will allow us to assess the robustness of the proposed rule hierarchy to improvements in semantic role prediction and to quantify the extent to which rule effectiveness is dependent on SRL quality versus intrinsic structural relevance in the news domain.

Ethics Statement

This work relies on previously available NLP tools, including syntactic parsing and SRL rule-based strategies, and linguitic resources, particularly news-domain corpora with gold-standard syntactic and semantic annotations. No personal or sensitive data were intentionally collected or processed. As this work involves NLP tools and rule-based AMR parsing, potential biases and error propagation effects originating from upstream syntactic and SRL may propagate to the generated AMR structures. Nevertheless, we expect the proposed methodology to contribute positively to the expansion of semantic resources and symbolic parsing strategies for Portuguese, a language still underrepresented in AMR research.

Disclosure Regarding the Use of AI Tools

Using GenAI only to correct the grammar, without changing the tone, style or adding any new information.

References

Afonso, S., Bick, E., Haber, R., and Santos, D. (2002). Floresta sintá(c)tica: A treebank for portuguese. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands, Spain. European Language Resources Association (ELRA).

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, page 487–499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Alva-Manchego, F. E. and Rosa, J. L. G. (2012). Semantic role labeling for brazilian portuguese: A benchmark. In *Advances in Artificial Intelligence – IBERAMIA 2012*, volume 7637 of *LNCS*, pages 481–490, Berlin, Heidelberg. Springer.

Anchiêta, R. T. and Pardo, T. A. S. (2018). A rule-based amr parser for portuguese. In Simari, G. R., Fermé, E., Gutiérrez Segura, F., and Rodríguez Melquiades, J. A., editors, *Advances in Artificial Intelligence - IBERAMIA 2018*, pages 341–353.

Anchiêta, R. T. and Pardo, T. A. S. (2022). Abstract meaning representation parsing for the brazilian portuguese language. In Pinheiro, V., Gamallo, P., Amaro, R., Scarton, C., Batista, F., Silva, D., Magro, C., and Pinto, H., editors, *Computational Processing of the Portuguese Language*, pages 429–434, Cham. Springer International Publishing.

Bai, X., Chen, Y., and Zhang, Y. (2022). Graph pre-training for AMR parsing and generation. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6001–6015, Dublin, Ireland. Association for Computational Linguistics.

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.

Bevilacqua, M., Blloshmi, R., and Navigli, R. (2021). One spring to rule them both: Symmetric amr semantic parsing and generation without a complex pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12564–12573.

Bick, E. (2000). *The Parsing System Palavras: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press, Aarhus.

Cai, S. and Knight, K. (2013). Smatch: an evaluation metric for semantic feature structures. In Schuetze, H., Fung, P., and Poesio, M., editors, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Ceregatto, G. (2025). Representação formal de significado: o caso dos tweets do mercado financeiro. Dissertação (mestrado em linguística), Universidade Federal de São Carlos, São Carlos.

Damonte, M., Cohen, S. B., and Satta, G. (2017). An incremental parser for abstract meaning representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 536–546.

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

Duran, M. S. and Aluísio, S. M. (2012). Propbank-br: a Brazilian treebank annotated with semantic role labels. In Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1862–1867, Istanbul, Turkey. ELRA.

Ettinger, A., Hwang, J., Pyatkin, V., Bhagavatula, C., and Choi, Y. (2023). "you are an expert linguistic annotator": Limits of LLMs as analyzers of Abstract Meaning Representation. In Bouamor, H., Pino, J., and Bali, K., editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8250–8263, Singapore. Association for Computational Linguistics.

Freitas, C. (2024). Anotação de papéis semânticos no corpus porttinari-base: procedimentos, resultados e análises. Technical Report 450, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos-SP.

Freitas, C. and Pardo, T. A. S. (2024). Propbank e anotação de papéis semânticos para a língua portuguesa: O que há de novo? In *Proceedings of the 15th Symposium in Information and Human Language Technology (STIL)*, pages 118–128, Belém-PA, Brazil.

