



# XPTA: um *parser* AMR para o português baseado em uma abordagem entre línguas

**XPTA: an AMR parser for Portuguese based on cross-lingual approach**

Eloize Rossi Marques Seno ✉   
Instituto Federal de São Paulo

Helena de Medeiros Caseli ✉   
Universidade Federal de São Carlos

Marcio Lima Inácio ✉   
Universidade de São Paulo

Rafael Torres Anchiêta ✉   
Instituto Federal do Piauí

Renata Ramisch ✉   
Redação Nota 1000

## Resumo

O crescente interesse pelo processamento semântico automático, especialmente por parte dos pesquisadores de Compreensão e de Geração de Língua Natural, tem levado a muitas pesquisas relacionadas ao desenvolvimento de *parsers* semânticos. E, nesse contexto, a AMR (*Abstract Meaning Representation*) é um dos formalismos de representação semântica que tem recebido mais atenção recentemente, devido à sua maneira relativamente simples de capturar o significado de uma sentença. A construção de *parsers* AMR é em grande parte baseada em cópulas de referência anotados por humanos. Contudo, esse recurso é ainda bastante escasso para muitas línguas como o português. Por esse motivo, várias pesquisas têm explorado o uso de abordagens entre línguas (*cross-lingual*), que partem de cópulas e *parser* existentes em uma língua fonte, para o desenvolvimento de recursos semânticos para outras línguas alvo. Dado esse contexto, este artigo descreve o XPTA, um *parser* AMR para o português (PT) que se baseia na abordagem entre línguas (*cross-lingual*, X). O XPTA parte de *parser* AMR existente para o inglês e de vários recursos linguísticos-computacionais bilíngues inglês-português e mapeia o conhecimento semântico disponível no inglês para a representação do significado equivalente em português. Uma avaliação automática do XPTA mostrou que a abordagem adotada é promissora e os valores obtidos para *Smatch* (66%, no melhor caso) apontaram que o modelo tem potencial para competir com os resultados apresentados na literatura para outros idiomas. Além da análise automática, uma análise qualitativa dos grafos gerados possibilitou identificar e categorizar os principais erros do modelo e suas possíveis causas.

## Palavras chave

representação abstrata de significado, analisador semântico, abordagem entre línguas, Português

## Abstract

The growing interest in automatic semantic processing, especially by researchers in Natural Language Understanding and Natural Language Generation, has led to several researches related to the development of semantic parsers. In this context, the semantic representation formalism of AMR (Abstract Meaning Representation) has received the most attention lately, due to its relatively simple way of capturing the meaning of a sentence. The development of AMR parser is mainly based on human-produced reference corpus. However, this resource is still quite scarce for many languages such as Portuguese. For this reason, several works have explored cross-lingual approaches, which make use of corpora and parsers available for a source language, to develop semantic resources to other target languages. Given this context, this paper describes XPTA, an AMR parser for Portuguese which is based on a cross-lingual approach. XPTA makes use of an existing parser for English and several English-Portuguese bilingual resources to map the semantic knowledge available in English to equivalent meaning representation in Portuguese. An automatic evaluation of XPTA showed that the adopted approach is promising and the results obtained for *Smatch* (66% in the best case) suggest that the model has the potential to compete with the results presented in the literature to others idioms. In addition to the automatic analysis, a qualitative analysis of the graphs produced by the parser allowed to identify and categorize the main mistakes of the model and their possible causes.

## Keywords

abstract meaning representation, semantic parser, cross-lingual approach, Portuguese



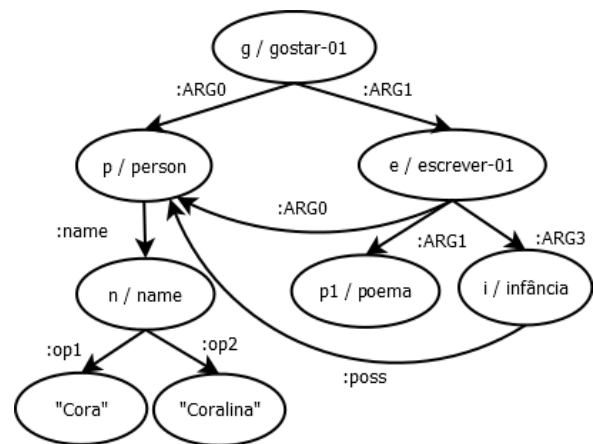
## 1. Introdução

Uma tarefa crucial do processamento inteligente da língua natural consiste em compreender a mensagem contida em um trecho de texto (uma sentença, por exemplo) a fim de derivar conhecimento, tomar uma decisão ou produzir uma saída esperada (tradução, resposta a uma pergunta, etc.). Para tanto, nos métodos tradicionais de aprendizado de máquina, esse processo envolve a obtenção de uma representação conceitual (semântica) dos textos, capaz de abstrair as escolhas lexicais e as características morfológicas e sintáticas, e ao mesmo tempo resolver as ambiguidades. Nesse contexto, muitos formalismos de representação do significado já foram explorados como: redes semânticas (Lehmann, 1992), interlíngua UNL (*Universal Network Language*) (Uchida et al., 2006), Lógica de Primeira Ordem (Jurafsky & Martin, 2009) e, mais recentemente, a AMR (*Abstract Meaning Representation*) (Banarescu et al., 2013).

Atualmente, a AMR tem ganhado bastante popularidade na área (Damonte et al., 2017; van Noord & Bos, 2017; Peng et al., 2017; Lyu & Titov, 2018; Vilares & Gómez-Rodríguez, 2018; Anchiêta, 2020). Uma das razões se deve à sua forma relativamente simples de representar o significado das sentenças, que se baseia em grafos acíclicos dirigidos com raiz, nos quais os nós representam os conceitos e as arestas indicam as relações semânticas entre esses conceitos. Por ser uma representação simbólica, uma das vantagens da AMR é ser facilmente compreendida pelos humanos, sendo, ao mesmo tempo, de fácil manipulação pela máquina. Outra razão para a popularidade atual da AMR está na possibilidade de realizar a avaliação automática com o uso de métricas como a *Smatch* (Cai & Knight, 2013) e a SEMA (Anchiêta et al., 2019), que se baseiam em medidas clássicas de precisão, cobertura e medida-*F*, calculadas com base em cópulas de referência.

De acordo com Bos (2016), quando comparada a outras representações formais do significado, as estruturas AMR são também mais fáceis de serem construídas, uma vez que fazem uso de recursos pré-existentes e seguem as propriedades da teoria dos grafos.

A representação AMR é baseada na estrutura argumental — papéis semânticos — dos verbos de uma sentença, geralmente fornecida pelo PropBank (Palmer et al., 2005), sendo o nó raiz representado pelo predicado principal da sentença.



**Figura 1:** Grafo AMR para a sentença: *Cora Coralina gostava de escrever poemas sobre sua infância.*

Na Figura 1 é ilustrado um grafo AMR que tem como raiz o verbo “gostar” (*g / gostar-01*)<sup>1</sup>. As estruturas argumentais dos verbos em português são fornecidas pelo repositório Verbo Brasil<sup>2</sup> (Duran & Aluísio, 2012). Os conceitos na AMR são apresentados na sua forma canônica.

Como se pode notar na Figura 1, o predicado principal tem como argumentos o experienciador (relação indicada por *:ARG0*), isto é, uma entidade mencionada do tipo pessoa (*p / person*) com o nome (*n / name*) de “Cora Coralina”, e o tema (relação indicada por *:ARG1*), representando o objeto apreciado, no caso, “escrever” (*e / escrever-01*). O conceito (*e / escrever-01*) possui três argumentos (*:ARG0*, *:ARG1* e *:ARG3*) que representam, respectivamente: o agente (isto é, o escritor (*p / person*)), o tema (isto é, a coisa escrita (*p1 / poema*)) e o assunto (*i / infância*). A aresta “:poss”, conectando os conceitos “*i / infância*” e “*p / person*”, indica um relação de posse entre eles.

O desenvolvimento de *parsers* AMR é fundamentalmente baseado em cópulas de referência anotados manualmente por especialistas, uma tarefa laboriosa e demorada<sup>3</sup>. Devido a isso, esse tipo de recurso é, ainda, bastante escasso para boa parte das línguas. Enquanto o inglês dispõe de um cópula AMR com cerca de 59 mil sentenças<sup>4</sup>, para o português, mais especificamente, tem-se conhecimento apenas do cópula do Pequeno Príncipe (Anchiêta & Pardo, 2018a),

<sup>1</sup>Cada conceito no grafo é associado a uma variável (*g*, *p*, *e*, *n*, *p1* e *i*), usada para fazer referência ao conceito.

<sup>2</sup><http://143.107.183.175:21380/verbobrasil/> (acessado em 29/08/2021).

<sup>3</sup>Banarescu et al. (2013) relatam que gastaram entre 7 e 10 minutos para anotar cada sentença do inglês.

<sup>4</sup><https://amr.isi.edu/download.html> (acessado em 29/08/2021).

composto por 1.527 sentenças, do cópús jornalístico desenvolvido por Sobrevilla Cabezudo & Pardo (2019), com 299 sentenças e o OpiSums-PT-AMR (Inácio, 2021), um cópús de opiniões contendo 481 sentenças<sup>5</sup>.

Se por um lado há carência de recursos AMR para muitas línguas, por outro lado há a disponibilidade de cópús expressivo para o inglês, o que tem proporcionado o desenvolvimento de vários *parsers* AMR (Flanigan et al., 2014; Zhou et al., 2016; van Noord & Bos, 2017; Peng et al., 2017; Lyu & Titov, 2018; Vilares & Gómez-Rodríguez, 2018), e tem ajudado a alavancar a criação desses recursos para as línguas mais carentes. Damonte & Cohen (2018), por exemplo, usaram *parser* e cópús do inglês, para o desenvolvimento de *parsers* AMR para o italiano, espanhol, alemão e chinês.

É neste contexto que o trabalho aqui descrito se insere. Mais especificamente, este artigo descreve o XPTA, um *parser* AMR para o português baseado em uma abordagem entre línguas, que parte de um *parser* AMR existente para a língua inglesa e de recursos bilíngues inglês-português e mapeia o conhecimento semântico disponível no inglês para a representação do significado equivalente em português. Para o mapeamento são utilizados diversos recursos linguísticos-computacionais bilíngues como alinhamentos lexicais e conceituais, vetores de palavras (*word embeddings*), dicionário (léxico) de tradução e repositório de verbos e estruturas argumentais, além de regras definidas com base em análise de cópús.

Os valores obtidos para *Smatch* (Cai & Knight, 2013) em nossos experimentos (58% antes da normalização e 66% após a normalização, no melhor caso) mostram que a abordagem adotada neste trabalho, embora simples, é bastante promissora, especialmente quando se observam os resultados reportados na literatura para outros idiomas, usando modelos mais sofisticados. Sheth et al. (2021), por exemplo, obtiveram um *Smatch* de 67,9% para o espanhol (melhor caso), com uma abordagem baseada em projeção de anotação, que projeta a anotação AMR disponível no inglês para as línguas alvos com base em alinhamentos de palavras contextualizadas entre sentenças paralelas. Cai et al. (2021) obtiveram um valor de 67,3% também para o espanhol (melhor caso), usando um modelo baseado em Transformadores (*Transformers*) (como será explicado na seção 3).

