

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Sumarização de Opinião com base em Abstract Meaning Representation

Marcio Lima Inácio

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-C²MC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Marcio Lima Inácio

Sumarização de Opinião com base em Abstract Meaning Representation

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Thiago Alexandre Salgueiro Pardo

USP – São Carlos
Setembro de 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

I35s Inácio, Marcio Lima
 Sumarização de Opinião com base em Abstract
Meaning Representation / Marcio Lima Inácio;
orientador Thiago Alexandre Salgueiro Pardo. -- São
Carlos, 2021.
 149 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2021.

 1. Sumarização Automática. 2. Mineração de
Opiniões. 3. Semântica. 4. Abstract Meaning
Representation. I. Pardo, Thiago Alexandre
Salgueiro, orient. II. Título.

Marcio Lima Inácio

Opinion Summarization based on Abstract Meaning
Representation

Dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Master in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Thiago Alexandre Salgueiro Pardo

USP – São Carlos
September 2021

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 88887.355305/2019-00. Também houve o apoio do projeto OPINANDO (*Opinion Mining for Portuguese*) e do grupo NLP2 do C4AI (*Center for Artificial Intelligence*).

RESUMO

INÁCIO, M. L. **Sumarização de Opinião com base em Abstract Meaning Representation**. 2021. 149 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

A área de Mineração de Opiniões visa a processar automaticamente textos subjetivos que emitem a crítica do autor a respeito de alguma entidade (como produtos ou serviços). Essa área vem crescendo devido às grandes quantidades de dados produzidos na *web*, tendo como uma de suas aplicações a Sumarização de Opiniões, em que um sistema gera automaticamente um resumo dos principais comentários a respeito da entidade avaliada. Alguns trabalhos mais recentes propõem abordagens baseadas em análises semânticas mais profundas através do uso de representações semânticas, argumentando que essas análises produzem melhores resultados capazes de lidar com sentimentos e informações implícitas no texto. Entre as representações semânticas existentes na literatura, a *Abstract Meaning Representation* (AMR) vem ganhando notoriedade por se basear em recursos bem consolidados, como o PropBank, e por ter apresentado bons resultados em diversas tarefas, como a Sumarização Automática. Neste trabalho de mestrado, propõe-se fazer o uso da representação AMR na Mineração de Opinião, mais especificamente aplicada à Sumarização de Opiniões, abordando diferentes frentes de trabalho, como a anotação de um corpus de opiniões em AMR e uma análise dos resultados dessa anotação em comparação a textos jornalísticos, bem como o desenvolvimento de novos métodos de sumarização automática de opiniões baseados em AMR. Como principal resultado, foi possível verificar que o uso das representações semânticas explícitas auxiliou na seleção de informações para os resumos. Além disso, alguns dos novos métodos desenvolvidos se mostraram melhores quando comparados às técnicas baseadas em AMR já existentes na literatura.

Palavras-chave: Sumarização Automática, Mineração de Opiniões, Semântica, Abstract Meaning Representation.

ABSTRACT

INÁCIO, M. L. **Opinion Summarization based on Abstract Meaning Representation**. 2021. 149 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Opinion Mining aims at automatically processing subjective texts which carry the sentiment of the author towards some entity (such as products or services). This area has been growing due to large amounts of data produced on the Web, having as one of its application Opinion Summarization, in which a system automatically generates a summary of the main commentaries concerning the evaluated entity. Some more recent works propose approaches based on deeper semantic analysis through the use of semantic representations, arguing that these analyses produce better results capable of dealing with implicit sentiments and information throughout the text. Among the semantic representations in the literature, Abstract Meaning Representation (AMR) has been gaining notoriety for being based on well-consolidated resources, such as the PropBank, and for having presented good results in many tasks, among which Automatic Summarization. In this project, we propose to make use of the AMR representation in Opinion Mining, more specifically applied to Opinion Summarization, addressing different work fronts, such as the annotation of an opinion corpus in AMR and a comparative analysis of these annotation results to news texts, as well as the development of new AMR-based Automatic Opinion Summarization methods. As a main result, we could verify that including explicit semantic representations improved content selection for the summaries. Furthermore, some of our novel methods outperformed other AMR-based techniques from the literature.

Keywords: Automatic Summarization, Opinion Mining, Semantics, Abstract Meaning Representation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de resumo extrativo	21
Figura 2 – Exemplo de resumo abstrativo	22
Figura 3 – Arquitetura para sistemas de sumarização automática	22
Figura 4 – Exemplo de resumo extrativo baseado em aspectos	24
Figura 5 – Exemplo de sentença representada em AMR	28
Figura 6 – Modelo de implementação CLSA	36
Figura 7 – Exemplo de resumo de opiniões gráfico com respeito ao sentimento	39
Figura 8 – Exemplo de resumo de opiniões gráfico com respeito à frequência de termos	40
Figura 9 – Exemplo de resumo de opiniões gráfico contrastivo	40
Figura 10 – Exemplo de resumo de opiniões textual contrastivo	41
Figura 11 – Exemplo de <i>frameset</i> do Verbo Brasil	42
Figura 12 – Exemplo de sentença representada em AMR	43
Figura 13 – Exemplo de sentença com negação representada em AMR	44
Figura 14 – Exemplo de sentença com pergunta representada em AMR	44
Figura 15 – Exemplo de sentença com verbalização de preposição anotada em AMR	45
Figura 16 – Exemplo de sentença com verbalização de substantivo anotada em AMR	45
Figura 17 – Exemplo de sentença com sujeito elíptico anotado em AMR	47
Figura 18 – Exemplo de sentença com verbos modais anotados em AMR	48
Figura 19 – Exemplo de grafo do sistema Opinosis	61
Figura 20 – Linha do tempo contendo os principais trabalhos da área	66
Figura 21 – Exemplo de sentença com verbos modais anotados em AMR	69
Figura 22 – Exemplo de anotação de metáfora	71
Figura 23 – Exemplo de anotação com marcador discursivo	72
Figura 24 – Exemplo de anotação com expressão multipalavra unida por hífen	73
Figura 25 – Exemplo de anotação com intervalo de tempo	73
Figura 26 – Exemplo de anotação com diminutivo semântico	74
Figura 27 – Exemplo de anotação com diminutivo pragmático	74
Figura 28 – Número de sentenças em cada grupo de acordo com o número de <i>tokens</i>	77
Figura 29 – Comparação da distribuição das relações entre comentários e resumos	80
Figura 30 – Comparação da distribuição dos conceitos entre comentários e resumos	80
Figura 31 – Grafos AMR de sentenças a serem mescladas em um único grafo multi-sentencial	84
Figura 32 – Grafo AMR multi-sentencial final	85

Figura 33 – Forma lógica correspondente a um grafo AMR referente à sentença “O menino quer brincar”	92
Figura 34 – Forma lógica correspondente à sentença “O menino quer o bilboquê”	92
Figura 35 – Expansão de um grafo usando sua Árvore Geradora Mínima para garantir conectividade	95
Figura 36 – Ilustração de funcionamento do método de <i>Simulated Annealing</i>	100
Figura 37 – Exemplo de Árvore de Decisão	103
Figura 38 – Exemplo de <i>Multilayer Perceptron</i>	103
Figura 39 – Exemplo de grafo de Levi	104
Figura 40 – Processo de criação automática do corpus de treinamento	108
Figura 41 – Visualização dos pesos otimizados no experimento <code>score_optimization</code>	116
Figura 42 – Importância dos atributos para cada técnica de AM utilizada	118
Figura 43 – Exemplo de grafo AMR e resumo para o produto “Iphone 5”	148
Figura 44 – Exemplo de grafo AMR e resumo para o produto “Galaxy SIII”	149

LISTA DE QUADROS

Quadro 1 – Exemplo de textos em que é feito o uso de <i>meta nodes</i>	55
Quadro 2 – Atributos utilizados por Liu <i>et al.</i> (2015) e Liao, Lebanoff e Liu (2018) para representar conceitos AMR	87
Quadro 3 – Atributos utilizados por Liu <i>et al.</i> (2015) e Liao, Lebanoff e Liu (2018) para representar relações entre pares de nós AMR	88
Quadro 4 – Exemplo de triplas extraídas pela ferramenta DptOIE	94

LISTA DE TABELAS

Tabela 1 – Resultados da sumarização (ROUGE-1) obtidos por Liu <i>et al.</i> (2015)	53
Tabela 2 – Resultados da seleção de nós e arestas (Medida F) obtidos por Liu <i>et al.</i> (2015)	53
Tabela 3 – Resultados da sumarização (ROUGE-1) obtidos por Dohare, Gupta e Karnick (2018)	56
Tabela 4 – Influência da medida de similaridade nos grafos sumarizados	57
Tabela 5 – Resultados da sumarização (ROUGE-01) obtidos por Liao, Lebanoff e Liu (2018)	57
Tabela 6 – Resultados obtidos por Carenini, Ng e Pauls (2006)	60
Tabela 7 – Resultados da sumarização (ROUGE-1) obtidos por Ganesan, Zhai e Han (2010)	62
Tabela 8 – Resultados da sumarização (ROUGE-01) obtidos por Condori e Pardo (2017)	64
Tabela 9 – Resultados da análise linguística realizada por Condori e Pardo (2017)	65
Tabela 10 – Estatísticas do cópuz OpiSums-PT	68
Tabela 11 – Resultados da anotação	70
Tabela 12 – Resultados do cálculo de concordância	70
Tabela 13 – Relações mais frequentes nos cópuz jornalístico e opinativo	75
Tabela 14 – <i>Framesets</i> mais frequentes no cópuz jornalístico e opinativo	76
Tabela 15 – Frequências (%) das 15 relações mais frequentes em textos de opinião de acordo com o tamanho das sentenças (em <i>tokens</i>)	78
Tabela 16 – Frequências (%) dos 10 <i>framesets</i> mais frequentes em textos de opinião de acordo com o tamanho das sentenças (em <i>tokens</i>)	79
Tabela 17 – Medida F obtida por cada alinhamento de variáveis possível	93
Tabela 18 – Resultado dos experimentos com a configuração LiuEtAl2015	110
Tabela 19 – Resultado dos experimentos com a configuração LiaoEtAl2018	111
Tabela 20 – Resultado dos experimentos com a configuração DohareEtAl2018	113
Tabela 21 – Resultado dos experimentos com a configuração DohareEtAl2018_TF	114
Tabela 22 – Conceitos selecionados nos experimentos DohareEtAl2018 e DohareEtAl2018_TF	114
Tabela 23 – Resultado dos experimentos com a configuração <code>score_optimization</code>	115
Tabela 24 – Resultado dos experimentos com a configuração <code>machine_learning</code>	117

Tabela 25 – Resultado dos experimentos com a configuração <code>machine_learning_levi</code>	120
Tabela 26 – Taxa de compressão dos diferentes métodos explorados	121
Tabela 27 – Melhores e piores resultados de cada modelo de sumarização semântica	122
Tabela 28 – Resultado dos experimentos com métodos de sumarização de opiniões .	124

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
AMR	<i>Abstract Meaning Representation</i>
BSU	<i>Basic Semantic Unit</i>
CLSA	<i>Concept-Level Sentiment Analysis</i>
ILP	<i>Integer Linear Programming</i>
ILP	<i>Integer Linear Programming</i>
InIt	<i>Information Item</i>
LCS	<i>Longest Common Subsequence</i>
MLP	<i>Multilayer Perceptron</i>
MO	Mineração de Opiniões
PAS	<i>Predicate Argument Structure</i>
PLN	Processamento de Língua Natural
TF-IDF	<i>Term frequency-inverse document frequency</i>
UCCA	<i>Universal Conceptual Cognitive Annotation</i>
UNL	<i>Universal Networking Language</i>
VSM	<i>Vector Space Models</i>

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização e Motivação	24
1.2	Objetivo e Hipóteses	29
1.3	Contribuições	30
1.4	Organização do Texto	31
2	CONCEITOS FUNDAMENTAIS	33
2.1	Mineração de opiniões	33
2.1.1	<i>Concept-Level Sentiment Analysis</i>	35
2.2	Sumarização de opiniões	37
2.3	Representação semântica	41
2.3.1	<i>Abstract Meaning Representation</i>	42
2.3.1.1	<i>Adaptações para a Língua Portuguesa</i>	46
2.4	Anotação de córpus	47
2.5	Considerações finais	50
3	TRABALHOS RELACIONADOS	51
3.1	Sumarização com representações semânticas	51
3.1.1	<i>Sumarização automática baseada em AMR</i>	52
3.1.2	<i>Outros trabalhos</i>	57
3.2	Sumarização de opiniões	58
3.2.1	<i>Condori e Pardo (2017)</i>	63
3.3	Considerações finais	65
4	ANOTAÇÃO DO CÓRPUS	67
4.1	Estrutura do córpus	67
4.2	Metodologia de anotação	68
4.3	Resultados de anotação	69
4.3.1	<i>Concordância de anotação</i>	69
4.3.2	<i>Dificuldades de anotação</i>	70
4.3.3	<i>Comparação entre textos opinativos e jornalísticos</i>	74
4.3.4	<i>Comparação entre diferentes tamanhos de sentenças</i>	77
4.4	Comparação entre comentários e resumos	79

5	MÉTODOS DE SUMARIZAÇÃO AUTOMÁTICA	83
5.1	Pré-processamento	83
5.2	Métodos de Sumarização Semântica	85
5.2.1	<i>Sumarização baseada em Programação Linear Inteira</i>	<i>86</i>
5.2.2	<i>Sumarização baseada em regras</i>	<i>93</i>
5.2.3	<i>Sumarização baseada em Aprendizado de Máquina</i>	<i>101</i>
5.3	Medidas de avaliação	105
5.4	Experimentos	106
5.4.1	<i>Treinamento</i>	<i>107</i>
5.4.2	<i>Geração textual</i>	<i>108</i>
5.5	Resultados	109
5.5.1	<i>Sumarização de opiniões</i>	<i>123</i>
6	CONSIDERAÇÕES FINAIS	127
6.1	Conclusões	127
6.1.1	<i>Contribuições</i>	<i>129</i>
6.2	Trabalhos Futuros	130
	REFERÊNCIAS	133
APÊNDICE A	EXEMPLOS DE GRAFOS AMR	147

INTRODUÇÃO

A área de mineração de textos tem como objetivo principal, assim como a mineração de dados, a extração de conhecimento a partir de bases de dados, focando-se especificamente em informações textuais (FELDMAN; SANGER, 2006). Como é indicado por Aggarwal e Zhai (2012), algumas aplicações da área de mineração de textos são: recuperação de informação, categorização textual e, em especial, sumarização, foco deste projeto.

A finalidade dos sistemas de sumarização automática é extrair as informações mais importantes de um documento, ou conjunto de documentos, para análise pelo usuário (MANI, 2001; AGGARWAL, 2018). Normalmente, os resumos gerados são extrativos, isto é, em que exclusivamente trechos de interesse nos textos originais são selecionados e apresentados. Um exemplo de resumo extrativo pode ser visto na Figura 1. O extrato contém sentenças retiradas de uma série de comentários de usuários acerca de um *smartphone*.

Figura 1 – Exemplo de resumo extrativo

A tela com certeza é algo que impressiona, possuindo 4 polegadas, uma resolução de 1136 x 640 pixels, e ainda com um vidro de alta resistência. Com certeza trata-se de um iOS muito avançado, e apesar de ter tudo e mais um pouco que qualquer pessoa deseja em um celular, não é pesado e nem muito grande, além de ser extremamente fino com apenas 7.6mm, o que facilita na hora de carregá-lo para todos os lugares. Muito prático, com boa aparência, e ótima qualidade. DESDE O DESIGN, AO PESO, TAMANHO, NITIDEZ, VELOCIDADE, CAPACIDADE, OTIMAS FOTOS, TELA DE RETINA (NUNCA VI IGUAL). Simplesmente fantástico! Eu recomendo!

Fonte: Adaptado do corpus OpiSums-PT (CONDORI *et al.*, 2015)

Diferentemente dos métodos extrativos, os métodos abstrativos de sumarização têm como função conectar as informações e tópicos extraídos em novas sentenças e for-

mulações que não necessariamente estariam presentes de forma literal no documento original (MANI, 2001, p. 129). A Figura 2 apresenta um exemplo de resumo abstrativo referente ao mesmo produto (*smartphone*) da Figura 1.

Figura 2 – Exemplo de resumo abstrativo

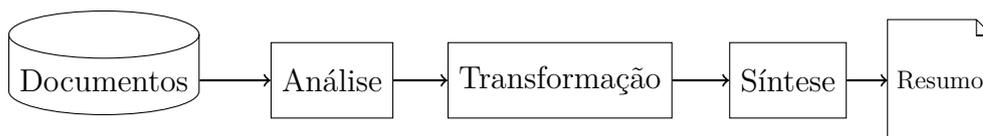
Todos os usuários gostaram e recomendam o Iphone 5. Algumas das frases que podem descrever o produto são: ótimo e sonho de consumo. Algumas das características que os usuários salientam são que possui uma tela ótima e a qualidade de imagens e vídeo é muito boa. Também, a fácil portabilidade, já que é leve; a longa durabilidade da bateria, a facilidade de uso, a velocidade de internet é muito rápida, é muito bom se for usado para videoconferência e finalmente, possui uma grande quantidade de funcionalidades que fazem como se fosse uma extensão do computador de escritório. A única coisa que foi colocada como ruim foi o alto preço.

Fonte: Córpus OpiSums-PT (CONDORI *et al.*, 2015)

É importante notar que métodos extrativos tendem a ser mais simples, por utilizarem mais atributos referentes à superfície do texto, como frequências de palavras, identificação de palavras-chave, tamanho das sentenças e suas posições nos documentos (NENKOVA; MCKEOWN, 2012). Eles também estão sujeitos ao formato, organização e possíveis desvios presentes nos textos originais. Como indicado por Mani (2001, p. 165), extratos tendem a ser mais incoerentes, sendo necessário um procedimento posterior de revisão do texto. Resumos abstrativos, por sua vez, tendem a se basear mais em conhecimento semântico e, com isso, eles buscam uma melhor organização das informações, principalmente devido aos métodos de geração textual utilizados. Essa abordagem possui, porém, a desvantagem de tender a ser específica de determinado domínio além de ser mais sensível a erros durante o processo, podendo apresentar informações imprecisas ou incoerentes, com possíveis problemas de escrita, ao usuário (MANI, 2001, p. 165).

Dependendo do tipo de sumarização a ser realizada, os processos aplicados podem ser mais ou menos sofisticados. Mani (2001) indica uma arquitetura genérica para sistemas de sumarização automática, apresentada na Figura 3.

