Towards a General Abstract Meaning Representation Corpus for Brazilian Portuguese

Marco Antonio Sobrevilla Cabezudo and Thiago Alexandre Salgueiro Pardo
Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
São Carlos/SP, Brazil
msobrevillac@usp.br, taspardo@icmc.usp.br

Abstract

Abstract Meaning Representation (AMR) is a recent and prominent semantic representation with good acceptance and several applications in the Natural Language Processing area. For English, there is a large annotated corpus (with approximately 39K sentences) that supports the research with the representation. However, to the best of our knowledge, there is only one restricted corpus for Portuguese, which contains 1,527 sentences. In this context, this paper presents an effort to build a general purpose AMR-annotated corpus for Brazilian Portuguese by translating and adapting AMR English guidelines. Our results show that such approach is feasible, but there are some challenging phenomena to solve. More than this, efforts are necessary to increase the coverage of the corresponding lexical resource that supports the annotation.

1 Introduction

In recent years, there has been renewed interest in the Natural Language Processing (NLP) community in language understanding and dialogue. Thus, the issue of how the semantic content of language should be represented has reentered into the NLP discussion. In this context, several semantic representations, like Universal Networking Language (UNL) (Uchida et al., 1996), the semantic representation used in the Groningen Meaning Bank (Basile et al., 2012), Universal Conceptual Cognitive Annotation (UCCA) (Abend and Rappoport, 2013), and, more recently, the Abstract Meaning Representation (AMR) (Banarescu et al., 2013), have emerged.

Abstract Meaning Representation is a semantic formalism that aims to encode the meaning of a sentence with a simple representation in the form of a directed rooted graph (Banarescu et al., 2013). This representation includes information about semantic roles, named entities, wiki entities, spatial-temporal information, and co-references, among other information. AMR may be represented using logic forms (see (a) in Figure 1), PENMAN notation (see (b) in Figure 1), and graphs (see (c) in Figure 1). AMR has gained relevance in the research community due to its easiness to be read by computers and humans (as it could be represented using graphs or first-order logic, which are representations that are more familiar to computers and humans, respectively), its attempt to abstract away from syntactic idiosyncrasies (making the tasks to focus only on semantic processing) and its wide use of other comprehensive linguistic resources, such as PropBank (Bos, 2016).

In relation to its attempt to abstract away from syntactic idiosyncrasies, it may be seen that AMR annotation in Figure 1 could be generated from the sentences "The boy wants the girl to believe him." and "The boy wants to be believed by the girl.", which are semantically similar, but with different syntactic realizations. Regarding the use of linguistic resources, AMR annotation in Figure 1 shows information provided by PropBank, as the framesets "want-01" and "believe-01", and some semantic roles that they require.

The available AMR-annotated corpora for English are large, containing approximately 39,000 sentences. Some efforts have been performed for using AMR as an interlingua and building corpus for Non-English languages, taking advantage of the alignments and the parallel corpora that exist (Xue et al., 2014; Damonte and Cohen, 2018). Other works tried to adapt the AMR guidelines to other languages (Migueles-Abraira et al., 2018), considering its cross-linguistic potential.

It is unnecessary to stress the importance of corpus creation for other languages. Annotated corpora provide qualitative and reusable data for building or improving existing methods and ap-

Proceedings of the 13th Linguistic Annotation Workshop, pages 236–244
Florence, Italy, August 1, 2019.