Como principais contribuições deste trabalho destacam-se:

- O primeiro *parser* AMR para o português baseado em abordagem entre línguas: o XPTA;
- Um cópús AMR de referência para o português do domínio de divulgação científica, contendo 200 grafos;

O restante deste artigo está organizado como segue. Na Seção 2, apresenta-se, de forma breve, o formalismo de representação semântica AMR. Na Seção 3, são apresentados os principais trabalhos da literatura relacionados à este. Na Seção 4, é apresentado o *parser* AMR proposto e desenvolvido para o português, que parte de uma representação AMR do inglês e gera a equivalente em português. Na Seção 5, são apresentados o cópús de trabalho, os experimentos realizados na avaliação do modelo proposto e uma análise qualitativa dos grafos gerados pelo modelo. Por fim, na Seção 6, apresentam-se as principais conclusões deste trabalho.

## 2. Abstract Meaning Representation (AMR)

A AMR é um formalismo de representação semântica proposto por Banarescu et al. (2013). Segundo os autores, a AMR captura o significado de uma sentença abstraindo informações sintáticas como a função gramatical, as características morfossintáticas e a ordem de constituição das palavras. Essas decisões foram tomadas, de acordo com os seus criadores, de forma a facilitar e acelerar o processo de criação de bancos de anotação semântica (*sembanks*), baseando-se em recursos e ferramentas já existentes na área. Similarmente, Hovy & Lavid (2010) discutem que existe um equilíbrio necessário entre a profundidade da teoria linguística a ser utilizada e a estabilidade do processo de anotação, com isso, os autores defendem que é necessário realizar a “neutralização” (“*neutering*”) da teoria linguística para tornar a anotação viável e pragmática, com o foco no objetivo que se espera alcançar – no caso da AMR, a criação de ferramentas linguístico-computacionais. Há, porém, dentro da comunidade interessada nessa representação semântica, críticos a algumas decisões feitas originalmente (por exemplo, os trabalhos de Donatelli et al. (2018) e Bonial et al. (2018)).

O significado de uma sentença em AMR é capturado a partir da sua estrutura predicado-argumentos (Palmer et al., 2005), e pode ser representado como um grafo direcionado com

<sup>5</sup>Todos estão disponíveis em: <https://github.com/nilc-nlp/AMR-BP> (acessado em 13/09/2021).

raiz, no qual os nós representam os conceitos e as arestas representam as relações entre eles (Figura 1). Ao abstrair a morfologia e a sintaxe, sentenças com variações linguísticas (por exemplo, paráfrases e sinônimos) têm a mesma representação na AMR.

Os conceitos na AMR podem representar entidades, eventos, propriedades e estado, podendo ser uma palavra na sua forma lexicalizada (exemplo, “infância”), um *frameset* de verbo (exemplo, “gostar-01”), ou conceitos-chaves especiais como *date-entity*, *percentage-entity*, *temporal-quantity*, entre muitos outros descritos no manual da AMR<sup>6</sup>. As relações, por sua vez, podem ser argumentos de um verbo (:ARG0, :ARG1, etc.), relações semânticas gerais (:age, :destination, :location, :name, etc.), relações que indicam quantidades (:unit, :scale, :quant, :volume-quantity, etc.), relações para datas (:day, :month, :year, :time, :decade, etc.), relações para listas (:op1, :op2, etc.) ou relações discursivas (:cause, :purpose, :concession, :manner, etc.). No total, são fornecidas cerca de 100 relações pré-definidas.

Além da representação por meio de grafos, outras representações possíveis são a Lógica de Primeira Ordem e a anotação de PENMAN (Matthiessen & Bateman, 1991). Por exemplo, a Figura 2 apresenta a estrutura AMR para a sentença da Figura 1 na notação PENMAN.

```
(g / gostar-01
  :ARG0 (p / person
    :name (n / name
      :op1 ‘‘Cora’’
      :op2 ‘‘Coralina’’))
  :ARG1 (e / escrever-01
    :ARG0 p
    :ARG1 p1 / poema
    :ARG3 i / infância
    :poss p))
```

**Figura 2:** Notação PENMAN para a sentença: *Cora Coralina gostava de escrever poemas sobre sua infância.*

A AMR prevê também representações para diversos fenômenos linguísticos como correferência, modalidade, negação, cópula, relações inversas (como :ARG1-of, :location-of), conjunções, entidades mencionadas, entre outros. Exemplos de correferência podem ser observados na Figura 2, onde a variável *p* faz referência ao conceito *person*, remetendo à entidade “Cora Coralina”, que representa o agente (:ARG0) de “escrever-01”, e ao mesmo tempo se relaciona com o conceito

(*i* / infância) por meio da relação “:poss”, que se refere à infância de Cora.

Representações abstratas do significado como a apresentada na Figura 2 são muito úteis para diversas aplicações que processam sentenças em língua natural. Por exemplo, na literatura há aplicações da AMR na Sumarização Automática (Liu et al., 2015; Dohare & Karnick, 2017; Liao et al., 2018), na Tradução Automática (Song et al., 2019), na Extração de Informações (Garg et al., 2016) e em Sistemas de Perguntas e Respostas (Mitra & Baral, 2016).

Song et al. (2019), por exemplo, utilizam representações AMR de sentenças fontes como conhecimento adicional em um modelo de tradução neural (*Neural Machine Translation*), o paradigma de Tradução Automática considerado o estado da arte. Em experimentos reportados pelos autores, eles relatam um ganho de 2 pontos na medida BLEU (Papineni et al., 2002) para a tradução de inglês para alemão quando a AMR foi incorporada ao modelo. Trabalhos como esse mostram o impacto positivo do uso de AMR em aplicações de PLN. Contudo, para que seja possível manipular as representações semânticas das sentenças, as AMRs correspondentes precisam ser fornecidas por especialistas humanos ou geradas automaticamente por meio de um *parser*.

A seção seguinte descreve alguns trabalhos da literatura que investigam o mapeamento semântico entre línguas, o que se convencionou chamar de abordagem entre línguas (*cross-lingual approach*, no inglês).

### 3. Trabalhos Relacionados

Embora a AMR não tenha sido projetada para ser uma língua universal (Banarescu et al., 2013), como é o caso da interlíngua UNL (Uchida et al., 2006), ao abstrair a função morfológica, gramatical e a ordem das palavras de uma sentença, ela também abstrai diversas idiosincrasias linguísticas que representam as principais diferenças entre as línguas. Nesse sentido, a AMR se assemelha a uma interlíngua.

Xue et al. (2014) corroboram em parte essa afirmação ao verificar, por meio da comparação de grafos AMR do inglês com grafos do tcheco e do chinês, que há bastante compatibilidade estrutural, principalmente entre o inglês e o chinês. Segundo os autores, muitas das divergências encontradas decorreram de diferentes interpretações dos anotadores humanos no nível sintático e de traduções divergentes. Defendem que um refinamento nos padrões de anotação poderia resolver muitas dessas diferenças.

<sup>6</sup><https://github.com/amrisi/amr-guidelines/blob/master/amr.md> (acessado em 29/07/2021).



	Abordagem	Alemão	Italiano	Espanhol	Chinês
Damonte & Cohen (2018)	Proj. Anot (alinh SMT)	39,0%	43,0%	42,0%	35,0%
Blloshmi et al. (2020)	LSTM	53,0%	58,1%	58,0%	41,5%
Cai et al. (2021)	Transformer	<b>64,0%</b>	65,4%	67,3%	<b>56,5%</b>
Sheth et al. (2021)	Proj. Anot (alinh contexto)	62,7%	<b>67,4%</b>	<b>67,9%</b>	-

**Tabela 1:** Valores para medida- $F$  obtidos com *Smatch* reportados na literatura para outros idiomas.

Com base na suposição de que a representação AMR pode ser mapeada (e é a mesma) em qualquer idioma, diversos trabalhos recentes têm explorado as propriedades que são preservadas entre as línguas na AMR visando, principalmente, minimizar os esforços despendidos na criação de recursos linguísticos baseados em semântica para as línguas que carecem desses recursos. As abordagens propostas partem de cópulas paralelos bilíngues e *parsers* AMR existentes na língua inglesa, a fim de obter a representação AMR equivalente na língua alvo.

Damonte & Cohen (2018) foram os precursores nessa área e treinaram *parsers* AMR para as línguas alvo italiano, espanhol, alemão e chinês. A proposta dos autores é baseada em um modelo de projeção de anotação que parte de cópulas paralelos bilíngues e projeta a anotação AMR fornecida por um *parser* do inglês para a representação correspondente nas línguas alvo. Inicialmente, as sentenças paralelas foram alinhadas lexicalmente (daqui em diante, alinhamentos lexicais), usando um modelo de alinhamento de palavras tradicionalmente usado em *Statistical Machine Translation – SMT* (Dyer et al., 2013), e as palavras de cada sentença em inglês foram alinhadas aos respectivos conceitos no grafo AMR (daqui em diante, alinhamentos conceituais), usando o alinhador JAMR (Flanigan et al., 2014). Depois de obter os alinhamentos lexicais e conceituais, os autores projetaram o alinhamento entre cada sentença na língua alvo e o grafo AMR correspondente à sua tradução no inglês. Por fim, esses alinhamentos foram usados para o treinamento de *parsers* nas línguas alvo. O trabalho descrito neste artigo também usa alinhamentos conceituais e alinhamentos lexicais entre sentenças paralelas, a fim de mapear/traduzir a anotação AMR do inglês para o português. Porém, diferente do trabalho de Damonte & Cohen (2018), o modelo aqui proposto não requer treinamento. Ao invés disso, ele faz uso de um conjunto de regras e de diversos recursos linguísticos-computacionais bilíngues, para a tradução conceitual para o português. O XPTA é inovador no sentido de utilizar outros recursos bilíngues no mapeamento conceitual de uma AMR-fonte (em inglês) para uma AMR-alvo (em português), não sendo de-

pendente do alinhamento lexical como em Damonte & Cohen (2018).

De forma similar a proposta por Damonte & Cohen (2018), Sheth et al. (2021) também treinaram *parsers* AMR para várias línguas alvo, a partir da projeção da anotação de grafos AMR do inglês. Contudo, a projeção da anotação AMR-fonte para a AMR-alvo é baseada no alinhamento de palavras que compartilham contextos similares, fornecidos por um modelo pré-treinado de vetores de palavras multilíngues, o XLM-RoBERTa (Conneau et al., 2020), baseado em Transformadores (*Transformers*) (Vaswani et al., 2017). Os alinhamentos conceituais foram obtidos pelo JAMR (Flanigan et al., 2014) e pelo alinhador proposto por Pourdamghani et al. (2014). Com base nos alinhamentos, os autores projetaram o alinhamento entre os grafos AMR-fonte e as sentenças alvo e, posteriormente, treinaram os *parsers* nas línguas alvo, usando o modelo de *parsing* baseado em transição com Transformadores de Pilha (*Stack-Transformers*) proposto por Fernandez Astudillo et al. (2020).

Ao contrário dos trabalhos citados até o momento, a abordagem de Blloshmi et al. (2020) não usa alinhamentos. Mais especificamente, com o intuito de eliminar a necessidade dos alinhamentos, os autores empregaram um modelo de aprendizado por transferência, o seq2seq (*sequence-to-sequence*) (Zhang et al., 2019). A ideia por trás desse modelo é aproveitar a anotação AMR existente para o inglês para construir modelos de aprendizado que sejam capazes de generalizar e reproduzir esse tipo de anotação para outras línguas. A abordagem dos autores é baseada em duas etapas: identificação de conceitos e identificação de relações. A primeira etapa é realizada pelo modelo seq2seq, que usa uma rede neural BiLSTM (*Bidirectional Long-Short Term Memory*) como *encoder* e outra unidirecional LSTM como *decoder*. A partir das sentenças de entrada, o modelo produz uma lista de conceitos (ou nós). Para a identificação de relações (segunda etapa), os autores utilizaram um classificador *bi-affine* desenvolvido por Dozat & Manning (2017), que cria relações entre os nós identificados na etapa anterior, buscando sempre pela relação com maior pontuação entre as arestas possíveis.