Figura 3 – Arquitetura para sistemas de sumarização automática



Fonte: Adaptada de Mani (2001).

Essa arquitetura possui três etapas básicas: Análise, Transformação e Síntese. A Análise tem como objetivo construir uma representação computacional dos documentos de entrada. A etapa de Transformação, também chamada de Refinamento, transforma a representação da entrada na representação do resumo a ser produzido. Essa etapa é reali-

zada a partir de três operações básicas: seleção, em que é feita a filtragem dos elementos textuais relevantes; agregação, em que elementos são mesclados; e generalização, em que elementos são substituídos por outros mais abstratos ou gerais. Por fim, a Síntese converte a representação do sumário para língua natural, de forma a permitir a interpretação do usuário (GATT; KRAHMER, 2018).

Além da sumarização, os trabalhos em mineração de texto também têm aplicações na área de análise de sentimentos – também chamada de mineração de opinião – em que o conteúdo textual é classificado com relação ao sentimento ou opinião do autor sobre o assunto tratado. Comumente, os sentimentos são classificados em “positivo”, “neutro” ou “negativo”. Esses métodos podem ser utilizados em diversos âmbitos, como sistemas de recomendação e ferramentas de inteligência de negócios (PANG; LEE, 2008; PEREIRA, 2021).

Como intersecção, tem-se o campo de estudos da sumarização de opiniões, em que técnicas de sumarização automática são aplicadas no contexto da análise de sentimentos. Isso se dá, segundo Liu (2012), devido ao fato de opiniões serem essencialmente subjetivas, sendo necessária a análise de múltiplas visões para que o usuário tome alguma decisão, de forma que uma sumarização das principais opiniões se faz relevante.

Os trabalhos realizados na área de sumarização de opiniões, em sua maioria, têm como foco a geração de resumos extrativos (KIM *et al.*, 2011). Os trabalhos também se concentram em resumir informações com relação a diferentes aspectos (por exemplo, “bateria”, “tela”, “durabilidade”) presentes no texto sobre o objeto comentado (um *smartphone*, por exemplo). Um exemplo desse tipo de resumo pode ser observado na Figura 4, em que as opiniões relativas ao aspecto “imagem” de um produto são sumarizadas. O extrato apresenta a quantidade de opiniões referentes a cada polaridade (positivo e negativo), bem como exemplos de comentários.

Sistemas de sumarização de opiniões normalmente se baseiam em três etapas: a extração dos aspectos, a predição de sentimento e a geração do resumo (KIM *et al.*, 2011). A relação desse modelo com a arquitetura de Mani (2001), anteriormente apresentada, está, principalmente, na etapa de Análise: a extração de informações sobre os aspectos e sentimentos podem ser consideradas como parte dessa etapa. Por sua vez, a geração final do resumo é obtida seguindo os próximos passos da arquitetura de Mani (2001) (Transformação e Síntese). Métodos baseados em agrupamento, frequência e análise da forma de escrita das sentenças (posição no texto, integridade gramatical, posição dos termos na sentença) podem ser aplicados para a definição da relevância de cada frase.

A geração dos resumos também pode ser feita de maneira abstrativa, como ilustra o exemplo já apresentado na Figura 2, em que se mostra o resumo dos comentários sobre um *smartphone*. Como indicado por Mani (2001, p. 129), a criação de resumos abstrativos se baseia na interpretação dos tópicos presentes no texto, de forma a criar novas formulações.

Figura 4 – Exemplo de resumo extrativo baseado em aspectos

Aspecto: imagem
Positivo: 12
<ul style="list-style-type: none">• Em geral, essa é uma boa câmera com uma nitidez de imagem realmente boa.• As imagens são simplesmente incríveis – a câmera captura até os mínimos detalhes.• Depois de quase 800 fotos eu descobri que essa câmera tira fotos incríveis.
...
Negativo: 2
<ul style="list-style-type: none">• As imagens ficam desfocadas se a suas mãos tremem mesmo que um momento durante o processo de tirar uma foto.• Focando em um expositor a mais ou menos 6 metros de distância em uma sala bastante iluminada durante o dia, as imagens feitas por essa câmera ficaram desfocadas e com uma tonalidade alaranjada.

Fonte: Adaptada de [Condori e Pardo \(2017\)](#).

A etapa de geração do resumo final também é bastante importante, dependendo dos trabalhos na área de Geração de Língua Natural, sendo necessário se atentar aos desafios postos por esse tipo de tarefa, como a estruturação do texto, escolha de palavras e expressões, realização linguística, entre outros. Os trabalhos focados na sumarização abstrativa comumente utilizam modelos de sentenças pré-definidos para serem preenchidos com as informações extraídas, apesar de, com os métodos de aprendizado profundo recentes, a geração textual venha despertando mais interesse da comunidade ([HUANG; WU; WANG, 2020](#)).

Nesse contexto, este projeto de mestrado tem como foco a Sumarização de Opiniões, especialmente a etapa de Transformação, ou seja, a condensação das informações presentes nos textos opinativos. A seguir, apresentamos a motivação para este trabalho e seu contexto mais específico.

1.1 Contextualização e Motivação

Os métodos de sumarização podem ser classificados com relação à natureza do conhecimento aplicado durante o processo. [Mani \(2001\)](#) apresenta uma distinção entre métodos superficiais (*shallow*) e profundos (*deep*) de acordo com o nível de informação semântica usada. Métodos superficiais são aqueles que se baseiam principalmente na superfície textual, incorporando nenhum ou pouco conhecimento semântico, limitando-se apenas à semântica lexical, como o uso de ontologias ou vetores de palavras. Métodos profundos, por outro lado, tratam o texto em um maior grau de abstração, utilizando-se de representações semânticas em nível sentencial ou até mesmo textual, além de também

poder incorporar informações discursivas. Esse conhecimento é formulado de maneira a permitir a interpretação de quais conceitos estão presentes no texto e como eles se relacionam, bem como de que maneira os diferentes trechos textuais se conectam para formar um documento completo e coerente.

Alguns trabalhos na área de sumarização automática defendem que o uso de estratégias profundas podem produzir melhores sumários, visto que as representações utilizadas permitem uma interpretação mais completa do texto – ou conjunto de textos – processado. Tanto [Leskovec, Milic-Frayling e Grobelnik \(2005\)](#) quanto [Dohare, Gupta e Karnick \(2018\)](#) argumentam que o processo humano de sumarização, especialmente a criação de resumos abstrativos, é realizado reescrevendo, parafraseando e, principalmente, interpretando os textos. O processo de interpretação necessário, segundo os autores, tem como equivalente a criação das representações semânticas profundas dos textos em questão. Similarmente, [Camargo et al. \(2013\)](#) citam que métodos mais linguisticamente informados podem ser mais adequados para essa tarefa, utilizando-se dos diversos dispositivos linguísticos usados pelos humanos durante a composição de um resumo.

[Dohare, Gupta e Karnick \(2018\)](#) também apontam que métodos superficiais têm maior dificuldade para lidar com aspectos sutis da língua, como o tratamento de negações ou correferência. Esses fenômenos podem ser, porém, analisados a partir de uma perspectiva semântica mais profunda. [Leskovec, Milic-Frayling e Grobelnik \(2005\)](#), assim como [Khan, Salim e Farman \(2016\)](#), também mencionam que o uso da semântica profunda pode ajudar a lidar com o fato de que os conceitos estão espalhados pelas sentenças do texto, permitindo a conexão entre as diferentes menções a um determinado conceito.

Outra limitação dos métodos superficiais é apontada por [Greenbacker \(2011\)](#) ao argumentar que eles podem não ser ideais para a sumarização multimodal, já que as informações dos diferentes modos se relacionam em um nível conceitual. A mesma ideia é seguida por [Sherry e Bhatia \(2015\)](#) ao citarem que a sumarização multilíngue requer o uso de representações semânticas unificadas.

Apesar de já existirem técnicas de sumarização abstrativa e geração de texto sem a utilização de representações semânticas profundas (baseados em redes neurais e modelos de língua) ([GEHRMANN; DENG; RUSH, 2018](#); [LIU; LAPATA, 2019](#)), esses métodos ainda possuem limitações. Como indicado por [Bender e Koller \(2020\)](#), modelos de língua atuais são ainda muito semanticamente restritos, limitando-se a uma mera manipulação da forma textual de maneira bastante eficiente. Os autores argumentam que esse tipo de ferramenta tem sim a capacidade de capturar informação relevante para diversas tarefas de PLN subsequentes, porém, não são capazes de capturar significado em si.

[Cao et al. \(2018\)](#) também mostram que métodos de sumarização abstrativa usando redes neurais *sequence-to-sequence* às vezes fabricam fatos que não estavam nos textos originais, prejudicando, assim, a confiabilidade das informações apresentadas. Segundo

os autores, foi observado que 30% dos sumários produzidos apresentaram problemas de confiabilidade.

Observações questionando a confiabilidade de alguns sumários obtidos através de métodos neurais (sem o uso de semântica profunda) também foram feitas por [See, Liu e Manning \(2017\)](#). Os autores observaram a presença de informações falsas ou faltantes nos resumos. Também foi analisada a taxa de palavras copiadas diretamente do resumo, ou seja, o quão extrativo foi o processo de sumarização; foi reportado que 35% das sentenças geradas pelo sistema eram apenas trechos retirados diretamente dos textos originais, enquanto que nos resumos de referência (criados por humanos) essa taxa é de 1.3%. Portanto, se o objetivo é aproximar-se de uma sumarização humana, é necessário lidar com truncamento de sentenças, paráfrases, sinonímia, trechos desnecessários, entre outros fenômenos.

Mais recentemente, [Huang, Wu e Wang \(2020\)](#) mostram que a inclusão de informações por meio de representações semânticas profundas pode ajudar a reduzir esses problemas, gerando resumos mais informativos e confiáveis.

Diversos autores indicam que, para uma sumarização abstrativa ser satisfatória, ela precisa ser feita por métodos que utilizam conhecimento semântico profundo, visto que esse tipo de resumo requer a reorganização das informações presentes nos textos, bem como o tratamento de redundâncias e, possivelmente, a incorporação de conhecimento de mundo, extratextual ([MANI, 2001](#); [GENEST](#); [LAPALME, 2011](#); [LI, 2015](#); [HUANG; WU; WANG, 2020](#)). [Genest e Lapalme \(2011\)](#) desenvolvem mais essa ideia, apontando que a introdução de conhecimento semântico profundo produz textos mais focados, ricos em informação e menos redundantes.

Já na área de mineração de opiniões, [Cambria \(2013\)](#) apresenta que as técnicas comumente utilizadas tratam o texto como uma unidade única. Isso pode resultar em dificuldades para lidar com diferentes fenômenos presentes nesse tipo de texto, como sarcasmo, ironia e opiniões contraditórias – ou complementares – espalhadas ao longo do documento. Além disso, uma única sentença pode conter múltiplas opiniões diferentes sobre diversos aspectos de uma mesma entidade. O autor também argumenta que os métodos tradicionais tendem a ser semanticamente fracos, dependendo de textos em que a opinião do autor está explicitamente presente na superfície textual (por meio de palavras-chave que exprimem sentimentos). Além de palavras-chave, as opiniões também são expressas implicitamente nos textos, sendo necessário um entendimento mais profundo de como as palavras se relacionam e seus significados. Um exemplo pode ser visto na frase “É um livro que prendeu minha atenção do começo ao fim e me fez refletir bastante sobre o assunto proposto por o autor, que a humanidade não é boa por natureza.” (retirado do *córpus OpiSums-PT* de [Condori et al. \(2015\)](#)). Na frase, o sentimento geral com relação ao produto (um livro) é positivo, mesmo expressando sentimentos negativos a outros elemen-

tos (a humanidade). Também é importante notar que o trecho indicativo do sentimento positivo quanto ao livro (“É um livro que prendeu minha atenção do começo ao fim”) não possui adjetivo ou palavra-chave específica que indique este sentimento, que é evocado através da interpretação completa do trecho.

Por conta dessas dificuldades envolvidas na análise dos textos, Cambria (2013) propôs uma nova abordagem para a mineração de opinião chamada *Concept-Level Sentiment Analysis* (CLSA), tendo como objetivo gerar melhores análises dos textos a partir das relações entre os conceitos presentes nos mesmos. Essas relações, obtidas através de ontologias e redes semânticas, permitem a incorporação de informações implícitas associadas aos conceitos em língua natural.

A aplicação de CLSA também permite a incorporação de conhecimento de mundo à análise de sentimentos, extraindo o sentimento associado a partes do texto em um maior nível de abstração, focando-se não apenas nas palavras em si, mas nos seus significados, como eles estão relacionados e qual a sua percepção segundo o senso comum. Por exemplo, é possível identificar a presença de um sentimento negativo para a característica de “espaço pequeno”, enquanto “fila pequena” pode ser associado a um sentimento positivo em seus devidos contextos.

Por se basear na semântica dos conceitos existentes no texto, sistemas baseados em CLSA dependem, em grande parte, das técnicas de análise semântica capazes de identificar os conceitos e suas relações. Na área de Processamento de Língua Natural (PLN), esse tipo de informação é representado formalmente por meio de representações semânticas computacionalmente manipuláveis.

Como indicado por Specia e Rino (2002), diversas formas de representação já foram propostas ao longo da história, como o uso de lógica proposicional clássica, gramáticas de casos, *Universal Networking Language*, etc. Segundo Lappin (2003), essas representações são, em sua maioria, baseadas em uma interpretação de como os componentes sintáticos de uma sentença se relacionam para compor seu significado.

Cada representação possui suas maneiras específicas de indicar o significado dos conceitos de uma língua, podendo ser mais ou menos detalhadas nessas interpretações. Além disso, cada representação pode se basear, ou não, em recursos diferentes, como léxicos, dicionários ou ontologias.

Neste cenário de múltiplas formas de se representar significado computacionalmente, surge, mais recentemente, a linguagem de representação semântica *Abstract Meaning Representation* (AMR) (BANARESCU *et al.*, 2013), que se mostrou bastante promissora para diversas tarefas.

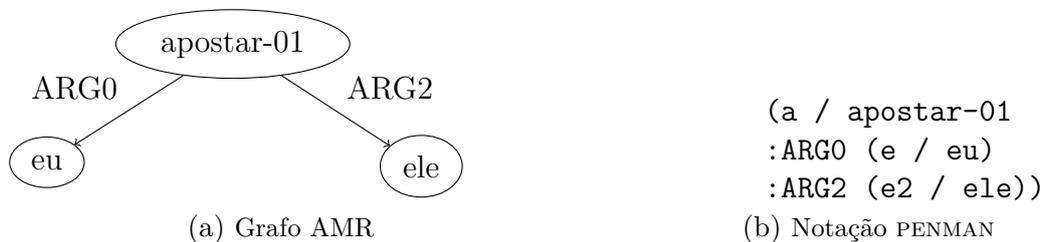
Sentenças em AMR são representadas como grafos em que os vértices exprimem os conceitos principais da sentença, obtidos do recurso PropBank (KINGSBURY; PALMER,

2002), e as arestas correspondem às relações entre tais conceitos.

A Figura 5 apresenta um exemplo de sentença em AMR, tanto na representação gráfica de grafo quanto na notação PENMAN indicada por [Matthiessen e Bateman \(1992\)](#) apud [BANARESCU *et al.*, 2013](#)). A sentença em questão transmite a informação “Eu apostava nele” ou “Eu apostei nele” e até “Eu aposto nele”, visto que o tempo verbal é descartado na representação, pois apenas os conceitos lexicais são importantes no contexto da AMR.

Na figura, pode-se ver três conceitos: “apostar-01”, “eu” e “ele”; também é possível identificar 2 relações entre esses conceitos: “ARG0” e “ARG2”. Essas relações são argumentos do conceito verbal “apostar-01”, referindo-se ao apostador e à aposta respectivamente. O sentido verbal, assim como seus argumentos, foram extraídos do recurso Verbo Brasil ([DURAN; ALUÍSIO, 2015](#)), como indicado por [Anchiêta e Pardo \(2018b\)](#). Esse recurso é o correspondente, para o Português brasileiro, ao PropBank ([KINGSBURY; PALMER, 2002](#)), um repositório de predicados com seus sentidos e argumentos juntamente com indicações de seus papéis semânticos. O Verbo Brasil é, porém, mais restrito quando comparado ao PropBank, visto que apenas contém verbos, enquanto o recurso inglês possui qualquer tipo de predicado.

Figura 5 – Exemplo de sentença representada em AMR



Fonte: Elaborada pelo autor.

O trabalho de [Banarescu *et al.* \(2013\)](#) focou no desenvolvimento da representação para o Inglês, visto que a representação de uma sentença AMR, segundo os autores, não é independente de língua. Sendo assim, [Anchiêta e Pardo \(2018b\)](#) propuseram as primeiras adaptações necessárias para transpor a representação AMR para a língua portuguesa.

Com relação a aplicações, alguns trabalhos já foram desenvolvidos com o objetivo de utilizar a linguagem AMR no contexto da sumarização automática, focados, contudo, em textos literários e jornalísticos ([LIU *et al.*, 2015](#); [LIAO; LEBANOFF; LIU, 2018](#); [DOHARE; GUPTA; KARNICK, 2018](#)). Porém, há muito a se avançar na área, principalmente para o Português do Brasil, visto que a maioria dos trabalhos são focados na língua inglesa. Esses avanços também podem se dar na direção de se investigar o uso da AMR na área de sumarização de opiniões especificamente, que pode se beneficiar do conhecimento representado para suprir as demandas já discutidas da tarefa.

Tendo em vista esses caminhos de investigação em aberto, este projeto propõe a aplicação da abordagem CLSA, por meio da representação AMR, na sumarização de opiniões a fim de produzir melhores sumários abstrativos para o Português do Brasil. Desse modo, tem-se como possível aplicação dos métodos desenvolvidos neste trabalho e em outros semelhantes, auxiliar usuários a tomar suas decisões levando em consideração as opiniões de terceiros com um texto coerente, coeso e de fácil leitura.

1.2 Objetivo e Hipóteses

Como indicado na [Figura 3](#), o processo de sumarização é composto por diferentes fases de processamento, cada uma com suas especificidades e desafios. Este trabalho tem como foco a fase de Transformação, apoiando-se em ferramentas de Análise e Síntese já existentes ou em desenvolvimento por outros pesquisadores. Portanto, o objetivo deste trabalho foi investigar métodos de sumarização de opiniões em Português do Brasil incorporando conhecimento semântico, alinhando-se com o *framework* CLSA.

