

Semantic Representation of Relative Clauses in Lexicalized Abstract Meaning Representation

Jorge Baptista

U. Algarve - FCHS / Portugal
INESC-ID Lisboa - HLT / Portugal
jbaptis@ualg.pt

Sónia Reis

U. Algarve / Portugal
AE Albufeira / Portugal
INESC-ID Lisboa - HLT / Portugal
smreis@ualg.pt

Abstract

This paper analyzes the semantic parsing of relative clauses, with focus on Portuguese, in two meaning representation frameworks: Abstract Meaning Representation (AMR) and Lexicalized Meaning Representation (LMR). While both treat relatives as noun modifiers, AMR fails to distinguish restrictive from appositive clauses—an important traditional grammatical distinction. We argue for explicitly encoding this difference. The study draws on annotated translations of *The Little Prince* (Saint-Exupéry, 1943) in Brazilian and European Portuguese, highlighting issues in the Brazilian AMR annotations.

1 Introduction

Abstract Meaning Representation (AMR) is a popular semantic formalism designed to capture the meaning of sentences through rooted, directed, acyclic graphs, in which nodes represent concepts and edges express semantic relationships (Banarescu et al., 2013; Schneider et al., 2015; Banarescu et al., 2019). Unlike syntactic representations, such as dependency or constituency trees, AMR abstracts away from surface linguistic forms and concentrates on the underlying semantics – essentially, *who is doing what to whom*. It adopts a PENMAN directed acyclic graph representation without a root node, where nodes are the concepts (semantic predicates and their arguments or circumstants), while arcs correspond to the semantic relations between them. Figure 1 is a visual representation of the PENMAN graph shown in Figure 2 and both corresponds to the meaning of sentences like *The boy wants the girl to believe in him*.

By prioritizing meaning over structure, AMR enables a unified representation for semantically equivalent sentences. This makes it particularly valuable in natural language processing tasks like machine translation (Li and Flanigan, 2022), summarization (Dohare et al., 2017), and text genera-

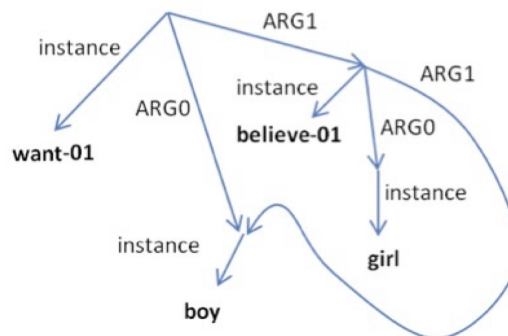


Figure 1: Caption: Visual representation of AMR-graph (from (Banarescu et al., 2013))

```
(w / want-01 :ARG0 (b / boy) :ARG1 (b2 /  
believe-01 :ARG0 (g / girl) :ARG1 b))
```

Figure 2: Caption: Penman graph representing the same AMR (from (Banarescu et al., 2013))

tion (Vlachos and Hardi, 2018; Wang et al., 2020), fact-checking (Qiu et al., 2024), to name a few, where direct access to semantic content helps systems handle linguistic variation more effectively (see (Mansouri, 2025), for a recent overview). AMR has been adapted for several languages, including Brazilian Portuguese (Anchiêta and Pardo, 2018a; Anchiêta, 2020; Cabezudo and Pardo, 2020; Inácio et al., 2023), for which specific tools, such as a parser (Anchiêta and Pardo, 2018a) and a novel semantic evaluation metric, SEMA (Anchiêta et al., 2019), have been developed. More recently, Lexicalized Meaning Representation (LMR) (Baptista et al., 2024a,b, 2025; Baptista, 2025)¹ has been proposed as another framework for semantic parsing annotation. Directly inspired by the AMR formalism, LMR differs from it by assuming itself as a text annotation framework and by emphasizing the direct annotation of surface elements in the text. Unlike AMR, it does not require the lemmatiza-

¹For further detail, see LMR Guidelines: <https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt>

tion of all predicative elements into verbal forms, nor the substitution of certain grammatical elements with abstract semantic relations. Instead, LMR preserves these elements to explicitly encode their semantic specificity in the resulting graph. Furthermore, it integrates several theoretical constructs—such as support-verb constructions (SVCs) (Harris, 1991; Baptista and Mamede, 2020b; Baptista et al., 2022), and related phenomena—, and adapts AMR notational system in order to be capable of coherently representing verbal, nominal, adjectival and adverbial predicative constructions. While AMR resources to OntoNotes (Pradhan and Ramshaw, 2017) as a catalog of senses for predicative elements, which basically lemmatizes concepts as verbs, LMR relies on language-specific (Portuguese) lexical resources for both full verbs and predicative nouns (Baptista and Mamede, 2020a,b). Finally, as an annotation framework dedicated to the semantic parsing of sentences, LMR significantly streamlines the identification of multiword expressions and the representation of named entities—including temporal expressions (Hagège et al., 2010)—by shifting the cognitive focus of the annotation towards the encoding of semantic predicates, their arguments, circumstantial elements, and the semantic relations that structure them.