A proposta de Cai et al. (2021) também é baseada no modelo seq2seq, porém, durante o treinamento, os autores fizeram uso de entradas bilíngues (isto é, sentenças na língua fonte concatenadas às respectivas traduções na língua alvo) e de uma tarefa auxiliar, que visava a predição da sentença original de entrada (em inglês). Segundo os autores, as entradas bilíngues, juntamente com a tarefa auxiliar, favoreceram a predição de conceitos AMR mais precisos. Outra diferença em relação ao trabalho de Blloshmi et al. (2020) está no uso de um modelo de Transformadores no *encoder-decoder* (Vaswani et al., 2017), enquanto que em Blloshmi et al. (2020) foi utilizado um modelo baseado em Redes Neurais Recorrentes (LSTM).

A Tabela 1 sumariza os resultados alcançados pelos modelos de Damonte & Cohen (2018), Blloshmi et al. (2020), Cai et al. (2021) e (Sheth et al., 2021) para medida- $F$  calculada pela *Smatch* (conforme será explicada na seção 5.2.1). Essas quatro estratégias foram avaliadas utilizando o conjunto de teste do corpus LDC2017T10<sup>7</sup>, que dispõe de sentenças em inglês e as respectivas traduções para o italiano, espanhol, alemão e mandarim chinês, sendo 1.371 sentenças para cada língua. Sheth et al. (2021), entretanto, fizeram essa avaliação somente para os três primeiros idiomas.

Conforme se pode observar na Tabela 1, o desempenho de cada estratégia varia de acordo com o idioma. Por exemplo, para o alemão e o chinês o modelo baseado em Transformadores de Cai et al. (2021) obteve os melhores resultados, enquanto que para o italiano e o espanhol, o modelo de projeção de anotação que usa alinhamentos contextuais (Sheth et al., 2021) obteve o melhor desempenho.

Apesar das divergências linguísticas existentes entre as línguas e, sobretudo, as diferenças estruturais e conceituais que frequentemente decorrem do processo de tradução de uma língua fonte para uma língua alvo, os trabalhos aqui mencionados mostraram que as abordagens entre línguas têm potencial para superar muitas dessas divergências.

#### 4. O *parser* XPTA

O XPTA parte de um *parser* AMR existente na língua inglesa e, com base em recursos bilíngues inglês-português, mapeia/traduz o conhecimento semântico disponível na representação do significado da sentença na língua fonte (inglês) para

a representação equivalente na língua alvo (português). Em outras palavras, trata-se de um modelo de tradução conceitual no qual os conceitos (nós) dos grafos AMR-fonte são traduzidos (mapeados) para os conceitos dos grafos do AMR-alvo. A Figura 3 traz uma ilustração das etapas implementadas no XPTA. O mapeamento é baseado em duas etapas de processamento em *pipeline*, a saber: (1) **alinhamento conceitual** entre grafos e sentenças em inglês e (2) **tradução conceitual e lexicalização** do grafo AMR-fonte para o grafo AMR-alvo.

O **alinhamento conceitual** (etapa 1) consiste em identificar as correspondências entre palavras/conceitos da sentença em inglês com um fragmento do grafo, representado por um ou mais nós. Essa etapa é bastante relevante, pois permite obter as correspondências conceituais também entre o grafo AMR-fonte e a sentença paralela na língua alvo (se esta estiver disponível), por meio do alinhamento lexical entre as sentenças paralelas.

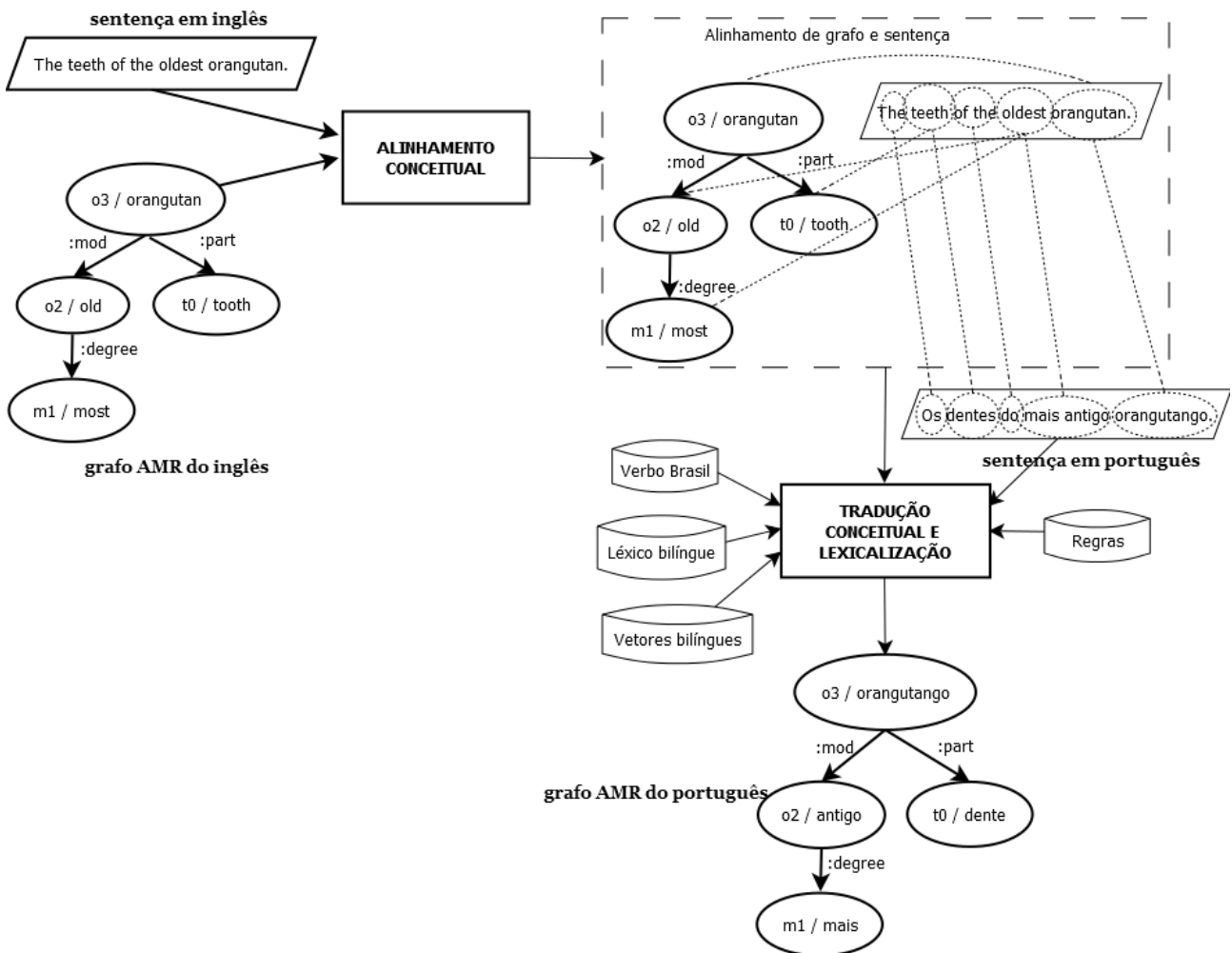
O alinhamento conceitual é realizado pelo alinhador proposto por Flanigan et al. (2014), o qual se baseia em um conjunto de regras definidas manualmente. Nos experimentos realizados pelos autores, esse alinhador apresentou um desempenho de 92% de precisão, 89% de cobertura e 90% de medida- $F$ .

A Figura 4 apresenta a sentença em inglês e seu respectivo grafo (notação PENMAN), ilustrado na Figura 3, no formato de entrada do alinhador conceitual. Esse arquivo de entrada é fornecido pelo *parser* AMR do inglês<sup>8</sup>. A saída do alinhador é um arquivo semelhante ao da entrada, acrescido dos alinhamentos entre um ou mais *tokens* da sentença e um ou mais nós do grafo, como no exemplo apresentado na Figura 5. Os seguintes alinhamentos foram gerados pelo algoritmo para o exemplo da Figura 3: “o3 / orangutan” com *orangutan* (isto é, alinhamento 5-6|0), “o2 / old” e “m / most” com *oldest* (alinhamento 4-5|0.0+0.0.0) e “t0 / tooth” com *teeth* (alinhamento 1-2|0.1).

Na **tradução conceitual e lexicalização** (segunda etapa), o alinhamento conceitual produzido na etapa 1 e os diversos recursos bilíngues são usados como âncora para a tradução conceitual e a lexicalização, que consiste na escolha da unidade lexical que melhor representa cada conceito da AMR-fonte na língua alvo. A entrada para esta etapa do processamento consiste na saída gerada pelo alinhador conceitual (con-

<sup>7</sup><https://catalog.ldc.upenn.edu/LDC2017T10> (acessado em 24/03/2022)

<sup>8</sup>Outras anotações também fornecidas pelo *parser*, como informações de *Part-of-Speech* e NER, foram omitidas na figura por questão de espaço.



**Figura 3:** Etapas do XPTA, com destaque para a tradução conceitual de grafo AMR-fonte (inglês) para o AMR-alvo (português).

```

# ::snt    The teeth of the oldest orangutan
# ::tok    The teeth of the oldest orangutan
# ::node   o2    old      4-5
# ::node   m1    most     3-4
# ::node   t0    tooth    1-2
# ::node   o3    orangutan 5-6
# ::edge   old   :mod-of  orangutan  o2    o3
# ::edge   most  :degree-of old      m1    o2
# ::edge   tooth :part-of orangutan  t0    o3
(o3 / orangutan
 :mod (o2 / old
      :degree (m1 / most))
 :part (t0 / tooth))

```

**Figura 4:** Sentença em inglês e seu respectivo grafo (notação PENMAN) no formato de entrada do alinhador conceitual.

forme Figura 5) e nos alinhamentos entre as palavras da sentença em inglês e as palavras da sentença paralela em português. Na Figura 6 é apresentado um exemplo de alinhamento lexical para as sentenças paralelas “The teeth of the oldest orangutan” e “Os dentes do mais antigo orangu-

tango” no formato de entrada esperado pelo tradutor conceitual. No exemplo, os seguintes alinhamentos foram produzidos: “the” e “o”, “teeth” e “dentes”, “of” e “do”, “oldest” e “mais antigo” e “orangutan” e “orangutango”.

```

# ::snt    The teeth of the oldest orangutan
# ::tok    The teeth of the oldest orangutan
# ::node   o2      old 4-5
# ::node   m1      most 3-4
# ::node   t0      tooth 1-2
# ::node   o3      orangutan 5-6
# ::edge   old     :mod-of orangutan o2 o3
# ::edge   most    :degree-of old m1 o2
# ::edge   tooth   :part-of orangutan t0 o3
# ::alignments 5-6|0 1-2|0.1 4-5|0.0+0.0.0 ::annotator
Aligner v.03 ::date 2020-10-08T14:53:40.073
# ::node   0      orangutan 5-6
# ::node   0.0    old 4-5
# ::node   0.0.0  most 4-5
# ::node   0.1    tooth 1-2
# ::root   0      orangutan
# ::edge   old     degree   most    0.0 0.0.0
# ::edge   orangutan mod     old     0     0.0
# ::edge   orangutan part    tooth   0     0.1
(o3 / orangutan
 :mod (o2 / old
 :degree (m1 / most))
 :part (t0 / tooth))

```

**Figura 5:** Saída fornecida pelo alinhador conceitual para a sentença e o grafo apresentados na Figura 3.

```

os dentes do mais antigo orangotango
NULL ( ) the ( 1 ) teeth ( 2 ) of ( 3 ) the ( ) oldest ( 4 5 ) orangutan ( 6 )

```

**Figura 6:** Alinhamentos lexicais entre sentenças paralelas inglês-português.

A tradução conceitual e lexicalização ocorre por meio de 3 passos, nesta ordem: (1) mapeamento de verbos por meio da consulta ao Verbo Brasil, (2) mapeamento dos conceitos via aplicação das regras e (3) mapeamento dos conceitos usando os diversos recursos linguísticos-computacionais bilíngues.

