Lexicalized Meaning Representation (LMR)

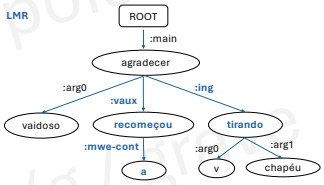
Jorge Baptista [1,2], Sónia Reis [1,2], João Dias [1,2,3], Pedro A. Santos [2,4]
[1]University of Algarve - FCHS/FCT, [2]INESC-ID Lisboa, [3]CISCA, [4]Univ. Lisbon - IST
{jbaptis, smreis, jmdias}@ualg.pt, pedro.santos@tecnico.ulisboa.pt

Abstract

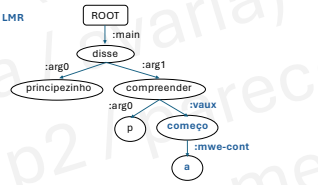
This paper introduces Lexicalized Meaning Representation (LMR), an adaptation of Abstract Meaning Representation (AMR) for European Portuguese. LMR addresses language-specific grammar challenges and linguistic complexities not adequately handled by AMR. It also simplifies aspects like multi-word expressions and named entities while ensuring compatibility with AMR standards, making it suitable for several Natural Language Processing (NLP) tasks.

Abstract Meaning Representation (AMR) (Banarescu *et al.* 2013)	**Lexicalized Meaning Representation (LMR)**
Catalog of senses (semantic predicates) for verbs: OntoNotes, Weischedel, R. et al., 2013. Other categories (nouns and adjectives) are represented by verbal predicates.	Catalog of senses: Lexico-Grammar of Portuguese. Verbs: ViPEr, Baptista (2012, 2013); Dictionary of Portuguese Verb Grammar, Baptista & Mamede (2020a). Predicative nouns: SNIPER, Baptista & Mamede (2020b).
Directed acyclic graphs without root (arc :TOP in loop over the main predicative element node of the sentence).	Directed acyclic graphs but with a root (ROOT) ROOT is connected to the main predicative element (:MAIN), which is the node to which elements with scope over the entire sentence are connected.
Reconstruction of reduced elements.	There is no reconstruction of reduced elements.
Graph representation 'bets' on the text, without a direct relationship with the text forms.	Explicit relationship between text and its representation, treating text forms as nodes of the graph.
Substitution of predicative elements in the text by lemmas (especially verbs represented in OntoNotes).	Maintenance of predicative elements of the text (association of lemmas and construction in post-processing phase).
Substitution of textual elements (especially grammatical) by the semantic relations they express (e.g., conjunctions, prepositions, etc.).	Maintenance of textual elements, explicating semantic relations (e.g., conjunctions, prepositions, etc.).
Does not consider auxiliary verbs, copulative verbs, or support verbs (or light verbs).	Considers all types of auxiliary verbs: verbal auxiliaries – temporal, modal, and aspectual; adjectival auxiliaries – copulative verbs; nominal auxiliaries – support verbs; auxiliaries of passive constructions; also considers constructions with operator verbs – causative operator verb, linking verb, agentive verb (M. Gross 1981, 1998; Baptista 2005).
Representation of multi-word expressions (MWE) of varying complexity; sophisticated representation of mentioned entities (ME), particularly temporal and quantification expressions.	(Very) simplified representation of multi-word expressions (MWE), mentioned entities (ME), and temporal and quantification expressions. Identification of MWE and ME in pre-processing phase and integration as nodes in the LMR graph.
Representation of intra-phrasal anaphoric relations. Extension of notation (O'Gorman et al. 2018) for trans-phrasal anaphoric relations through coreference chains at the text level.	Representation of intra-phrasal anaphoric relations only between explicit elements in the text; anaphora resolution as a post-processing task (trans-phrasal anaphoric relations not yet considered).
Very distinct treatment of verbal predicates (standard representation) and adjectival (:DOMAIN); nominal constructions represented by verbal constructions (if present in OntoNotes).	Homologous representation of arguments of verbal, nominal, and adjectival predicates, corresponding to the standard representation: PRED (:ARG0, :ARG1, ...)

Auxiliary verbs, subordination, gerund



O vaidoso recomeçou a agradecer, tirando o chapéu.
`The vain person started to thank again, tipping his hat.' [TLP id=620]



-- Começo a compreender, disse o principezinho.
`I begin to understand, said the little prince.' [TLP id=1080]

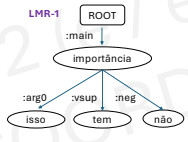
Relative subclauses

```
(g / girl
   :ARG0-of (a / adjust-01
      :ARG1 (m / machine)))
(g / girl
   :ARG0-of (a / adjust-01
      :ARG0 (w / who))
      :ARG1 (m / machine)))
```

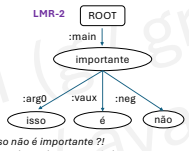
AMR
LMR

The girl who adjusted the machine

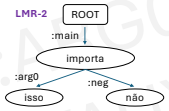
Equivalent representation of Verbal, Nominal and Adjectival predicates



Isso não tem importância.
`That has not importance.'