This paper addresses the semantic parsing of relative clauses, a major subordinating device in Portuguese grammar, with a particular focus on the traditional distinction between restrictive relatives, illustrated in (1), and appositive (or explicative) relatives, exemplified in (2):

- (1) *The boy **that believed** cried.*
- (2) *The boy, **who believed**, cried.*

In the first case (1), the restrictive force of the clause introduced by *that* is understood as delimiting the reference of the noun *boy*, implicitly allowing for the interpretation that there may be other boys. The relative clause contributes to the identification of the referent, and it is, therefore, considered an integral part of the noun phrase (Quirk et al., 1985; Huddleston and Pullum, 2002), while its function can be defined as equivalent to an adjectival modifier. In the second case (2), the appositive relative clause introduced by *who* provides supplementary information about an already identified referent, and it is syntactically and semantically external to the noun phrase, serving instead to elaborate on its content (Demonte, 1999). Prosodically,

restrictive relative clauses are typically produced without a pause between the antecedent noun and the relative pronoun *that*, whereas appositive relatives are marked by a clear intonational break preceding the relative pronoun *who* (Huddleston and Pullum, 2002; Selkirk, 2005).

We argue that this distinction between restrictive and appositive relative clauses is highly relevant for representing sentence meaning, and we aim to demonstrate, on the one hand, that AMR fails to capture it adequately, while, on the other hand, we present the alternative descriptive proposal offered by LMR, along with several issues that its adoption can help resolve.

This paper is structured as follows. First, in Section 2, we present the AMR representation of relative clauses and highlight several issues observed in the treatment of this construction. Then, in Section 3, we compare the descriptive solutions proposed by LMR and show how they address the issues raised in the previous section. Subsequently, in Section 4, the paper examines the representation of relative subclauses in the AMR annotation of the Brazilian Portuguese translation of the novella *The Little Prince*, by Antoine de Saint-Exupéry (1943), to better contextualize the notational solutions adopted by LMR in the European Portuguese translation of the same work. A short Conclusion (Section 5) presents the main findings.

2 Relative clauses in AMR: Inverse role representation

AMR adopts a semantic approach by linking the noun modified by the relative clause directly to the predicate inside the clause through inverse semantic roles – such as :ARG0-OF, :ARG1-OF, etc., where :ARG i represents the semantic role of an entity in relation to its respective predicate and the -OF suffix denotes this inversion. This means that the head noun (e.g., *boy*) becomes the main node in the AMR subgraph of the relative clause, and the verb or adjective from the relative clause is attached via an inverse relation, showing the role played by that noun in the event or property described by that element. Thus, the base sentence (3), which is represented by the PENMAN graph given below (to streamline the presentation, the PENMAN graphs are provided without indentations; AMR graphs are shown in green, while LMR code is shown in blue):

- (3) *The boy believes.*

→ (b1 / believe-01 :ARG0 (b2 / boy))

while the same representation is produced for both types of relative clauses (repeated below; the relative and its representation are highlighted in bold typeface; the numbering of nodes *boy* and *believe* in (4) was kept the same as in (3), for clarity):

- (4) a. *The boy **that believed** cried.*
b. *The boy, **who believed**, cried.*

→ c / cry-01 :ARG0 (b2 / boy :ARG0-OF (b1 / believe-01))

It is worth noticing that the appositive nature of the relative subclause in (4-b) is not addressed, nor, for that matter, is the grammatical construct of apposition in general.

Next, we observe that adjectival sentences—those whose main predicate is an adjective—are typically represented using the relation :DOMAIN, omitting the copula verb, as shown in example (5). In contrast, the corresponding relative subclauses are annotated with the inverse relation :DOMAIN-OF. Otherwise, and as stated in the AMR guidelines, this relation is “often shortened”² to :MOD (from *modifier*). For example, consider (6). Notice that both sentences are represented by the same graph, and in the second graph :MOD replaces :DOMAIN-OF:

- (5) *The boy is tall.*

→ (t / tall :DOMAIN (b / boy))

- (6) a. *The boy **that is tall** cried.*
b. *The boy, **who is tall**, cried.*

→ (c / cry :ARG0 (b / boy :DOMAIN-OF (tall)))

or

→ (c / cry :ARG0 (b / boy :MOD (tall)))

On the other hand, predicative nouns and support verb constructions (SVC) are usually represented by the corresponding verb in the OntoNotes ontology (Hovy et al., 2006).

3 Relative clauses in LMR: Apposition

The Lexicalized Meaning Representation (LMR) framework (Baptista et al., 2024a) offers a detailed account of the annotation of relative clauses – both restrictive and appositive – and also supports the representation of other syntactic structures with

²<https://github.com/amrиси/amr-guidelines/blob/master/amr.md#focus>

comparable semantic behavior, such as reduced subclauses or embedded interrogatives. The LMR representation of restrictive relative clauses, example given in (7-a), largely mirrors that of AMR. Key differences include the introduction of a ROOT node, the use of the :MAIN relation to link this node to the top-level semantic predicate, and the explicit encoding of the copula and auxiliary verbs like *be* via the :VAUX relation. (To simplify, we use the :VAUX label to represent copula and auxiliary verbs alike. In future versions of the guidelines a :VCOP label will be included to specifically denote copula verbs, similarly to what is already done for support verbs, :VSUP).