Em um primeiro momento, os predicados verbais são mapeados usando os *framesets* fornecidos pelo Verbo Brasil (Duran & Aluísio, 2012), os quais estão alinhados aos *framesets* do PropBank (Palmer et al., 2005). Os verbos modais (representados na AMR por *possible-01*, *obligate-01*, *permit-01*, etc.) não são mapeados para o português, uma vez que não estão no Verbo Brasil ou não estão alinhados aos seus correspondentes no inglês. Vale dizer também que os verbos de cópula não são representados na AMR e, portanto, não aparecem no grafos AMR-fonte.

Após o mapeamento de verbos, os conceitos restantes são mapeados com base na aplicação de 5 regras que foram propostas a partir da análise do corpus:

#### – REGRA 1: Entidades Nomeadas

Nos subgrafos com raiz representando uma entidade nomeada (por exemplo, *location*, *coun-*

*try*, *state*, *person*, etc.), a raiz e o seu descendente direto (isto é, o conceito *name*) são preservados e somente os descendentes de *name*, que carregam o nome da entidade e se relacionam com esse por meio de arestas do tipo *:op1*, *:op2*, etc., são mapeados para o português e colocados em ordem. A Figura 7-a traz um exemplo de aplicação desta regra no qual a entidade nomeada “Southern Hemisphere” é mapeada e lexicalizada para “Hemisfério Sul”.

#### – REGRA 2: Advérbios

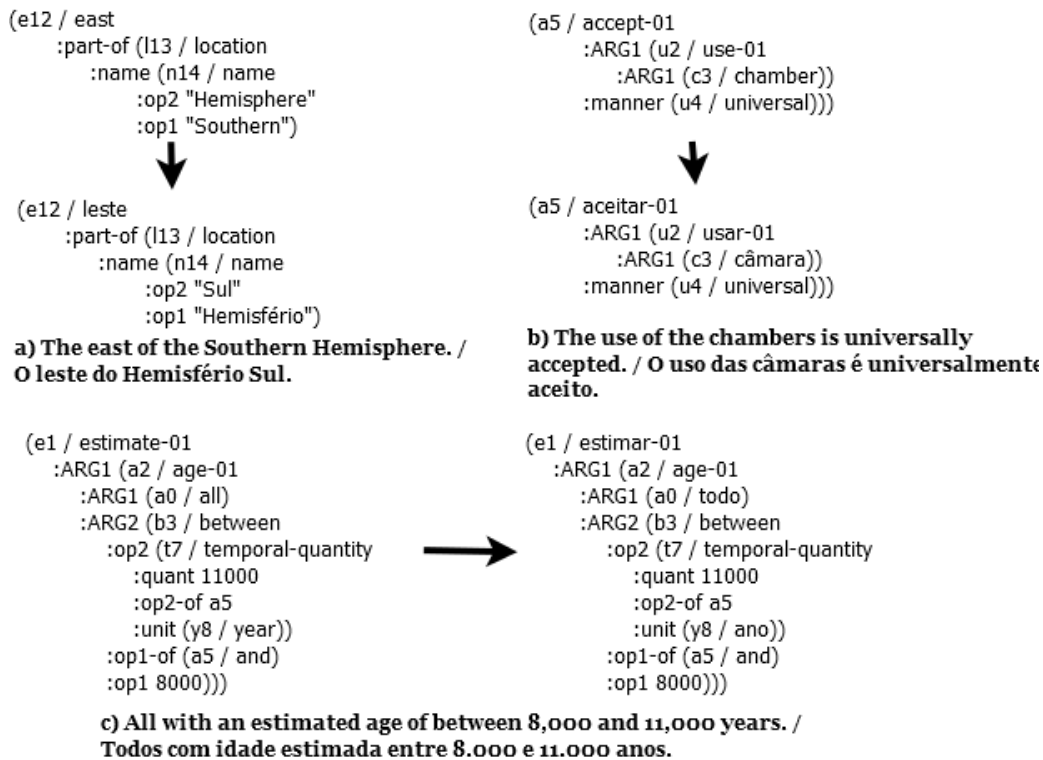
Seguindo as *guidelines* da AMR<sup>9</sup>, advérbios terminados em “ly” (por exemplo, *universally*, *incredibly*, *extremely*, etc.), cujas traduções no português terminam em “mente”, são mapeados para a sua forma adjetiva no português (por exemplo, *universalmente* → *universal*, *incrivelmente* → *incrível*, *extremamente* → *extremo*, etc.). A Figura 7-b ilustra a aplicação desta regra para o advérbio “universal”.

#### – REGRA 3: Frames especiais e reificação

Conceitos representando *frames* especiais da AMR (*have-rel-role-91* e *have-org-role-91*) ou reificação (por exemplo, *have-purpose-91*, *be-*

<sup>9</sup><https://github.com/amrisi/amr-guidelines/blob/master/amr.md> (acessado em 29/08/2021).





**Figura 7:** Exemplos de tradução conceitual de grafos AMR (formato PENMAN).

from-91, age-01, etc.) são preservados em inglês. A Figura 7-c traz um exemplo de mapeamento em que o conceito reificado “age-01” é mantido na representação AMR-alvo.

#### – REGRA 4: Conceitos-chaves

Conceitos da AMR representando quantidades (como *monetary-quantity*, *temporal-quantity*, *distance-quantity*, etc.) ou entidades como datas (*date-entity*), porcentagens (*percentage-entity*), urls (*url-entity*), entre outras, são preservados em inglês. A Figura 7-c ilustra o mapeamento da quantidade de tempo (*temporal-quantity*) do inglês para o português, no qual apenas o conceito “year” foi alterado para “ano”, preservando-se o restante da estrutura e seus valores.

#### – REGRA 5: Conjunções

Conjunções como *and*, *or*, *contrast-01*, *either* e *between* são preservadas em inglês, como ocorre com “between” e “and” no exemplo da Figura 7-c.

O mapeamento de conceitos gerais, ou seja, para os quais nenhuma regra se aplica, é realizado com base nos recursos bilíngues. Neste trabalho, investigou-se o uso de três recursos bilíngues, aplicados nesta ordem: (1) alinhamento lexical entre a sentença fonte e a sentença alvo, (2) vetores bilíngues fonte-alvo e (3) léxico bilíngue fonte-alvo.

### 1. Alinhamento lexical

Para o alinhamento lexical fonte-alvo, utilizou-se o alinhador GIZA++<sup>10</sup> (Och & Ney, 2004). Os alinhamentos lexicais permitem identificar as palavras (conceitos) correspondentes entre duas sentenças paralelas: a original (fonte) e a correspondente tradução (alvo). Um exemplo de alinhamento lexical obtido via GIZA++ pode ser visto na Figura 6. Embora bastante útil para o processo de tradução e lexicalização, nem sempre as sentenças paralelas estão disponíveis. Neste caso, os demais recursos se tornam bastante úteis.

### 2. Vetores bilíngues

Os vetores de palavras (*word embeddings*) bilíngues usados neste trabalho são os disponibilizados pelo MUSE<sup>11</sup>. Esses vetores foram treinados para o par de línguas inglês-português a partir de cópulas multilíngue (e não paralelo) da Wikipedia.

### 3. Léxico bilíngue

O léxico bilíngue (dicionário de tradução) utilizado neste trabalho é o disponível no PORTAL<sup>12</sup> (Vieira & Caseli, 2011). Esse léxico foi gerado a partir de alinhamentos lexicais ob-

<sup>10</sup><https://github.com/moses-smt/giza-pp> (acessado em 11/08/2021).

<sup>11</sup><https://github.com/facebookresearch/MUSE/blob/master/README.md> (acessado em: 11/08/2021).

<sup>12</sup><http://www.lalic.dc.ufscar.br/portal/> (acessado em 11/08/2021).

tidos pelo GIZA++ para o cópuz paralelo inglês-português da FAPESP (Aziz & Specia, 2011).

```
(o3 / orangutango
  :mod (o2 / antigo
    :degree (m1 / mais))
  :part (t0 / dente))
```

**Figura 8:** Grafo AMR gerado pelo XPTA para a sentença “Os dentes do mais antigo orangutango.”.

Após encontrar o melhor correspondente lexical fonte para um conceito alvo, seguindo a ordem de aplicação de recursos apresentada anteriormente, cada conceito é, então, mapeado para a sua forma canônica usando o lematizador do UDPipe<sup>13</sup>.

A saída do processo de tradução conceitual e lexicalização é o grafo AMR em português correspondente à entrada em inglês na notação PENMAN, conforme ilustrado na Figura 8.

## 5. Experimentos e resultados

O XPTA foi avaliado automaticamente em um conjunto de 200 sentenças. A avaliação foi realizada intrinsecamente comparando os grafos AMR produzidos pelo *parser* proposto com grafos de referência criados por anotadores humanos. Para essa avaliação, foram usadas três medidas automáticas conhecidas na área: *Smatch* (Cai & Knight, 2013), SEMA (Anchieta et al., 2019) e SemBleu (Song & Gildea, 2019). Além da avaliação automática, os grafos gerados automaticamente foram manualmente analisados pelos anotadores, com o propósito de categorizar e contabilizar os principais erros cometidos pelo modelo.

A subseção a seguir descreve o cópuz utilizado nos experimentos, bem como a construção do cópuz de referência. A avaliação automática e a análise manual serão descritas nas subseções 5.2 e 5.3, respectivamente.

### 5.1. Preparação do cópuz

Para os experimentos realizados neste trabalho foi adotado o cópuz paralelo FAPESP (Aziz & Specia, 2011), composto por textos de divulgação científica escritos originalmente em português e suas respectivas traduções para o inglês, os quais

foram extraídos da revista científica Pesquisa FAPESP<sup>14</sup>.

Para as avaliações do XPTA foram usados 200 pares de sentenças paralelas, conforme será explicado na subseção 5.1.1.

Os grafos AMR correspondentes às sentenças na língua fonte (inglês, neste caso) foram obtidos pelo *parser* proposto por Lyu & Titov (2018). Baseado em um modelo neural, esse *parser* representa o estado da arte para a língua inglesa com medida-*F* reportada pelos autores de 74,4%.

Como a construção de cópuz AMR é uma tarefa bastante laboriosa, optou-se por gerar as referências a partir da pós-edição de grafos produzidos automaticamente, reduzindo, assim, a complexidade da tarefa. Assim, para a construção do cópuz AMR de referência na língua alvo, foram utilizados como ponto de partida os grafos AMR do português gerados pelo XPTA (vide seção 4).

A subseção a seguir descreve a construção do cópuz de referência.

#### 5.1.1. Construção do cópuz AMR de referência

Para a construção do cópuz AMR de referência na língua alvo (português), foram selecionados os 200 pares de sentenças paralelas do cópuz FAPESP melhores ranqueados de acordo com o *score* de alinhamento produzido pelo Giza++ (Och & Ney, 2004). Esse critério de seleção das sentenças está fundamentado no fato do modelo de tradução conceitual ter como passo inicial o uso dos alinhamentos lexicais, de modo que erros nesse processo impactariam negativamente o desempenho do XPTA. Além do mais, as sentenças com melhores alinhamentos tendem a ser mais curtas, o que também simplifica a produção do grafo AMR-fonte.