Os objetivos específicos desse trabalhos são:

- Anotação e adaptação de um cópús de textos opinativos (OpiSums-PT de [Condori et al. \(2015\)](#));
- Implementação dos métodos da literatura que utilizam a representação semântica AMR na sumarização automática;
- Proposta de melhorias aos métodos existentes, bem como o desenvolvimento de novos métodos;
- Realização de experimentos para efeitos de comparação entre os métodos e avaliação das melhorias propostas.

As hipóteses do trabalho dizem respeito às suas várias etapas de desenvolvimento, permeando a anotação do cópús OpiSums-PT ([CONDORI et al., 2015](#)) em AMR e também os métodos de sumarização automática.

Com relação à anotação do cópús, como primeira hipótese, sobre o gênero textual, acredita-se que a AMR ainda possui limitações para representar alguns fenômenos que podem ocorrer em textos opinativos gerados por usuário, visto que esta foi criada a partir de textos literários e científicos.

Similarmente, como hipótese sobre a língua, acredita-se que existam nuances da língua portuguesa, não abordadas em trabalhos anteriores ([ANCHIÊTA; PARDO, 2018b](#); [CABEZUDO; PARDO, 2019](#)), que ainda precisam de uma codificação formal em AMR,

como os diferentes sentidos do diminutivo (ALVES, 2006) e a existência de sujeito indeterminado.

Apesar dessas possíveis limitações, há a possibilidade de elas não afetarem o processo de sumarização, ou seja, a representação pode ser, na sua forma atual, suficientemente robusta para a criação automática de resumos.

Já, a respeito dos métodos de sumarização automática, a hipótese principal é a de que o uso de representações semânticas explícitas produzem melhores resumos de acordo com medidas presentes na literatura. Os resultados podem ser comparados a métodos, tanto extrativos, quanto abstrativos, que não utilizam esse conhecimento semântico profundo para a criação dos resumos.

Também tem-se a hipótese de que existem informações (conceitos e relações específicas) na representação sistematicamente mais importantes do que outras para efeitos de sumarização. Por exemplo, a relação *:mod*, que conecta conceitos a seus modificadores (como adjetivos), poderia ser, sistematicamente, mais importante se comparada à relação *:medium*, que indica o meio de transmissão de uma informação.

Visto que os métodos baseados em conhecimento semântico profundo, existentes até o momento, não abordam textos opinativos, tem-se como hipótese que seus resultados podem ser melhorados a partir da incorporação de informação proveniente da Análise de Sentimentos, como a polaridade – positiva, negativa ou neutra – expressa por um conceito e a identificação de aspectos (características ou partes específicas de um produto sobre as quais o usuário opina).

1.3 Contribuições

A contribuição principal deste trabalho foi o desenvolvimento de métodos de sumarização de opiniões utilizando AMR. Alguns métodos da literatura foram adaptados (tanto com relação à língua quanto ao gênero dos textos) e novas variações foram propostas e avaliadas. Como resultado prático, uma ferramenta, chamada SemOpinionS, foi implementada e disponibilizada¹.

Este trabalho também contribui ao fazer uma análise comparativa entre as diferentes abordagens de sumarização implementadas sobre um mesmo corpus. Dessa forma, pode-se ter uma melhor dimensão de como os diferentes métodos se comportam em um mesmo cenário de textos opinativos.

Também foram realizadas e disponibilizadas as anotações do corpus OpiSums-PT² em AMR. As anotações, porém, não contemplam o corpus inteiro. Além disso, também

¹ Disponível em: <<https://github.com/Superar/SemOpinionS>>. Acesso em: 14 jan. 2021.

² Disponível em: <<https://github.com/nilc-nlp/AMR-BP>>. Acesso em: 14 jan. 2021.

está disponibilizada uma lista de verbos não presentes no Verbo Brasil (DURAN; ALUÍSIO, 2015), para que seja facilitada uma futura extensão do recurso.

Além da anotação, foi realizada uma análise comparativa dos fenômenos capturados pela representação semântica nos textos opinativos em relação a textos jornalísticos. Também fez-se uma comparação das representações dos textos opinativos no que diz respeito a diferentes tamanhos de sentença, bem como uma comparação entre os comentários e os resumos, identificando pontos de relevância que são comumente mantidos ou suprimidos na construção dos sumários.

Todas as contribuições teóricas e práticas serão detalhadas no decorrer desta dissertação.

1.4 Organização do Texto

O [Capítulo 2](#) apresenta os conceitos fundamentais, permeando as áreas de Mineração de Opiniões, Sumarização de Opiniões, Representações Semânticas e Anotação de *Cópus*. Em seguida, os principais trabalhos relacionados são apresentados no [Capítulo 3](#), focando-se em trabalhos de sumarização automática com uso de representações semânticas e sumarização de opiniões para o Português do Brasil. A anotação do *cópus* utilizado e os resultados desse processo são apresentados, em seguida, no [Capítulo 4](#). Mais adiante, o [Capítulo 5](#) apresenta os métodos desenvolvidos e aplicados ao longo do projeto, bem como os resultados obtidos e suas análises. Por fim, algumas outras discussões e perspectivas de continuação de pesquisa são abordadas no [Capítulo 6](#).

CONCEITOS FUNDAMENTAIS

Este trabalho se baseia em diferentes conceitos ao longo de sua execução: análise de sentimentos, sumarização de opiniões, representação semântica e anotação de corpus. Esses conceitos são apresentados mais profundamente de modo a permitir um melhor entendimento do trabalho proposto.

2.1 Mineração de opiniões

A área de Mineração de Opiniões (MO), também chamada de Análise de Sentimentos, tem como objetivo a investigação de opiniões, sentimentos e avaliações de pessoas a respeito de determinado conceito, como produtos, serviços, eventos e tópicos de discussão. Como esses conteúdos estão formulados, em sua maioria, no formato textual, muitas técnicas de Mineração de Textos e PLN são aplicadas na área de MO (LIU, 2012).

Diferentemente de trabalhos clássicos de Mineração de Textos, os trabalhos em MO precisam levar em consideração uma característica essencial aos textos opinativos: a subjetividade. Os textos opinativos exprimem a opinião de determinada pessoa a partir de suas próprias visões subjetivas. Dessa forma, foi necessário criar novos métodos e técnicas que levassem em consideração essa característica (LIU, 2012).

Por outro lado, similarmente à área de mineração de dados, as principais frentes de trabalho referentes à mineração textual são: pré-processamento, algoritmos de descoberta de padrões e técnicas de visualização (FELDMAN; SANGER, 2006).

As etapas de pré-processamento têm como objetivo principal identificar e extrair atributos representativos dos documentos textuais, de forma a transformar os dados não estruturados em representações intermediárias melhor estruturadas e computacionalmente manipuláveis. Esses atributos podem ser diversos, atuando para diferentes propósitos e em diferentes níveis de abstração do texto.

Liu (2012) indica os diferentes níveis de análise mais comumente utilizados na área de MO. O primeiro nível apresentado é o de documentos, assumindo-se o documento como uma unidade única de informação, tendo como foco o sentimento geral expresso no texto e não em pequenas nuances de opinião. A análise dos textos também pode ser feita no nível sentencial, permitindo um maior detalhamento da análise.

O maior nível de detalhamento, porém, é obtido a partir da análise de sentimentos baseada em aspectos, em que atributos (aspectos) da entidade sendo comentada são identificados e as opiniões expressas no texto são associadas a eles. Na sentença “A câmera desse celular é muito boa, mas a bateria deixa a desejar”, por exemplo, as opiniões seriam associadas aos aspectos do produto (“celular”) a que elas se referem, isto é, a expressão “muito boa” se relaciona ao aspecto “câmera”, enquanto “deixa a desejar” está ligada a “bateria”. Dessa maneira, Liu (2012) representa, formalmente, uma opinião como sendo uma quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ em que o autor h_k expressa um sentimento s_{ijkl} sobre o aspecto a_{ij} da entidade e_i em um determinado tempo t_l .

Segundo Feldman e Sanger (2006), as representações textuais geradas a partir do pré-processamento dos textos puros são utilizadas na análise de padrões, em que algoritmos e heurísticas são aplicados para extrair padrões e informações relevantes sobre os textos processados e os conceitos neles contidos.

Existem duas abordagens principais para se realizar a Mineração de Opiniões: baseadas em léxico e baseadas em Aprendizado de Máquina. Os métodos baseados em léxico, como os desenvolvidos por Huettner e Subasic (2000), Avango (2015) e outros, partem da comparação, através de heurísticas, das palavras dos textos com léxicos – listas de termos e o nível de correlação do termo com um conjunto de emoções (felicidade, medo, tristeza, euforia, etc.) – a fim de determinar o sentimento predominante no texto. Métodos baseados em Aprendizado de Máquina, por sua vez, utilizam vetores de atributos (*features*) para representar os textos. Esses atributos são utilizados para treinar automaticamente modelos de Aprendizado de Máquina (supervisionados ou não) capazes de identificar o sentimento expresso no texto. Esses atributos podem ser diversos, como a frequência das palavras no texto (PANG; LEE; VAITHYANATHAN, 2002; DAVE; LAWRENCE; PENNOCK, 2003), etiquetas morfossintáticas (PANG; LEE; VAITHYANATHAN, 2002; KAMAL, 2013), as próprias palavras representadas por meio de vetores (DONG *et al.*, 2020 apud PEREIRA, 2021) e outros.

Por último, é importante destacar a importância das técnicas de visualização no contexto da mineração de textos. Ferramentas e métodos são desenvolvidos para facilitar a navegação e exploração dos usuários em meio aos resultados e padrões obtidos a partir dos dados. Essas técnicas aplicadas se baseiam em diversas representações, principalmente gráficas, para expressar dados e relações complexas.

Como indicado por Pang e Lee (2008), os métodos de visualização se aproximam

dos trabalhos de sumarização automática de opiniões porque realizam uma seleção de sentenças representativas do sentimento geral no texto que, em seguida, são combinadas para gerar um resumo (textual ou gráfico). Mais informações sobre sumarização automática de opiniões são apresentadas na [Seção 2.2](#).

Como limitação dos trabalhos tradicionais em MO, [Cambria \(2013\)](#) defende que os documentos tratados podem não dizer respeito a um único tópico e podem também apresentar opiniões contrárias ou complementares relativas a determinado objeto, o que também foi verificado por [Dave, Lawrence e Pennock \(2003\)](#). Além disso, os métodos tradicionais baseados em aprendizado de máquina se mostram insuficientes para o tratamento de textos pequenos.

Nesse cenário, [Cambria et al. \(2015\)](#) apresentam um novo *framework* para MO, chamado CLSA (*Concept-Level Sentiment Analysis*), que “foca na análise semântica textual através do uso de ontologias ou redes semânticas, o que permite a agregação de informações conceituais e afetivas associadas a opiniões em língua natural” ([CAMBRIA et al., 2015](#), tradução do autor)¹. Essa abordagem permite incorporar mais conhecimento linguístico à MO, como a relação entre os termos das sentenças analisadas e o sentimento geral do contexto em que ocorrem.

Mais recentemente, modelos de língua e modelos de semântica distribucional têm ganhado bastante popularidade na área de MO ([LI et al., 2019](#); [KE et al., 2020](#)). Apesar de incorporarem certo grau de informação semântica ao processo da análise de sentimentos, esse tipo de método não está completamente alinhado ao CLSA, visto que o foco do *framework* é explicitar os conceitos e relações nos textos para além da semântica lexical. Além disso, tem sido discutido se esses modelos de língua realmente têm a capacidade de aprender significado, como já apresentado anteriormente na [Seção 1.1](#).

2.1.1 *Concept-Level Sentiment Analysis*

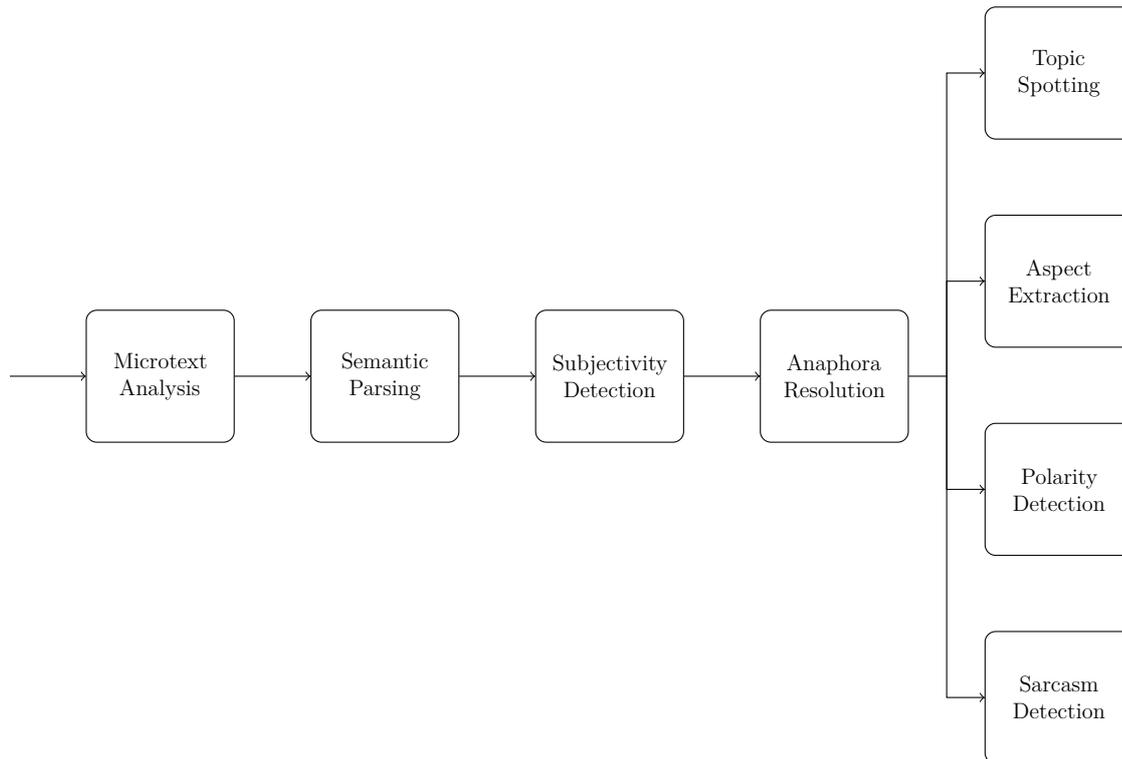
Um modelo de implementação de sistemas de análise de sentimentos em nível conceitual é apresentado por [Cambria et al. \(2015\)](#), constituído por 8 módulos principais, como é mostrado na [Figura 6](#).

O passo de normalização textual (*Microtext Analysis*) tem como objetivo tratar erros e ruídos na escrita, de forma a estruturar melhor o texto de acordo com a norma padrão do idioma. Os ruídos tratados podem ser, por exemplo, erros de ortografia, abreviações, acrônimos, *internetês*, *hashtags* e *emoticons* ([BERTAGLIA; NUNES, 2016](#)).

A etapa de análise semântica (*Semantic Parsing*), foco do modelo de CLSA, baseia-

¹ “*Concept-level sentiment analysis focuses on the semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation of the conceptual and affective information associated with natural language opinions.*”

Figura 6 – Modelo de implementação CLSA



Fonte: Adaptada de Cambria *et al.* (2015).

se no uso de ontologias e redes semânticas de forma a permitir o uso de informações conceituais e subjetivas presentes no texto. Dessa forma, relações implícitas entre os conceitos em língua natural podem ser incorporadas à análise de sentimentos e gerar melhores resultados.

O modelo proposto também prevê a incorporação de técnicas de detecção de subjetividade (*Subjectivity Detection*), como os trabalhos conduzidos por Kamal (2013) e Moraes *et al.* (2016), apresentados anteriormente.

A resolução de anáfora (*Anaphora Resolution*) também é indicada como um passo a ser realizado durante o processamento do texto para encontrar correferências e quais os termos a que se referem. Na frase “Todo mundo fala que esse livro é ótimo, mas na verdade ele é um lixo.”, por exemplo, é necessária a informação de que “ele” se refere a “livro” (o produto sendo avaliado) de forma a interpretar a característica “lixo” corretamente.

Modelos de identificação de tópicos (*Topic Spotting*) podem ser incorporados ao processo de MO, a fim de associar partes do texto a uma ou mais categorias, contextualizando, assim, conceitos presentes na entrada de texto (como “corpo”, “aroma”, “cor”) com relação ao tópico (“vinho”).

A etapa de extração de aspectos (*Aspect Extraction*) foca na relação entre características específicas do produto ou serviço avaliado e a polaridade indicada no texto. Na

frase “Porque apesar de ser um livro com uma história muito boa e interessante, ele é extremamente monótono.”, por exemplo, é importante identificar a relação entre as características “boa” e “interessante” com o aspecto da história do livro sobre o qual se está falando.

A detecção de polaridade (*Polarity Detection*) é, segundo Cambria *et al.* (2015), a tarefa de MO mais popular, tendo como objetivo classificar um trecho de texto como positivo ou negativo. Os trabalhos de Pang, Lee e Vaithyanathan (2002), Dave, Lawrence e Pennock (2003), Kamal (2013) e Avanço (2015), apresentados anteriormente, são exemplos de aplicações focadas nesta tarefa.

A detecção de sarcasmo (*Sarcasm Detection*) também é indicada como um passo importante na análise de sentimentos, pois a sua presença pode alterar completamente o significado de uma frase, ou até do texto inteiro.

Os autores defendem que a aplicação do modelo proposto compreende todas as tarefas de processamento necessárias para a melhor extração das informações opinativas do texto. O sistema será capaz de normalizar as informações, detectar os conceitos presentes no texto, detectar sentenças subjetivas de interesse, identificar referências entre os conceitos no decorrer do texto, identificar as polaridades e inverter as mesmas para sentenças sarcásticas e irônicas. Além disso, o sistema seria capaz de contextualizar as opiniões relativas a tópicos específicos e a determinados aspectos (características) do objeto. É possível notar, porém, que o modelo CLSA não compreende explicitamente o aspecto temporal da emissão das opiniões, como definido pela formalização em quintupla de Liu (2012).

2.2 Sumarização de opiniões

A área de sumarização de opiniões surgiu devido às limitações dos sistemas tradicionais de sumarização automática em cenários com diferentes tipos de texto, dado que esses sistemas são comumente criados para a sumarização de textos jornalísticos. A sumarização de opiniões precisa levar em conta aspectos de subjetividade do autor, além de lidar, na maioria das vezes, com uma linguagem mais informal e com erros (MITHUN; KOSSEIM, 2009). Também é importante lidar, nesse cenário, com a quantificação das opiniões, isto é, quantas pessoas possuem as mesmas opiniões sobre o produto; dessa forma é possível ter uma ideia mais abrangente de como a opinião geral se organiza.