Freitas, C. and Pardo, T. A. S. (2025). Propbanks e representações semânticas: o que temos, o que queremos e o que podemos. *LinguaMÁTICA*, 17(2):1–29.

Hartmann, N., Fonseca, E., Shulby, C., Treviso, M., Silva, J., and Aluísio, S. (2017). Portuguese word embeddings: Evaluating on word analogies and natural language tasks.

In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*, pages 122–131.

Hartmann, N. S., Duran, M. S., and Aluísio, S. M. (2016). Automatic semantic role labeling on non-revised syntactic trees of journalistic texts. In *International Conference on Computational Processing of the Portuguese Language*, pages 202–212.

Heinecke, J. (2023). metamorphosed, a graphical editor for abstract meaning representation. In *Proceedings of the 19th Joint ACL-ISO Workshop on Interoperable Semantics (ISA-19)*, pages 27–32, Nancy, France. Association for Computational Linguistics.

Hermjakob, U. (2013). Amr editor: A tool to build abstract meaning representations. `https://amr.isi.edu/papers/amr-editor-ulf2013a.pdf`. Accessed: 25 Aug 2024.

Ho, S. H. (2025). Evaluation of finetuned llms in amr parsing. *arXiv preprint arXiv:2508.05028*. 27 pages, 32 figures.

Inácio, M. L., Cabezudo, M. A. S., Ramisch, R., Di-Felippo, A., and Pardo, T. A. S. (2023). The amr-pt corpus and the semantic annotation of challenging sentences from journalistic and opinion texts. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 39(3):1–31.

Jurafsky, D. and Martin, J. H. (2026). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*. 3rd edition. Online manuscript released January 6, 2026.

Li, Y. and Fowlie, M. (2025). Gpt makes a poor amr parser. *Journal for Language Technology and Computational Linguistics*, 38(2):43–76.

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.

Mansouri, B. (2025). Survey of abstract meaning representation: Then, now, future.

Meyes, R., Lu, M., de Puiseau, C. W., and Meisen, T. (2019). Ablation studies in artificial neural networks. *CoRR*, abs/1901.08644.

Oliveira, S., Loureiro, D., and Jorge, A. (2021). Improving portuguese semantic role labeling with transformers and transfer learning. In *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–9.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Sadeddine, Z., Opitz, J., and Suchanek, F. (2024). A survey of meaning representations – from theory to practical utility. In Duh, K., Gomez, H., and Bethard, S., editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2877–2892, Mexico City, Mexico. Association for Computational Linguistics.

Sanches 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. Association for Computational Linguistics.

Sobrevilla Cabezudo, M. A. and Pardo, T. (2019). Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In Friedrich, A., Zeyrek, D., and Hoek, J., editors, *Proceedings of the 13th Linguistic Annotation Workshop*, pages 236–244, Florence, Italy. Association for Computational Linguistics.

Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: Pretrained bert models for brazilian portuguese. In *Proceedings of the 9th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 403–417, Cham. Springer.

Straka, M., Straková, J., and Gamba, F. (2024). ÚFAL LatinPipe at EvaLatin 2024: Morphosyntactic analysis of Latin. In Sprugnoli, R. and Passarotti, M., editors, *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)*, pages 207–214, Torino, Italia. ELRA and ICCL.

Vasylenko, P., Huguet Cabot, P. L., Martínez Lorenzo, A. C., and Navigli, R. (2023). Incorporating graph information in transformer-based AMR parsing. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1995–2011, Toronto, Canada. Association for Computational Linguistics.

Wang, C., Xue, N., and Pradhan, S. (2015). Boosting transition-based AMR parsing with refined actions and auxiliary analyzers. In Zong, C. and Strube, M., editors, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 857–862, Beijing, China. ACL.

Wein, S. and Opitz, J. (2024). A survey of AMR applications. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6856–6875, Miami, Florida, USA. Association for Computational Linguistics.

Zhou, J., Naseem, T., Fernandez Astudillo, R., Lee, Y.-S., Florian, R., and Roukos, S. (2021). Structure-aware fine-tuning of sequence-to-sequence transformers for transition-based AMR parsing. In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6279–6290, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.