A partir dos grafos AMR produzidos para o inglês para essas 200 sentenças, foram gerados os grafos paralelos no português, usando a configuração completa do XPTA descrita na seção 4. Posteriormente, os grafos na língua alvo (português) foram pós-editados por 4 anotadores nativos do português, todos pesquisadores da área de PLN com experiência em anotação de cópuz AMR.

Com o objetivo de estabelecer um padrão de revisão e de correção dos grafos, os anotadores participaram de duas sessões de treinamento. Na primeira sessão, todos editaram conjuntamente 20 grafos<sup>15</sup>. Na segunda sessão, outros 20 gra-

<sup>14</sup><http://revistapesquisa.fapesp.br/> (acessado em 11/08/2021).

<sup>15</sup>O cópuz de referência e as diretrizes da anotação/pós-edição estão disponíveis em: <https://github.com/TakeLab/spacy-udpipe> (acessado em 11/08/2021).

<sup>13</sup><https://github.com/TakeLab/spacy-udpipe> (acessado em 11/08/2021).

fos foram dados aos 4 anotadores para realizar a edição de forma individual, o que possibilitou calcular a concordância entre eles e aprimorar a tarefa por meio de discussões sobre os principais erros cometidos.

Após as sessões de treinamento, os 160 grafos restantes do corpus foram divididos em 4 conjuntos de 40 grafos cada, que foram pós-editados em 4 etapas. Durante cada etapa eram formadas 2 duplas de anotadores e cada dupla era responsável por revisar metade dos grafos (isto é, 20) daquele conjunto. Os dois membros de cada dupla revisavam, separadamente, os 20 grafos daquela etapa. Posteriormente, a concordância entre os membros de cada dupla era calculada. A fim de garantir maior homogeneidade na tarefa de pós-edição de grafos, as duplas eram refeitas a cada nova etapa não havendo, portanto, repetição da mesma dupla para conjuntos diferentes de grafos.

A concordância entre os anotadores foi calculada usando a medida *Smatch* (Cai & Knight, 2013), conforme será explicada na seção 5.2.1. Essa medida tem sido amplamente usada na literatura (Banarescu et al., 2013; Damonte & Cohen, 2018; Anchiêta et al., 2019), tanto no cálculo da concordância entre anotadores como na comparação de grafos AMR produzidos automaticamente com grafos de referência.

A concordância entre todos os anotadores foi calculada a partir do valor *Smatch* obtido por cada dupla de anotadores, considerando todas as combinações de pares possíveis. Durante o treinamento, a média desses valores foi de 0,77, sendo que o menor e o maior valor de *Smatch* obtidos foram 0,72 e 0,86, respectivamente. Após o treinamento, a concordância média entre as duplas na pós-edição do corpus foi de 0,78, sendo 0,70 e 0,87 o menor e o maior valor de *Smatch* alcançados, respectivamente.

Os valores de concordância obtidos neste trabalho são superiores aos reportados por outros trabalhos da literatura. Banarescu et al. (2013), por exemplo, alcançaram valores entre 0,70 e 0,80 (0,71 em média) na construção de um corpus AMR para o inglês, enquanto Sobrevilla Cabezudo & Pardo (2019) obtiveram valores de *Smatch* entre 0,70 e 0,77 (0,72 em média) na construção de um corpus AMR de notícias em português. Uma explicação possível para a alta concordância obtida entre os anotadores neste trabalho pode ser o fato de que as duplas tinham como ponto de partida da anotação o mesmo grafo gerado automaticamente pelo XPTA.

A *Smatch*, no entanto, não considera a equivalência entre grafos que apresentam variações de conceitos e/ou de relações, mas que são considerados equivalentes de acordo com a AMR, e atribui um baixo valor quando essas variações ocorrem. As duas estruturas AMR exibidas na Figura 9, por exemplo, são equivalentes, sendo que a (a) utiliza reificação (isto é, a relação de causa é representada por meio de um conceito), enquanto a (b) não utiliza.<sup>16</sup>

(s / sair-01	(s / sair-01
:ARG1 (m / menino)	:ARG1 (m / menino)
:ARG1-of (c / cause-01	:cause (c / chegar-01
:ARG0 (c1 / chegar-01	:ARG0 (g / garoto)))
:ARG0 (g / garoto)))))	
<b>a) Com reificação</b>	<b>b) Sem reificação</b>

**Figura 9:** Grafos AMR equivalentes para a sentença: *A menina saiu porque o garoto chegou* (Formato PENMAN).

Com o intuito de tornar essa comparação mais justa, Goodman (2019) propôs um normalizador de grafos AMR, o Norman, que, entre outras transformações, converte um grafo não reificado para um grafo reificado, garantindo, assim, que grafos similares sejam avaliados como equivalentes. O Norman foi aplicado aos grafos pós-editados pelos anotadores e o *Smatch* foi novamente calculado. A concordância média obtida durante a fase de treinamento passou de 0,77 para 0,81, após a normalização, sendo que os valores obtidos entre cada dupla de anotadores ficaram entre 0,77 e 0,89. A concordância média obtida pelas duplas para o corpus todo (depois do treinamento), por sua vez, passou de 0,78 para 0,81, sendo o menor e o maior valor obtidos 0,74 e 0,90, respectivamente.

Exemplos de grafos de referência podem ser vistos na seção 5.3, nas Figuras 13 e 14.

## 5.2. Avaliação Automática

Com o objetivo de medir a qualidade dos grafos do português, produzidos a partir do mapeamento de grafos de sentenças paralelas em inglês, esses foram comparados automaticamente com os grafos de referência, gerados conforme descrito na subseção 5.1.1. A subseção seguinte descreve as

<sup>16</sup>Reificação é quando uma relação AMR é representada por meio de um conceito. Esse recurso é usado quando se deseja enfatizar a importância de uma relação. Por exemplo, a relação “:cause” ao ser substituída pelo conceito “cause-01” é enfatizada no grafo. Para uma explicação mais detalhada, sugere-se consultar as especificações da AMR em <https://github.com/amrisi/amr-guidelines/blob/master/amr.md> (acessado em 15/08/2021).

medidas de avaliação usadas nessa comparação. Os resultados da avaliação, por sua vez, são apresentados e discutidos na subseção 5.2.3.

### 5.2.1. Medidas de Avaliação

Três medidas de avaliação automática de grafos AMR disponíveis na literatura foram usadas neste trabalho, a saber: *Smatch* (Cai & Knight, 2013), SEMA (Anchieta et al., 2019) e SemBleu (Song & Gildea, 2019).

A *Smatch* calcula a Precisão, a Cobertura e a medida-*F* de conceitos e relações comuns entre dois grafos AMR. Mais especificamente, cada grafo é representado como uma conjunção de proposições lógicas, nomeadas de triplas, e a Precisão (*P*) e a Cobertura (*C*) são calculadas sobre essas triplas, seguindo as Equações 1 e 2, respectivamente. Nessas equações, *M* representa o total de triplas corretas segundo algum grafo AMR de referência, *N* representa o número total de triplas produzidas por outro anotador humano ou por um *parser* e *T* representa o total de triplas do grafo AMR de referência. A medida-*F*, por sua vez, representa a média harmônica entre a Precisão e a Cobertura (Equação 3).

$$P = \frac{M}{N} \quad (1)$$

$$C = \frac{M}{T} \quad (2)$$

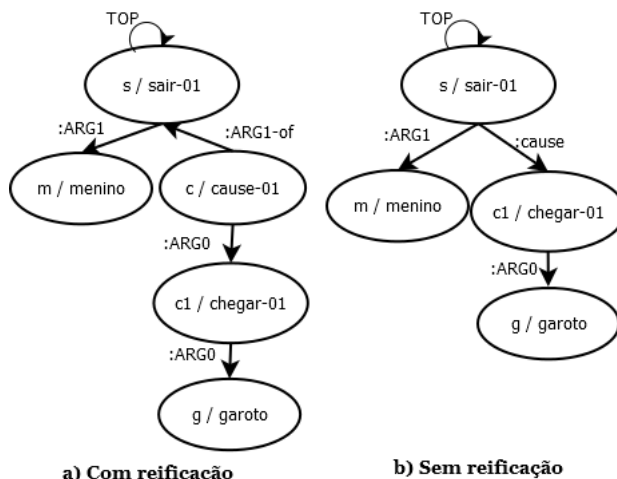
$$F = \frac{2 \times P \times C}{P + C} \quad (3)$$

Ao fazer a comparação entre duas representações AMR, a *Smatch* cria uma relação “TOP” para cada grafo, indicando o nó que representa a raiz. Ao avaliar, por exemplo, o grafo (a) da Figura 9 (teste) comparando-o com o grafo (b) (referência), reproduzidos na Figura 10 em formato de diagrama, a *Smatch*<sup>17</sup> retorna os seguintes valores:

- *M* = 7 uma vez que há 7 triplas corretas: “sair-01”, “menino”, “chegar-01”, “garoto”, “TOP”, “:ARG1” entre “sair-01” e “menino” e “:ARG0” entre “chegar-01” e “garoto”;
- *N* = 10 uma vez que há 10 triplas no total: “sair-01”, “menino”, “cause-01”, “chegar-01”, “garoto”, “TOP”, “:ARG1” entre “sair-01” e “menino”, “:ARG1” entre “cause-01” e “sair-01”, “:ARG0” entre “cause-01” e “chegar-01” e “:ARG0” entre “chegar-01” e “garoto”) e

- *T* = 8 uma vez que há 8 triplas no grafo AMR de referência: “sair-01”, “menino”, “chegar-01”, “garoto”, “TOP”, “:ARG1” entre “sair-01” e “menino”, “:cause” entre “sair-01” e “chegar-01” e “:ARG0” entre “chegar-01” e “garoto”.

Portanto, *P* é igual a 70% ( $7 \div 10$ ), *C* é igual a 88% ( $7 \div 8$ ) e *F* é igual a 78%.



**Figura 10:** Grafos AMR equivalentes para a sentença: *A menina saiu porque o garoto chegou.*

A SEMA é uma extensão da *Smatch* que, ao contrário desta, leva em consideração a dependência entre os nós dos dois grafos no cálculo da Precisão e da Cobertura (Equações 1 e 2). Em outras palavras, a SEMA considera o antecedente de um nó ao fazer a comparação de triplas, enquanto *Smatch* considera apenas a contagem de triplas em comum entre dois grafos. Outra diferença é que a SEMA não inclui a relação “TOP” na raiz do grafo. Ao comparar os grafos da Figura 10 usando a SEMA tem-se *P* igual a 67% (isto é,  $6 \div 9$ , desconsiderando a relação “TOP”), *C* igual a 86% (ou seja,  $6 \div 7$  ao desconsiderar “TOP”) e *F* igual a 75%.

SemBleu, por sua vez, é uma variante da BLEU (Papineni et al., 2002), uma medida amplamente usada na avaliação de tradutores automáticos, que calcula a correlação entre um texto (ou sentença) traduzido automaticamente e um texto de referência produzido por um tradutor humano. De maneira similar, SemBleu calcula a correlação entre um grafo AMR produzido automaticamente (*a*) e outro grafo de referência (*r*). Diferentemente de *Smatch* e SEMA, SemBleu considera a coocorrência de *n*-gramas (*strings* idênticas) entre dois grafos, ao invés de usar triplas lógicas.

<sup>17</sup>[https://www.isi.edu/cgi-bin/div3/mt/text\\_shu\\_v03.cgi](https://www.isi.edu/cgi-bin/div3/mt/text_shu_v03.cgi) (acessado em 15/08/2021).