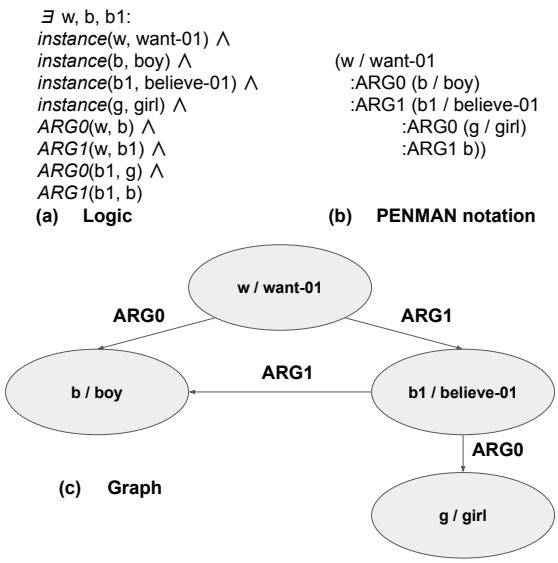

Figure 1: AMR examples

plications, as well as for serving as benchmarks to compare different approaches. In the case of Portuguese language, to the best of our knowledge, there is an unique AMR-annotated corpus, composed by the sentences of the "The Little Prince" book (Anchiêta and Pardo, 2018). The lexical resource they used to annotate some concepts was the Verbo-Brasil (Duran and Aluísio, 2015), which replicates the PropBank experience for Portuguese.

One difficulty related to the above corpus is its unusual writing style (since it is a tale) and its restricted vocabulary, which make the creation or adequacy of general purpose tools a more difficult task. More than this, the corpus is too small, hindering the development or adaptation of methods for tasks that require semantics. In this context, this work intends to show the extension process of the AMR annotation on a general purpose corpus (which covers a wide vocabulary and several domains) using the current AMR guidelines and some adaptations for Portuguese.

This paper is organized as follows. Section 2 briefly introduces some previous work that tried to build AMR corpora for Non-English languages. The corpus in Portuguese is described in Section 3. The annotation methodology and evaluation are described in Section 4 and 5, respectively. The current state of the annotation is reported in Section 6, and, finally, some concluding remarks are presented in Section 7.

2 Related Work

One of the first works that tried to build an AMR-annotated corpus for a Non-English language was proposed by Xue et al. (2014). The main goal of this work was to evaluate the potentiality of AMR to work as an interlingua. In order to achieve this goal, the authors annotated 100 English sentences of the Penn Treebank using AMR and then translated them to Czech and Chinese, which were annotated with AMR as well. Their main finding was that the level of compatibility of AMR between English and Chinese was higher than between English and Czech.

In other research line, Vanderwende et al. (2015) proposed an AMR parser to convert Logic Form representations into AMR for English. The authors also built an AMR-annotated corpus for French, German, Spanish, and Japanese.

Damonte and Cohen (2018) developed an AMR parser for English and used parallel corpora to learn AMR parsers for Italian, Spanish, German, and Chinese. The main results showed that the new parsers overcame structural differences between the languages. The authors also proposed a method to evaluate the parsers that does not need gold standard data in the target languages.

In the case of Spanish, Migueles-Abraira et al. (2018) performed a manual AMR annotation of the book "The Little Prince" using the guidelines of the AMR project. The main goal was to analyze the guidelines and to suggest some adaptions in order to cover the relevant linguistic phenomena in Spanish.

For Portuguese, Anchiêta and Pardo (2018) built the first AMR-annotated corpus taking advantage of the alignments between the book "The Little Prince" for English and Portuguese languages. Thus, the strategy consisted of importing the corresponding AMR annotation for each sentence from the English annotated corpus and revising the annotation to adapt it to Portuguese.

3 The Corpus for Brazilian Portuguese

As mentioned, the AMR-annotated corpus for Brazilian Portuguese was composed by sentences of the "The Little Prince" book (Anchiêta and Pardo, 2018). In order to broaden the annotation to other domains and text genres, our proposal focused on annotating news in several domains.

The news texts were extracted from RSS[1] from *Folha de São Paulo* news agency[2], one of the mainstream agencies in Brazil. The selected news came from different sections/domains: "daily news", "world news", "education", "environment", "sports", "science", "balance and health", "*ilustrada*", "*ilustríssima*", "power", and "technology". Additionally to these sentences, sentences of the PropBank.Br[3] (Duran and Aluísio, 2012) were collected in order to enrich the corpus (PropBank.Br already contains semantic role annotation, which makes the AMR annotation task much easier). It is important to note that PropBank.Br sentences are also from news texts.