Isso não é importante ?!
`That is not important?!' [id=348]



Isso não importa.
'That does_not matter.'

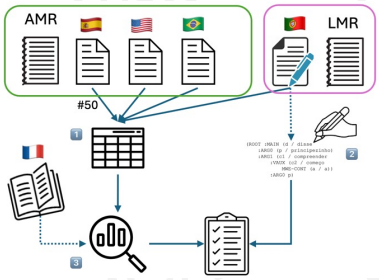
```
(t / think-01                                  AMR-1
   :ARG0 (y / you)
   :ARG1 (t2 / that
      :ARG1-of (i / important-01
         :polarity -)))
```

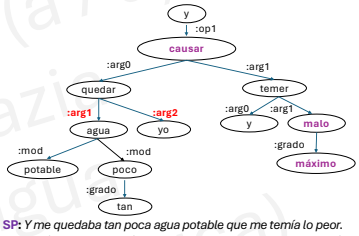
You think that is not important ! [id=348]

```
(i / important-01 :polarity -  AMR-2
   :domain (t / that))
```

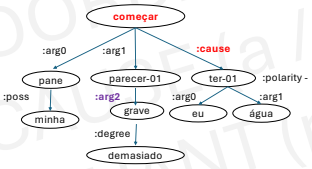
That is not important !

Contrastive analysis

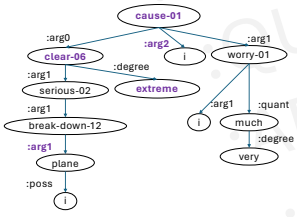


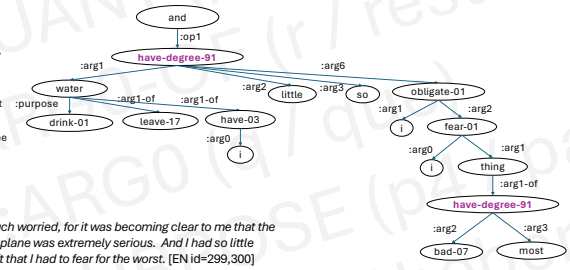


SP: *Y me quedaba tan poca agua potable que me temía lo peor.*
'And so little water remained for me that I feared the worst' [SP id=15]

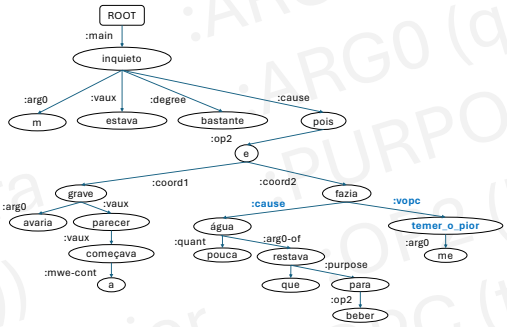


BR: *Minha pane começava parecer demasiado grave, e em, missed content e a água para beber.* 'My breakdown began to look too_much serious and soon [I] would not have water to drink.' [BR id=299;300]





EN: *I was very much worried, for it was becoming clear to me that the breakdown of my plane was extremely serious. And I had so little drinking-water left that I had to fear for the worst.* [EN id=299,300]



PT: *Estava bastante inquieto, pois a avaria começava a parecer grave, e a pouca água que restava para beber fazia-me temer o pior.* '[I] was very worried, for the break-down started to look serious and the little water that remained for drinking made me fear the worst' [PT id=300]

Conclusions

- This paper underscored the challenges inherent in implementing standard AMR directives.
- Discrepancies arise from variations in original versions or translator choices
- But *inconsistencies* in applying AMR directives (particularly pronounced in Spanish and Brazilian Portuguese translations).
- Proposes LMR annotation scheme: anchors annotation directly onto the text, consistent operator-argument relations, while highly compatible with standard AMR.

LMR offers a promising solution; a representation closer to the text and less susceptible to the inherent inconsistencies in abstracting the text's meaning.

Future work

- Expand the annotated texts in LMR, completing the annotation of *O Principezinho* (The Little Prince)
- incorporate texts from various genres and domains, including more legal texts.
- develop tools to facilitate faster and more efficient annotation implementation, including:

(a) a *lemmatizer*: associate text forms with lemmas and frames' unique identifiers in the Lexicon-Grammar);
(b) a *LMR graph builder*: instantiate argument slots based on Lexicon-Grammar information; mark anaphors for anaphora resolution: ensure overall formal consistency;
(c) *PENMAN graphs convertor* into graphic format, to facilitate interpretation;

Future work (cont.)

(d) a tool for *comparing annotations*: assessing agreement between annotators, and across translations of the same text in different languages.

Aim: build a LMR parser for automatic semantic representation, with the potential for several NLP applications.

Acknowledgments
This work was partially supported by Portuguese national funds through FCT - Fundação para a Ciência e a Tecnologia (DOI: 10.54499/UIDB/50021/2020) and by European funds through the European Commission (Project: iRead4Skills, Grant number: 1010094837, Topic: HORIZON-CL2-2022-TRANSFORMATIONS-01-07, DOI: 10.3030/101094837).

iRead4Skills
Funded by the European Union
inesc id lisboa