Kim *et al.* (2011) dividem os trabalhos de sumarização de opiniões entre duas categorias: baseadas em aspectos ou não. Os trabalhos focados na sumarização baseada em aspectos são mais comuns e têm como objetivo dividir o texto em aspectos, também chamados pelos autores de tópicos, e apresentar o sentimento geral do autor com relação a cada tópico. O processo de sumarização, independentemente da inclusão de aspectos,

consiste principalmente de duas etapas: predição de sentimentos e geração do resumo.

Os métodos baseados em aspectos possuem, ainda, uma etapa anterior de identificação desses aspectos. Esse passo de identificação tem como objetivo encontrar os tópicos mais importantes do texto para serem resumidos. Isso pode ser feito de diferentes formas, com base em algoritmos de mineração de associação (HU; LIU, 2004; BLAIR-GOLDENSOHN *et al.*, 2008), aprendizado não-supervisionado (LU; ZHAI; SUNDARESAN, 2009; TITOV; MCDONALD, 2008) e supervisionado (BLAIR-GOLDENSOHN *et al.*, 2008), bem como modelos neurais mais recentemente (ANGELIDIS; LAPATA, 2018; PECAR, 2018).

Métodos mais simples, usando listas de palavras-chave especializadas, também podem ser utilizados para a identificação dos aspectos (ZHUANG; JING; ZHU, 2006; AKHTAR *et al.*, 2017). Além disso, métodos baseados em análises linguísticas mais aprofundadas – utilizando léxicos, relações ontológicas e resolução de correferência – também foram propostos (VARGAS; PARDO, 2018a).

A próxima etapa no processo de sumarização diz respeito à classificação dos textos com relação ao sentimento (positivo ou negativo) do autor. No caso de sumários baseados em aspecto, essa etapa é realizada sobre cada aspecto extraído anteriormente. A classificação de sentimentos utiliza técnicas de Análise de Sentimentos, como indicado na Seção 2.1.

A última etapa indicada por Kim *et al.* (2011) é a de geração do resumo. Esse passo combina os resultados das etapas anteriores de modo a gerar um resumo das informações de maneira concisa e de fácil entendimento pelo usuário. Esse passo, assim como nos trabalhos tradicionais de Sumarização Automática, pode gerar sumários tanto extrativos (em que um conjunto de sentenças mais informativas são extraídas diretamente dos textos e apresentadas, como são, ao usuário) quanto abstrativos (em que as informações mais relevantes das opiniões são combinadas em um novo texto, podendo conter palavras não existentes nos textos originais).

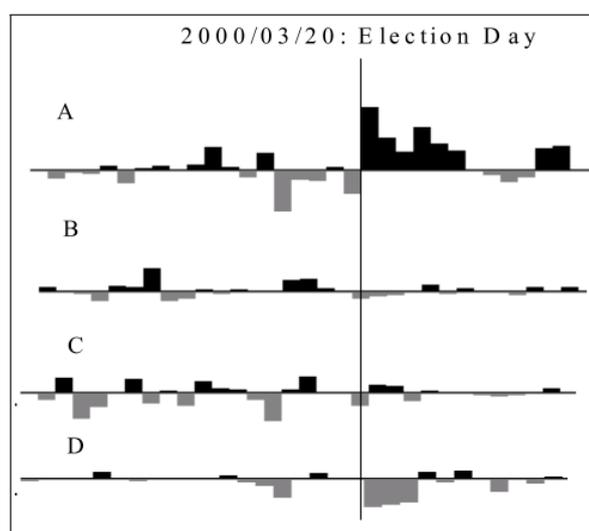
Os trabalhos de sumarização de opiniões tendem a ser extrativos (HU; LIU, 2004; LU; ZHAI; SUNDARESAN, 2009; CONDORI; PARDO, 2017; AKHTAR *et al.*, 2017; LOVINGER; VALOVA; CLOUGH, 2019), porém existem alguns trabalhos que realizam a sumarização abstrativa (GANESAN; ZHAI; HAN, 2010; WANG; LING, 2016; CONDORI; PARDO, 2017).

A apresentação das informações em um resumo de opiniões pode assumir diferentes formatos, cada um requerendo suas especificidades na seleção das informações relevantes. A maioria dos resumos extrativos são apresentados na forma de um resumo estruturado, similar ao exemplo apresentado anteriormente na Figura 4. Esse tipo de resumo foca na apresentação das estatísticas referentes à classificação de sentimento dos textos, bem como

exemplos das sentenças mais relevantes. No caso da sumarização baseada em aspectos, esse processo é repetido para cada aspecto.

Resumos também podem ser gráficos, como os gerados por [Mishne e Rijke \(2006\)](#) e [Ku, Liang e Chen \(2006\)](#), em que diagramas são criados a partir das informações extraídas. Esses gráficos podem dizer respeito à mudança do sentimento geral dos usuários referente a algum tópico, como mostrado na [Figura 7](#). A figura mostra a mudança no sentimento a respeito de 4 candidatos à presidência de Taiwan (A, B, C, D) ao longo do tempo, em que as barras cinzas representam a tendência de um sentimento negativo e as barras pretas a de um sentimento positivo. Quanto maior uma barra, mais forte é o sentimento identificado.

Figura 7 – Exemplo de resumo de opiniões gráfico com respeito ao sentimento



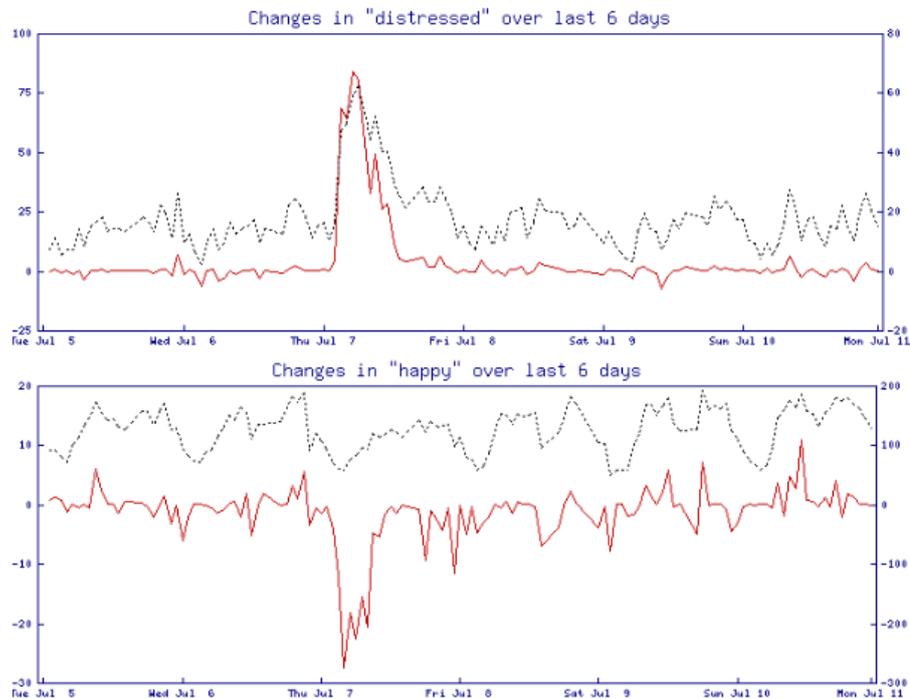
Fonte: [Ku, Liang e Chen \(2006\)](#).

Esse tipo de sumário também pode apresentar o uso (ocorrências nos textos) de determinados termos importantes ao longo do tempo, permitindo a análise de como acontecimentos influenciam o sentimento geral com relação a determinado tópico. Um exemplo pode ser visto na [Figura 8](#), em que o número absoluto de usos do termo é apresentado pela linha pontilhada preta, enquanto que a taxa de mudança no uso da palavra é apresentado na linha contínua vermelha.

Os resumos também podem ser apresentados de forma contrastiva ([LIU; HU; CHENG, 2005](#); [KIM; ZHAI, 2009](#)). Esse tipo de sumarização apresenta as informações de modo a comparar dois produtos similares, ajudando o usuário a realizar sua escolha.

Resumos contrastivos podem ser tanto gráficos quanto textuais. Um exemplo de resumo gráfico, comparando os aspectos de duas câmeras digitais, pode ser visto na [Figura 9](#). O resumo apresenta, usando barras, a taxa de comentários positivos e negativos relativos a cada aspecto de cada câmera. Dessa forma, o usuário pode ver os aspectos de

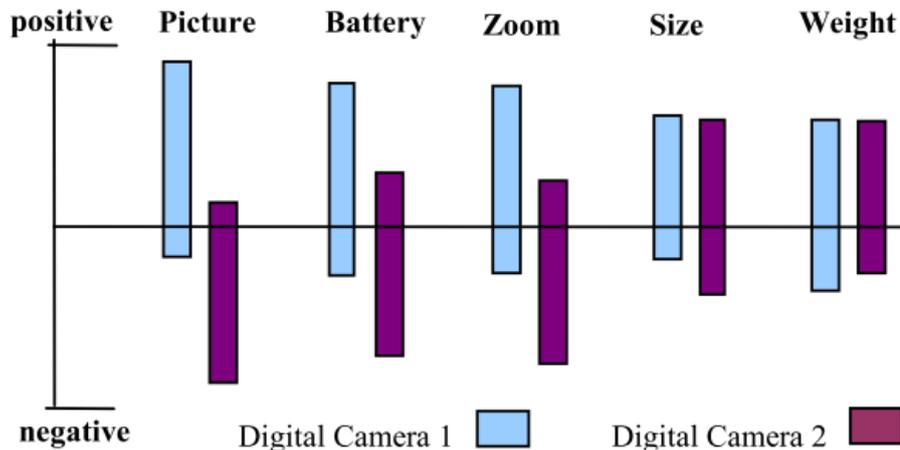
Figura 8 – Exemplo de resumo de opiniões gráfico com respeito à frequência de termos



Fonte: Mishne e Rijke (2006).

cada produto que foram melhor avaliados e, assim, tomar sua decisão.

Figura 9 – Exemplo de resumo de opiniões gráfico contrastivo



Fonte: Liu, Hu e Cheng (2005).

Os resumos contrastivos também podem ser textuais. Assim sendo, as opiniões são apresentadas diretamente ao usuário, lado a lado. Um exemplo de resumo textual contrastivo pode ser visto na [Figura 10](#).

Os resumos, como discutido, podem possuir um formato textual ou não. Os formatos textuais se encaixam perfeitamente na classificação padrão entre extrativos e abstrativos; os outros formatos (tabelas de estatísticas ou gráficos), porém, não aparentam

Figura 10 – Exemplo de resumo de opiniões textual contrastivo

Celular A	Celular B
Smartphone com custo-benefício muito bom. Tirando esses detalhes, o celular é excelente! Moto g 5 , não indico pra ninguém. Destaque para o desempenho em jogos e camera. Um bom aparelho porém baixo desempenho. Bonito e leve. A câmera não é das melhores, nem o design. Boa durabilidade da bateria e câmeras perfeitas. A bateria mal da pra um dia completo. Muito bom o manuseio, configuração ótima. A tv não tem boa recepção. A resolução da câmera é perfeita, o audio é muito bom. Desligou sozinho e parou de funcionar. Sempre inovando e como sempre, Androide puro. Ótimo custo benefício.	O aparelho poderia ser um pouco mais barato. A melhor experiência com smartphone flagship. O aparelho é excelente, o desempenho me surpreendeu muito. A câmera e suas funcionalidades estão me surpreendendo. Por isso acabei devolvendo o aparelho para tentar adquirir outro. Aparelho bonito, processador poderoso, ótima tela. Ótimo Aparelho. Porém Caríssimo, Bateria meia boca. Rápido e com boa duração de bateria. O leitor de digital é muito rápido e eficiente. DEVERIA SE CHAMAR SAMSUNG TRINCA RAPIDO. A cor prata é bem enjoativa. Top de linha, excelente custo benefício. Complicado de encontrar películas compatíveis com a tela toda. Câmera espetacular.

Fonte: [Silva \(2019\)](#).

ser adequados a essas categorias. Portanto, eles podem ser considerados como uma classe distinta do par extrativo-abstrativo.

2.3 Representação semântica

Como apresentado, os métodos de Mineração e Sumarização de Opiniões podem incorporar diferentes níveis de conhecimento linguístico, sendo um deles o da semântica. A Semântica é a área da linguística que estuda o significado das expressões da linguagem humana ([CHIERCHIA, 2003](#), p. 21).

Em PLN, o conhecimento semântico pode ser utilizado tanto na compreensão quanto na geração de línguas naturais. A compreensão é realizada de modo a representar o significado dos textos de acordo com estruturas e formalizações semânticas (também chamadas de formas lógicas ou representações conceituais). Por sua vez, a geração se ocupa em produzir expressões em língua natural a partir de suas representações semânticas ([SPECIA; RINO, 2002](#)).

Ao longo do tempo, diversas maneiras de se representar significado foram propostas, como a Gramática de Casos de [Fillmore \(1968 apud SPECIA; RINO, 2002\)](#), a Teoria da Dependência Conceitual de [Schank \(1975 apud SPECIA; RINO, 2002\)](#), a Teoria da Semântica Conceitual de [Jackendoff \(1990 apud SPECIA; RINO, 2002\)](#), *Universal Networking Language* (UNL) de [Uchida, Zhu e Senta \(1999 apud SPECIA; RINO, 2002\)](#)

e, mais recentemente, *Universal Conceptual Cognitive Annotation* (UCCA) de Abend e Rappoport (2013) e *Abstract Meaning Representation* (AMR) de Banarescu *et al.* (2013).

Para a execução deste trabalho, foi escolhida a representação AMR como instanci- ação do conhecimento semântico, visto que essa é uma representação mais recente e que tem se mostrado promissora em diversas aplicações, incluindo a sumarização automática (LIU *et al.*, 2015; LIAO; LEBANOFF; LIU, 2018; DOHARE; GUPTA; KARNICK, 2018).

2.3.1 Abstract Meaning Representation

A representação semântica AMR foi primeiramente proposta por Banarescu *et al.* (2013) com o objetivo de criar grandes bancos de dados, chamados *semlbanks*, com sentenças e suas respectivas representações semânticas.

Em AMR, sentenças são representadas como grafos direcionados com nós repre- sentando conceitos e arestas representando relações entre eles. A anotação utiliza diversas informações semânticas (como entidades nomeadas, correferência, relações semânticas, en- tidades temporais, etc.) além de recursos consolidados, como o PropBank (KINGSBURY; PALMER, 2002), um banco com palavras predicadoras e suas respectivas estruturas ar- gumentais.

Segundo os autores, AMR é uma representação fortemente influenciada pelo inglês, não podendo ser considerada uma interlíngua. Com isso em mente, Anchiêta e Pardo (2018b), seguido por Cabezudo e Pardo (2019), propuseram adaptações para permitir a anotação de sentenças em português em AMR. As diretrizes criadas têm como base principal o trabalho de Migueles-Abraira (2017) para o espanhol.

Uma das adaptações, por exemplo, foi a seleção do recurso Verbo Brasil² (DURAN; ALUÍSIO, 2015) como correspondente do PropBank para o português. Esse recurso é constituído exclusivamente de verbos e seus respectivos argumentos em uma estrutura chamada *frameset*. Verbos com diferentes significados possuem um *frameset* para cada sentido. Um exemplo pode ser visto na Figura 11.

Figura 11 – Exemplo de *frameset* do Verbo Brasil

Roleset id: aguardar.01 , esperar, vncls: 47.1-1, Mapeamento para o inglês: wait.01
Roles:
Arg0: Entidade que espera
Arg1: Entidade ou coisa esperada (vnrole: 47.1-1-theme)
Argm: período de tempo da espera

A figura apresenta o *frameset* para o sentido de número 01 para o verbo “aguardar”, com o significado de “esperar”. A classificação do verbo (*vncls*) de acordo com a VerbNet (KIPPER *et al.*, 2008) também é apresentada; esse é um recurso que mantém uma clas- sificação semântica de verbos seguindo a teoria de Levin (1993). Além disso, o *frameset*

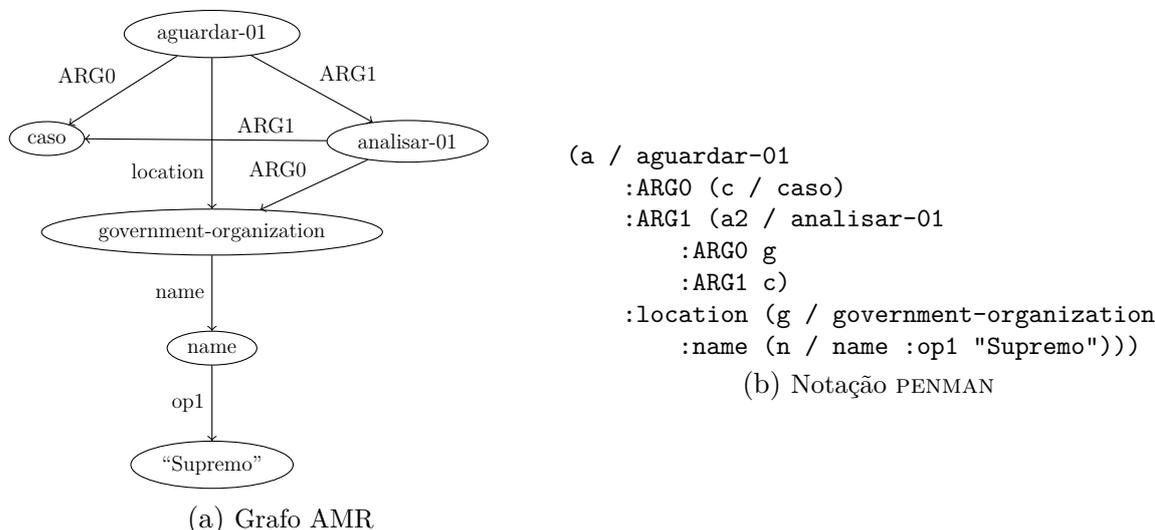
² Disponível em: <<http://143.107.183.175:21380/verbobrasil/>>. Acesso em: 16 out. 2019.

também comporta um mapeamento para o recurso original do PropBank para a língua inglesa. Em seguida, são apresentados os argumentos do verbo, seguindo a taxonomia do PropBank.

Esse esquema de anotação não pretende realizar um detalhamento profundo dos papéis semânticos anotados, como, por exemplo, a diferenciação de “agente”, “experenciador” e “causador”. Alternativamente, os argumentos são anotados com números de 0 a 5 e existe um mapeamento dos significados gerais anotados em cada argumento, por exemplo “ARG0” costuma representar o agente, causador ou experienciador da ação, enquanto o “ARG1” representa o paciente ou tema. Existem também argumentos modificadores (Argm), que representam informações adicionais sobre o verbo, como local, tempo ou duração.