The news download interval was from November 25th to November 28th, 2018. Overall, 249 news were collected from different domains, totalizing 7,643 sentences. The news distribution is presented in Table 1.

Section	# News	# Sentences	Avg. tokens by sentence	# Selected sentences
Daily news	48	1,521	22.94	848
World news	43	1,212	24.38	617
Education	13	426	23.72	222
Environment	4	98	25.40	45
Sports	29	875	20.93	531
Science	10	460	23.50	243
Balance and Health	6	159	23.15	88
Ilustrada	27	648	24.10	348
Ilustríssima	7	305	24.41	161
Power	51	1,677	19.93	1,121
Technology	11	262	22.55	149
Total	249	7,643	22.53	4,563

Table 1: News collection statistics

Due to the statistics observed in Table 1 and the difficulty that the task of semantic annotation carries, the scope of the work was focused on annotating only short sentences (but guaranteeing that different domains are covered). In order to define what a short sentence is, the average number of tokens by sentence was calculated and this value was used as threshold. Thus, sentences with a number of tokens below the average (in our case, it was 22.53 tokens) were selected, resulting in 4,563 sentences to be AMR annotated (indicated by the "Selected sentences" column in the table).

In relation to the PropBank.Br sentences (Duran and Aluísio, 2012), the same strategy for selecion was adopted. In total, 3,012 PropBank.Br sentences were added to our corpus.

[1]RSS stands for "Really Simple Syndication".

[2]Available at https://www.folha.uol.com.br/.

[3]PropBank.Br was the basis for the construction of the previously cited Verbo-Brasil.

4 Annotation Methodology

The proposed annotation methodology consisted of two main steps. The first step aimed to independently analyze and think about the sentence structure, while the second step counted with the aid of the AMR Editor tool (Hermjakob, 2013) to produce the AMR annotation in PENMAN format in order to export the annotation.

In relation to the first step, a sequence of actions need to be carried out in order to facilitate the second step. These actions are described as follows:

- To identify the kind of sentence to be analyzed (default, comparative, superlative, coordinate, subordinate, and others). This is useful to determine whether it is necessary to build two or more sub-graphs (in case of coordinate or subordinate sentences) and then to join them using a conjunction (usually coordinate sentences) or a concept of the main sub-graph (in the case of subordinate sentences).

- To identify concepts. Annotators must follow the AMR guidelines[4] in order to define a concept. Thus, they may identify general concepts, concepts from AMR Guidelines or concepts from Verbo-Brasil.

- To identify the main concept from the two previous steps. For example, the main verb could be the main concept in a default sentence.

- To identify the relations among the identified concepts[5].

An example of the execution of the actions is presented in Figure 2. The sentence to be analyzed is "*Ieltsin adotou outras medidas simbólicas para mostrar a perda de poderes do Parlamento.*"("Yeltsin took other symbolic measures to show the loss of Parliament's power."). This is the case of a subordinate sentence. Then, we need to identify the concepts. Thus, some words became general concepts, named-entities or Verbo-Brasil framesets. Then, it was necessary to identify the graph top (in this case, the verb "*adotar*" because

[4]Available at https://github.com/amrisi/amr-guidelines/blob/master/amr.md. Accessed on April 1st, 2019. The adopted version was the 1.2.5.

[5]The relations were extracted from Verbo-Brasil (for core relations) and AMR guidelines (for non-core relations).

it is the main verb of the main sentence "*Ieltsin adotou outras medidas simbólicas*"). Finally, the relations among all concepts were identified.