A medida SemBleu estende a BLEU redefinindo o coeficiente de penalidade, *Brevity Penalty - BP*, que multiplica a média geométrica dos valores de precisão ( $pn$ ) obtidos para cada  $n$ -grama, conforme mostrado na Equação 4.  $BP$  representa o tamanho de um grafo, dado pela soma do número de nós e de arestas que ele possui. A precisão ( $pn$ ) de cada  $n$ -grama é dada pelo total de  $n$ -gramas em comum entre ( $a$ ) e ( $r$ ), dividido pelo número de  $n$ -gramas de ( $a$ ).  $pn$  é ponderado por um peso positivo  $w_n = 1/3$ , onde 3 é o comprimento máximo de  $N$ . Em outras palavras, SemBleu considera apenas uni-gramas (nós), bigramas (pares de nós diretamente conectados) e trigramas (três nós diretamente conectados). Por exemplo, para o grafo  $b$  da Figura 10 tem-se como unigramas os nós “sair-01”, “menino”, “chegar-01” e “garoto”, como bigramas tem-se “sair-01 :ARG1 menino”, “sair-01 :cause chegar-01” e “chegar-01 :ARG0 garoto” e, por fim, como trigrama tem-se “sair-01 :cause chegar-01 :ARG0 garoto”.

$$\text{SemBleu} = BP \cdot \exp \sum_{n=1}^N w_n \log p_n \quad (4)$$

Na comparação entre os dois grafos da Figura 10, SemBleu é igual a 46%.

### 5.2.2. Configurações do modelo

Com o objetivo de avaliar a contribuição individual de cada recurso linguístico-computacional usado no mapeamento dos grafos AMR-fonte para o AMR-alvo, o XPTA foi executado usando diferentes configurações que são apresentadas a seguir:

- VB + ali + vet + lex** Com todos os recursos, isto é, Verbo Brasil (VB), alinhamentos lexicais e conceituais, vetores e léxico bilíngues;
- VB + ali + lex** Usando o VB, os alinhamentos lexicais e conceituais e o léxico bilíngue;
- VB + ali + vet** Usando o VB, os alinhamentos lexicais e conceituais e os vetores bilíngues;
- VB + vet + lex** Usando o VB, léxico e vetores bilíngues;
- VB + lex** Usando somente o VB e o léxico bilíngue;
- VB + vet** Usando apenas o VB e os vetores bilíngues;
- VB + ali** Usando somente o VB e os alinhamentos lexicais e conceituais.

Como o Verbo Brasil é o único recurso usado no mapeamento de verbos, ele foi mantido em todas as configurações.

### 5.2.3. Resultados da Avaliação Automática e Discussão

As Tabelas 2 e 3 apresentam os valores obtidos por cada configuração do XPTA para *Smatch*, SEMA e SemBleu, antes e depois da normalização dos grafos pelo Norman (Goodman, 2019), respectivamente.

Com base nos valores das três medidas de avaliação usadas é possível notar um melhora substancial dos valores apresentados na Tabela 2 após a normalização (conforme Tabela 3), independente da configuração do modelo. A melhora nos valores apresentados por *Smatch* e SemBleu foi em média de 8 pontos. Para SEMA, o ganho depois da normalização foi ainda maior: de 10 a 13 pontos, dependendo da configuração. Esses resultados apontam a dificuldade dessas medidas em lidar com variações de significado equivalentes presentes nas representações AMR, e ressaltam a importância da normalização para possibilitar uma avaliação mais justa do modelo.

De maneira geral, quando se compara as diferentes configurações testadas, observa-se pouca diferença nos valores apresentados por cada uma delas, independente da medida utilizada. Os piores valores, contudo, foram alcançados quando se usou apenas o Verbo Brasil (VB) e os alinhamentos (VB + ali). De fato, como se pode observar nos resultados obtidos pela configuração que usou todos os recursos menos os alinhamentos (VB + lex + vet) e também nos resultados apresentados pelas três últimas configurações (vide Tabela 2), o léxico e os vetores bilíngues parecem ter tido um papel mais importante no mapeamento conceitual do que os alinhamentos. Uma possível explicação para isso pode ser erros decorrentes do alinhamento lexical, fornecidos pelo Giza++, ou conceitual, fornecidos pelo JAMR.

Comparando o uso do léxico e dos vetores bilíngues, nota-se que ao aplicar o VB e os alinhamentos juntamente com o léxico (VB + ali + lex), o resultado foi ligeiramente superior ao obtido quando se aplicou o VB e os alinhamentos com os vetores (VB + ali + vet), para todas as medidas. O mesmo se observa quando se exclui os alinhamentos dessas duas configurações, isto é, quando se aplica apenas o VB e o léxico (VB + lex), nota-se uma pequena melhora (de 1 até 3 pontos dependendo da medida) em relação ao uso do VB com os vetores (VB + vet). Neste caso é relevante destacar que enquanto os léxicos

Modelo	Smatch			SEMA			SemBleu F-score
	P	C	F-score	P	C	F-score	
VB + ali + vet + lex	58%	57%	57%	<b>35%</b>	<b>34%</b>	<b>35%</b>	<b>26%</b>
VB + ali + lex	<b>59%</b>	<b>57%</b>	<b>58%</b>	<b>35%</b>	<b>34%</b>	<b>35%</b>	<b>26%</b>
VB + ali + vet	58%	57%	57%	<b>35%</b>	<b>34%</b>	<b>35%</b>	25%
VB + lex + vet	57%	56%	57%	34%	33%	33%	25%
VB + lex	57%	56%	57%	34%	33%	33%	25%
VB + vet	55%	54%	55%	32%	32%	32%	24%
VB + ali	55%	54%	54%	32%	31%	31%	24%

**Tabela 2:** Valores obtidos por cada configuração do modelo para *Smatch*, SEMA e SemBleu antes da normalização.

Modelo	Smatch			SEMA			SemBleu F-score
	P	C	F-score	P	C	F-score	
VB + ali + vet + lex	67%	64%	65%	47%	45%	46%	33%
VB + ali + lex	<b>67%</b>	<b>64%</b>	<b>66%</b>	<b>48%</b>	<b>46%</b>	<b>47%</b>	<b>34%</b>
VB + ali + vet	67%	64%	65%	47%	45%	46%	33%
VB + lex + vet	66%	64%	65%	47%	45%	46%	<b>34%</b>
VB + lex	66%	63%	65%	47%	45%	46%	<b>34%</b>
VB + vet	65%	62%	64%	44%	42%	43%	33%
VB + ali	64%	62%	63%	43%	41%	42%	31%

**Tabela 3:** Valores obtidos por cada configuração do modelo para *Smatch*, SEMA e SemBleu após a normalização pelo Norman.

foram gerados para o mesmo corpus (corpus FA-PESP), os vetores bilíngues são de domínio geral e foram gerados a partir de textos da Wikipedia. Entretanto, não é possível afirmar que essas diferenças são significativas, devido à falta de testes estatísticos.

É importante notar também que a configuração VB + lex mostrou-se equivalente à configuração que usou todos os recursos, ressaltando a contribuição do léxico bilíngue para o mapeamento conceitual. Vale ressaltar que o bom desempenho apresentado por recursos já disponíveis, como o léxico bilíngue do PorTAL e os vetores bilíngues do MUSE, demonstra a aplicabilidade do XPTA mesmo na ausência de uma versão paralela da sentença a ser processada.

Por fim, ressalta-se que a experimentação com diferentes configurações de recursos usados pelo XPTA foi realizada apenas a título de análise empírica do impacto de cada recurso no valor da medida de avaliação. Como o intuito não é estabelecer qual é a melhor configuração, uma vez que os recursos já existem e se complementam, nenhuma análise de significância estatística foi realizada.

Na comparação dos resultados apresentados pelas três medidas de avaliação, para todas as configurações do XPTA, os maiores valores foram obtidos com a *Smatch*. Por se tratar de uma

medida mais simples e que não considera a dependência entre os nós de um grafo AMR, esse resultado já era esperado de certa forma. Isso explica também a diminuição nos valores apresentados por SEMA, quando comparados aos obtidos com a *Smatch*, que, ao contrário dessa medida, considera a dependência entre os elementos do grafo ao fazer a comparação de triplas. Por sua vez, os menores valores foram apresentados por SemBleu, que tende a ser ainda mais rígida do que SEMA, ao comparar n-gramas.

Os resultados relatados com base apenas nas medidas de avaliação automática indicam que a abordagem adotada é factível e as representações AMR podem ser compartilhadas com êxito entre as línguas, apesar de suas diferenças linguísticas e estruturais. Contudo, para permitir uma análise mais detalhada do impacto dessas diferenças na representação AMR entre idiomas diferentes, realizou-se também uma análise qualitativa dos grafos AMR em português gerados pelo XPTA.

### 5.3. Análise Qualitativa

Com o propósito de identificar e categorizar os principais erros cometidos pelo XPTA, durante a construção do corpus AMR de referência do português (vide seção 5.1.1), os anotadores foram solicitados a registrar cada edição (correção) que

faziam nos grafos. Mais especificamente, cada anotador recebia um arquivo com os grafos a serem pós-editados e as respectivas sentenças em português. Para cada grafo os anotadores registravam:

- o total de *frameset* de verbos substituídos, ou seja, quando outro *frameset* disponível no Verbo Brasil era substituído pelo escolhido pelo modelo;
- o total de conceitos substituídos (excluindo-se os verbos), isto é, quando o anotador fazia outra escolha lexical para representar um conceito;
- o total de relações (arestas) substituídas, já que em muitos casos a mudança de *frameset* provocava também alterações nas relações que representavam seus argumentos;
- o total de *links* (conexões) alterados, isto é, quando um nó era conectado a outro antecedente diferente do escolhido pelo modelo;
- o total de conceitos novos incluídos (inclusive verbos), isto é, que não haviam sido representados no grafo antes;
- o total de relações incluídas uma vez que, normalmente, a inclusão de um novo conceito leva à inclusão de uma relação;
- o total de conceitos removidos, inclusive verbos;
- o total de relações excluídas já que a remoção de um conceito normalmente leva à remoção de uma relação também;
- o total de conceitos lematizados, ou seja, que apresentavam erros de lematização e eram corrigidos pelos anotadores.

O gráfico da Figura 11 resume o total de edições realizadas para cada uma dessas categorias. Foram realizadas 992 edições em todo o corpus. A porcentagem de grafos em que cada tipo de edição ocorreu é apresentada na Figura 12.

Como se pode observar na Figura 11, a substituição de relações foi a principal edição realizada pelos anotadores, seguida pela substituição de conceitos gerais (não incluindo verbos). A substituição de relações e de conceitos correspondem, respectivamente, a 19% e 18% das edições realizadas. Ambas ocorreram em 67,7% dos grafos (Figura 12). Contudo, quando se considera a substituição de conceitos e de verbos (*framesets*) juntas, elas representam 25,2% de todas as correções.

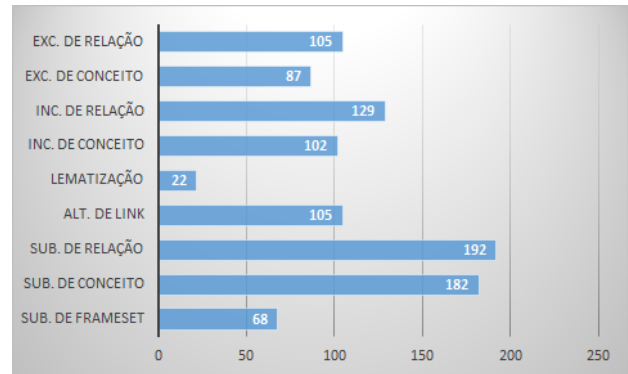


Figura 11: Total de edições realizadas por categoria.