Sentenças em AMR também podem ser representadas em notação PENMAN, uma notação textual parentética para os grafos (MATTHIESSEN; BATEMAN, 1992). Essa notação possui ainda a definição de variáveis referentes a cada conceito, de forma a permitir a referência de um mesmo conceito em diferentes partes do grafo. Um exemplo de grafo AMR com seu PENMAN correspondente pode ser visto na Figura 12. O grafo da figura representa a sentença “O caso aguarda análise no Supremo”.

Figura 12 – Exemplo de sentença representada em AMR



Fonte: Elaborada pelo autor.

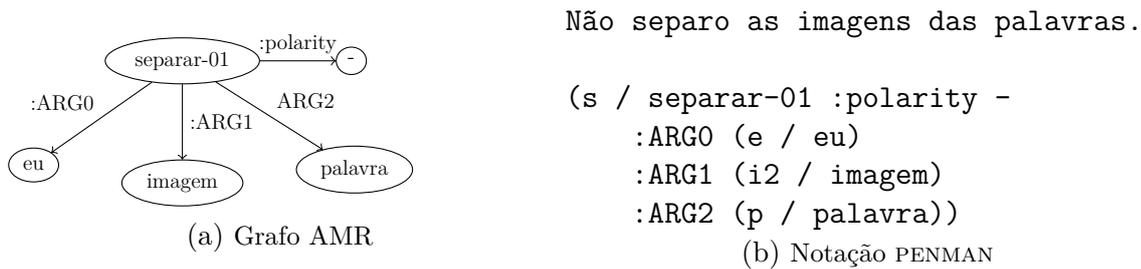
Como pode ser visto, nós (conceitos) podem ser tanto palavras originais do texto (“caso”, “Supremo”), indicações de *framesets* do Verbo Brasil (aguardar-01, analisar-01) ou palavras-chave específicas da própria AMR (*government-organization*, *name*). Já arestas possuem seus rótulos originários dos *framesets* dos verbos a que se referem (ARG0, ARG1) ou são relações específicas da representação (*name*, *op1*, *location*). Também é possível verificar na notação PENMAN que a variável *g* é utilizada como argumento do verbo

“analisar” para deixar explícita a sua relação com a entidade nomeada “Supremo”, o que é feito no grafo por meio de uma aresta.

Entidades nomeadas em AMR são anotadas de acordo com diversas categorias pré-definidas, como *government-organization*, *person*, *book*, *city*, entre outros ³.

AMR também possui mecanismos para representar outras informações semânticas, como perguntas e negações (polaridade). Negações são representadas utilizando a relação *:polarity -*, como pode ser visto na Figura 13.

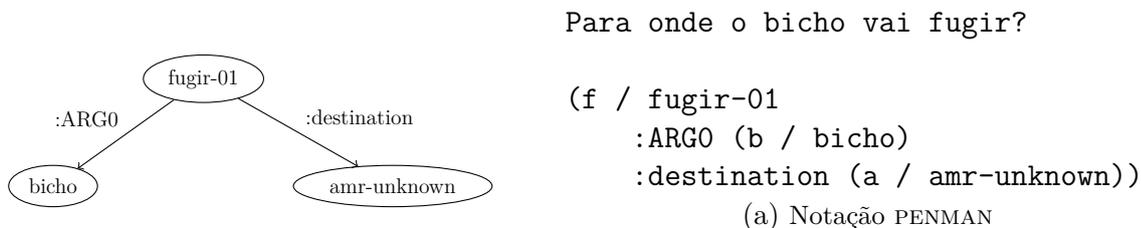
Figura 13 – Exemplo de sentença com negação representada em AMR



Fonte: Elaborada pelo autor.

Por sua vez, perguntas utilizam o conceito *amr-unknown* no lugar dos conceitos a que a pergunta se refere. No caso de perguntas do tipo sim ou não, esse conceito se refere à relação *polarity*. Um exemplo pode ser visto na Figura 14.

Figura 14 – Exemplo de sentença com pergunta representada em AMR

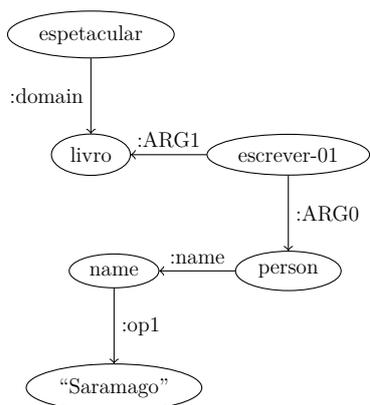


Fonte: Elaborada pelo autor.

Um dos fenômenos mais importantes para a anotação de sentenças em AMR é chamado de “verbalização” (do inglês, *verbalization*). AMR é uma representação focada em capturar principalmente os eventos que ocorrem na sentença. Tais eventos podem estar, na superfície textual, representados por palavras de diversas classes gramaticais diferentes; em AMR, porém, tais eventos devem ser representados a partir dos seus *framesets* correspondentes (normalmente verbos com seus argumentos). Um exemplo pode ser visto na Figura 15, em que “o livro de Saramago” é interpretado e anotado como “o livro escrito por Saramago”, pois a preposição “de” na superfície textual representa, na realidade, o evento de Saramago ter escrito o livro.

³ Uma lista completa dos tipos de entidades nomeadas pode ser acessada em <https://www.isi.edu/~ulf/amr/lib/ne-types.html>. Acesso em: 16 out. 2019.

Figura 15 – Exemplo de sentença com verbalização de preposição anotada em AMR



(a) Grafo AMR

O livro de Saramago é espetacular.

```
(e / espetacular
  :domain (l / livro
    :ARG1-of (e2 / escrever-01
      :ARG0 (p / person
        :name (n / name
          :op1 "Saramago"))))))
```

(b) Notação PENMAN

Fonte: Elaborada pelo autor.

Outro exemplo de verbalização pode ser visto na Figura 16, em que a sentença é interpretada como sendo “Ele cozinha de uma maneira ótima”, já que o fato de o sujeito ser um ótimo cozinheiro evoca o evento de esse mesmo sujeito cozinhar otimamente.

Figura 16 – Exemplo de sentença com verbalização de substantivo anotada em AMR



```
(c / cozinhar-01
  :ARG0 (e / ele)
  :manner (o / ótimo))
```

(a) Notação PENMAN

Fonte: Elaborada pelo autor.

Há diversas outras instruções para a anotação de sentenças em AMR espalhadas, principalmente, por dois meios diferentes: um manual⁴ e um dicionário de anotação⁵. Além desses dois recursos, há também diversos outros menores, como listas de entidades nomeadas, unidades de medidas, reificações (troca de uma relação por um conceito correspondente), verbalizações, entre outros. Essa falta de centralização das instruções de anotação em um único recurso é um dos fatores que contribuem para tornar esse processo bastante complexo.

Outro elemento que pode dificultar a anotação é a especificidade de alguns conceitos criados para contextos muito específicos, como o *binding-affinity-91*, que indica a força da ligação entre moléculas; ou *have-degree-of-resemblance-91*, representando o quão parecidos são dois conceitos ou entidades. Esses, e outros, são conceitos criados especificamente

⁴ O manual de anotação, para o inglês, está disponível em <https://github.com/amrisi/amr-guidelines>. Acesso em: 16 out. 2019.

⁵ Disponível em: <https://amr.isi.edu/doc/amr-dict.html>. Acesso em: 04 fev. 2021.

para a AMR e estão inclusos nas diretrizes de anotação. Isso expande consideravelmente a quantidade de informação que é preciso ter em mente durante o processo, deixando-o mais complexo. Essa mesma discussão pode ser aplicada a outros aspectos, como tipos de quantidades (*pressure-quantity*, *temporal-quantity*, *distance-quantity*, etc.), locais (*:location*, *:relative-position*, *:direction*, etc.), reificações (*be-located-at-91*, *have-purpose-91*, *be-from-91*, etc.).

Existem também algumas críticas acerca da perda de informações quando se realiza uma anotação em AMR. [Donatelli et al. \(2018\)](#) apontam que as diretrizes originais da AMR ignoram algumas informações semânticas expressas por elementos morfossintáticos, como aspecto (se a ação foi concluída ou não) e tempo verbal, por considerá-las menos importantes que outras (como negação e modos verbais), porém também afirmam que essa parece ser uma decisão deliberada, sem justificativa aparente.

[Bonial et al. \(2018\)](#), por sua vez, argumentam que, para se manter uma representação prática na área de PLN, a anotação em AMR teve como foco os fenômenos linguísticos mais frequentes e previsíveis, desconsiderando casos mais raros, essa é, contudo, uma afirmação fraca, visto que fenômenos como aspecto e tempo, citados anteriormente, são bastante frequentes. Os autores também indicam que ainda há muito a se expandir na representação para além da semântica composicional (em que o significado do todo é uma combinação dos significados das partes), pois, segundo eles há inconsistências na anotação de intensidades, quantidades e outras construções complexas.

2.3.1.1 Adaptações para a Língua Portuguesa

Para a Língua Portuguesa, especificamente, [Anchiêta e Pardo \(2018b\)](#) propuseram as primeiras adaptações para se representar sentenças em AMR, como o uso do recurso Verbo Brasil ou a anotação explícita de sujeitos ocultos. Os autores apresentaram o primeiro corpúsculo em português anotado em AMR, com 1.527 sentenças extraídas do livro “O Pequeno Príncipe” a partir de traduções alinhadas do corpúsculo, composto de sentenças do mesmo livro, em inglês.

Em seguida, [Cabezudo e Pardo \(2019\)](#) apresentaram um trabalho de anotação manual de 299 sentenças extraídas de textos jornalísticos, bem como mais algumas adaptações necessárias para se fazer a adaptação da representação para a Língua Portuguesa. As diretrizes criadas foram:

- Para a anotação de sujeitos ocultos, usar os pronomes “ele” ou “isso” no caso de sintagmas nominais elípticos em que o verbo esteja conjugado na terceira pessoa. A escolha entre os dois pronomes se dá de acordo com o verbo: se o argumento em questão deve ser uma entidade animada ou não, respectivamente, segundo a interpretação do anotador;

- Não usar nenhum pronome no caso de sujeitos indeterminados. Essa decisão foi tomada com a finalidade de não inserir nenhuma entidade (conceito) específica como ator da ação, já que o sujeito é indeterminado;
- Usar um equivalente pleno (com uma única palavra) de expressões multipalavras quando possível. Caso não haja um equivalente, criar um conceito concatenando as palavras da expressão separadas por hífen. Essa instrução está de acordo com o tratamento dessas expressões pelas *guidelines* originais da representação em inglês;
- Usar os *framesets* originais do AMR (em inglês) para verbos modais, visto que o recurso do Verbo Brasil não contempla inteiramente esse tipo de verbo.

No exemplo de sentença anotada em AMR apresentado na [Figura 17](#), o sujeito dos dois verbos “parecer” e “dar” é o mesmo e é elíptico. Nesse caso, o anotador interpretou que o sujeito em questão deve se tratar de um objeto inanimado, optando pelo uso do conceito “isso” para sua representação.

Figura 17 – Exemplo de sentença com sujeito elíptico anotado em AMR

Parece legal, mas dá sono.

```
(c / contrast-01
  :ARG1 (p / parecer-01
    :ARG1 (i / isso)
    :ARG2 (l / legal))
  :ARG2 (d / dar-12
    :ARG0 i
    :ARG1 (s / sono)))
```

Fonte: Elaborada pelo autor.

Outro exemplo de anotação seguindo essas diretrizes pode ser visto na [Figura 18](#), em que os conceitos *possible-01* e *recommend-01*, em inglês, foram utilizados para representar os verbos “poderia” e “deveria”.

2.4 Anotação de *córpus*

Todo método de SO e MO deve ter um conjunto de dados de entrada, chamado de *córpus*, para realizar o processamento. Um *córpus*, também chamado de *corpus* ou corpo, pode ser definido como um conjunto de textos (ou objetos linguísticos) acrescidos de informações (anotações ou marcações) a respeito de seus conteúdos. Uma característica muito importante para um *córpus* é sua finalidade, o propósito para qual ele é criado, diferenciando-o de uma simples coleção de textos ([SANTOS, 2008](#)).

Figura 18 – Exemplo de sentença com verbos modais anotados em AMR

Eu poderia e deviria ter lido coisa melhor.

```
(e / e
  :op1 (p / possible-01
        :ARG1 (l / ler-01
                :ARG0 (e2 / eu)
                :ARG1 (c / coisa
                        :ARG1-of (h / have-degree-91
                                  :ARG2 (b / bom)
                                  :ARG3 (m / mais))))))
  :op2 (r / recommend-01
        :ARG1 l))
```

Fonte: Elaborada pelo autor.

Segundo [Sardinha \(2000\)](#), um *córpus* é constituído necessariamente de textos autênticos em língua natural, além de ser representativo do domínio ou do fenômeno a ser estudado. *Córpus* também podem ser classificados de acordo com diversas características tanto dos textos quanto da origem dos dados.

Exemplos de características textuais podem ser: o modo de produção do texto (falado, escrito) e seu conteúdo (*córpus* especializados em um domínio, *córpus* dialetais, *córpus* multilíngues). Por sua vez, características externas ao texto podem dizer respeito à autoria (*córpus* de aprendizes ou *córpus* de língua nativa), ao período de tempo em que os textos foram produzidos (*córpus* sincrônicos, diacrônicos, históricos), entre outros.

Quando o *córpus* em questão é acrescido de outras formas de conhecimento linguístico, ele pode ser chamado de “*córpus* anotado” ([MCENERY, 2003](#)). Esse processo de anotação pode ser realizado manual ou automaticamente, de maneira a representar explicitamente determinado tipo de análise sistemática sobre os textos do *córpus*.

Como indicado por [Hovy e Lavid \(2010\)](#), o processo de criação e anotação de um *córpus* pode ser dividido em sete passos principais:

1. Identificar e selecionar os textos que compõem o *córpus*;
2. Modelar e instanciar uma teoria linguística a fim de criar um manual de anotação;
3. Anotar uma parcela do *córpus* para decidir se o esquema de anotação é apropriado;
4. Avaliar os resultados das anotações;
5. Definir o conceito de um resultado satisfatório. Se necessário, o processo deve ser reiniciado, a fim de garantir melhores resultados;

6. Após garantir bons resultados em uma pequena parcela do *corp*us, deve-se anotar o *corp*us final;
7. Por fim, deve-se aplicar o *corp*us para seu propósito e avaliar seu desempenho final.

Diversas questões devem ser levadas em conta durante o processo de criação de um *corp*us. É comum que se defina a necessidade de um *corp*us ser representativo e balanceado (HOVY; LAVID, 2010; PUSTEJOVSKY; STUBBS, 2013). É preciso, porém, ter em mente o fato de que podem existir fenômenos que ocorram raramente nos textos naturais e, conseqüentemente, acontecerem pouco – ou nunca acontecerem – no *corp*us. Apesar disso, também se deve levar em consideração que balancear artificialmente um *corp*us compromete a representatividade da frequência dos fenômenos em seus cenários de ocorrência natural.

A complexidade da teoria linguística em que a anotação se baseia também deve ser considerada. Teorias e esquemas de anotação muito complexos podem prejudicar a estabilidade da anotação, pois os anotadores podem divergir muito em suas decisões. Por esse motivo, Hovy e Lavid (2010) indicam que às vezes é necessário realizar uma “neutralização” da teoria, de forma a simplificar as instruções de anotação e garantir a qualidade do *corp*us criado.

Ainda existem outros pontos a serem considerados no processo de anotação de um *corp*us: seleção e treinamento dos anotadores, escolha da ferramenta de anotação, solução de discordâncias entre anotadores, disponibilização do *corp*us e outros.

Como exemplo de *corp*us em Língua Portuguesa anotados para Análise de Sentimentos, área deste trabalho, há o TweetSentBR⁶ (BRUM; NUNES, 2018), com 15.000 *tweets* anotados com suas respectivas polaridades (negativo, neutro e positivo). Outro exemplo é o *corp*us de Vargas e Pardo (2018b)⁷, com anotação de aspectos em opiniões sobre produtos. Também há o *corp*us criado por Vieira da Silva, Roman e Carvalho (2020)⁸, contendo *tweets* sobre o domínio do mercado financeiro com a anotação de diversas dimensões de sentimento, como medo e raiva, felicidade e tristeza, entre outros.

Especificamente para a sumarização de opiniões, existem o *corp*us OpiSums-PT⁹ (CONDORI; PARDO, 2017) e o *corp*us 4P¹⁰ (SILVA; PARDO, 2019). O primeiro contém comentários de usuários sobre produtos (livros e produtos eletrônicos) e seus respectivos

⁶ Disponível em: <<https://bitbucket.org/HBrum/tweetsentbr/src/master/>>. Acesso em: 10 set. 2021.

⁷ Disponível em: <<https://github.com/francielleavargas/Explicit-and-implicit-aspects>>. Acesso em: 10 set. 2021.

⁸ Disponível em: <<https://www.kaggle.com/fernandojvdasilva/stock-tweets-ptbr-emotions>>. Acesso em: 10 set. 2021.

⁹ Disponível em: <<https://sites.google.com/icmc.usp.br/opinando/>>. Acesso em: 10 set. 2021.

¹⁰ Disponível em: <<https://github.com/raphsilva/corpus-4p>>. Acesso em: 10 set. 2021.

resumos extrativos e abstrativos. O *cópus* 4P, por sua vez, possui comentários sobre produtos eletrônicos e resumos comparativos entre eles. Esse *cópus* também possui anotação de aspectos e polaridade. Para a realização deste trabalho em específico, foi escolhido o *cópus* OpiSums-PT como fonte dos textos a serem anotados, sumarizados e avaliados, visto que o 4P tem como foco a sumarização contrastiva de opiniões.

2.5 Considerações finais

A área de Sumarização de Opiniões é bastante diversa, abrangendo múltiplas formas de seleção e apresentação do conteúdo das opiniões: resumos gráficos e textuais, extrativos ou abstrativos, contrastivos ou não. As técnicas aplicadas também são várias, utilizando aprendizado tanto supervisionado quanto não supervisionado ou outras técnicas baseadas em extração de associações através de regras.

Essa área pode ser considerada como estando inserida na de Mineração de Opiniões, mais especificamente nas etapas seleção de conteúdo e sua apresentação ao usuário, ajudando-o, assim, a tomar uma decisão com base nas opiniões de outras pessoas sobre um mesmo ponto.