Similar to the work of Migueles-Abraira et al. (2018), our proposal tried to adapt the AMR guidelines to Brazilian Portuguese, making some modifications on it in order to deal with the specific linguistic phenomena. The general guideline used to annotate a sentence is described as follows:

- To use the framesets of Verbo-Brasil (Duran and Aluísio, 2015) to determine verb senses and the argument structure of verbs.
- To use the 3rd singular person ("*ele*") or the pronoun "that" ("*isso*") in case of NP Ellipsis, clitic or possessive pronouns. Differently from Migueles-Abraira et al. (2018), we propose to use ("*ele*") or "that" ("*isso*") as a default value. We decided to determine this guideline in order to keep some annotation pattern.
- In the case of indeterminate subject, not to use any pronoun.
- In the case of multi-word expression, to identify the one-word synonym of the expression and use it in the annotation, or define a one-word as the join of the words.
- To use the AMR framesets to annotate modal verbs, since Verbo-Brasil does not include that kind of verbs. In order to facilitate the identification of a modal verb, to try to replace by "*poder*" ("can") or "*dever*" ("should") verbs.
- In cases where the difference among two or more senses is subtle, to use the most frequent sense that satisfies the predicted argument structure.
- To use the AMR guidelines and dictionary[6] for the other cases.

The proposed annotation strategy consisted of annotating sentences of shorter size at the beginning and then increasing sentence size up to 22 tokens, according to the annotators' learning. Sentences that had verbs that were not included in the Verbo-Brasil repository were not annotated and the new verbs were put in a list in order to enrich the repository in the future.

Smatch score (Cai and Knight, 2013) was used to calculate the inter-annotator agreement. Unlike the work of Banarescu et al. (2013), which built a gold standard (using the total agreement between the annotators), the way to calculate the inter-annotator agreement consisted in comparing all annotations in an all-against-all configuration, obtaining the average of all inter-annotator agreements. Finally, the annotated versions of the sentences belonging to the agreement sample that were included in the final corpus were chosen by an adjudicator (since that more than one possible annotation exists).

5 Evaluation

In relation to the overview of the annotation process, it is important to know that the annotation team was originally composed of 14 annotators[7] that belong to the areas of Computer Science and Linguistics (all of them focused on Natural Language Processing). These annotators participated in two training sessions. In the first session, the task and the resources to be used were presented. The participants were trained by annotating sentences of PropBank.Br (Duran and Aluísio, 2012) in order to perceive the difficulty of the task. The second session aimed to answer questions about the annotation, show the inter-annotator agreement during the training stage, some common mistakes, and launch the annotation process.

5.1 Inter-annotator Agreement

The results of the inter-annotator agreement are presented in Table 2. During the training stage, the agreement was measured once in each week (with 4-5 sentences to annotate per week). Currently, the annotators are building AMR annotations for more sentences until they reach 100 sentences (as in the original AMR project) in order to have an adequate sample to measure the agreement.

In general, the Smatch was 0.72, with the minimum being 0.70 and the maximum 0.77. These results are similar to the obtained by the work of Banarescu et al. (2013) (between 0.70 and 0.80), although the number of sentences assessed in English was 100 (in our case, there were 34 sentences) and the number of annotators was 4 (we

[6] Available at https://www.isi.edu/~ulf/amr/lib/amr-dict.html. Accessed on April 1st, 2019.

[7] During the annotation process, some of the annotators gave up.

