

Inter-Sentential Relations in Enhanced Lexicalized Meaning Representation (E-LMR)

Jorge Baptista   

University of Algarve, Faro, Portugal
INESC-ID Lisboa, Portugal

Abstract

Abstract Meaning Representation (AMR) is a widely adopted framework for graph-based sentence-level semantics. However, its abstraction from surface form and its focus on intra-sentential structure limit its ability to capture discourse-level phenomena. Uniform Meaning Representation (UMR) addresses this limitation by introducing cross-sentential mechanisms, at the cost of increased representational complexity. We argue that inter-sentential relations can be modeled without abandoning lexical anchoring. Within Lexicalized Meaning Representation, we propose a programmatic extension grounded in three domains: (i) nominal and pronominal anaphora, (ii) temporal anchoring, and (iii) inter-sentential cohesion devices. We call this Enhanced LMR (E-LMR). We contend that these relations are not abstract add-ons, but are lexically realized and should be represented accordingly. Rather than a full annotation scheme, we introduce design principles and representation strategies illustrated with European Portuguese data. A lexically grounded approach improves interpretability, annotation consistency, and cross-linguistic robustness. We thus position E-LMR as a viable alternative within a modular semantic architecture where discourse relations are explicitly encoded while remaining tightly coupled to lexical form.

2012 ACM Subject Classification Computing methodologies → Discourse, dialogue and pragmatics

Keywords and phrases Inter-sentential relations, Lexicalized Meaning Representation, Nominal and Pronominal Anaphora Resolution, Temporal Anaphora Resolution, Inter-Sentential Cohesion Devices, European Portuguese

Digital Object Identifier 10.4230/OASICS.SLATE.2026.5

Supplementary Material *Dataset:* <https://doi.org/10.13140/RG.2.2.29402.50888>

Funding *Jorge Baptista:* Work supported by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT) under projects UID/50021/2025 (DOI: <https://doi.org/10.54499/UID/50021/2025>) and UID/PRR/50021/2025 (DOI: <https://doi.org/10.54499/UID/PRR/50021/2025>).

1 Introduction

Abstract Meaning Representation (AMR) has been central to the development of graph-based semantic representations, providing a structured and largely language-agnostic encoding of sentence-level meaning [1, 2, 22, 27]. In this framework, meaning is primarily located in the domain of predication, that is, in the relations between a predicative operator and its arguments. Core semantic relations are encoded by `:argi`-labeled edges, while non-core relations capture circumstantial information such as location, time, cause, or concession. However, its design principles – particularly its focus on intra-sentential structure and its abstraction from surface lexical material – limit its ability to capture discourse-level phenomena. Recent extensions, most notably Uniform Meaning Representation (UMR) [11, 12, 13, 14, 15, 28], address these limitations by introducing mechanisms for cross-sentential linking, including coreference, temporal anchoring, and discourse relations.

Within this evolving landscape, we argue for a complementary, lexically grounded approach to inter-sentential semantics within Lexicalized Meaning Representation (LMR) [7, 9, 10]. While LMR prioritizes lexical realization and sentence-level anchoring, it lacks a fully



© Jorge Baptista;
licensed under Creative Commons License CC-BY 4.0

15th Symposium on Languages, Applications and Technologies (SLATE 2026).

Editors: Fernando Batista, Eugénio Ribeiro, Ricardo Ribeiro, and André L. Santos; Article No. 5; pp. 5:1–5:18

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

articulated model for representing relations across sentence boundaries. We therefore propose a programmatic extension, Enhanced LMR (E-LMR), designed to capture inter-sentential relations while preserving lexical transparency and structural interpretability.

We focus on three domains: (i) nominal and pronominal anaphora, (ii) temporal anchoring and temporal anaphora [17, 21], and (iii) inter-sentential cohesion devices, including discourse adverbials and connective structures. Building on insights from UMR [13, 28], we provide preliminary evidence how these phenomena can be represented in a lexically anchored framework using PENMAN graphs [16], without excessive abstraction or loss of alignment with surface form.

In this framework, both in AMR and LMR, sentences are represented as PENMAN graphs, lexical units are represented as graph nodes, while syntactic-semantic relations are encoded as labeled edges. Predicative elements introduce argument structures, typically represented through `:argi` relations, while modifiers, possession, coordination, and co-reference are represented through dedicated edge labels. The representation remains closely aligned with surface lexical realization (see Fig. 1 for an illustration). Particular attention is given to lexical items, such as pronouns, temporal expressions, and discourse markers, as carriers of inter-sentential relations. The E-LMR extension to capture discourse relations also uses the same PENMAN graph formalism, making integration of sentence-level annotations straightforward.

Although the examples discussed here are drawn from European Portuguese, the proposed mechanisms are not language-specific. On the contrary, the explicit lexical anchoring adopted in E-LMR is expected to facilitate adaptation to morphologically rich languages, where reference and discourse relations are frequently encoded through inflectional or function-word material.

Rather than proposing a fully formalized annotation scheme, this paper advances a principled set of design hypotheses and representation strategies, illustrated with examples from Portuguese. We argue that a lexically grounded approach to discourse representation offers advantages in interpretability, annotation consistency, and cross-linguistic adaptability, especially for languages with rich morphological and syntactic variation, such as Portuguese.

We position E-LMR as a lexically grounded alternative within a modular semantic architecture, where sentence-level representations remain text-anchored and inter-sentential relations are explicitly encoded. This contributes to ongoing discussions on the balance between abstraction and lexical specificity and on the role of discourse-level semantics in NLP.

The remainder of the paper is structured as follows. Section 2 motivates and defines E-LMR, focusing on nominal and pronominal anaphora, temporal anchoring, and discourse relations. Section 3 presents the annotation methodology. Section 4 reports the main results. Section 5 concludes and outlines future work.

2 Discourse modeling: Multiple relations across sentences

A variety of linguistic devices contribute to the cohesion and coherence that turn a sequence of sentences into a discourse, that is, a meaningful and interpretable text [18, 19, 20]. These include utterance structure, dialogue organization, pragmatic functions, speakers' intentions, and cohesive mechanisms such as reference, ellipsis, conjunction, and lexical relations.

These dimensions operate at different levels of abstraction and with varying degrees of explicit linguistic realization. In this paper, we focus on inter-sentential cohesion devices, which provide a linguistically grounded entry point for modeling discourse relations within semantic representation frameworks.