Neste trabalho, com base em um *cópus* de opiniões anotado em AMR, é feita a investigação de métodos de seleção de conteúdo para a criação automática de sumários. Para tanto, se fez necessário o uso de diferentes conceitos e técnicas das áreas de Sumarização Automática e de Mineração de Opiniões, bem como os fundamentos da Semântica e suas formas de representação.

TRABALHOS RELACIONADOS

A proposta deste projeto está relacionada a trabalhos de sumarização automática baseada em representações semânticas (apresentados na [Seção 3.1](#)), bem como a pesquisas em sumarização de opiniões (apresentados na [Seção 3.2](#)), apresentados a seguir.

3.1 Sumarização com representações semânticas

O uso de representações semânticas na área de sumarização automática pode ser resgatado até os anos 1980, em que sistemas como o TOPIC ([REIMER; HAHN, 1988](#)) foram desenvolvidos sob o nome de “condensação textual baseada em conhecimento”. Mais adiante, a sumarização com base em representações semânticas foi revisitada com sistemas como o GISTexter ([HARABAGIU; LĂCĂTUȘU, 2002](#)) e o UNLSumm ([MARTINS; RINO, 2002](#)), com avanços nas pesquisas até os tempos atuais.

A maioria das representações utilizadas se baseiam em grafos como nos trabalhos de [Vanderwende, Banko e Menezes \(2004\)](#), [Rusu et al. \(2009\)](#), [Liu et al. \(2015\)](#), [Dohare, Gupta e Karnick \(2018\)](#) e diversos outros. Outros trabalhos são baseados em representações de triplas ator-ação-receptor chamadas de *Information Item* (InIt) por [Genest e Lapalme \(2011\)](#), *Basic Semantic Unit* (BSU) por [Li \(2015\)](#) ou *Predicate Argument Structure* (PAS) por [Khan, Salim e Kumar \(2015\)](#).

Existem também trabalhos que focam no uso de outras formas de conhecimento semântico, como papéis semânticos para o cálculo de pontuações de saliência de sentenças no texto ([WANG et al., 2008](#); [AKSOY et al., 2009](#)). Nesse contexto, este projeto tem como foco o estudo de representações semânticas explicitamente enunciadas e formalizadas (como grafos e triplas)¹.

¹ Esse forma de representação é conhecida na literatura como “representação semântica explícita” (“*explicit semantic representation*” em inglês), em contraste a outras formas chamadas de implícitas, como vetores de palavras e modelos de língua.

Especificamente usando a representação semântica escolhida no escopo deste trabalho, a AMR, há também algumas pesquisas focadas na sumarização automática de textos: os trabalhos de Liu *et al.* (2015), Dohare, Gupta e Karnick (2018) e Liao, Lebanoff e Liu (2018), brevemente revistos a seguir.

3.1.1 Sumarização automática baseada em AMR

Liu *et al.* (2015) propuseram o primeiro trabalho de sumarização automática baseada no uso da AMR como representação semântica. Seu método consiste de três passos principais: (i) Criação dos grafos AMR das sentenças nos documentos; (ii) Criação de um único grafo AMR multi-sentençae (iii) Seleção do subgrafo sumarizado. Apesar de não ser citado explicitamente pelos autores, acredita-se que esse processo pode ser utilizado tanto para sumarização mono quanto multidocumento, pois é realizado no nível conceitual dos grafos AMR.

A geração dos grafos AMR para cada sentença na coleção de documentos a ser sumarizada é realizada utilizando *parsers* já existentes, como o JAMR (FLANIGAN *et al.*, 2014). Em seguida, é necessário criar uma representação AMR única a ser sumarizada a partir da junção de todos os grafos AMR obtidos para cada sentença. Essa junção é feita mesclando os nós com os mesmos conceitos, mantendo as relações originais. Por fim, um nó raiz novo é adicionado, ao qual todas as raízes dos grafos originais são conectadas, de modo a garantir a conectividade do grafo criado.

Após a criação do grafo referente ao documento, os autores selecionam um subgrafo equivalente ao sumário dos documentos. Essa seleção é formalizada como um problema de Programação Linear Inteira (ILP, do inglês *Integer Linear Programming*) que associa uma pontuação (*score*) a cada subgrafo possível, de vértices V e arestas E , de acordo com a Equação 3.1.

$$\text{score}(V, E; \theta, \psi) = \sum_{v \in V} \theta^T \mathbf{f}(v) + \sum_{e \in E} \psi^T \mathbf{g}(e) \quad (3.1)$$

Cada vértice v é representado por um conjunto de atributos $\mathbf{f}(v)$, assim como as arestas e são representadas por atributos $\mathbf{g}(e)$. Os parâmetros θ e ψ são estimados através de aprendizado supervisionado, sendo x^T a representação da operação de transposição em Álgebra Linear. Esse aprendizado é realizado comparando os subgrafos possíveis com os grafos AMR criados a partir de sumários de referência (criados manualmente) de forma que os parâmetros sejam atualizados para associar valores maiores àqueles subgrafos mais próximos dos sumários manuais.

Por fim, o grafo com maior pontuação é escolhido como o resultado da sumarização automática pelo sistema. Algumas restrições são levadas em consideração durante a seleção para garantir que o grafo seja bem formado (se uma aresta é selecionada, seus

nós devem ser selecionados também) e conectado (deve ser possível acessar todos os nós a partir da raiz através das arestas). Existe também uma condição opcional para limitar, ao longo do processo de otimização, o tamanho do sumário produzido, medido em número de arestas.

Para avaliar o sistema, os autores utilizaram a subcategoria “*proxy report*” do corpus AMR Bank (KNIGHT *et al.*, 2014), constituído de notícias com diferentes metadados, incluindo seus respectivos resumos. O corpus consiste de um total de 298 documentos com uma média de 17,5 sentenças por documento. Os resumos possuem uma média de 1,5 sentença por documento.

A avaliação foi realizada considerando as representações AMR automáticas obtidas através do *parser* JAMR e com grafos AMR de referência criados por humanos. Os autores afirmam que, à época, não existiam geradores textuais de qualidade para a representação AMR, então utilizaram heurísticas que encontram o fragmento de texto mais comumente alinhado aos conceitos do grafo (obtidos através do *parser* JAMR) para gerar um *bag-of-words* dessas palavras no resumo, isto é, as palavras não são geradas em uma ordem específica. Isso, porém, não afeta os resultados da medida de avaliação utilizada: a ROUGE-1 (LIN, 2004), com indicação de precisão, cobertura e medida F. Os resultados são apresentados na Tabela 1.

Tabela 1 – Resultados da sumarização (ROUGE-1) obtidos por Liu *et al.* (2015)

<i>Parser</i>	Precisão (%)	Cobertura (%)	Medida F (%)
Referência	51,90	39,0	44,3
JAMR	51,32	40,0	44,7

Fonte: Adaptada de Liu *et al.* (2015).

Como é possível ver, os resultados demonstram que o uso dos grafos gerados automaticamente pelo *parser* não impactou negativamente a qualidade do sumário produzido em termos da medida aplicada. Apesar disso, os autores indicam que o uso desses grafos, em contraste aos grafos de referência, afeta de forma negativa os resultados do aprendizado para seleção dos nós e arestas mais importantes, como pode ser visto na Tabela 2.

Tabela 2 – Resultados da seleção de nós e arestas (Medida F) obtidos por Liu *et al.* (2015)

<i>Parser</i>	Nós (%)	Arestas (%)
Referência	58,7	39,0
JAMR	51,5	20,0

Fonte: Adaptada de Liu *et al.* (2015).

Mais tarde, Dohare, Gupta e Karnick (2018) apresentaram uma nova proposta de processamento para a sumarização automática de documentos utilizando AMR através

de um sistema de sumarização monodocumento. Os autores deixam claro, porém, que acreditam ser possível o uso do método para um cenário multidocumento. Similarmente a Liu *et al.* (2015), os autores propõem a criação de um único grafo AMR para representar o documento a ser sumarizado, mesclando os nós existentes nas representações das sentenças do documento.

Diferentemente de Liu *et al.* (2015), que realizaram a combinação dos grafos por simples padrão de *strings*, Dohare, Gupta e Karnick (2018) utilizam uma ferramenta de resolução de correferência neural (LEE *et al.*, 2017) para decidir quais nós devem ser combinados em um único. Esse método para a criação dos grafos no nível de documento permite lidar, por exemplo, com pronomes, o que o método de Liu *et al.* (2015) é incapaz de capturar.

Existem também casos em que duas palavras não se referem a uma mesma entidade (não podendo ser mescladas em um único nó), porém indicam o mesmo conceito abstrato, o que não é identificado pelo sistemas de resolução de correferência. Para garantir que esses nós atribuam uma importância maior um ao outro, sem juntá-los, os autores criaram o conceito de “*meta node*”.

Meta nodes são nós utilizados para aumentar a importância de substantivos comuns, não reconhecidos pelos sistemas de resolução de correferência. Cada *meta node* é conectado a todas as ocorrências do substantivo a que ele se refere. Todos os processos de ranqueamento dos nós serão feitos com relação aos *meta nodes*, porém eles não são utilizados para a geração final do sumário.

Dois exemplos de textos em que são utilizados *meta nodes* podem ser vistos no Quadro 1. O primeiro exemplo apresenta dois conceitos (*illegal* e *ban*) relacionados que devem reforçar um ao outro no processo de sumarização por meio de um *meta node*. Já o segundo exemplo mostra o caso de um substantivo comum (*opium*) que não é reconhecido pelos sistemas de resolução de correferência e que não necessariamente deve se relacionar ao mesmo referente, mas por representar o mesmo conceito abstrato, esses nós não devem se tornar um único, porém devem reforçar a importância do conceito através de um *meta node*.

É importante salientar que a formalização dos *meta nodes* foi apenas exposta por Dohare, Gupta e Karnick (2018), não sendo feita nenhuma implementação ou avaliação prática dessa técnica. Inclusive técnicas para se identificar os conceitos a serem relacionados através desses *meta nodes* são deixadas em aberto, sendo uma discussão puramente teórica.

Após a criação do grafo referente ao documento a ser sumarizado, é feita a seleção do subgrafo relativo a seu resumo. Primeiramente, os nós mais importantes são escolhidos. A importância de um vértice é dada pela medida *Term frequency-inverse document*

Quadro 1 – Exemplo de textos em que é feito o uso de *meta nodes*

<p>1. On 011006 The Citizen newspaper stated that it is illegal for South Africans to be involved in mercenary activity or to render foreign military assistance inside or outside of South Africa. The Citizen newspaper stated that the South African Foreign Ministry announced on 011005 that the South African government imposed the mercenary activity ban following reports that 1000 Muslims with military training have enlisted to leave South Africa for Afghanistan to fight for the Taliban against the United States.</p>
<p>2. Head of the U.N. drug office Antonio Maria Costa said that Afghanistan has produced so much opium in recent years that the Taliban are cutting back poppy cultivation and stockpiling raw opium in an effort to support prices and preserve a major source of financing for the insurgency. Costa said this to reporters last week as the U.N. Drug Office Office prepared to release its latest survey of Afghanistan’s opium crop.</p>

Fonte: Dohare, Gupta e Karnick (2018).

frequency (TF-IDF) de seu documento (JURAFSKY; MARTIN, 2019, p. 12). Os autores utilizaram o corpus CNN-Dailymail (HERMANN *et al.*, 2015), com 300 mil artigos de notícias, para os cálculos de frequência. Como citado anteriormente, os *meta nodes* foram levados em consideração neste processo ao invés dos conceitos a que se referem. Os n nós com maior relevância são selecionados para compor o sumário, sendo n um parâmetro definido de acordo com o tamanho do sumário desejado.

Após a seleção dos nós mais importantes, o método elege também as relações mais importantes entre cada par desses conceitos. Isso é feito escolhendo o caminho entre os dois nós mais próximos da raiz do grafo da sentença em que eles aparecem pela primeira vez em conjunto. Os autores argumentam que a relação mais importante entre dois conceitos comumente estará na primeira sentença em que eles ocorrem juntos.

Um último passo é a seleção de informações extras a partir de uma ferramenta de extração de informação chamada OpenIE (BANKO *et al.*, 2008), que extrai relações entre os conceitos de uma sentença a partir de sua árvore sintática de dependência. As relações escolhidas, assim como os conceitos a elas relacionados, são aquelas contidas no caminho selecionado no passo anterior.

Como citado anteriormente, os resultados relatados por Dohare, Gupta e Karnick (2018) foram obtidos a partir do uso do grupo “*proxy report*” do corpus AMR Bank, com textos gerados pelo gerador Neural AMR (KONSTAS *et al.*, 2017). Os resultados, de acordo com a medida ROUGE-1, são apresentados na Tabela 3. Os resultados na tabela comparam a influência do método de resolução de correferência automática utilizado (LEE *et al.*, 2017) com o mesmo processo feito manualmente. Como pode ser visto, não há grandes diferenças de valores ao se comparar os dois métodos, com um aumento percentual de apenas 3,5% quando utilizada a correferência manual. Os autores indicam que essa diferença pode se dar devido à dificuldade do sistema automático de lidar com pronomes catafóricos.

Tabela 3 – Resultados da sumarização (ROUGE-1) obtidos por Dohare, Gupta e Karnick (2018)

Método	ROUGE-1 (%)
Correferência automática	39,5
Correferência manual	40,9

Fonte: Adaptada de Dohare, Gupta e Karnick (2018).

Liao, Lebanoff e Liu (2018), por sua vez, continuaram os trabalhos desenvolvidos por Liu *et al.* (2015). O processo de sumarização se dá da mesma forma, através da seleção de um subgrafo sumarizado usando técnicas de ILP, mas com algumas melhorias.

Enquanto o método anterior realizava a junção dos nós a partir da comparação de suas *strings*, este trabalho utiliza, assim como Dohare, Gupta e Karnick (2018), métodos de resolução de correferência para a criação de um único grafo AMR referente aos documentos sendo sumarizados. Em seguida, um subgrafo é selecionado de maneira similar a Liu *et al.* (2015), maximizando uma pontuação, formalizada pela Equação 3.1, com respeito aos nós e arestas selecionados.

Diferentemente dos trabalhos anteriores, Liao, Lebanoff e Liu (2018) citam um foco exclusivo na sumarização multidocumento. Dessa forma, as sentenças dos documentos são agrupadas de acordo com seus conteúdos antes de passarem ao processo de sumarização AMR descrito anteriormente. O agrupamento é realizado de acordo com algumas medidas de similaridade comparadas no trabalho: (i) *Longest Common Subsequence* (LCS); (ii) *Vector Space Model* (VSM), usando similaridade de cosseno entre vetores representantes das sentenças; (iii) Medida *Smatch* (CAI; KNIGHT, 2013), calculada sobre os grafos AMR das sentenças; (iv) *Concept Coverage*, priorizando sentenças cujas representações AMR cubram os mesmos conceitos. Ao final do agrupamento, algumas sentenças representantes de cada grupo são selecionadas para continuar o processo de sumarização.

Os experimentos foram realizados pelos autores sobre os corpúscos DUC-04² e TAC-2011³. Os resultados – em termos de precisão, cobertura e medida F dos grafos de sumário quando comparados aos grafos dos sumários de referência – obtidos da comparação entre as diferentes métricas de similaridade podem ser vistos na Tabela 4. Como indicado na tabela, a seleção de nós se dá melhor quando comparada à previsão de arestas, tendo melhores resultados utilizando o método de *Concept Coverage* durante o agrupamento. Também é possível notar que o uso da medida de cosseno sobre o VSM gera melhores resultados referente à seleção de arestas. Os autores também indicam a necessidade de trabalhos futuros para a inclusão de outras arestas, dados os resultados mais baixos em suas seleções.

² Disponível em: <<https://duc.nist.gov/data.html>>. Acesso em: 9 dez. 2019

³ Disponível em: <<https://tac.nist.gov/data/index.html>>. Acesso em: 9 dez. 2019

Tabela 4 – Influência da medida de similaridade nos grafos sumarizados

Método	Nós (%)			Arestas (%)		
	Precisão	Cobertura	Medida F	Precisão	Cobertura	Medida F
LCS	16,7	26,7	19,9	6,7	8,0	6,9
VSM	24,0	38,6	28,8	9,6	11,3	9,8
<i>Smatch</i>	20,9	33,2	24,9	9,3	10,7	9,4
<i>Concept Cov.</i>	25,0	40,3	30,1	7,3	10,0	8,0

Fonte: Adaptada de [Liao, Lebanoff e Liu \(2018\)](#).

Os autores também apresentam os resultados da geração dos resumos em termos da medida ROUGE-1, como pode ser visto na [Tabela 5](#). O modelo escolhido pelos autores para comparação dos resultados utiliza a medida VSM durante o processo de agrupamento. Quando comparado com outros métodos de sumarização automática multidocumento, indicado em mais detalhes no trabalho original ([LIAO; LEBANOFF; LIU, 2018](#)), esse método se mostrou melhor com relação à medida ROUGE-1.

Tabela 5 – Resultados da sumarização (ROUGE-01) obtidos por [Liao, Lebanoff e Liu \(2018\)](#)

Córpus	Precisão (%)	Cobertura (%)	Medida F (%)
DUC-04	29,9	39,0	37,8
TAC-2011	40,1	42,3	41,1

Fonte: Adaptada de [Liao, Lebanoff e Liu \(2018\)](#).

3.1.2 Outros trabalhos

Os trabalhos em sumarização automática utilizando representações semânticas são diversos, variando tanto em relação à representação utilizada, quanto nos métodos aplicados.

Como apresentado anteriormente, as representações semânticas são, em sua maioria, baseadas em grafos, possuindo como característica comum a representação de conceitos abstratos como nós e relações entre eles como arestas. Alguns exemplos são as representações *Universal Networking Language* ([UCHIDA; ZHU; SENTA, 1999](#)), *Language-Neutral Syntax* ([CAMPBELL; SUZUKI, 2002](#)), *Rich Semantic Graph* ([MOAWAD; AREF, 2012](#)), *Event Semantic Link Network* ([LI; HE; ZHUGE, 2016](#)) e outros. Existem também outras representações que não se baseiam em grafos, mas em triplas agente-verbo-paciente, como as representações de *Information Items* ([GENEST; LAPALME, 2011](#)), *Basic Semantic Units* ([LI, 2015](#)) e *Predicate Argument Structure* ([KHAN; SALIM; KUMAR, 2015](#)).