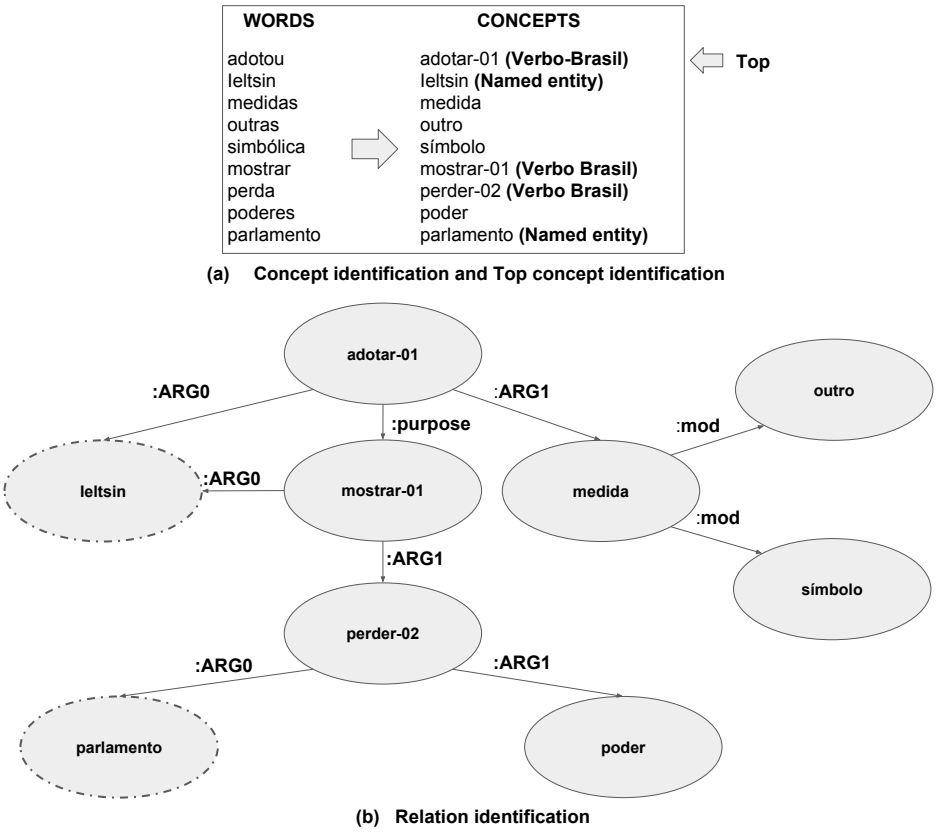

Figure 2: Example of the annotation steps

had from 5 to 7).

Week	# Annotators	# Sentences	Smatch
1	5	5	0.77
2	7	5	0.72
3	5	4	0.73
-	-	20	0.70
Total		34	0.72

Table 2: Annotation agreement

5.2 Disagreement Analysis

It is important to highlight some reasons that led to the occurring disagreements. One of the reasons was the difficulty identifying some kinds of verbs, as modal, copula, light and auxiliary verbs. Additionally, due to the use of English framesets for modal verbs, there were cases where the frameset to be used was difficult to be determined. For example, the sentence "*A quem podemos nos aliar?*" ("Who can we ally with?") was encoded as follows:

(r / **recommend-01**
  :ARG1 (a / aliar-01
    :ARG0 (n / nós)
    :ARG1 (a2 / amr-unknown)))

(p5 / **possible-01**
  :ARG1 (a8 / aliar-01
    :ARG1 (n3 / nós)
    :ARG2 (a9 / amr-unknown)))

As one may see, the modal verb "*poder*" was encoded as "recommend-01" and "possible-01", depending on the interpretation of the annotator. This problem occurred because a modal verb in Portuguese may be translated in different ways to English according to the context.

Another difficulty was the identification of verbs whose modality could not be easy to identify. For example, the verb "*conseguir*" (usually translated to "get") in the sentence "*Ele contou que conseguiu adquirir 20 entradas porque ofereceu Cr$ 5.000 ao bilheteiro.*" ("He said he was able to get 20 tickets because he offered Cr$ 5.000 to the ticket clerk.") was annotated using a Verbo-Brasil frameset (without modal verb) by some annotators and using the AMR frameset (for modal verb) by others. To solve this difficulty, the guidelines (adapted for Portuguese) suggested that they should try to substitute verbs for some modal verbs

as "*dever*" or "*poder*". In the previous sentence, the verb "*conseguir*" could be replaced by the verb "*poder*". This way, "*conseguir*" might be identified as a modal verb.