To this end, we argue for a programmatic extension of the Lexicalized Meaning Representation (LMR) framework [9], aimed at systematically capturing inter-sentential relations within a lexically grounded approach to discourse semantics. We call this *Enhanced-LMR* (E-LMR).

Originally inspired in Abstract Meaning Representation (AMR) [2] for encoding intra-sentential relations, LMR remains limited if confined to the sentence level. We contend that any adequate meaning representation must extend to inter-sentential structure, where reference, temporality, and discourse connectivity are constructed and resolved.

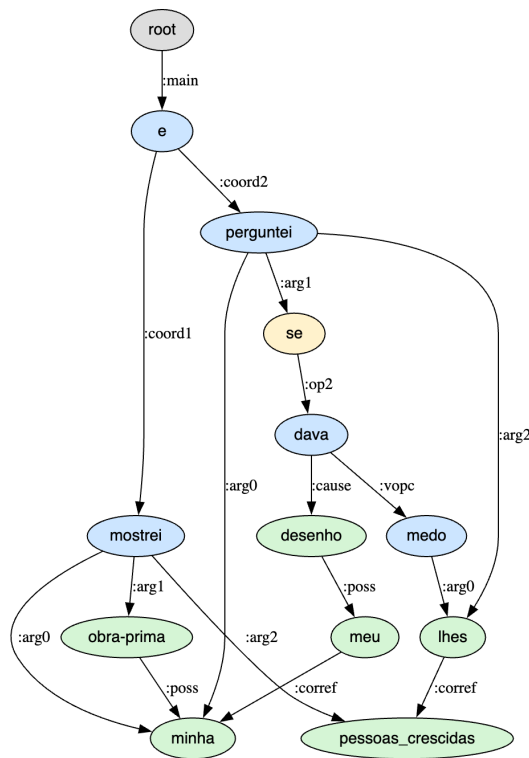
In this respect, the extension from intra- to inter-sentential representation is not merely desirable, but theoretically necessary. Still, while E-LMR relies on lexical anchoring, the relations encoded are not purely lexical associations. Rather, lexical items function as explicit access points to discourse-level semantic interpretation, including reference resolution, temporal anchoring, and discourse connectivity. Certain inferential relations requiring extensive world knowledge or pragmatic reasoning remain, however, outside the scope of the current framework.

This position aligns with recent developments extending AMR into Uniform Meaning Representation (UMR), which incorporate cross-sentential phenomena such as coreference, temporal anchoring, and discourse relations [14, 11, 12, 13, 15, 28]. However, while UMR advances discourse modeling through increasingly abstract and layered representations, it tends to weaken the link to surface lexical realization. In contrast, LMR maintains a principled commitment to lexical transparency and text anchoring. It therefore requires an alternative path to inter-sentential modeling – one that encodes discourse relations explicitly while preserving their lexical grounding. This motivates the approach proposed in this paper.

2.1 Inter-sentential, nominal and pronominal, anaphora resolution

The LMR guidelines [9] adopt a minimal annotation strategy aimed at reducing annotator cognitive load by encoding only elements explicitly realized in the sentence, almost without reconstructing zeroed material. Still, controlled reconstruction of linguistically recoverable zeroed material is proposed in [3]. Part-of-speech tagging and word-sense disambiguation are deferred to a post-processing stage unless required for specific annotation decisions. For word-sense disambiguation of predicative elements, we rely on the Lexicon-Grammar of European Portuguese [4, 5, 6, 7].

In both AMR and LMR, intra-sentential anaphora is resolved at the annotation level by assigning multiple semantic relations to a single variable, instantiated only once in the sentence graph. This is illustrated in Fig.1. (Unless otherwise indicated, examples are drawn from the European Portuguese translation of *O Príncipezinho*, based on the French original of Antoine de Saint-Exupéry’s *Le Petit Prince* (*The Little Prince*, 1943). Examples are identified by sentence id; English translations are our own.) In this Figure, several edges point to the node *pessoas crescidas* “grown-ups”. As in many null-subject languages, Portuguese does not systematically realize grammatical subjects at the surface level. This is the case for the first-person singular subjects of *mostrei* “showed” and *perguntei* “asked” (here referring to the speaker/narrator). In this sentence, however, the referent is overtly realized by the possessive pronoun *minha* “my_fem.sg.”, and the zeroed subjects are therefore linked (*arg0*) to this explicit node. The possessive *meu* “my_masc.sg.” co-refers to the same entity, yielding an explicit *:corref* relation. In both cases, these pronouns refer to the intra-diegetic narrator. Similarly, the pronoun *lhes* “to_them” is linked to *pessoas crescidas* “grown-ups” via a *:corref* relation. While narrator reference operates at the discourse level, the resolution of *lhes* remains local to the sentence.



■ **Figure 1** LMR graph of the sentence: *Mostrei a minha obra-prima às pessoas crescidas e perguntei-lhes se o meu desenho lhes dava medo* [opz-id-10b] “I showed my masterpiece to the grown-ups and asked them if my drawing scared them.”

Inter-sentential reference resolution is significantly more complex. In what follows, we outline the main components of this task.

We propose that each sentence-level LMR graph remains an autonomous representation, while being embeddable within a higher-level discourse graph expressed in PENMAN format. At the discourse level, a root node DISCOURSE provides a global anchor and introduces a small set of (default) metalinguistic parameters, to encode speaker (SPKR), addressee (ADRS), location (HERE), and utterance time (NOW), and serve as stable referential anchors across sentences.

Sentence-level graphs are inserted into the discourse graph in textual order, determined by their identifiers. At an initial stage, they are linked by a generic relation such as `:and`, encoding linear succession while leaving the precise discourse relation underspecified.

This structure enables a second processing layer in which the discourse graph is enriched with inter-sentential links. For example, a zeroed first-person subject can be resolved via a `:corref` relation linking it to SPKR. Similarly, pronominal, nominal, spatial, and temporal references can be connected to discourse anchors or antecedents introduced in previous sentences. Sentence-level annotation thus remains simple, while discourse-level interpretation is handled incrementally through graph enrichment.