Furthermore, the relative pronouns *that* or *who* are explicitly annotated and linked via an explicit coreference (:CORREF) relation to the variable representing its antecedent noun. (It should be noted that, similarly to AMR, LMR does not encode definite or indefinite articles in its annotation. Both schemes, however, represent demonstratives and quantifiers, as these elements are semantically relevant.)

In contrast, for appositive relative constructions, LMR introduces a generic subordinative mechanism: the :APPOS relation, which explicitly marks appositional content (7-b). Thus, the appositive relative clause in the example above receives the following representation (the relative and its representation are highlighted in bold typeface):

- (7) a. *The boy **that is tall** cried.*
→ (ROOT :MAIN (c / cried :ARG0 (b / boy :ARG0-OF (t1 / tall :VAUX (i / is) :ARG0 (t2 / that :CORREF b))))))
b. *The boy, **who is tall**, cried.*
→ (ROOT :MAIN (c / cried :ARG0 (b / boy :APPOS (t / tall :ARG0 b :VAUX (i / is) :ARG0 (t2 / that :CORREF b))))))

The introduction of the :APPOS relation into the notational framework is more than just a formal convenience: it also serves other functions in meaning representation, as it can be applied in other environments. Still, in this case, it preserves the distinction between the two interpretations of relative clauses. For example, in the next example (8), only the appositive interpretation of the relative subclause is adequate, which is made clear by the use of delimiting commas (The examples taken from the European Portuguese translation of *The Little Prince* are signaled by the prefix ‘opz-’ (*O Principezinho*), plus the sentence ID, while the

Brazilian Portuguese translation are indicated by the prefix ‘opp-’ (*O Pequeno Príncipe*). The English translations of the sentences are our own. They were made directly from the Brazilian Portuguese version of *O Pequeno Príncipe*, and may not correspond exactly to the original English wording of the novella. The same was done for the European Portuguese version, *O Príncipezinho*):

- (8) *O principezinho, que me fazia muitas perguntas, nunca parecia ouvir as minhas.* [opp-id=105]. ‘The little prince, who asked me many questions, didn’t even seem to hear mine.’

→(ROOT :MAIN (p1 / parecia :FREQUENCY (n / nunca :NEG +) :ARG0 (p2 / principezinho :DIMINUTIVE :APPOS (p3 / perguntas :QUANT (m1 / muitas) :ARG0 (q / que :CORREF p2) :ARG1 (m2 / me) :VSUP (f / fazia))) :ARG1 (o / ouvir :ARG0 p2 :ARG1 (p4 / [perguntas] :ARG0 (m3 / minhas :CORREF m2))))))

This example also illustrates the representation, within LMR, of a support verb construction (SVC), namely *fazer perguntas* ‘ask questions’ (Baptista and Mamede, 2020b). Note that the relative pronoun – functioning as the subject of the SVC – is explicitly represented, along with its coreference to the antecedent noun. In addition, the example demonstrates the reconstruction of a second instance of *perguntas* ‘questions’, which is zeroed in the surface form to avoid lexical repetition. This second occurrence, however, is not coreferential with the first, as indicated by the possessive determiner *minhas* ‘my’.

We note, however, that the orthographic (and syntactic) device of delimiting commas is not always applied consistently in texts. Nevertheless, this inconsistency does not impact the semantic representation, as shown in (9). In this example, the relative clause *que se inclinou . . .* ‘who leaned over’, although not set off by commas, still conveys an appositive meaning.

- (9) – *Oh! Mas eu já vi, disse o príncipe que se inclinou para dar ainda uma olhadela do outro lado do planeta.* [opp-id=574] ‘Oh! But I have already seen it, said the prince, who leaned over to take one more look from the other side of the planet.’

→ (ROOT :MOD (o1 / oh :MODE-EXCLAMATIVE :MAIN (d / disse :DIRECT-SPEECH + :ARG1 (v / vi MOD (m / mas) :ARG0 (e / eu :CORREF p1) :TIME (j / já)) :ARG0 (p1 / príncipe :APPOS (i / se-inclinou :ARG0 (q / que :CORREF p1) :PURPOSE (p2 / para :OP2 (o2 / olhadela :VSUP (d / dar) :TIME (a / ainda) :ARG0

p1 :LOCATIVE (d / de :OP2 (l / lado :MOD (o3 / outro) :PART-OF (p3 / planeta))))))

The interpretation of an appositive reading is derived from the broader discourse context, given that there is only one prince in the story (i.e., within the relevant narrative situation).³ Under this assumption, it can be shown that the relative clause conveys an appositive meaning. This analysis is supported by the following diagnostics: (i) commas may be inserted without altering the interpretation; (ii) the relative pronoun *que* ‘that’ can be replaced by *o qual* ‘who’; and (iii) in spoken language, a prosodic break typically precedes the clause introduced by *que*. Naturally, in the absence of formal clues or discourse context, an automatic parser would have to assign this sentence a restrictive reading.