As for the modal verbs, the annotation of auxiliary verbs also presented some difficulties. Some annotators used the Verbo-Brasil framesets and others omitted that verb annotation, being this last one the correct way to annotate. For example, this happens for the verb "*ficar*" in the sentence "*Eles ficaram aguardando o resultado da negociação.*" ("They were waiting for the outcome of the negotiation."), where the verb fulfills an auxiliary function, and, therefore, it should not be considered in the final AMR representation.

Another difficulty was related to the identification of the verb sense in the Verbo-Brasil repository. This identification was problematic in some cases. For example, the verb "*admitir*" in the sentence "*Ele não treinava como devia, o que não admito*" ("He did not train as he should, what I do not admit") was associated to the concept "*admitir-01*" (whose meaning is related to confess or acknowledge as truth) and to the concept *admitir-02* (whose meaning is related to agree, allow, or tolerate). In this case, i.e., when the verb sense is difficult to identify, the suggestion was to select the most frequent sense (usually the first in the sense list) that covers all the arguments in the sentence.

In a similar way, sometimes the identification of the argument labels and the relations between concepts presented challenges to the annotators. For example, the word "*porque*" in the sentence "*Ele contou que conseguiu adquirir 20 entradas porque ofereceu Cr$ 5.000 ao bilheteiro.*" was associated to the relation "*cause*". However, some annotators omitted this relation.

In relation to the reference annotation, we may highlight that the annotators had disagreements in some cases, mainly when they had to choose where the reference should be inserted. For example, in the sentence "*A empresa considera os equipamentos ultrapassados e quer adquirir modelos modernos.*" ("The company considers the equipment to be outdated and wants to acquire modern models.") represented in the two following ways), the concept "*empresa*" ("company") was used as reference for "querer-01" and "adquirir-01" by some annotators and as reference only for "*querer-01*" by others.

```
(e / and
   :op1 (c / considerar-01
      :ARG0 (e2 / empresa)
      :ARG1 (e3 / equipamento)
      :ARG2 (u / ultrapassado))
   :op2 (q / querer-01
      :ARG0 e2
      :ARG1 (a2 / adquirir-01
         :ARG0 e2
         :ARG1 (m / modelo
            :mod (m2 / moderno)))))
```

```
(e / and
   :op1 (c6 / considerar-01
      :ARG0 (e / empresa)
      :ARG1 (e12 / equipamento
      :ARG2 (u2 / ultrapassado)))
   :op2 (q / querer-01
      :ARG0 e
      :ARG1 (a12 / adquirir-01
         :ARG1 (m / modelo
            :mod (m2 / moderno)))))
```

In relation to part of speech tags, we remark that there were problems in the annotation of some adjectives and nouns. In the case of adjectives, there were some difficulties to nominalize some adjectives (pertainym adjectives). For example, the adjective "*tributária*" ("tributary") in the expression "*carga tributária*" ("Tax burden") refers to a type of "*carga*" ("charge"), therefore, the concept "*tributo*" ("tribute") should be used instead of "*tributária*". In the case of nouns, there were difficulties to convert some nouns into verbs and to deal with some nouns like executors of some action. For example, the word "*competivididade*" ("competitiveness") was encoded using the concept "*competivididade*" (wrong way) and using the concept "*competir-01*" (correct way). Another example is the word "*bilheteiro*" ("ticket clerk"), which was encoded using the concept "*bilheteiro*" by some annotators. However, the correct encoding was to interpret "*bilheteiro*" as "*pessoa que vende bilhetes*" ("person that sells tickets") and, thus, encoding it as follows:

```
(p / pessoa
   :ARG0-of (v / vender-01
      :ARG1 (b / bilhete)
```

Finally, another difficulty was associated to the

use of temporal expressions. For example, the expression "*até agora*" ("until now") was encoded in several ways by the annotators. In this case, this expression was treated as fixed, using the concept "até-agora".

5.3 Common Mistakes

Some of the frequent errors made in the annotation process include the following:

- No lemmatization: there were several cases where some annotators did not use the lemmas to represent the concepts. In this way, this decreased inter-annotator agreement and could harm the annotation quality. For example, the concept "*equipamento*" ("equipment") should be used instead of "*equipamentos*" ("equipments"), and the concept "*ele*" ("he") instead of "*eles*" ("they").