The following pseudo-code illustrates this architecture.

```
(d / discourse
  :SPKR (SPKR / speaker)
  :ADRS (ADRS / addressee)
  :HERE (HERE / here)
  :NOW (NOW / now)
  :AND (s1 / sentence
        :id "10a"
        :content ( ...graph for sentence 10a... ))
  :AND (s2 / sentence
        :id "10b"
        :content ( ...graph for sentence 10b... ))
  :AND (s3 / sentence
        :id "10c"
        :content ( ...graph for sentence 10c... )))
```

Once sentence id="10b" is integrated into the discourse graph, the first-person reference can be resolved by introducing a `:corref` relation linking it to SPKR, e.g., `(m / minhas :corref SPKR)`. This architecture provides a lexically grounded alternative to discourse-level representation, avoiding the need for abstract intermediate layers.

In the case of a simple zeroed subject, as illustrated in sentence (1):

(1) *Voei um pouco por todo o mundo.* [id=opz-22-1] "I flew all over the world."

the initial LMR annotation therefore contains only the overt material:

```
(root :main (v / voei
  :path (p / por :op2 (m / mundo
    :quant (t / todo
      :quant (up / um_pouco))))))
```

After post-processing, the graph is enriched with lemmatized, morphosyntactic, and word-sense information, yielding the E-LMR representation shown below:

```

(root
  :main (v / voei
    :lemma (lem1 / voar)
    :pos (pos1 / verb
      :tense (ten1 / past)
      :mode (mod1 / ind)
      :pers (per1 / 1)
      :nb (num1 / sg))
    :wsd (wsd1 / 37LD)
    :arg0 SPKR
    :path (prep1 / por
      :op2 (m1 / mundo
        :lemma (lem2 / mundo)
        :pos (pos2 / noun
          :subcat (sub1 / common)
          :gen (gen1 / m)
          :nb (num2 / sg))
        :quant (t1 / todo
          :lemma (lem3 / todo)
          :pos (pos3 / pron
            :subcat (sub2 / indef)
            :gen (gen2 / m)
            :nb (num3 / sg))))))
    :quant (up1 / um_pouco
      :lemma (lem4 / um_pouco)
      :pos (pos4 / adv
        :subcat (sub3 / mq-i))))))

```

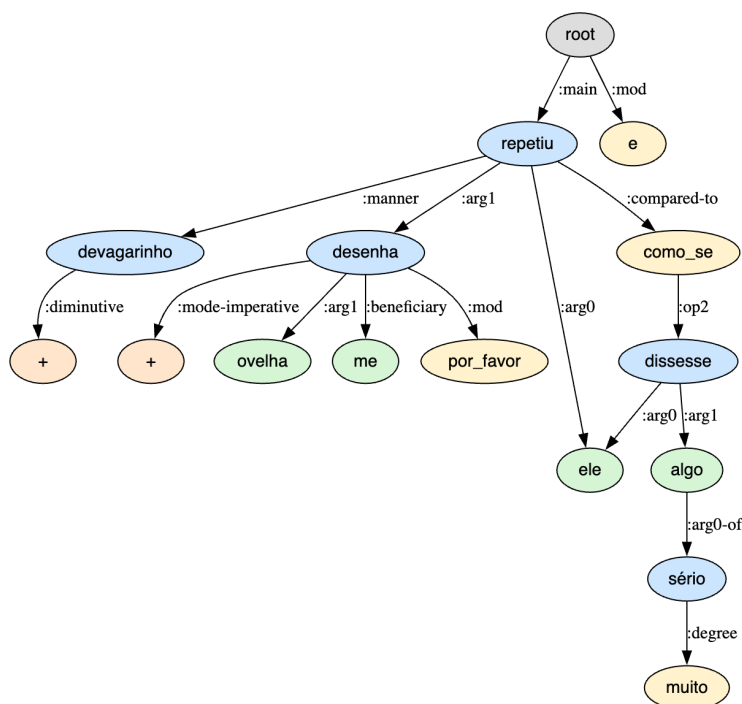
In this representation, the zeroed subject of *voar* “fly” is explicitly instantiated and resolved as coreferential with the discourse speaker (SPKR). Under the imperative mode, as in (2), the zeroed addressee (*você*s “You_pl.”) can be reconstructed by linking the verb to ADRES, e.g., (*i* / *imaginem* :arg0 ADRES). Similarly, the first-person singular (*eu* “I”) zeroed passive agent is linked to the narrator, e.g., (*d* / *despertado* :arg1 SPKR).

- (2) *Assim, **imaginem** [*você*s] a minha surpresa quando, ao amanhecer, [*eu*] fui **despertado** por uma vozinha muito curiosa.* [id=opz-43] “So, imagine my surprise when, at daybreak, I was woken up by a very curious little voice.”

We further exploit direct-speech annotation to distinguish narrator-level utterances from character-level utterances, which may occur in either reported (indirect) or direct speech. Consider the next example (3) and the corresponding LMR graph, presented in Fig.2.

- (3) *E ele repetiu, devagarinho, como se dissesse algo muito sério: – **Por favor... desenha-me uma ovelha.**...* [id=opz-60] “And he repeated, softly, as if he were saying something very serious: – Please... draw me a sheep...”

Since the embedded clause is in direct speech (here in the imperative) the beneficiary is interpreted as coreferential with the subject of the “verb of saying” (*verbum dicendi*) *repetir* “repeat”. This licenses the insertion of the edge (*m* / *me* :corref *e2*), where *e2* denotes *ele* “he”. As for the subject of *desenhar* “draw”, a subsequent annotation module reconstructs the implicit argument slot, yielding (*d* / *desenha* :arg0 SPKR).



■ **Figure 2** LMR graph of the sentence OPZ-ID-60: *E ele repetiu, devagarinho, como se dissesse algo muito sério: – Por favor... desenha-me uma ovelha... “And he repeated, softly, as if he were saying something very serious: – Please... draw me a sheep...”*

The more general case of inter-sentential *pronominal anaphora* resolution is illustrated in (4)–(5). In such cases, the antecedent of the pronoun *ele* in (5) typically occurs in a nearby sentence, in this case, *rapazinho* in (4). Unresolved anaphors are those lacking an explicit **:corref** relation. An automatic anaphora resolution module [8, 23, 26] may be applied at this stage, with subsequent manual correction.

- (4) *E vi um **rapazinho** extraordinário a fitar-me com ar sério.* [id=opz-50] “And I saw an extraordinary little boy looking at me seriously.”
- (5) *Este é o melhor retrato que, mais tarde, consegui fazer dele.* [id=opz-51] “This is the best portrait [that] I was able to make of him later.”