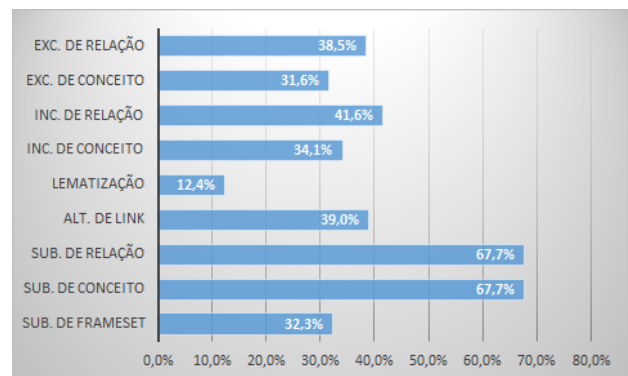


Figura 12: Porcentagem de grafos em que cada tipo de edição foi realizada.

A inclusão de novas relações também foi comum, ocorrendo em 41,6% dos grafos. Ao contrário do que se poderia esperar, a inserção de uma relação nem sempre está relacionada à inclusão de um novo conceito, conforme se observa nos gráficos das Figuras 11 e 12. A correferência é um exemplo típico de inclusão de uma relação, que não envolve a adição de um novo conceito.

A exclusão de relações e as alterações de conexões (*links*) representam, cada uma, 10,5% do total de edições. A primeira acontece quando há alguma remoção de conceito (nó) ou quando um nó deixa de ser interno e passa a ser a raiz do grafo. As alterações de conexões, por sua vez, estão associadas às mudanças na estrutura dos grafos.

Com menos frequência observa-se alterações de *framesets* de verbos (representando 6,8% do total de edições) e lematização de conceitos (representando 2,2%).

O número médio de edições realizadas pelos anotadores, em cada grafo, foi de 6. Contudo, 34,7% dos grafos tinham entre 0 e 3 erros apenas, sendo que desses 10,0% estavam corretos (nenhum erro), 7,4% com 1 erro, 9,3% com 2 erros somente e 8,0% continham exatamente 3 erros.

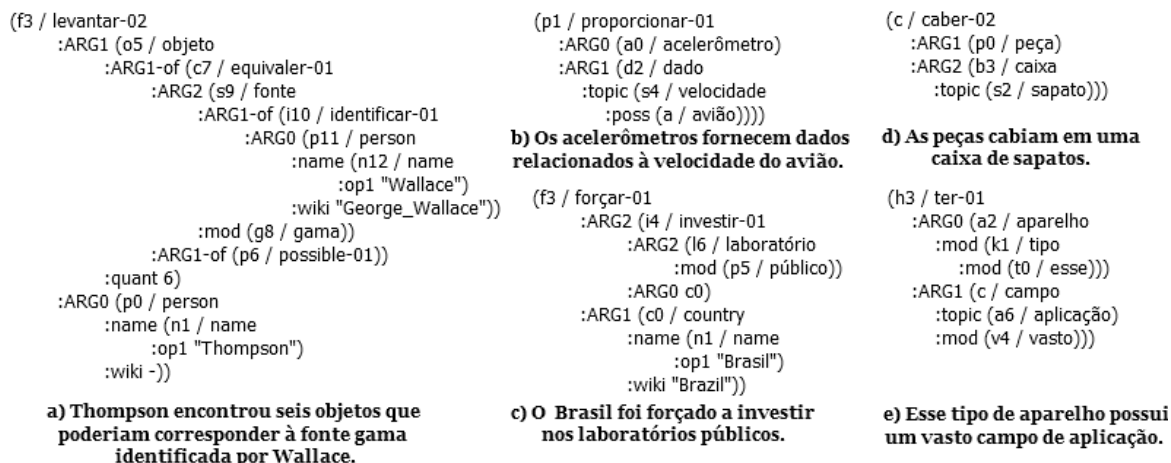


Figura 13: Exemplos de grafos corretos gerados pelo XPTA (formato PENMAN).

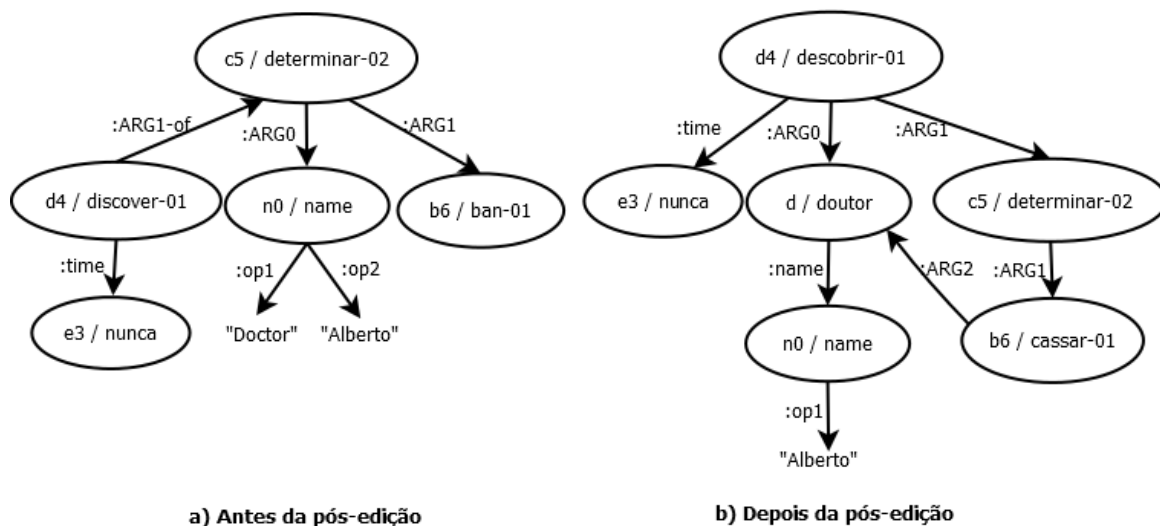


Figura 14: Exemplo de grafo gerado pelo XPTA antes e depois da pós-edição pelos humanos.

A Figura 13 apresenta 5 exemplos de grafos gerados pelo XPTA (seção 4), julgados corretos pelos anotadores. É importante dizer que o mapeamento de relações “:wiki” não foi abordado na versão atual do XPTA e, portanto, quando presentes nos grafos gerados a partir do inglês, essas relações eram simplesmente ignoradas pelos anotadores<sup>18</sup>.

Um exemplo de grafo antes e depois da pós-edição dos anotadores pode ser visto na Figura 14. No exemplo ocorreram duas substitui-

ções de *framesets* (“discover-01” → “descobrir-01” e “ban-01” → “cassar-01”), uma troca de conceito (“Doctor” → “Alberto”), uma exclusão de conceito (“Alberto”) e, conseqüentemente, da relação que este estabelecia com o seu antecedente (ou seja, “:op2”), uma inclusão de conceito (“doutor”), duas inclusões de relação (“:name” e a correferência indicada por “:ARG2” entre “cassar-01” e “doutor”) e, por fim, uma mudança de conexão (*link*) do nó “name”, que tinha como antecedente “determinar-02” e passou a conectar-se ao nó “doutor”, por meio da relação “:name”.

Os erros observados nos grafos gerados pelo XPTA têm várias origens distintas, uma vez que cada recurso linguístico-computacional adotado pode causar algum tipo de erro no processo. Por exemplo, as falhas relacionadas aos *framesets*

<sup>18</sup>A relação “:wiki” é usada pela AMR para representar a forma canônica de entidades nomeadas, por exemplo, “NY” é representada por “New York City”. O *parser* AMR (Lyu & Titov, 2018) usa a Wikipédia do inglês para recuperar a forma canônica de entidades nomeadas. Quando ele não encontra, a relação é representada por “:wiki -”.



estão normalmente associadas à falta de um correspondente no Verbo Brasil ou, ainda, à falta de um mapeamento entre o *frameset* do inglês e o seu correspondente no português. Em outras palavras, embora haja um correspondente de um *frameset* qualquer do inglês no Verbo Brasil, este nem sempre está associado ao *frameset* do PropBank. Esse é o caso de “discover-01” e “ban-01” (Figura 14), que apesar de terem um correspondente no Verbo Brasil, isto é, “descobrir-01” e “cassar-01”, respectivamente, falta uma associação entre eles. Devido à isso, não foi possível fazer o mapeamento automaticamente. Essa dependência do PropBank, no caso do inglês, e do Verbo Brasil, no caso do português, se impõe como uma das limitações da própria AMR.

Há também erros advindos do próprio grafo AMR-fonte, que foram derivados do processo de construção do mesmo pelo *parser* ou, ainda, de alguma ferramenta de pré-processamento usada por ele. Esse deve ser o caso, por exemplo, do conceito “Doctor” (na Figura 14-a), que foi incluído no grafo como um componente de uma entidade nomeada. Outro exemplo é a relação de correferência não identificada pelo *parser* do inglês, ou seja, entre “cassar-01” e a subárvore que tem com raiz “doutor”.

Outro tipo de erro propagado no modelo proposto é derivado do lematizador do UDPipe. Conforme mostrado na Figura 12, os erros de lematização representam 12,4% dos erros encontrados no cópulo de avaliação.

Por fim, os erros nos alinhamentos lexicais e/ou conceituais também podem causar ruídos, especialmente quando se utiliza apenas os alinhamentos no processo de mapeamento conceitual. Porém, ao combinar outros recursos como léxico e vetores bilíngues, os conceitos e palavras não alinhados geralmente são mapeados com base nesses recursos.

Embora os resultados com as medidas de avaliação automática tenham sido promissores, a partir do que foi exposto com a discussão da análise manual apresentada nesta seção, conclui-se que há, ainda, muito espaço para se aprimorar os resultados apresentados neste trabalho, conforme será discutido na seção 6.

## 6. Conclusões

Este artigo descreveu o processo de construção de um *parser* AMR para o português baseado em uma abordagem entre línguas: o XPTA. O XPTA parte de *parser* AMR existente para o inglês e de recursos linguísticos-computacionais bilíngues, além de regras definidas a partir de

análise de cópulo, e mapeia (traduz) o conhecimento semântico disponível na língua fonte (inglês) para gerar uma representação AMR equivalente na língua alvo (português). Alguns recursos como vetores bilíngues, léxico bilíngue e alinhamentos lexicais e conceituais podem ser empregados de forma independente pela abordagem ou conjuntamente (conforme visto na seção 5.2.2).

Enquanto algumas abordagens da literatura são completamente dependentes de cópulo paralelos bilíngues (vide, por exemplo, Damonte & Cohen (2018) e Sheth et al. (2021)), uma avaliação automática do XPTA usando diferentes configurações dos recursos empregados pelo modelo mostrou a sua aplicabilidade mesmo na ausência da versão paralela da sentença a ser processada. Mais especificamente, ao adotar apenas o repositório Verbo Brasil e os vetores de palavras multilíngues como recursos de base para o mapeamento conceitual da anotação AMR do inglês para o português, o modelo atingiu um *Smatch* de 64%. Todavia, o maior valor para *Smatch*, ou seja, 66%, foi obtido usando alinhamentos lexicais entre sentenças paralelas juntamente com o Verbo Brasil e um léxico bilíngue. Como as sentenças usadas na avaliação foram selecionadas tendo como critério os melhores *scores* de alinhamento fornecidos pelo Giza++, é possível que as traduções sejam bastante fiéis às sentenças originais. Investigações futuras serão necessárias para verificar o impacto de se utilizar traduções divergentes nas configurações do modelo que usam os alinhamentos paralelos como recurso para o mapeamento conceitual.