- Specific characters for Portuguese: the AMR Editor tool was developed for annotating English sentences. Thus, this tool does not work well when a sentence to be annotated includes words with characters used in Portuguese like "*â*" or "*ç*". To solve this problem, it was suggested that annotators omit these characters when using the editor (replacing by one general character like "*a*" and "*c*") and then restore the correct characters as a post-editing step. However, these errors occurred, impairing the agreement.

- Variable errors or format errors: some annotators opted not to use the AMR Editor tool to build the AMR graphs, resulting in mistakes related to the number of parenthesis of the PENMAN notation and the variable declaration repetition. For example, the concept "*correr*" ("run") was represented by the variable "c" and the concept "*coelho*" ("rabbit") was also represented by the same variable, producing an error in the graph representation.

5.4 Annotation Challenges

During the annotation process (after the training stage), several challenges emerged. In what follows, some of these challenges are briefly discussed.

- Expressions or short sentences. Although the length of the sentences (or expressions) were tiny (3-5 words), expressions like "*nada demais?*", "*De quem é a culpa?*", "*Não, em hipótese alguma.*" were difficult to annotate. In some cases, it happened due to lack of context. In other cases, to identify which concepts should be included in the representation and how these concepts should be related was a hard task. This representation problem may be reflected in the inter-annotator agreement decay down to 0.70 (in comparison with the previous agreement).

- Multi-word expressions (MWE). Expressions like "*toda hora*", "*todo mundo*", or "*estar na moda*" in the sentence "*Academias especializadas estão na moda.*" were examples of multi-word expressions that annotators could not represent as a 1-word synonym (as the guideline indicates). In these cases, annotators join the words (for example, "*toda-hora*" is described as AMR dictionary suggests) or tried to separate the concepts in the graph. Another problem was the MWE identification. Expressions like "*na moda*" could be difficult to identify as a MWE and bring some challenges into the annotation.

- Particularities of Portuguese. Some expressions are specific for Portuguese or similar languages. For example, we may see a double negation in the sentence "***Não** temos **nenhuma** intelectualidade pronta.*", which does not naturally occur in English. Thus, annotators omitted one of the negations to preserve the meaning of the sentence.

- Indeterminate subjects. In some cases, the subject was indeterminate and the annotators did not annotate the reference. For example, in the sentence "*bebe-se*", the particle "*se*" did not show who is the subject, so, it was not marked in the representation.

6 Current State of the Annotation

Currently, the corpus is composed by 299 AMR-annotated sentences (considering the inter-annotator agreement sample), which include 907 concepts and 711 relations (excluding "instance", "name", and "op" relations). It is important to notice that there are 26 verbs (or verb senses) that did not appear in the Verbo-Brasil and it is necessary

to analyze them in order to increase the coverage of the repository in the future.

Table 3 and Table 4 show the statistics about the concepts and the top 10 most frequent relations annotated in the corpus. For comparison purposes, Table 4 also shows the top 10 most frequent relations annotated in the AMR-annotated corpus based on "The Little Prince" book for Brazilian Portuguese.

One point to remark in relation to Table 4 is that both corpora keep the same proportion in the first relations (the top 5); then, both show slightly different distributions. In the case of "The Little Prince", relations like "degree" and "poss" are more frequent. One reason to explain this is that tales use intensifiers like "more" or "less" and possessives like "mine" or "his" in their vocabulary. On the other hand, news texts, and the sentences and expressions contained in it, describe facts and usually use numbers to report quantities ("quant" relation). More than this, some expressions collected until now (due to their short size) describe imperatives like "*arranje!*" ("get it"). Thus, the imperative mode is frequent in the corpus. It is expected that, when the news corpus grows, these relation will change a bit.