Lexical anaphora, where a noun refers to a previously introduced entity, poses a greater challenge. In (6), the noun *aparição* “apparition” refers to *rapazinho*, also in (4).

- (6) *Olhei para aquela **aparição** com os olhos arregalados de espanto.* [id=opz-55] “I looked at that apparition with wide eyes of astonishment.”

Such cases typically require world knowledge and contextual interpretation, and are therefore less amenable to automatic resolution, demanding human intervention. This case is then annotated as (a / **aparição :corref id-50_r**), where the compound reference **id-50_r** stands for the previous occurrence of *rapazinho*, and consists of the sentence id and the variable instantiating the antecedent noun.

Finally, certain omitted arguments are reconstructed from lexical information available in the Lexicon-Grammar entries of the relevant predicates. This occurs in the two consecutive sentences shown in Examples (7)–(8), where the unexpressed subject of *representar* “represent”

in (8) refers to the same discourse entity as the subject of the previous occurrence of that verb in (7). To represent this dependency explicitly, an artificial [zero] node is introduced in the corresponding argument position and linked to its antecedent through a coreference relation.

- (7) *O meu **desenho** não representava um chapéu.* [id=opz-13] “My drawing did not depict a hat.”
- (8) *Representava uma jiboia a digerir um elefante.* [id=opz-14] “It depicted a boa constrictor digesting an elephant.”

```
(root :main (r / representava
  :arg0 (z / [zero] :corref id-13_d)
  :arg1 (j / jiboia
    :arg0-of (d / digerir
      :arg0 j
      :arg1 (e / elefante))))))
```

At this stage, we envisage an annotation interface that performs: (i) automatic lemmatization, PoS tagging, and word-sense disambiguation (with human-in-the-loop when needed); (ii) insertion of placeholder nodes and edges based on Lexicon-Grammar argument structure; (iii) resolution of references to discourse anchors (SPKR, ADRS, HERE, NOW); and (iv) manual resolution of remaining anaphora.

The proposed framework is intended as a semi-automatic enrichment architecture. Several components are amenable to automatic processing, including lemmatization, morphosyntactic tagging, argument-slot instantiation based on Lexicon-Grammar resources, temporal expression recognition, and candidate antecedent detection for anaphora resolution. The resulting discourse graph can therefore be incrementally enriched through a pipeline combining rule-based modules, statistical resolvers, and human validation.

2.2 Inter-sentential temporal anchoring and temporal anaphora

Temporal anaphora involves identifying events, analyzing tense and mode, handling clause structure, and interpreting temporal modification [21, 29, 30, 31, 32].

In LMR, temporal expressions are annotated as single nodes with basic semantic types: *date* (temporal location), *duration*, and *frequency* (combining modulo and duration values). Examples, taken from the novella, include *ao amanhecer* “at dawn”, *mais tarde* “later”, *em 1920* “in 1920” (date); *para sempre* “forever” (duration); and *às vezes* “sometimes” (frequency). For temporal anaphora, *date* expressions are the most relevant. A temporal named-entity recognition (NER) system should be able to identify and annotate such expressions with their corresponding types and normalized values [17].

Consider the temporal expression *ao amanhecer* “at dawn”, already appearing in (2). It is annotated as `NE_TEMPO_T-DATE_T-REF-ENUNC(a_o amanhecer)`, indicating a referential-relative temporal expression anchored to the time of enunciation. The event *despertar* “wake” is thus located at a time of day whose precise reference depends on its relation to the time of enunciation. Since the temporal expression and the event occur in the same sentence, the annotation reduces to encoding its semantic type and reference.

```
... :time (qd / quando :op2 (d / despertado
  :vaux (f / fui)
  :time (aa / ao_amanhecer
    :lemma (lem1 / ao_amanhecer)
```

```

      :pos (pos1 / adv
        :type (type / date)
        :ref (ref / enunc)))
    ...

```

When the antecedent of a temporal expression occurs in a previous utterance, the annotation links it to the corresponding discourse segment. For example, in (9), the expression *nesse momento* “at that moment” refers to a preceding stretch of dialogue and is encoded via a *t-ref* relation to that segment.

- (9) *Nesse momento, um clarão iluminou o mistério da presença dele, o que me fez perguntar bruscamente: – Então vieste de outro planeta?* [id=opz-123] “At that moment, a flash of insight illuminated the mystery of his presence, which made me abruptly ask: – So, you came from another planet?”

```

(root :main (i / iluminou
  :time (nm / nesse_momento
    :lemma (lem1 / nesse_momento)
    :pos (pos1 / adv
      :type (type / date)
      :ref (ref / enunc :t-ref id-121-id-122)))
  ...

```

Consider now (10) and the temporal expression *mais tarde* “later”, modifying the event *fazer um retrato* “make a portrait”.

- (10) *Este é o melhor retrato que, **mais tarde**, consegui fazer dele.* [id=opz-51] “This is the best portrait [that] I was able to make of him later.”

The event expressed by the support-verb construction *fazer retrato* “make portrait” in (4) is located *after* some narrated events and *before* the completion of the narration as a whole, values that are encoded by *t-after* and *t-before*.

```

(root :main (r1 / retrato
  ...
  :vsup-of (f / fazer
    :vaux (c / consegui)
    :time (mt / mais_tarde
      :lemma (lem1 / mais_tarde)
      :pos (pos1 / adv
        :type (type / date)
        :ref (ref / enunc :t-after id-50 :t-before NOW)))
  ...

```

While *t-before* is straightforward, determining the scope of *t-after* is more difficult. By default, it is linked to the immediately preceding sentence.

Absolute dates, such as *em 1920* “in 1920”, are mapped to normalized temporal values (simplified here as [value]).

- (11) *O astrónomo repetiu a sua apresentação **em 1920**, vestido com um fato muito elegante.* [id=opz-158] “The astronomer repeated his presentation in 1920, dressed in a very elegant suit.”

```
(root :main (a1 / apresentação
  :vsup (r / repetiu)
  :arg0 (a / astrónomo)
  ...
  :time (d / em_1920
    :type (type / date)
    :ref (ref / absol :t-ref [value])))
```

These mechanisms provide evidence that temporal anchoring can be modeled within a lexically grounded framework through explicit linking, without requiring abstract temporal layers.