Com relação aos métodos utilizados, eles podem ser classificados em dois grupos: métodos baseados no algoritmo de PageRank ([BRIN; PAGE, 1998](#)) e métodos baseados em

cálculo de pontuação. Os algoritmos baseados em PageRank têm como objetivo selecionar os nós mais importantes em uma representação de grafo, de forma a associar uma maior importância aos nós com mais conexões. Exemplos de trabalhos que utilizam esse tipo de método são os de [Vanderwende, Banko e Menezes \(2004\)](#), [Greenbacker \(2011\)](#) e [Khan *et al.* \(2018\)](#).

Outro método utilizado para a seleção de conteúdo dos resumos é o cálculo de pontuação para estruturas (triplas, nós, arestas, subgrafos). A pontuação pode ser calculada levando em consideração diversas características, como a frequência dos termos nos textos ([SORNLERTLAMVANICH; POTIPITI; CHAROENPORN, 2001](#); [GENEST; LAPALME, 2011](#)), atributos específicos ([LI, 2015](#); [LI; HE; ZHUGE, 2016](#)) e estruturas da WordNet ([KHAN; SALIM; KUMAR, 2015](#)).

Algumas outras técnicas também já foram exploradas mais esparsamente, como o uso de Aprendizado de Máquina Supervisionado para seleção de conteúdo relevante ([LESKOVEC; GROBELNIK; Milic-Frayling, 2004](#)) ou heurísticas para remoção de informações redundantes ([MARTINS; RINO, 2001](#); [MOAWAD; AREF, 2012](#)).

Dentro do campo das representações semânticas implícitas – em que o significado de uma palavra, sentença ou texto está codificado em um vetor numérico – há também diversos trabalhos, focados, principalmente, no uso de Aprendizado Profundo e redes neurais ([SEE; LIU; MANNING, 2017](#); [CAO *et al.*, 2018](#); [LIU; LAPATA, 2019](#)).

3.2 Sumarização de opiniões

A importância de se distinguir os sistemas de sumarização para o contexto de textos opinativos começou a ser discutida nos anos 2000 ([BEINEKE *et al.*, 2004](#); [HU; LIU, 2004](#)), com o avanço dos sistemas de *e-commerce* e trabalhos em análise de sentimentos. Foi levantada a importância de se levar em conta a subjetividade no contexto da sumarização, aspecto até o momento não abordado, visto que os sistemas tradicionais eram criados para textos jornalísticos e impessoais.

Desde então, diversas técnicas de sumarização foram exploradas, como mineração de associação ([HU; LIU, 2004](#)), aprendizado não-supervisionado ([LU; ZHAI; SUNDARESAN, 2009](#)) e supervisionado ([BLAIR-GOLDENSOHN *et al.*, 2008](#)). Apesar disso, pouco tem sido feito para a língua portuguesa, visto que esses métodos foram desenvolvidos sobre corpus em inglês. O único método de sumarização de opiniões para a língua portuguesa até o momento foi desenvolvido por [Condori e Pardo \(2017\)](#).

Também há poucos indícios de trabalhos que tenham utilizado conhecimento semântico expresso explicitamente por meio de alguma representação formal para a sumarização de opiniões especificamente. Os trabalhos que utilizam conhecimento semântico em

algum nível atêm-se ao uso de classificações de polaridade, reconhecimento de entidades nomeadas, entre outros.

Um dos primeiros trabalhos em sumarização abstrativa de opiniões foi realizado por Carenini, Ng e Pauls (2006). Os autores argumentam que, para se obter um resumo de qualidade, é necessário a extração de, pelo menos, informações sobre os aspectos dos produtos avaliados, suas respectivas polaridades com indicação de intensidade.

O sistema MEAD*, desenvolvido a partir da adaptação do sistema MEAD de Radev *et al.* (2003), incorpora as informações de polaridade calculando uma pontuação a cada sentença segundo a Equação 3.2. A pontuação de uma sentença s é dada pelos valores de polaridade e referentes a todas as avaliações sobre todos os aspectos presentes na sentença ($eval$). Os valores de sentimento variam de -3 a 3 , sendo -3 extremamente negativo e 3 extremamente positivo.

$$\text{score}(s) = \sum_{e \in \text{eval}(s)} |e| \quad (3.2)$$

A seleção das sentenças mais importantes foi realizada segundo os aspectos. As sentenças foram agrupadas de acordo com os aspectos sobre os quais elas versam, podendo uma sentença estar em mais de um grupo. Então, a sentença com maior pontuação é selecionada de cada grupo, iniciando pelo grupo com mais sentenças, sem selecionar a mesma sentença duas vezes.

Para a geração do resumo abstrativo, os aspectos (juntamente com seus modificadores) mais importantes são primeiramente escolhidos de forma a priorizar aspectos que sejam muito citados ou que tenham opiniões fortes. Em seguida, o conteúdo do resumo é organizado por meio da indicação de relações discursivas entre as diferentes informações. Por fim, a geração textual foi realizada com o uso da ferramenta GEA (*Generator of Evaluative Arguments*).

Para avaliação, os autores contaram com 28 participantes. Cada participante recebeu 20 comentários extraídos aleatoriamente de um *córpus* de produtos eletrônicos (câmera e DVD *player*), bem como o resumo gerado, devendo avaliar o resumo seguindo algumas qualidades linguísticas: gramaticalidade, ausência de redundância, clareza, foco, estrutura e coerência. Alguns aspectos referentes à seleção do conteúdo também foram avaliados pelos participantes:

- Cobertura: O sumário contém as informações que o participante teria incluído
- Precisão: O sumário não contém informações que o participante não teria incluído
- Acurácia: A informação expressa no sumário corresponde à informação expressa nos textos originais

Todas as avaliações foram feitas seguindo uma escala de 1 a 5, sendo 1 a menor qualidade e 5 a maior. Os resultados referentes aos resumos abstrativos são apresentados na [Tabela 6](#). A tabela apresenta a comparação do método abstrativo citado (SEA), o método extrativo (MEAD*), o sistema original de [Radev et al. \(2003\)](#) (MEAD) e o sumário criado por um humano (Referência).

Tabela 6 – Resultados obtidos por [Carenini, Ng e Pauls \(2006\)](#)

	SEA	MEAD*	MEAD	Referência
Gramaticalidade	3,43	2,71	3,14	4,29
Ausência de redundância	3,14	3,86	3,57	4,43
Clareza	3,86	4,00	3,00	4,71
Foco	4,14	3,71	2,29	4,14
Estrutura e coerência	2,29	3,00	1,86	4,43
Cobertura	2,33	2,57	1,57	3,57
Precisão	4,17	3,50	2,17	3,86
Acurácia	4,00	3,57	2,57	4,29

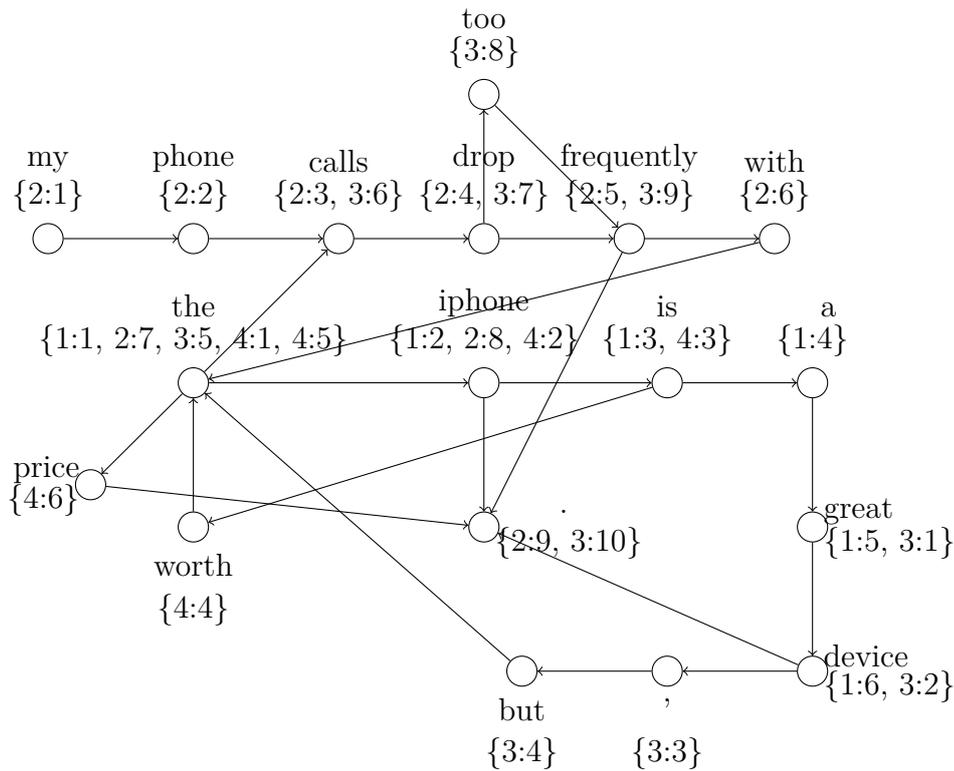
Dos resultados, é possível ver que os sumários abstrativos possuem maior pontuação referente à gramaticalidade (3,43) quando comparado ao método extrativo (2,71). Os autores também indicam que os resultados do método SEA para a avaliação de estrutura e coerência apontam falhas no processo de organização do conteúdo, sendo necessárias melhorias.

Com relação à seleção do conteúdo, os resultados apontados pelo conceito de cobertura indicam que, em comparação ao sumário de referência, ainda há bastante informação considerada importante sendo ignorada pelos métodos automáticos. Os piores resultados foram obtidos pelo sistema MEAD, indicando que o gênero de textos opinativos possui peculiaridades, não previstas pelo sistema, que precisam ser tratadas.

Posteriormente, o trabalho de [Ganesan, Zhai e Han \(2010\)](#) propõe um sistema chamado Opinosis para a sumarização de textos opinativos a partir de uma representação em grafo. A representação possui nós, representando palavras únicas, e arestas direcionadas indicando a estrutura das sentenças, ou seja, a ordem das palavras. Como uma mesma palavra pode estar em mais de uma sentença, identificadores das sentenças que contêm a palavra são também armazenadas no nó, assim como a sua posição na respectiva frase. Um exemplo pode ser visto na [Figura 19](#).

Na figura, as sentenças “*The iPhone is a great device*” (com identificador 1), “*My phone calls drop frequently with the ihpone*” (2), “*Great device, but the phone calls drop too frequently*” (3) e “*The iPhone is worth the price*” são representadas. O identificador da sentença e a posição do *token* são indicados segundo a notação: {Id. da sentença: Posição do *token*}. O token “*drop*”, por exemplo, está presente em duas sentenças (de identificadores 2 e 3) nas posições 4 e 7, respectivamente.

Figura 19 – Exemplo de grafo do sistema Opinosis



Fonte: Adaptada de [Ganesan, Zhai e Han \(2010\)](#).

Como indicado pelos autores, esse tipo de grafo possui algumas propriedades importantes:

1. Captura de redundância: informações redundantes nas sentenças são naturalmente capturadas pela lista de ocorrências do *token* referente ao nó. A sequência “*the phone calls drop (too) frequently*”, por exemplo, é redundante entre duas sentenças representadas;
2. Captura de sequências com lacunas: a representação permite identificar elementos presentes em sequências de *tokens* que podem ser ignorados, capturando, assim, diferentes formas de se afirmar a mesma informação. Um exemplo na sequência “*drop too frequently*” é a palavra “*too*”, que pode ser removida, já que há um caminho direto possível entre “*drop*” e “*frequently*”;
3. Estruturas colapsáveis: é possível encontrar estruturas frequentes e fixas, que ocorrem com frequência entre as sentenças. Essas estruturas podem ser colapsadas em poucos nós. A sequência de *tokens* “*the iPhone is*” está presente em várias sentenças, sendo o nó “*is*” um ponto de conexão dessa estrutura com as outras partes do grafo. Segundo os autores, essa expressão inteira poderia ser resumida a um único nó que concentra as conexões de entrada e de saída dos nós que a compõe.

O processo de sumarização do sistema Opinois tem como objetivo gerar apenas sentenças tidas como plausíveis. Para tanto, é necessário a seleção de um nó inicial plausível, um caminho válido e um nó final plausível.

Um nó inicial é considerado plausível se a média das posições em que ele aparece for menor que um determinado parâmetro pré-definido, ou seja, se o *token* costuma aparecer no início das sentenças que o contêm. Um nó final, por sua vez, é plausível se for uma pontuação ou uma conjunção coordenativa.

Um caminho válido deve ser iniciado por um nó inicial plausível, terminado por um nó final plausível e a sequência dos *tokens* deve satisfazer algumas condições referências às suas etiquetas morfossintáticas. Um exemplo de restrição, para a língua inglesa, é a de que adjetivos devem ocorrer antes de substantivos e depois de advérbios.

A partir da definição de quais caminhos são válidos, a seleção do melhor caminho é feita a partir do cálculo de um escore referente à quantidade de informação contida sobre as opiniões. São apresentadas diferentes maneiras de se calcular essa pontuação, porém todas levam em consideração o nível de redundância no caminho (quantas sentenças expressam aquelas mesmas informações).

Durante a geração do resumo, também existem mecanismos para combinar sentenças que possuem um mesmo prefixo, nome dado pelos autores à sequência de palavras iniciais da sentença, contanto que esse prefixo termine, necessariamente, em um verbo.

Para avaliação, os autores utilizam um conjunto de opiniões sobre hotéis, carros e produtos, com aproximadamente 100 sentenças por entidade avaliada. Os resultados foram primeiramente avaliados utilizando a medida ROUGE, como pode ser visto na Tabela 7. O Opinois foi comparado com o sistema MEAD (RADEV *et al.*, 2003).

Tabela 7 – Resultados da sumarização (ROUGE-1) obtidos por Ganesan, Zhai e Han (2010)

Método	Precisao (%)	Cobertura (%)	Medida F (%)
Opinois	44,82	28,31	32,71
MEAD	9,16	49,32	15,15

Fonte: Adaptada de Ganesan, Zhai e Han (2010).

Como é possível ver, o sistema Opinois obteve melhores resultados com respeito à precisão, quando comparado ao MEAD, porém este possui um alto valor de cobertura. Os autores argumentam que isso ocorre devido ao fato de o MEAD ser extrativo, selecionando as palavras exatas do texto original, o que aumenta esse valor referente à medida ROUGE, porém as informações selecionadas não são necessariamente significantes, como pode ser visto pelo valor de precisão.

É importante salientar que os autores reconhecem como limitação de seu trabalho a escolha da representação utilizada, pois ela se baseia primariamente na superfície textual

e ordem das sentenças, não sendo realizada uma análise semântica profunda dos dados. Dessa forma, os autores indicam que a inclusão de conhecimento semântico explícito pode melhorar os resultados obtidos.

3.2.1 Condori e Pardo (2017)

Condori e Pardo (2017) realizam a comparação de diversos métodos de sumarização automática de opiniões, tanto extrativos quanto abstrativos. Por se focarem na geração dos resumos, os autores utilizaram um corpus já com os aspectos e sentimentos manualmente anotados, evitando assim erros que possam ser inseridos por ferramentas automáticas.

Além dos trabalhos comparativos, Condori e Pardo (2017) também propuseram dois novos métodos de sumarização de opiniões: *Opizer-E* (extrativo) e *Opizer-A* (abstrativo). Todo o trabalho se focou na criação de resumos baseados em aspectos.

O método extrativo se baseia em duas fases principais: o agrupamento de sentenças, com relação ao aspecto e polaridade, e o ranqueamento das mesmas, conforme a Equação 3.3.

$$\text{Imp}(s_i) = \alpha \times W_{\text{pos}}(s_i) + (1 - \alpha) \times W_{\text{pro}}(s_i) \quad (3.3)$$

A importância (*Imp*) de determinada sentença s_i é calculada a partir de uma pontuação dada à sua posição no documento a que pertence (W_{pos}) e à proximidade (W_{pro}) entre os aspectos e seus qualificadores (segmento de texto que indica a polaridade com relação ao aspecto). Os valores são combinados seguindo um parâmetro α .

A pontuação referente à posição da sentença s_i no documento r é dada, por sua vez, pela Equação 3.4. Nela, $\text{len}(r)$ indica o número de sentenças no documento r , enquanto $\text{pos}(s_i)$ indica a posição da sentença s_i no documento (quanto mais próxima do início do documento, maior sua importância).

$$W_{\text{pos}}(s_i) = \frac{\text{len}(r) - \text{pos}(s_i)}{\text{len}(r)} \quad (3.4)$$

A medida de proximidade dos aspectos e suas polaridades é dada pela Equação 3.5, sendo q_j cada qualificador referente ao aspecto a .

$$W_{\text{pro}}(s_i) = 1 - \min_{\forall j \in a} \left(\frac{|\text{pos}(a) - \text{pos}(q_j)|}{\text{len}(s_i)} \right) \quad (3.5)$$

Após o agrupamento e ranqueamento das sentenças, as sentenças mais importantes são selecionadas para inclusão no resumo extrativo.

O método *Opizer-A* (abstrativo), por sua vez, realiza, primeiramente, um agrupamento com o intuito de identificar as informações mais representativas de cada aspecto.

Essa etapa de agrupamento é realizado sobre conjuntos de n-gramas (sequências de palavras), não levando em conta sentenças inteiras.

Em seguida, outro agrupamento é realizado com respeito à polaridade (positiva e negativa) de cada uma das informações sobre cada aspecto indicado no passo anterior. Por fim, um último agrupamento tem o objetivo de separar essas informações com relação ao seu conteúdo: qual foi a informação exata referente àquele aspecto com aquela polaridade. Esse último agrupamento seria capaz de, por exemplo, distinguir a sentença “A bateria dura muito pouco” de “A bateria esquenta demais”, pois apesar de o aspecto (bateria) e a polaridade (negativa) serem os mesmos, o conteúdo em si é diferente.

A partir dos grupos, o método seleciona aquele com mais elementos como sendo o mais representativo de um determinado aspecto e polaridade, ou seja, esta seleção é feita para cada combinação dessas duas informações. Dado o grupo selecionado, o elemento com maior pontuação, calculada seguindo a [Equação 3.6](#), é escolhido para compor o sumário.

$$T_{seg} = \max_{\forall s \in g} \left(\frac{\sum_{i=1}^{n_s} \text{TF-IDF}(s_i)}{n_s} \right) \quad (3.6)$$

A pontuação leva em consideração os valores TF-IDF de cada palavra s_i de cada fragmento de texto s no grupo g , sendo n_s o número de palavras no fragmento s . O fragmento de texto, em cada grupo, com maior pontuação é selecionado para compor o resumo.

Após selecionadas as informações relevantes, o método de sumarização abstrativa apresentado por [Condori e Pardo \(2017\)](#) combina essas informações (aspectos, qualificadores e grau de polaridade) com *templates* pré-estabelecidos para gerar as sentenças finais do resumo.