By explicitly acknowledging the syntactic device of apposition, LMR also enables the representation of other related constructions, such as adjective- or participial-reduced subclauses. Example (10) illustrates such an appositive, adjective-reduced relative clause with the adjective *roxo* ‘purple’:

- (10) — *Eu conheço um planeta onde há um sujeito vermelho, quase roxo.* [opp-id=332] ‘— I know a planet where there’s a red fellow, almost purple.’

→ (ROOT :MAIN (c / conheço :DIRECT-SPEECH + :ARG0 (e / eu) :ARG1 (p / planeta :LOCATION-OF (h / há :LOCATION (o / onde) :ARG1 (s / sujeito :MOD (v / vermelho) :APPOS (r / roxo :ARG0 s :DEGREE (q / quase))))))

In this sentence, two adjectives modify the noun *sujeito* ‘fellow’: the first, *vermelho* ‘red’ is a (restrictive) modifier of the noun; while the second, *roxo* ‘purple’, is in apposition to the first. Notice, also, the relative subclause introduced by the relative adverb *onde* ‘where’ and the inverted relation :LOCATION-OF. Conversely, the AMR representation (below) merely juxtaposes the two adjectives:

→ (c / conhecer-01 :ARG0 (e / eu) :ARG1 (p / planeta :LOCATION-OF (s / sujeito :MOD (v / vermelho) :MOD (r / roxo :DEGREE (q / quase))))

Notice that the corresponding sentence in European Portuguese (11) does not show this apposition:

³We thank the anonymous reviewer for drawing our attention to this point.

- (11) — *Sei de um planeta onde há um senhor muito vermelho.* [opz-id=332] ‘— I know of a planet where there is a very red-faced gentleman.’

→ (ROOT :MAIN (s1 / sei :ARG1 (p / planeta :LOCATION-OF (h / há :ARG1 (s2 / senhor :ARG0-OF (v / vermelho :DEGREE (m / muito)))) :LOCATION (o / onde))))))

4 Comparison AMR-LMR Relatives

We conducted a systematic survey of the relative clauses occurring in *O Pequeno Príncipe*, the Brazilian Portuguese translation of the novella, as documented in (Anchiêta and Pardo, 2018b; Anchiêta, 2020), and examined them against the corresponding AMR representations of the same sentences.

With respect to the annotation methodology adopted in this paper, an important distinction must be made. The AMR annotations for the Brazilian Portuguese version of *O Pequeno Príncipe* were derived from a corpus of AMR graphs that originated through alignment with the original English version of *The Little Prince*. In many cases, the Brazilian Portuguese graphs were obtained by transferring the English AMR annotations of aligned sentences and subsequently revising them manually to accommodate the Brazilian Portuguese text (Anchiêta and Pardo, 2018b).

By contrast, the LMR annotations for the European Portuguese version of *O Príncipezinho* were manually constructed from scratch on the basis of the European Portuguese text itself, without relying on automatic transfer procedures or cross-linguistic alignment with a pre-existing AMR version in another language. It should also be emphasized that both the Brazilian Portuguese and the European Portuguese translations were produced directly from the original French source text, *Le Petit Prince*, and not from the English version.

A total of 87 instances of relative clauses introduced by the pronoun *que* ‘that’, ‘who’, ‘which’ were identified, including occurrences where *que* is preceded by a preposition (e.g., *em* ‘in, on’). It is worth noting that in Portuguese, the same relative pronoun *que* can introduce both restrictive and appositive relative clauses. In addition, we recorded a single occurrence of *a qual* ‘which[_f.sg.]’, one of *as quais* ‘which[_f.pl.]’, and another one of *os quais* ‘which[_m.pl.]’. This is a special type of relative pronoun, exclusively used in appositive rela-

tives, and that shows gender and number agreement with its antecedent (as signaled inside brackets). Finally, we observed 1 instance of the so-called relative *quem* ‘who/whom’ and 5 instances of the relative adverb *onde* ‘where’. In total, 95 relative clauses were identified.

In this survey, the subordinate conjunction *que* ‘that’ was ignored. The so-called *cleft* relatives (e.g., *E foi desse modo que eu viajei conhecimento, um dia, com o pequeno príncipe.* [opp-id=102>] ‘And it was in this way that I became acquainted, one day, with the little prince.’), which we construe as a form of focusing, were also not considered. Finally, we discarded interrogative pronouns, either partial-interrogatives in direct speech (e.g., *Sobre quem reinaria o rei?* [opp-id=521] ‘Over whom would the king reign?’) or as interrogative sub-clauses in indirect speech (e.g. *Se . . . , adivinharão quem é.* [opp-id=1560] ‘If . . . , you will guess who it is.’)