2.3 Other inter-sentential cohesion devices

Conjunctive adverbs are a major device of inter-sentential cohesion [24, 25], e.g., *assim* “thus”, *consequentemente* “consequently”, *contudo* “nevertheless”, *então* “then”, *no entanto* “however”, and *portanto* “therefore”. In LMR, such sentence-external modifiers are represented as modifiers of the **root** node, enabling inter-sentential representation. Each adverb is associated with a semantic tag drawn from 13 categories, including *additive*, *concessive*, *conclusive*, *consecutive*, *continuative*, *contrastive*, *exemplificative*, and *temporal*.

Consider the conjunctive adverbs *mesmo assim* “even so/yet”, *pois* “indeed”, and *no entanto* “however”, illustrated in the following dialogue, examples (12)–(14):

- (12) – *As pessoas do teu planeta – disse o príncipezinho – cultivam cinco mil rosas num único jardim... e **mesmo assim** não encontram aquilo que procuram...* [id=opz-1307] ‘ – The people on your planet – said the little prince – grow five thousand roses in a single garden... and yet they do not find what they are looking for...’
- (13) – ***Pois**, não encontram... – repliquei.* [id=opz-1308;1309] ‘ – Indeed, they do not find it... – I replied.’
- (14) – ***No entanto**, o que procuram pode ser encontrado numa só rosa ou num pouco de água...* [id=opz-1310] ‘ – And yet, what they are looking for can be found in a single rose or in a little water...’

These adverbs are represented in LMR as modifiers of the **root** node or of the highest predicate in the clause they modify. Crucially, however, this representation does not encode any explicit link between the current sentence and preceding discourse segments.

In E-LMR, by contrast, discourse relations are explicitly encoded through the combination of the adverb’s syntactic-semantic class (**advconj**), its semantic category, and dedicated discourse-linking relations. Specifically, a **:d-ref** edge connects the current sentence (or predicate) to the relevant element in the preceding discourse, while **:d-ref-type** specifies the nature of the semantic relation.

```
(root :main (d / disse ...
  :arg1 (e1 / e :direct-speech +
    :coord1 (c / cultivam ...
    :coord2 (e2 / encontram
      :mod (ma / mesmo_assim
        :lemma (lem1 / mesmo_assim)
        :pos (pos1 / adv
          :subcat (sub1 / advconj
            :d-ref-type (type1 / concessive)
            :d-ref c )))
    ...
```

In this example, representing (12), the adverb *mesmo assim* “even so/yet” links the main predicate of the second coordinated clause (*encontram* “find”) to that of the first (*cultivam* “grow”), establishing a *concessive* relation. The adverb thus operates within a single syntactic sentence.

Considering example (13) and the adverb *pois* “indeed”:

```
(root :main (r / repliquei ...
  :arg1 (e / encontram :direct-speech +
    :mod (p / pois
      :lemma (lem1 / pois)
      :pos (pos1 / adv
        :subcat (sub1 / advconj
          :d-ref-type (type1 / justifier)
          :d-ref id-1307-e2)))
    ...
```

In this example, the adverb *pois* links *encontram* “find” to its previous occurrence in the preceding sentence, identified by the sentence id-1307 and the variable e2 within.

Finally, regarding example (14) and the adverb *no entanto* “yet/however”:

```
(root :mod (ne / no_entanto
  :lemma (lem1 / no_entanto)
  :pos (pos1 / adv
    :subcat (sub1 / advconj
      :d-ref-type (type1 / contrastive)
      :d-ref id-1307)))
  ...
```

The adverb *no entanto* “yet/however”, represented as a modifier of the **root** node, links the entire sentence to a preceding utterance of the same speaker, establishing a *contrastive* relation.

In these examples, adverbs establish relations between distinct utterances. While *pois* links consecutive sentences, *no entanto* connects the current sentence to a non-adjacent one.

Another major cohesive device is the use of conjunctions such as *e* “and” (15) and *mas* “but” (16) at the beginning of sentences, where they function similarly to conjunctive adverbs. We therefore reclassify these elements as conjunctive adverbs and assign them discourse functions, namely *continuative* and *contrastive*.

(15) *E a geografia, é verdade, ajudou-me muito.* [id=opz-22-2] “And geography, it’s true, helped me a lot.”

(16) *Mas, tal como os anteriores, foi rejeitado.* [id=opz-85] “But, like the previous ones, it was rejected.”

In this case, we reclassify these conjunctions as adverbs and assign them the corresponding discourse functions, namely *continuative* and *contrastive*, respectively.

These cases indicate that discourse relations can be directly encoded through lexical items and their distribution, supporting a lexically grounded approach to inter-sentential semantics.

3 Evaluation

To evaluate the proposed E-LMR extension, we consider four dimensions: interpretability, lexical grounding, discourse coverage, and annotation feasibility. These dimensions were assessed qualitatively through manual inspection of the enriched graphs and quantitatively

through counts of introduced relations, reconstructed arguments, and discourse-level links. Rather than benchmarking against existing frameworks, this preliminary study aims to determine whether the proposed mechanisms can be systematically implemented within a lexically grounded representation. Given the absence of comparable UMR annotations and mature annotation tools, we conducted an exploratory annotation experiment focused on assessing the feasibility of lexically grounded discourse enrichment.

To this end, we selected the first two chapters of the European Portuguese version of *O Príncipezinho*. This small corpus comprises 102 sentences, including both narrative passages and dialogue. We leveraged Lexicon-Grammar resources for verbs and predicative nouns to perform word-sense disambiguation and to instantiate zeroed argument slots (e.g., subjects and essential complements).

The annotation process followed a staged enrichment procedure. Annotators first produced sentence-level LMR graphs based on overt lexical material.¹ In a second stage, these graphs were enriched with lexical, morphosyntactic, and discourse-level information, including reconstructed argument slots, co-reference relations, temporal anchoring, and discourse links.²

We then enriched the LMR graphs to comply with the E-LMR guidelines described in Section 2, introducing new nodes and edges for inter-sentential features, including co-reference :**corref**, temporal reference :**t-ref**, and discourse relations **d-ref**. Grammatical features were not the focus, as they can be added mechanically.

This setup allows us to assess how E-LMR captures inter-sentential structure while preserving lexical anchoring, and to evaluate its feasibility on a linguistically rich, albeit small, corpus.

The annotation was performed independently by two linguists familiar with LMR, using task-specific guidelines. Given the exploratory nature of the study and the ongoing refinement of the guidelines, disagreements were resolved through adjudication and inter-annotator agreement is left for future work.