Concepts	Frequency
General concepts	504
Verbo-Brasil concepts	235
Named entities	66
Modal verbs	20
Amr-unknown	33
Other entities and special frames	49

Table 3: Statistics of concepts in the corpus

Current corpus			"The Little Prince" corpus		
Relation	Freq.	%	Relation	Freq.	%
ARG1	173	24.33	ARG1	1,734	25.88
ARG0	140	19.69	ARG0	1,520	22.69
polarity	70	9.85	mod	678	10.12
mod	69	9.70	ARG2	454	6.78
ARG2	53	7.45	polarity	295	4.40
domain	35	4.92	time	246	3.67
quant	25	3.52	domain	211	3.15
time	23	3.23	degree	194	2.90
manner	20	2.81	manner	187	2.79
mode	17	2.39	poss	162	2.42

Table 4: Ten most frequent relations in the news corpus and in the "The Little Prince" corpus

7 Concluding Remarks

This paper showed the process of the AMR annotation on a general purpose corpus using the current AMR guidelines and some adaptations for Portuguese. In general, most of the guidelines could be translated to Portuguese. However, there were some cases that needed improvements, as the use of modal verbs and multi-word expressions. On the other hand, the adopted PropBank-like lexical resource (Verbo-Brasil) needs to increase its coverage.

As future work, besides extending Verbo-Brasil, we plan to try back-translation strategies to accelerate the annotation process.

More details about the corpus and the related ongoing work may be found at the OPINANDO project webpage[8].

Acknowledgments

The authors are grateful to CAPES and USP Research Office for supporting this work and to the several corpus annotators that have collaborated with this research.

References

Omri Abend and Ari Rappoport. 2013. Ucca: A semantics-based grammatical annotation scheme. In *Proccedings of the 10th International Conference on Computational Semantics*, pages 1–12, Potsdam, Germany. Association for Computer Linguistics.

Rafael Anchiêta and Thiago Pardo. 2018. Towards AMR-BR: A SemBank for Brazilian Portuguese Language. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pages 974–979, Miyazaki, Japan. European Language Resources Association (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 and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 3196–3200, Istanbul, Turkey. European Language Resource Association (ELRA).

[8]Available at https://sites.google.com/icmc.usp.br/opinando/

Johan Bos. 2016. Expressive power of abstract meaning representations. *Computational Linguistics*, 42(3):527–535.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Marco Damonte and Shay B. Cohen. 2018. Cross-lingual abstract meaning representation parsing. In *Proceedings of the 16th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1146–1155, New Orleans, Louisiana. Association for Computational Linguistics.

Magali Sanches Duran and Sandra Maria Aluísio. 2012. Propbank-br: a brazilian treebank annotated with semantic role labels. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 1862–1867, Istanbul, Turkey. European Language Resources Association (ELRA).

Magali Sanches Duran and Sandra Maria Aluísio. 2015. Automatic generation of a lexical resource to support semantic role labeling in portuguese. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 216–221, Denver, Colorado. Association for Computational Linguistics.

Ulf Hermjakob. 2013. Amr editor: A tool to build abstract meaning representations.

Noelia Migueles-Abraira, Rodrigo Agerri, and Arantza Diaz de Ilarraza. 2018. Annotating abstract meaning representations for spanish. In *Proceedings of the 11th Language Resources and Evaluation Conference*, pages 3074–3078, Miyazaki, Japan. European Language Resource Association (ELRA).

Hiroshi Uchida, M Zhu, and T Della Senta. 1996. UNL: Universal networking language–an electronic language for communication, understanding, and collaboration. *Tokyo: UNU/IAS/UNL Center*.

Lucy Vanderwende, Arul Menezes, and Chris Quirk. 2015. An amr parser for english, french, german, spanish and japanese and a new amr-annotated corpus. In *Proceedings of the 13th Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 26–30, Denver, Colorado. Association for Computational Linguistics.

Nianwen Xue, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of english amrs to chinese and czech. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 1765–1772, Reykjavik, Iceland. European Language Resources Association (ELRA).