As expected, in the AMR representation of *O Pequeno Príncipe*, no formal distinction is made between restrictive and appositive relative clauses, mostly introduced by *que* ‘that’. In most cases, the relative is represented by an inverted -OF relation and, more rarely, by :MOD. For example, in sentence (12), represented in LMR in (8) above and in AMR below, the appositive relative clause is represented through the :ARG-OF relation linking *príncipezinho* ‘little prince’ to *perguntar* ‘ask’:

- (12) *O príncipezinho, que me fazia milhares de perguntas, não parecia sequer escutar as minhas.* [opp-id=105] ‘The little prince, who asked me thousands of questions, didn’t even seem to listen to mine.’

→ (p / parecer-01 :ARG1 (e / escutar-01 :POLARITY - :ARG0 (p1 / príncipezinho :ARG0-OF (p2 / perguntar-01 :ARG2 (e1 / eu) :QUANT (m1 / milhares)))) :ARG1 (m2 / meu)))

Note that the lemmatization of the support verb construction (SVC) *fazer perguntas* ‘make questions’ is performed by reducing it to the main verb *perguntar* ‘ask’, thereby omitting the support verb, as prescribed by the AMR guidelines. However, this reduction conflicts with the quantifier *milhares* ‘thousands’ and the possessive *minhas* ‘mine’, which syntactically and semantically select for a nominal head. Notice also that the possessive is represented without any reference to its nominal antecedent (*perguntas* ‘questions’). Additionally,

the negation adverb (represented by :POLARITY – is incorrectly attached: instead of modifying the verb *parecer* ‘seem’, as it is in the sentence, it is made to modify *escutar* ‘listen’. The adverb *sequer* ‘even’, which functions as a *focus adverb* with negative polarity and it is associated with *escutar* ‘listen’, is entirely omitted from the graph – a pattern frequently observed in AMR parsing of adverbs, particularly with focus adverbs (Molinier and Lévrier, 2000; Baptista et al., 2025).

Moving now to the relative pronoun *o qual* ‘which’, only used in appositive relative clauses in Portuguese, a typical example is given in (13):

- (13) *É preciso que a gente se conforme em arrancar regularmente os baobás logo que se distingam das roseiras, com as quais muito se parecem quando pequenos.* [opp-id=243] ‘We must resign ourselves to regularly pulling out the baobabs as soon as they can be told apart from the rosebushes, which they closely resemble when small.’

→ (p / precisar-01 :ARG0 (a / a_gente) :ARG1 (c / conformar-01 :ARG0 a :ARG1 (a1 / arrancar-01 :ARG0 a :ARG1 (b / baobá) :MANNER (r / regularmente) :TIME (d / distinguir-01:ARG1 b))))

In this sentence, besides other inaccuracies – such as the pronoun *a gente* ‘we’ being parsed as the subject of *precisar*, the entire relative clause has not been represented in the AMR graph, which ends abruptly after *distinguir* ‘tell apart’.

For the corresponding sentence in European Portuguese, the appositive relative is LMR-represented in (14); the variable *e* stands for *embondeiros* ‘baobabs’:

- (14) *Tenho de me obrigar a arrancar os embondeiros assim que os distingo das roseiras, com as quais se confundem antes de crescer.* [opz-id=243] ‘I have to force myself to pull out the baobabs as soon as I can tell them apart from the rosebushes, which they are easily confused with before they grow.’

→ (ROOT :MAIN (o / obrigar :VAUX (t / tenho :MWE-CONT (d / de)) :ARG1 (m / me) :ARG2 (d / distingo :ARG0 m :ARG1 e :ARG2 (r / roseiras :APPOS (c / se-confundem :ARG1 e :ARG2 (aq / as_quais :CORREF r) :TIME (ad / antes_de :OP2 (c / crescer :ARG0 e))))))

In this sentence, we assume that the the symmetric construction of verb *confundir* ‘confound’

(Baptista and Mamede, 2020a, class 36S1) is here used in a *se*-passive (or median-passive) alternation, hence, the annotation of :ARG1 *e* and the fact that the reflexive pronoun *-se* is not annotated.

The following example (15) is more complex, as it involves a relative clause whose pivot is a noun complement (v.g. *um exemplar dessa flor = da qual* ‘an example of that flower = of which’ (for brevity, we only show the LMR parse from the conditional subclause onwards, in bold):

- (15) *Corou um pouco, e continuou em seguida: – Se alguém ama uma flor da qual só existe um exemplar em milhões e milhões de estrelas, isso basta para que seja feliz quando a contempla.* [opp-id=349] ‘He blushed a little and then continued: If someone loves a flower of which there is only one example among millions and millions of stars, that is enough to make him happy when he looks at it.’