Differences arose mainly from: (i) omitted argument slots; (ii) missing :**corref** relations; (iii) uncertainty in connective’s semantics; (iv) ambiguity in anaphoric or discourse targets; and (v) formal inconsistencies (e.g., indentation, spacing, typos).

Given the exploratory nature of this study and the consensus achieved between annotators, inter-annotator agreement is not reported at this stage. This will be addressed in future work once the guidelines and annotation tools have reached a more stable level of maturity.

4 Results

We now present the main results of the annotation task, followed by a discussion of selected problematic cases.

Arguments. The LMR corpus contains 306 argument relations, including 38 :**arg_i-of** relations (relative subclauses and adjectival modifiers). We introduced 66 additional relations, yielding a total of 372 argument slots. Of the instantiated argument slots, 55 are assigned to SPKR, 8 to previously introduced variables, and 3 to ADRS. These figures suggest that a substantial portion of the semantic connectivity in the text emerges from referential links extending across sentence boundaries, forming a dense discourse structure that is not fully

¹ <https://doi.org/10.13140/RG.2.2.15980.73607>

² <https://doi.org/10.13140/RG.2.2.29402.50888>

captured by sentence-level representations alone. These results also provide preliminary evidence that inter-sentential relations can be systematically encoded within a lexically grounded framework.

Co-reference. Regarding coreference, 31 nodes already carried `:corref` relations, while 89 were newly introduced (32 to SPKR, 57 to previous nodes). The predominance of SPKR-oriented links reflects the narrative structure of the text, in particular its first-person perspective and is likely to differ across genres. Argument omission at the surface level is, then, relatively frequent in this corpus. When a coreferential anaphor is overtly realized, omitted arguments are already represented in LMR. Otherwise, artificial **zero** nodes are introduced (22 cases).

Temporal reference. Temporal reference is annotated only when triggered by explicit temporal expressions, primarily temporal adverbs and temporal named entities. Temporal interpretation derived from tense and mood is not addressed at this stage and is left for future work. The corpus contains 14 `:time` relations: 5 involving adverbs and 9 involving phrases or subclauses introduced by temporal prepositions or conjunctions. No absolute referential dates were found. All relative temporal expressions are anchored to the enunciation, typically to the main event of the clause.

Discourse relations. Discourse relations are annotated primarily when signalled by explicit connective elements. A total of 19 `:d-ref-type` relations were introduced: 1 conclusive, 5 consecutive, 8 continuative, and 5 contrastive. In addition, 6 instances of `:direct-speech` + mark clauses realized in direct speech, typically under a *verbum dicendi*, and 2 instances of `:direct-speech-cont` + indicate continuation within the same discourse turn.

These results provide preliminary evidence that discourse relations can be systematically encoded through overt lexical cues, reinforcing a lexically grounded approach to inter-sentential semantics.

The annotation process also revealed systematic limits of text-internal discourse representation, particularly in cases involving underspecified deixis, extra-textual reference, and world-knowledge-dependent interpretation.

Indefinites. One noteworthy case concerns argument positions with *indefinite* reference (4 instances), as exemplified in (17). Although verbal inflection licenses the reconstruction of a first-person plural subject *nós* “we”, the recovered argument does not refer to a specific plural entity or group. Rather, it receives a generic interpretation, denoting an indefinite set of individuals that includes the speaker as a representative member.

- (17) *Elas precisam sempre que [nós] lhes explicuemos as coisas.* [id=opz-16] “They always need us to explain things to them.”

This is the same generic use of first-person plural *nós* “we” that appears overtly in sentences such as (18).

- (18) *Quando um mistério nos deixa demasiado espantados, não nos atrevemos a desobedecer.* [id=opz-61] “When a mystery leaves us too astonished, we do not dare to disobey.”

In other cases, the indefinite referent corresponds to a zeroed argument (19), recovered from the predicate syntactic-semantic represented in the Lexicon-Grammar.

- (19) *Responderam-me: “Porque é que um chapéu haveria de dar medo [a alguém]?”*
[id=opz-11-12] “They replied to me, ‘Why would a hat be scary?’”

This leads us to introduce an INDEF node in the annotation scheme: (z / [zero] :corref INDEF), for the zeroed arguments (17) and (19), and (n / nos :corref INDEF) for the overt pronouns of (18).

These cases constitute evidence that indefiniteness can be encoded through morphosyntactic cues without requiring explicit discourse anchors.

Not-defined. In a number of cases (12), no clear antecedent could be assigned to an explicit anaphor, highlighting limits of strictly text-internal reference resolution.

- (20) *Que era assim:* [id=opz-10a] “It went like this.”
(21) *Aqui fica uma cópia dessa gravura.* [id=opz-4] “Here is a copy of that illustration.”

In these cases, deictic expressions refer to extra-textual objects/entities (e.g., illustrations) rather than to discourse-internal nodes. These cases are represented by a **not-defined** node.

A more complex situation is illustrated in (22):

- (22) *Quando finalmente consegui falar, perguntei-lhe: – Mas... que estás a fazer aqui?*
[id=opz-59] “When I finally managed to speak, I asked him: – But... what are you doing here?”

Although the previously introduced noun *deserto* “desert” could in principle serve as an antecedent for the adverb *aqui* “here”, its reference remains underspecified and does not clearly map to the discourse-level variable *HERE*.

In other cases, the pronoun *isso* in expressions such as *por isso* “because of that/thus” cannot be straightforwardly mapped onto a specific discourse node, segment or discourse span, reflecting a form of underspecified or diffuse reference, as in example (23).

- (23) *Por isso, tive de escolher outra profissão, e aprendi a pilotar aviões.* [id=opz-21]
“So, I had to choose another profession and learned to fly airplanes.”

In this context, the pronoun refers to the reaction of the *peessoas crescidas* “grown-ups” to the narrator’s attempts to pursue a drawing career, which had been elaborated across an extended stretch of the preceding discourse.

Similarly, some reconstructed **zero** nodes lack a clear referential target, like in example (24):

- (24) *Ele analisou com toda a atenção e disse: – Não!* [id=opz-76] “He examined it very carefully and said: ‘No!’”

In this instance, the object of *analisar* “analyse” can be interpreted as referring to a drawing, even though no suitable antecedent is found in the immediately preceding discourse. Instead, this interpretation is inferred from the overall dialogue context, where the Little Prince repeatedly requests drawings and the narrator successively produces a series of figures. These cases are therefore encoded as **not-defined**, deferring resolution to a later stage.