Embora não seja possível fazer uma comparação direta com outros trabalhos da literatura, ao projetar os grafos AMR para o italiano, o espanhol, o alemão e o chinês, a partir dos grafos AMR de sentenças paralelas do inglês usando somente os alinhamentos lexicais e conceituais, Damonte & Cohen (2018) alcançaram um *Smatch* de 43% para o italiano, 42% para o espanhol, 39% para o alemão e 35% para o chinês, ao comparar os grafos produzidos pelo modelo com grafos de referência.<sup>19</sup> Ao avaliar a aplicabilidade do modelo de Damonte & Cohen (2018) para o português do Brasil, usando o cópulo do Pequeno Príncipe (Anchieta & Pardo, 2018a), Anchieta & Pardo (2018b) observaram um *Smatch* de apenas 37%.

<sup>19</sup>Os autores relatam apenas resultados para os grafos sem normalização.

A abordagem adotada aqui é passível de ser reproduzida para outras línguas alvo, desde que disponham dos recursos e ferramentas linguístico-computacionais necessários (conforme seção 4).

Além da avaliação automática, uma análise manual dos grafos produzidos pelo XPTA permitiu categorizar os erros cometidos e identificar suas origens. A análise mostrou, por exemplo, que os erros relacionados à escolha das relações que conectam dois conceitos (192 no total, conf. Figura 11) e também relacionados à estruturação dos grafos propriamente dita (105 erros conf. Figura 11) estão entre os principais problemas identificados. Esses dois tipos de erros juntos representam cerca de 30% de todos os problemas encontrados no corpus, ou seja, 297 erros de um total de 992, de acordo com a Figura 11. Esses erros são advindos do *parser* AMR do inglês, que forneceu os grafos AMR-fonte. É importante dizer que o XPTA aborda apenas o mapeamento conceitual, preservando as relações e a estrutura dos grafos AMR-fonte.

Como investigações futuras, elencam-se duas que trariam maior benefício para o modelo aqui apresentado: (1) investigar o mapeamento de relações e (2) melhorar a associação entre os *framesets* do Verbo Brasil e o do PropBank, incluindo o mapeamento entre *frames* equivalentes já existentes no repositório. Além dessas, outra possibilidade seria investigar o mapeamento de conceitos e relações “wiki”, não abordados na versão atual do XPTA.

Como continuação deste trabalho, está em desenvolvimento uma interface *web* que permitirá a comunidade de PLN o acesso rápido e gratuito ao XPTA.

## Referências

- Anchiêta, Rafael & Thiago Pardo. 2018a. Towards AMR-BR: A SemBank for Brazilian Portuguese language. Em *11<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*, 974–979.
- Anchiêta, Rafael T. 2020. *Abstract meaning representation parsing for the Brazilian Portuguese language*: Universidade de São Paulo — ICMC. Tese de Doutorado.
- Anchiêta, Rafael T., Marco Antonio Sobrevilla Cabezudo & Thiago A. S. Pardo. 2019. SEMA: an extended semantic evaluation metric for AMR. *CoRR* abs/1905.12069. arXiv.
- Anchiêta, Rafael Torres & Thiago Alexandre Salgueiro Pardo. 2018b. A rule-based AMR parser for Portuguese. Em *16<sup>th</sup> Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, 341–353. doi:10.1007/978-3-030-03928-8\_28.
- Aziz, Wilker & Lúcia Specia. 2011. Fully automatic compilation of a Portuguese–English parallel corpus for statistical machine translation. Em *8<sup>th</sup> Brazilian Symposium in Information and Human Language Technology (STIL)*, 234–238.
- Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer & Nathan Schneider. 2013. Abstract meaning representation for Sembanking. Em *7<sup>th</sup> Linguistic Annotation Workshop and Interoperability with Discourse*, 178–186.
- Blloshmi, Rexhina, Rocco Tripodi & Roberto Navigli. 2020. XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques. Em *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2487–2500. doi:10.18653/v1/2020.emnlp-main.195.
- Bonial, Claire, Bianca Badarau, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Tim O’Gorman, Martha Palmer & Nathan Schneider. 2018. Abstract meaning representation of constructions: The more we include, the better the representation. Em *11<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*, em linha.
- Bos, Johan. 2016. Squib: Expressive power of Abstract Meaning Representations. *Computational Linguistics* 42(3). 527–535.
- Cai, Shu & Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. Em *51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*, 748–752.
- Cai, Yitao, Zhe Lin & Xiaojun Wan. 2021. Making better use of bilingual information for cross-lingual AMR parsing. Em *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 1537–1547. doi:10.18653/v1/2021.findings-acl.134.
- Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer & Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. Em *58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 8440–8451. doi:10.18653/v1/2020.acl-main.747.

- Damonte, Marco & Shay B. Cohen. 2018. Cross-lingual abstract meaning representation parsing. Em *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 1146–1155. doi 10.18653/v1/N18-1104.
- Damonte, Marco, Shay B. Cohen & Giorgio Satta. 2017. An incremental parser for abstract meaning representation. Em *15<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 536–546.
- Dohare, Shibhansh & Harish Karnick. 2017. Text summarization using abstract meaning representation. *CoRR* abs/1706.01678. arXiv.
- Donatelli, Lucia, Michael Regan, William Croft & Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. Em *Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions*, 96–108.
- Dozat, Timothy & Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. Em *5<sup>th</sup> International Conference on Learning Representations*, 24–26.
- Duran, Magali Sanches & Sandra M. Aluísio. 2012. Propbank-Br: a Brazilian treebank annotated with semantic role labels. Em *8<sup>th</sup> Brazilian Symposium in Information and Human Language Technology (STIL)*, em linha.
- Dyer, Chris, Victor Chahuneau & Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. Em *Conference of the North American Chapter of the Association of Computational Linguistics (NAACL)*, 644–648.
- Fernandez Astudillo, Ramón, Miguel Ballesteros, Tahira Naseem, Austin Blodgett & Radu Florian. 2020. Transition-based parsing with stack-transformers. Em *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1001–1007. Online: Association for Computational Linguistics.
- Flanigan, Jeffrey, Sam Thomson, Jaime Carbonell, Chris Dyer & Noah A. Smith. 2014. A discriminative graph-based parser for the abstract meaning representation. Em *52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics*, 1426–1436. doi 10.3115/v1/P14-1134.
- Garg, Sahil, Aram Galstyan, Ulf Hermjakob & Daniel Marcu. 2016. Extracting biomolecular interactions using semantic parsing of biomedical text. Em *30<sup>th</sup> AAAI Conference on Artificial Intelligence*, 2718–2726. doi 10.1609/aaai.v30i1.10337.
- Goodman, Michael W. 2019. AMR normalization for fairer evaluation. Em *33<sup>rd</sup> Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, 37–46.
- Hovy, Eduard & Julia Lavid. 2010. Towards a ‘science’ of corpus annotation : A new methodological challenge for corpus linguistics. *International Journal of Translation Studies* 22(1). 13–36.
- Inácio, Marcio L. 2021. *Sumarização de opinião com base em abstract meaning representation*: Universidade de São Paulo — ICMC. Tese de Mestrado. doi 10.11606/D.55.2021.tde-13092021-141741.
- Jurafsky, Dan & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*. Prentice Hall.
- Lehmann, Fritz. 1992. *Semantic networks in artificial intelligence*. Elsevier Science Inc.
- Liao, Kexin, Logan Lebanoff & Fei Liu. 2018. Abstract meaning representation for multi-document summarization. Em *27<sup>th</sup> International Conference on Computational Linguistics (COLING)*, 1178–1190.
- Liu, Fei, Jeffrey Flanigan, Sam Thomson, Norman Sadeh & Noah A. Smith. 2015. Toward abstractive summarization using semantic representations. Em *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 1077–1086. doi 10.3115/v1/N15-1114.
- Lyu, Chunchuan & Ivan Titov. 2018. AMR parsing as graph prediction with latent alignment. Em *56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 397–407. doi 10.18653/v1/P18-1037.
- Matthiessen, Christian & John A Bateman. 1991. *Text generation and systemic-functional linguistics: experiences from english and japanese*. Pinter Publishers.
- Mitra, Arindam & Chitta Baral. 2016. Addressing a question answering challenge by combining statistical methods with inductive rule learning and reasoning. Em *30<sup>th</sup> AAAI Conference on Artificial Intelligence*, 2779–2785. doi 10.1609/aaai.v30i1.10354.

- van Noord, Rik & Johan Bos. 2017. Neural semantic parsing by character-based translation: Experiments with abstract meaning representations. *Computational Linguistics in the Netherlands Journal* 7. 93–108.
- Och, Franz J. & Hermann Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics* 30(4). 417–449. doi 10.1162/0891201042544884.
- Palmer, Martha, Daniel Gildea & Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics* 31(1). 71–106. doi 10.1162/0891201053630264.
- Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of Machine Translation. Em *40<sup>th</sup> Annual Meeting on Association for Computational Linguistics*, 311–318. doi 10.3115/1073083.1073135.
- Peng, Xiaochang, Chuan Wang, Daniel Gildea & Nianwen Xue. 2017. Addressing the data sparsity issue in neural AMR parsing. Em *15<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 366–375.
- Pourdamghani, Nima, Yang Gao, Ulf Hermjakob & Kevin Knight. 2014. Aligning English strings with abstract meaning representation graphs. Em *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 425–429. doi 10.3115/v1/D14-1048.
- Sheth, Janaki, Young-Suk Lee, Ramón Fernández Astudillo, Tahira Naseem, Radu Florian, Salim Roukos & Todd Ward. 2021. Bootstrapping multilingual AMR with contextual word alignments. Em *16<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 394–404. doi 10.18653/v1/2021.eacl-main.30.
- Sobrevilla Cabezudo, Marco Antonio & Thiago Pardo. 2019. Towards a general abstract meaning representation corpus for Brazilian Portuguese. Em *13<sup>th</sup> Linguistic Annotation Workshop*, 236–244. doi 10.18653/v1/W19-4028.
- Song, Linfeng & Daniel Gildea. 2019. SemBleu: A robust metric for AMR parsing evaluation. Em *57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 4547–4552. doi 10.18653/v1/P19-1446.
- Song, Linfeng, Daniel Gildea, Yue Zhang, Zhiguo Wang & Jinsong Su. 2019. Semantic neural machine translation using AMR. *Transactions of the Association for Computational Linguistics* 7. 19–31. doi 10.1162/tacl\_a\_00252.
- Uchida, Hiroshi, Meiyang Zhu & Tarcisio Della Senta. 2006. *UNL: Universal networking language*. UNDL Foundation, International Environment House.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. Em *31<sup>st</sup> International Conference on Neural Information Processing Systems*, 5998–6008.
- Vieira, Thiago L. & Helena M. Caseli. 2011. PORTAL: Recursos e ferramentas de tradução automática para o Português do Brasil. Em *Brazilian Symposium in Information and Human Language Technology (STIL)*, 179–183.
- Vilares, David & Carlos Gómez-Rodríguez. 2018. A transition-based algorithm for unrestricted AMR parsing. Em *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 142–149. doi 10.18653/v1/N18-2023.
- Xue, Nianwen, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova & Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. Em *9<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)*, 1765–1772.
- Zhang, Sheng, Xutai Ma, Kevin Duh & Benjamin Van Durme. 2019. AMR parsing as sequence-to-graph transduction. Em *57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 80–94. doi 10.18653/v1/P19-1009.
- Zhou, Junsheng, Feiyu Xu, Hans Uszkoreit, Weiguang Qu, Ran Li & Yanhui Gu. 2016. AMR parsing with an incremental joint model. Em *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 680–689. doi 10.18653/v1/D16-1065.