→ ... :CONDITION (s1 / se :OP2 (a1 / ama :ARG0 (a2 / alguém) :ARG1 (f / flor :APPOS (e1 / existe :FOCUS (s2 / só) :ARG0 (e2 / exemplar :QUANT 1 :MOD (aq / a qual CORREF f) :LOCATION (e / em :OP2 (e / estrelas :QUANT (mm / milhões e milhões))))))

In the AMR representation, however, a very different description of the sentence is provided (this is given in full below, with the relative subclause highlighted):

→ (e / e :OP1 (c / corar-01 :DEGREE (p / pouco)) :OP2 (c1 / continuar-01 :ARG1 (b / bastar-01 :ARG1 (a / amar-01 :ARG0 (a1 / alguém) :ARG1 (f / flor :MOD (e1 / existir-01 :ARG2 (e2 / exemplar-01 :QUANT 1 :LOCATION (e3 / Estrela :QUANT (m / multiplo :OP1 1000000)))))) :ARG2 (f1 / feliz) :TIME (c2 / contemplar-01 :ARG1 f))))

Firstly, the conditional subordinate clause is simply ignored. The content of the conditional, which is anaphorically resumed in the sentence by *isso* ‘that’, appears as the :ARG1 of *bastar* ‘to be enough’, while, as its subject, should be captured by :ARG0. The relative clause is represented by a :MOD relation linking *flor* ‘flower’ to *existir* ‘to exist’; however, in standard AMR, relative clauses should always have been represented instead by an inverse relation (v.g., ARGi-OF). In spite of these inconsistencies in the AMR representation, it is clear that the appositive nature of the relative clause in the text has not been adequately captured in the AMR representation.

Notice that the European Portuguese version

does not show this phenomenon (16) (for lack of space, some annotations are not explained here, please refer to the LMR Guidelines: (Baptista, 2025)):

- (16) *Já corado, prosseguiu. – Se amamos uma flor única entre milhões e milhões de estrelas, não precisamos de mais para ficarmos felizes quando as observamos.* [opz-id=349] ‘Already blushing, he went on: – If we love a flower that is unique among millions and millions of stars, we need nothing more to be happy when we look at them.’

→ ((ROOT :MAIN (d / [dizendo] :VAUX (p1 / prosseguiu) :ARG0 (p2 / [princepezinho] :DIMINUTIVE + :APPOS (c / corado :MOD (j / já))) :ARG1 (p3 / precisamos :DIRECT-SPEECH :NEG (n / não) :ARG1 (m / mais) :PURPOSE (p4 / para :OP2 (f1 / felizes :VAUX (f2 / ficarmos) :TIME (qd / quando :OP2 (o / observamos :ARG1 (as / as)))))) :CONDITION (se / se :OP2 (a / amamos :ARG1 (f3 / flor :QUANT 1 :ARG0-OF (u / única :LOCATION (e1 / entre :OP2 (e2 / estrelas :QUANT (mm / milhões e milhões))))))))))

We conclude with a striking example (17), which highlights the limitations of current abstract meaning representation frameworks, both AMR (below) and LMR:

- (17) *Uma rosa e três vulcões que me dão pelo joelho, um dos quais extinto para sempre.* [opp-id=1040] ‘A rose and three volcanoes that reach up to my knee, one of which is extinct forever.’

→ (d / dar-12 :ARG0 (e / e :OP1 (r / rosa) :OP2 (v / vulcão :QUANT 3)) :ARG1 (j / joelho))

This sentence includes two relative clauses, a restrictive one introduced by *que* and an appositive one, introduced by *os quais* ‘which[_m.pl.]’. Furthermore, this later pronoun is part of a partitive determination, *um [vulcão] desses três vulcões* ‘one volcano from/out of those three volcanoes’. The AMR representation, given above, adequately coordinates the two nouns (*rosa* ‘rose’ and *vulcões* ‘volcanoes’), though it simply ignores the appositive relative and only represents the main clause, rooted by a specific grammatical construction of *dar* lit. ‘give’. The corresponding European Portuguese sentence (18) exhibits the same overall structure, and the LMR representation, given below, successfully captures the two distinct relative clauses.

- (18) *Ela e os três vulcões que me chegam aos joelhos, um dos quais talvez esteja extinto para sempre, não fazem de mim um grande príncipe...* [opz-id=1040] ‘She and the three volcanoes that reach up to my knees, one of which may be extinct forever, do not make me a great prince...’

→ (ROOT :MAIN (f / fazem :DIRECT-SPEECH + :NEG (n / não) :ARG1 (m1 / mim) :VOP (p / príncipe :MOD (g / grande)) :OP1 (e / e :COORD1 (e / ela) :COORD2 (v / vulcões :QUANT 3 :ARG0-OF (c / chegam :ARG0 (q / que :CORREF v) :ARG1 (j / joelhos :MOD (m2 / :CORREF m1))) :APPOS (e / extinto :MOD (t / talvez) :VAUX (e / esteja) :ARG0 (u / um :MOD (oq / os quais :CORREF v)) :DURATION (ps / para_sempre))))))

(Though it is not the focus of this paper, the verb *fazer* ‘make’ in this example is parsed as a (causative) operator-verb (VOP); see LMR Guidelines (Baptista, 2025).)