These phenomena highlight the limits of strictly text-anchored representations, as some discourse relations can only be resolved by drawing on contextual or world knowledge that is not explicitly encoded in the text.

The results of the annotation experiment provide preliminary support for the four evaluation dimensions introduced above. In terms of *interpretability*, the use of lexically grounded nodes and explicit relations (e.g., `:corref`, `:d-ref`, `:t-ref`) yields representations that remain closely aligned with the surface text, facilitating human inspection and manual validation.

Regarding *lexical grounding*, the analysis confirms that a substantial portion of inter-sentential relations – particularly anaphora and discourse connectivity – can be directly anchored in overt lexical material, such as pronouns, temporal expressions, and conjunctive adverbs.

With respect to *discourse coverage*, the enrichment process suggests that key inter-sentential phenomena, including reference resolution, temporal anchoring, and discourse relations, can be systematically encoded within the proposed framework, albeit with some limitations in cases requiring extra-textual knowledge or underspecified reference.

Finally, in terms of *annotation feasibility*, the observed disagreements between annotators highlight the main sources of difficulty, namely, omitted arguments, underspecified anaphora, and conjunctive adverbs’ semantic interpretation; while also suggesting that these issues can be resolved through clearer guidelines and annotation tool support.

Taken together, these results indicate that E-LMR may provide a viable basis for extending lexically grounded semantic representations to the discourse level while maintaining a balance between expressive power and annotation tractability.

5 Conclusion and Future Work

This paper has argued that inter-sentential relations constitute a necessary extension of semantic representation frameworks and that such an extension can be achieved without abandoning lexical anchoring. Building on the Lexicalized Meaning Representation (LMR) framework, we proposed Enhanced LMR (E-LMR) as a programmatic approach to capturing discourse-level phenomena, focusing on nominal and pronominal anaphora, temporal anchoring, and inter-sentential cohesion devices. E-LMR is conceived not as a purely manual annotation formalism, but as a modular semantic enrichment architecture supporting incremental automation.

The annotation experiment, conducted on a small corpus of European Portuguese narrative text, provides initial evidence that these relations can be systematically encoded while preserving the core principles of LMR, namely lexical transparency and structural interpretability. The results show that a significant proportion of semantic information, particularly argument realization, coreference, and discourse connectivity, emerges only when moving beyond the sentence level. At the same time, the analysis highlights the complexity of certain phenomena, such as lexical and underspecified anaphora, temporal anchoring across discourse spans, and the interpretation of connective elements, which often require contextual or world knowledge.

Rather than proposing a fully specified annotation scheme, this work advances a set of design principles and representation strategies that suggest the feasibility and potential of a lexically grounded approach to inter-sentential semantics. In contrast to more abstract, layered approaches such as UMR, E-LMR maintains a close alignment with surface realization, offering advantages in interpretability and annotation consistency, particularly in morphologically rich languages, such as Portuguese.

Several directions for future work follow naturally from this study. First, the development of a dedicated annotation interface is essential to support the semi-automatic enrichment process envisioned here, including modules for word-sense disambiguation, argument slot

instantiation, and discourse-level linking. Second, a more systematic evaluation is required, including the measurement of inter-annotator agreement and the comparison with alternative frameworks such as UMR. Third, the current treatment of temporal relations should be extended to incorporate tense, aspect, and modality, enabling a more complete model of temporal anchoring. Fourth, the representation of underspecified or extra-textual reference (e.g., **not-defined** nodes) calls for further theoretical and computational refinement, potentially integrating external knowledge sources.

Finally, the generality of the proposed approach should be tested across different genres and languages, in order to assess its robustness and adaptability. In this respect, E-LMR is best viewed not as a fixed formalism, but as a modular and extensible framework, contributing to a broader ecosystem of interoperable meaning representations. Large language models may also play a role in candidate generation for discourse relations, anaphora resolution, and underspecified reference interpretation. In this respect, E-LMR could provide an interpretable semantic layer for validating or constraining LLM-generated discourse inferences.

References

- 1 Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse (ACL 2013)*, pages 178–186, Sofia, 2013. ACL. URL: <https://aclanthology.org/W13-2322.pdf>.
- 2 Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract Meaning Representation (AMR) 1.2.6 Specification, 2019. URL: <https://github.com/amrisi/amr-guidelines/blob/master/amr.md>.
- 3 Jorge Baptista. Reconstruction of Zeroed Words in Lexicalized Meaning Representation. *Diacrítica*, 2026. Accepted for publication (in press).
- 4 Jorge Baptista and Nuno Mamede. *Dicionário Gramatical de Verbos do Português Europeu*. Universidade do Algarve, Faro, Portugal, novembro 2020. URL: <https://sapiencia.ualg.pt/handle/10400.1/15605>.
- 5 Jorge Baptista and Nuno Mamede. Syntactic Transformations in Rule-Based Parsing of Support Verb Constructions: Examples from European Portuguese. In Alberto Simões, Pedro Rangel Henriques, and Ricardo Queirós, editors, *9th Symposium on Languages, Applications and Technologies (SLATE 2020)*, volume 83 of *OpenAccess Series in Informatics (OASICS)*, pages 11:1–11:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/OASICS.SLATE.2020.11.
- 6 Jorge Baptista, Nuno Mamede, and Sónia Reis. Support Verb Constructions across the Ocean Sea. In *18th Workshop on Multiword Expressions @ LREC2022*, pages 26–36. ACL, 2022. URL: <https://hal.science/hal-03705534>.
- 7 Jorge Baptista, Izabela Müller, and Sónia Reis. Semantic Representation of Adverbs in the Lexicalized Meaning Representation (LMR) Framework. In Jorge Baptista and José Barateiro, editors, *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, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/OASICS.SLATE.2025.9.
- 8 Jorge Baptista, Simone Pereira, and Nuno Mamede. ZAC – Zero Anaphora Corpus: A Corpus for Zero Anaphora Resolution in Portuguese. In Hilário Leal Fontes and António Branco, editors, *Proceedings of the Workshop on Corpora and Tools for Processing Corpora, (co-located with 12th International Conference Computational Processing of the Portuguese Language – PROPOR’2016)*, pages pp. 38–45. PROPOR 2016/Portuguese Language Department, Directorate-General for Translation, European Union, 2016.