However, since the coordinated subject could also serve as the antecedent of the relative pronouns – not just *vulcões* (‘volcanoes’) – the current representation framework imposes a choice (in this case, obvious) between two possible and alternative interpretations.

The results of the analysis show that, out of 95 instances of relative clauses, only 18 were adequately represented as such, using the inverted -OF relation. This corresponds to a low precision of 18.9%. In many cases, the relative clause is simply not annotated. In a few instances, relative clauses were parsed as adverbial subordinate clauses, marked with :CAUSE, for instance. For a detailed analysis, Section A in the Appendix lists the IDs of these sentences along with an assessment of their corresponding AMR parses, including both correct and incorrect cases.

5 Conclusion

This study has shown that the semantic representation of relative clauses – particularly the distinction between restrictive and appositive clauses – remains a challenge for frameworks such as AMR. This model does not explicitly capture the syntactic and semantic differences introduced by relative clauses, which are crucial for accurate sentence interpretation. Our analysis of the AMR annotations in the Brazilian Portuguese version of *The Little Prince* uncovered several omissions and inconsistencies — particularly in the classification of relative clauses as restrictive or appositive. In fact,

the rather low precision (18.9%) suggests there is room for improvement in this respect.

In contrast, the LMR framework proved more effective by incorporating specific formal mechanisms – such as the :APPOS relation – that systematically distinguish appositive relative constructions from restrictive modifiers. Moreover, LMR preservation of surface lexical elements, and the explicit handling of support and auxiliary verb constructions further enhance its adequacy for annotating these complex syntactic-semantic phenomena.

We argue that integrating the distinction between restrictive and appositive relatives into semantic representation frameworks contributes not only to a more nuanced analysis of sentence meaning, but could also be beneficial in natural language processing tasks that rely on precise semantic interpretation, such as machine translation, information extraction, summarization, and text generation.

Acknowledgments

This work was supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (FCT) under projects UID/50021/2025 (DOI: [10.54499/UID/50021/2025](https://doi.org/10.54499/UID/50021/2025)) and UID/PRR/50021/2025 (DOI: [10.54499/UID/PRR/50021/2025](https://doi.org/10.54499/UID/PRR/50021/2025)) and by the European Commission (Project: iRead4Skills, Grant number: 1010094837, Topic: HORIZON-CL2-2022-TRANSFORMATIONS-01-07, DOI: [10.3030/101094837](https://doi.org/10.3030/101094837)).

References

- Rafael Anchiêta. 2020. *Abstract Meaning Representation Parsing for the Brazilian Portuguese Language*. Ph.D. thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- Rafael Anchiêta, Marco Cabezudo, and Thiago Pardo. 2019. *Sema: An extended semantic evaluation metric for AMR*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pages 1–13, Florence. ACL. ArXiv preprint.
- Rafael Anchiêta and Thiago Pardo. 2018a. *Rule-Based AMR Parsing for Brazilian Portuguese*. In *Brazilian Conference on Intelligent Systems (BRACIS 2018)*, pages 490–495, São Paulo. SBC.
- Rafael Anchiêta and Thiago Pardo. 2018b. *Towards amr-br: A sembank for brazilian portuguese language*. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki. ELRA.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. *Abstract Meaning Representation for Sembanking*. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Dis-course (ACL 2013)*, pages 178–186, Sofia. ACL.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2019. *Abstract Meaning Representation (AMR) 1.2.6 Specification*.
- Jorge Baptista. 2025. *LMR Directives*. Technical Report Version 2025-09-03, University of Algarve & INESC-ID Lisboa.
- Jorge Baptista and Nuno Mamede. 2020a. *Dicionário Gramatical de Verbos do Português Europeu*. Universidade do Algarve, Faro, Portugal.
- Jorge Baptista and Nuno Mamede. 2020b. *Syntactic Transformations in Rule-Based Parsing of Support Verb Constructions: Examples from European Portuguese*. In *9th Symposium on Languages, Applications and Technologies (SLATE 2020)*, volume 83 of *OpenAccess Series in Informatics (OASICs)*, pages 11:1–11:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- Jorge Baptista, Nuno Mamede, and Sónia Reis. 2022. *Support Verb Constructions across the Ocean Sea*. In *18th Workshop on Multiword Expressions @ LREC2022*, pages 26–36. ACL.
- Jorge Baptista, Izabela Müller, and Sónia Reis. 2025. *Semantic Representation of Adverbs in the Lexicalized Meaning Representation (LMR) Framework*. In *14th Symposium on Languages, Applications and Technologies (SLATE 2025)*, volume 135 of *Open Access Series in Informatics (OASICs)*, pages 9:1–9:18, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- Jorge Baptista, Sónia Reis, João Dias, and Pedro A. Santos. 2024a. *Lexicalized Meaning Representation (LMR)*. In *Proceedings of the Fifth International Workshop on Designing Meaning Representations (DMR) @ LREC-COLING 2024*, pages 101–111, Torino, Italy. ELRA and ICCL.
- Jorge Baptista, Sónia Reis, Pedro A. Santos, and João Dias. 2024b. *Comparação de anotações AMR entre línguas: problemas e soluções*. *Revista da Associação Portuguesa de Linguística*, 1(11):27–54.
- Marco Cabezudo and Thiago Pardo. 2020. *Towards Building an AMR Corpus for Brazilian Portuguese*. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 236–244, Marseille. ELRA.
- Violeta Demonte. 1999. *Relative Clauses in Spanish: A Comparative View*. In Georges Rebuschi and L. Tuller, editors, *The Grammar of Focus*, pages 45–76. John Benjamins, Amsterdam.

- Shibhansh Dohare, Harish Karnick, and Vivek Gupta. 2017. [Text Summarization Using Abstract Meaning Representation](#). ArXiv preprint.
- Caroline Hagège, Jorge Baptista, and Nuno Mamede. 2010. [Caracterização e Processamento de Expressões Temporais em Português](#). *Linguamática*, 2(1):63–76.
- Zellig S. Harris. 1991. *A Theory of Language and Information: A Mathematical Approach*. Clarendon Press, Oxford.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. [Ontonotes: The 90% solution](#). In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60, New York. ACL.
- Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.
- Marcio Inácio, Marco Sobrevilla Cabezudo, Renata Ramisch, Ariani Di Felippo, and Thiago Pardo. 2023. [The AMR-PT Corpus and the Semantic Annotation of Challenging Sentences from Journalistic and Opinion Texts](#). *D.E.L.T.A.*, 39(3):1–31.
- Changmao Li and Jeffrey Flanigan. 2022. [Improving Neural Machine Translation with the Abstract Meaning Representation](#). In *2nd Workshop on Deep Learning on Graphs for Natural Language Processing (DLG4NLP 2022)*, pages 12–21, Seattle. ACL.
- Behrooz Mansouri. 2025. [Survey of Abstract Meaning Representation: Then, Now, Future](#). *Preprint*, arXiv:2505.03229. ArXiv preprint.
- Christian Molinier and Françoise Lévrier. 2000. *Grammaire des Adverbes: Description des Formes en -ment*. Droz, Genève.
- Sameer Pradhan and Lance Ramshaw. 2017. [OntoNotes: Large Scale Multi-Layer, Multi-Lingual, Distributed Annotation](#). In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 521–554. Springer, Dordrecht.
- Haoyi Qiu, Kung-Hsiang Huang, Jingnong Qu, and Nanyun Peng. 2024. [AMRFact: Enhancing Summarization Factuality Evaluation with AMR-Driven Negative Samples Generation](#). eprint=2311.09521. PrimaryClass=cs.CL.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Nathan Schneider, Jeff Flanigan, and Tim O’Gorman. 2015. [The Logic of AMR](#). In *Proceedings of the 2015 Conference of the Association for Computational Linguistics (ACL 2015)*, pages 4–5, Beijing. ACL.
- Elisabeth Selkirk. 2005. [The Prosodic Structure of Function Words](#). In Sónia Frota, Marina Vigário, and Maria João Freitas, editors, *Prosodies: With Special Reference to Iberian Languages*, pages 149–184. Mouton de Gruyter, Berlin.
- Andreas Vlachos and Hemant Hardi. 2018. [Guided Neural Language Generation for Abstractive Summarization Using Abstract Meaning Representation](#). arXiv:1808.09160. ArXiv preprint.
- Tianming Wang, Xiaojun Wan, and Hanqi Jin. 2020. [AMR-to-Text Generation with Graph Transformer](#). *Transactions of the Association for Computational Linguistics*, 8:19–33.

A Sentence IDs and AMR annotation of relative subclauses in *O Pequeno Príncipe* ‘The Little Prince’

Correctly parsed relative clauses: 3, 87, 105, 184, 294, 295, 332, 359, 643, 805, 827, 837, 905, 1008, 1175, 1176, 1382, 1424.

Incorrectly parsed relative clauses: 29, 37, 37, 40, 50, 51, 66;67, 119, 126, 152, 156, 178, 182, 185, 243, 256, 257, 283, 292;293, 344, 347, 375, 385, 424, 433, 518, 547, 547, 552;553;554;555, 559, 563, 574, 641, 671, 683, 686, 689, 722, 723, 747, 748, 763, 819, 820, 820, 821, 862, 863, 867, 905, 934, 936, 980, 980, 1008, 1022, 1039, 1040, 1045, 1064, 1101, 1111, 1118, 1272, 1296, 1336, 1380, 1391, 1417, 1428, 1444, 1445, 1454, 1461, 1548, 1548, 1548.

The original ID numbering from (Anchiêta, 2020) is preserved. In some cases, a single ID corresponds to multiple sentences in the original AMR annotation of the English version of *The Little Prince*. These IDs are joined with semicolon (;). Moreover, some sentences contain more than one relative clause, and it is therefore possible for the same sentence to include both a correctly parsed and an incorrectly parsed relative clause.