- 9 Jorge Baptista, Sónia Reis, João Dias, and Pedro A. Santos. Lexicalized Meaning Representation (LMR). In Claire Bonial, Julia Bonn, and Jena D. Hwang, editors, *Proceedings of the Fifth International Workshop on Designing Meaning Representations (DMR) @ LREC-COLING 2024*, pages 101–111, Torino, Italy, May 2024. ELRA and ICCL. doi:10.34626/6080.
- 10 Jorge Baptista, Sónia Reis, Pedro A. Santos, and João Dias. Comparação de anotações AMR entre línguas: problemas e soluções. *Revista da Associação Portuguesa de Linguística*, 1(11):27–54, 2024. doi:10.26334/2183-9077/rapln11ano2024a2.
- 11 Julia Bonn, Claire Bonial, Matt Buchholz, Hsiao-Jung Cheng, Alvin Chen, Ching-wen Chen, Andrew Cowell, William Croft, Lukas Denk, Ahmed Elsayed, et al. Uniform Meaning Representation 2.0, 2025. LINDAT/CLARIAH-CZ Digital Library.
- 12 Julia Bonn, Matthew J. Buchholz, Jayeol Chun, Andrew Cowell, William Croft, Lukas Denk, Sijia Ge, Jan Hajič, Kenneth Lai, James H. Martin, et al. Building a Broad Infrastructure for Uniform Meaning Representations. In *Proceedings of LREC-COLING*, pages 2537–2547, 2024.
- 13 Julia Bonn, Ching-wen Chen, Andrew Cowell, William Croft, Lukas Denk, Jan Hajič, Kenneth Lai, Martha Palmer, Alexis Palmer, James Pustejovsky, et al. Uniform Meaning Representation, 2023. LINDAT/CLARIAH-CZ Digital Library.
- 14 Julia Bonn, Skatje Myers, Jens E. L. Van Gysel, Lukas Denk, Meagan Vigus, Jin Zhao, Andrew Cowell, William Croft, Jan Hajič, James H. Martin, Alexis Palmer, Martha Palmer, James Pustejovsky, Zdenka Urešová, Rosa Vallejos, and Nianwen Xue. Mapping AMR to UMR: Resources for adapting existing corpora for cross-lingual compatibility. In Daniel Dakota, Kilian Evang, Sandra Kübler, and Lori Levin, editors, *Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023)*, pages 74–95, Washington, D.C., 2023. Association for Computational Linguistics. URL: <https://aclanthology.org/2023.tlt-1.8/>.
- 15 Jayeol Chun and Nianwen Xue. Uniform meaning representation parsing as a pipelined approach. In *Proceedings of TextGraphs*, pages 40–52, 2024.
- 16 Michael Wayne Goodman. Penman: An open-source library and tool for AMR graphs. In Asli Celikyilmaz and Tsung-Hsien Wen, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 312–319, Online, 2020. Association for Computational Linguistics. doi:10.18653/v1/2020.acl-demos.35.
- 17 Caroline Hagège, Jorge Baptista, and Nuno Mamede. Caracterização e Processamento de Expressões Temporais em Português. *Linguamática*, 2(1):63–76, 2010. URL: <https://www.linguamatica.com/index.php/linguamatica/article/download/47/60/>.
- 18 M. A. K. Halliday and Ruqaiya Hasan. *Cohesion in English*. Longman, London, 1976.
- 19 M. A. K. Halliday and Christian M. I. M. Matthiessen. *Halliday's Introduction to Functional Grammar*. Routledge, London, 4 edition, 2014.
- 20 Zellig S. Harris. Discourse Analysis. *Language*, 28(1):1–30, 1952.
- 21 Erhard Hinrichs. Temporal Anaphora in Discourses of English. *Linguistics and Philosophy*, pages 63–82, 1986.
- 22 Behrooz Mansouri. Survey of Abstract Meaning Representation: Then, Now, Future, 2025. arXiv preprint. doi:10.48550/arXiv.2505.03229.
- 23 João Silvestre Marques. Anaphora resolution in portuguese: A hybrid approach. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, November 2013. MSc dissertation.
- 24 Christian Molinier and Françoise Lévrier. *Grammaire des Adverbes: Description des Formes en -ment*. Droz, Genève, 2000. URL: https://books.google.com/books/about/Grammaire_des_adverbes.html?id=HSHMkiPy4pgC.
- 25 I. Müller, N. Mamede, and Jorge Baptista. Adverbios conjuntivos multipalabra en el portugués brasileño: Un enfoque metodológico basado en el léxico-gramática. *Studia Linguistica Romanica*, 2026. Accepted for publication (in press).

- 26 Simone Pereira. ZAC.PB: An Annotated Corpus for Zero Anaphora Resolution in Portuguese. In *Student Research Workshop in conjunction with RANLP-09*, pages 53–59, September 2009. URL: <https://aclanthology.org/R09-2010/>.
- 27 Nathan Schneider, Jeff Flanigan, and Tim O’Gorman. The Logic of AMR. In *Proceedings of the 2015 Conference of the Association for Computational Linguistics (ACL 2015)*, pages 4–5, Beijing, 2015. ACL. doi:10.3115/v1/N15-4003.
- 28 Jens Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O’Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, et al. Designing a Uniform Meaning Representation for Natural Language Processing. *KI – Künstliche Intelligenz*, 35(3):343–360, 2021. doi:10.1007/S13218-021-00722-W.
- 29 Jiarui Yao, Haoling Qiu, Bonan Min, and Nianwen Xue. Annotating Temporal Dependency Graphs via Crowdsourcing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5368–5380, 2020. doi:10.18653/v1/2020.emnlp-main.434.
- 30 Yuchen Zhang and Nianwen Xue. Neural ranking models for temporal dependency structure parsing. *arXiv preprint*, 2018. arXiv:1809.00370.
- 31 Yuchen Zhang and Nianwen Xue. Structured Interpretation of Temporal Relations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3222–3229, Miyazaki, Japan, 2018. European Language Resources Association (ELRA). URL: <https://aclanthology.org/L18-1490>.
- 32 Jin Zhao, Nianwen Xue, Jens E. L. Van Gysel, and Jinho D. Choi. UMR-Writer: A Web Application for Annotating Uniform Meaning Representations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 160–167, Online, 2021. Association for Computational Linguistics. doi:10.18653/v1/2021.emnlp-demo.19.