A avaliação dos métodos foi realizada sobre o corpus OpiSums-PT ([CONDORI et al., 2015](#)) de produtos (livros e produtos eletrônicos). Os resultados medidos automaticamente segundo a medida ROUGE são apresentados na [Tabela 8](#). É possível perceber que os resultados obtidos pelo método abstrativo foram inferiores, mas isso se dá pelo fato de métodos abstrativos gerarem textos novos e não apenas uma seleção dentre as sentenças originais.

Tabela 8 – Resultados da sumarização (ROUGE-01) obtidos por [Condori e Pardo \(2017\)](#)

Método	Precisão (%)	Cobertura (%)	Medida F (%)
<i>Opizer-E</i>	39,3	39,2	39,0
<i>Opizer-A</i>	15,5	28,0	18,9

Fonte: Adaptada de [Condori e Pardo \(2017\)](#).

Para fazer uma análise mais profunda, [Condori e Pardo \(2017\)](#) realizaram uma análise segundo características linguísticas, assim como [Carenini, Ng e Pauls \(2006\)](#). A

análise foi feita por humanos com respeito a 3 produtos do cópulus original. Os resultados são apresentados na [Tabela 9](#).

Tabela 9 – Resultados da análise linguística realizada por [Condori e Pardo \(2017\)](#)

	<i>Opizer-E</i>	<i>Opizer-A</i>	Referência
Gramaticalidade	3,763	3,141	4,647
Ausência de redundância	3,731	3,936	4,692
Clareza	4,045	3,981	4,763
Foco	3,750	3,910	4,705
Estrutura e coerência	3,423	3,115	4,660

Fonte: Adaptada de [Condori e Pardo \(2017\)](#).

É importante destacar o método *Opizer-A* como a única iniciativa de sumarização abstrativa para o Português do Brasil, que ainda se apresenta bastante limitada, principalmente durante a geração do texto do resumo. Os resumos abstrativos gerados são inferiores quanto à estrutura e coerência textual, e à gramaticalidade. Estas falhas se dão, principalmente, segundo os autores, pela rigidez imposta pelo uso de *templates* durante a geração do texto.

3.3 Considerações finais

A partir dos trabalhos analisados, é possível perceber que ainda há melhorias a serem feitas na área de sumarização de opiniões, principalmente em termos de gramaticalidade, estrutura e coerência dos textos, como indicado por [Condori e Pardo \(2017\)](#). O uso de representações semânticas, que também foi indicada como possibilidade de melhoria por [Ganesan, Zhai e Han \(2010\)](#), pode ser um caminho para se produzir melhores resumos, por se beneficiar de sistemas de geração textual, por exemplo, além de incorporarem informações em um nível mais profundo de análise.

O uso desse tipo de representação já foi explorado na área de sumarização automática, inclusive com a representação AMR mais recente, como apresentado na [Seção 3.1](#). Como indicado em todos os trabalhos, é necessário fazer a junção dos grafos AMR, visto que a representação é feita em nível sentencial, podendo ser feito por padrão de *strings* ([LIU et al., 2015](#)) ou resolução de correferência ([DOHARE; GUPTA; KARNICK, 2018](#); [LIAO; LEBANOFF; LIU, 2018](#)). Pode-se perceber que a qualidade do *parser* utilizado, consequentemente dos grafos, também afeta os resultados obtidos ([LIU et al., 2015](#)).

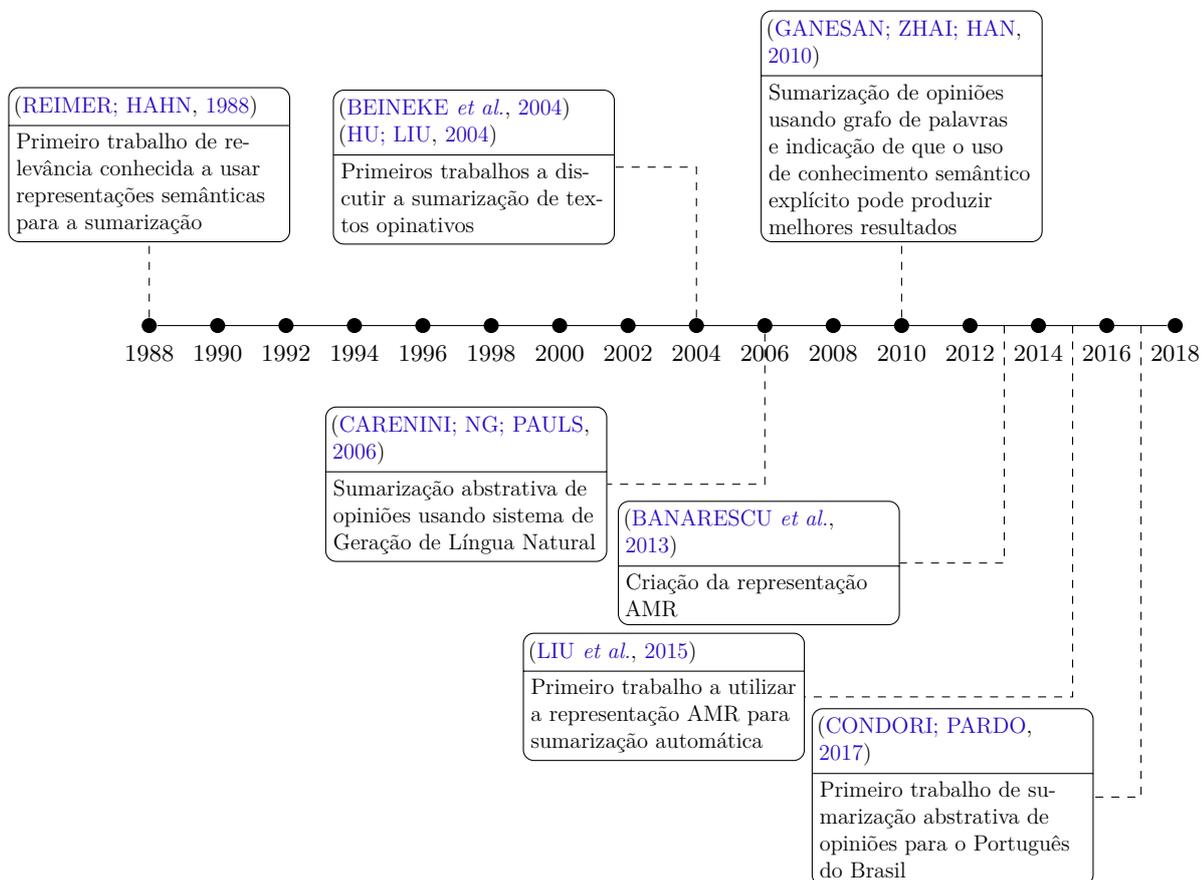
Apesar de já explorado, o uso de representações semânticas explícitas não foi abordado para o tratamento de textos opinativos, apenas jornalísticos. Como indicado por [Carenini, Ng e Pauls \(2006\)](#), o gênero de textos opinativos possui características que de-

vem ser consideradas, o que impacta negativamente o resultado de sistemas criados para textos de outros gêneros.

Da análise dos trabalhos, pode-se fazer a observação de que os métodos desenvolvidos a partir de representações semânticas tendem a não ter uma taxa de compressão, elemento indicado por [Mani \(2001\)](#) como sendo um dos principais na construção de um sistema de sumarização clássica. Dessa maneira, uma análise do poder de compressão das técnicas avaliadas neste projeto foi realizada e será apresentada juntamente com os resultados no [Capítulo 5](#).

Por fim, para se ter um melhor panorama da evolução da área, a [Figura 20](#) apresenta uma linha do tempo com os principais marcos indicados em ordem cronológica.

Figura 20 – Linha do tempo contendo os principais trabalhos da área



Fonte: Elaborada pelo autor.

ANOTAÇÃO DO CÓRPUS

Como discutido no [Capítulo 2](#), é necessária, para o andamento deste trabalho, a criação de um *córpus* de sumarização automática de opiniões com suas respectivas representações AMR. Portanto, foi realizada a adaptação do *córpus* OpiSums-PT ([CONDORI *et al.*, 2015](#)), um *córpus* de opiniões sobre produtos (livros e produtos eletrônicos) com resumos de referência criados manualmente. Desse modo, a anotação foi realizada seguindo as diretrizes originais da representação escolhida ([BANARESCU *et al.*, 2013](#)) e suas devidas adaptações para a língua portuguesa ([CABEZUDO; PARDO, 2019](#)).

A anotação do *córpus* também deve contemplar questões importantes para a área, como as indicadas por [Hovy e Lavid \(2010\)](#), permeando a instanciação do conhecimento linguístico (neste caso através da representação semântica escolhida), bem como a definição de seus limites e possíveis simplificações para se garantir a qualidade e estabilidade da anotação. A escolha dos anotadores e das ferramentas de anotação também são pontos importantes, bem como o treinamento necessário para se realizar a anotação.

4.1 Estrutura do *córpus*

O *córpus* OpiSums-PT¹ consiste de um total de 17 produtos (13 livros e 4 produtos eletrônicos), com 10 comentários, de no máximo 300 *tokens* cada. Os comentários sobre livros foram selecionados aleatoriamente do *córpus* ReLi ([FREITAS *et al.*, 2014](#)). As opiniões a respeito dos produtos eletrônicos foram, por sua vez, extraídas do *site* de comércio eletrônico Buscapé².

Para cada produto (com 10 documentos cada), o *córpus* apresenta 10 resumos diferentes (5 extrativos e 5 abstrativos) criados manualmente por 14 participantes. Os

¹ Disponível em: <<https://sites.google.com/icmc.usp.br/opinando>>. Acesso em: 16 fev. 2021.

² Disponível em: <<https://www.buscape.com.br>>. Acesso em: 16 fev. 2021.

anotadores receberam a instrução de que os resumos, tanto abstrativos quanto extrativos, tivessem aproximadamente 100 *tokens* cada. As estatísticas completas sobre o *córpus* podem ser vistas na [Tabela 10](#).

Tabela 10 – Estatísticas do *córpus* OpiSums-PT

	Comentários	Resumos extrativos	Resumos abstrativos
Documentos	170	85	85
Sentenças	860	534	430
<i>Tokens</i>	16.127	9.722	9.763
<i>Types</i>	3.595	6.525	6.413
Média de <i>tokens</i> por documento	94,86	114,38	114,86
Média de <i>tokens</i> por sentença	18,75	18,21	22,70
Média de <i>types</i> por documento	61,36	76,76	75,45
Média de <i>types</i> por sentença	12,13	12,22	14,91

Fonte: Dados da pesquisa.

4.2 Metodologia de anotação

A anotação foi realizada por três anotadores com experiência prévia em AMR, visto que participaram também da anotação realizada por [Cabezudo e Pardo \(2019\)](#). O processo se deu de forma interativa, com discussões constantes depois de determinados períodos de anotação, permitindo, assim, uma adequação das diretrizes e resolução de dificuldades de anotação.

Para o processo de anotação, foram escolhidas apenas as sentenças que compõem os comentários dos usuários sobre os produtos, sendo ignorados, neste momento, os resumos abstrativos criados manualmente por [Condori et al. \(2015\)](#), principalmente por uma questão de viabilidade do projeto, visto que a anotação semântica é bastante custosa.

Como pré-processamento, as sentenças foram normalizadas automaticamente pelo Enelvo ([BERTAGLIA; NUNES, 2016](#)), visto que as diretrizes da AMR indicam que erros de escrita devem ser corrigidos durante a anotação e essa ferramenta realiza a correção de problemas de ortografia e abreviações comuns na linguagem da internet. No *framework* CLSA, essa etapa recebe o nome de *Microtext Analysis*. Caso os anotadores encontrassem algum problema de normalização, o erro seria marcado em um campo específico de comentários.

As 860 sentenças dos comentários no *córpus* foram divididas entre os anotadores, sendo 70 delas aleatoriamente selecionadas e duplicadas entre todos. Essas sentenças duplicadas foram utilizadas no cálculo da concordância de anotação. Dessa forma, essas sentenças foram marcadas e os anotadores instruídos a não discutirem entre si sobre elas e os fenômenos que nelas ocorressem, especificamente.

Assim como na anotação de Cabezudo e Pardo (2019), sentenças com verbos inexistentes no repositório do Verbo Brasil (DURAN; ALUÍSIO, 2015) não foram anotadas e o verbo ausente foi registrado em uma lista para, futuramente, ser feita a expansão do recurso. Apesar disso, essa diretriz foi adaptada de forma a permitir a anotação dos casos em que não foi necessário ter as estruturas argumentais dos verbos; nesses casos, ainda adicionou-se o verbo em questão na lista de verbos inexistentes no repositório.

Um exemplo em que o verbo não existe no recurso, mas ainda foi anotado, pode ser visto na Figura 21. No caso, as diretrizes da AMR pedem que o substantivo “diálogo” seja verbalizado para o verbo “dialogar”, porém não há *frameset* para ele no Verbo Brasil. Como a anotação da sentença não pede diretamente os argumentos desse verbo, ela foi anotada independentemente. O mesmo ocorreu com outros verbos (“empolgar”, “aconselhar”) e expressões (“falar sério”, “ficar para trás”).

Figura 21 – Exemplo de sentença com verbos modais anotados em AMR

Aliás. A escrita dele está toda boa e bastante diálogo (gosto disso).

```
(e / e
  :op1 (e2 / e
    :op1 (e3 / escrever-01
      :ARG0 (e4 / ele)
      :mod (b / bom))
    :op2 (t / ter-01
      :ARG0 e3
      :ARG1 (d / dialogar
        :quant (b2 / bastante)))
    :ARG1-of (g / gostar-01
      :ARG0 (e5 / eu))))
```

Fonte: Elaborada pelo autor.

4.3 Resultados de anotação

A anotação resultou em 481 sentenças analisadas (55,93% do cópulus original), dentre as quais 404 (49,98%) foram efetivamente anotadas, isto é, possuem um grafo AMR correspondente, como pode ser visto na Tabela 11. Dentre as sentenças analisadas, 83 (9,65%) possuem verbos novos não previstos no Verbo Brasil, sendo 6 delas efetivamente anotadas apesar disso.

4.3.1 Concordância de anotação

Dentre as 70 sentenças selecionadas para o cálculo de concordância, apenas 17 foram efetivamente anotadas por todos os anotadores. Os resultados do cálculo de concordância, utilizando a medida *Smatch* (CAI; KNIGHT, 2013), para as 17 sentenças, é

Tabela 11 – Resultados da anotação

Sentenças analisadas	481
Sentenças efetivamente anotadas	404
Sentenças com verbos novos	83

Fonte: Dados da pesquisa.

apresentado na [Tabela 12](#). Dado que a medida utilizada compara o nível de similaridade entre dois grafos AMR apenas, o cálculo é feito para cada par de anotadores, sendo o valor médio de similaridade entre os grafos produzidos pelos dois anotadores. Mais detalhes sobre a medida *Smatch* serão apresentados no [Capítulo 5](#).

Tabela 12 – Resultados do cálculo de concordância

Par de anotadores		<i>Smatch</i>
Anotador 1	Anotador 2	0,90
Anotador 1	Anotador 3	0,89
Anotador 2	Anotador 3	0,92
Média		0,90

Fonte: Dados da pesquisa.

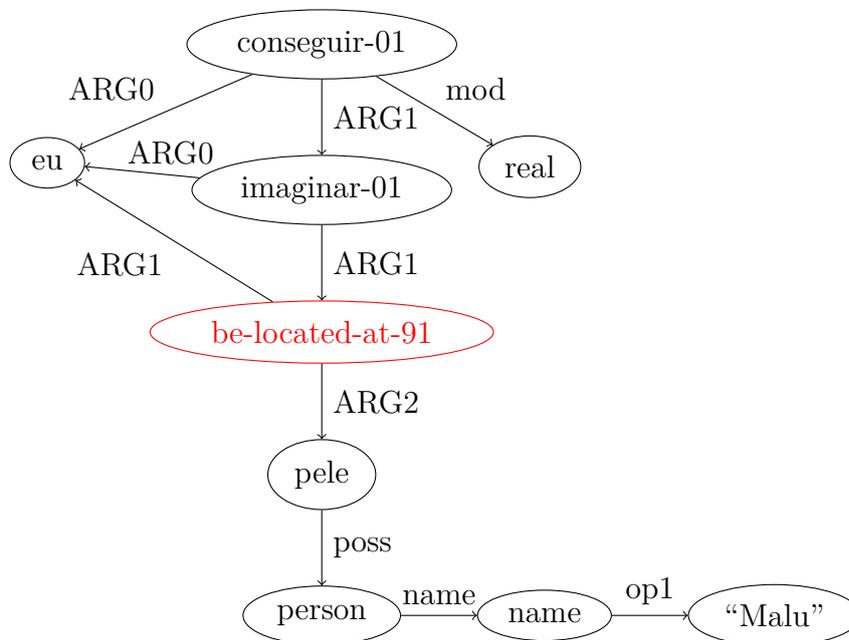
A média de *tokens* nas 17 sentenças utilizadas no cálculo de concordância é de 6, ou seja, elas são menores quando comparadas à média do *córpus* (18,75, como indicado na [Tabela 10](#)). Isso pode ser uma causa para os altos valores de concordância obtidos ([Cabezudo e Pardo \(2019\)](#) obtiveram uma concordância de 0,72) visto que sentenças menores podem ser consideradas mais fáceis de anotar. [Cabezudo e Pardo \(2019\)](#) trabalharam, também, com um grupo maior com 14 anotadores, o que influencia nos valores de concordância obtidos. É esperado que, caso mais sentenças sejam anotadas, esses valores diminuam. É importante salientar também que a medida *Smatch* pode ser considerada bastante otimista, penalizando pouco as discordâncias entre os anotadores ([ANCHIÊTA; CABEZUDO; PARDO, 2019](#)).

4.3.2 Dificuldades de anotação

Durante a anotação, foram notadas algumas dificuldades, sendo necessária a tomada de algumas decisões importantes para o processo. A primeira dificuldade diz respeito à anotação semântica de “metáforas”. Como a interpretação do significado de metáforas depende do contexto comunicativo em que determinado texto foi produzido ([LEGROSKI, 2010](#)), esses fenômenos estão sujeitos à análise subjetiva dos anotadores, podendo comprometer, assim, a uniformidade da anotação. Dessa forma, esses casos foram neutralizados (simplificados, afastando-se do rigor e precisão linguísticas, com o objetivo de tornar a anotação mais estável, segundo [Hovy e Lavid \(2010\)](#)), sendo, então, anotados de maneira

literal.³ Um exemplo para a sentença “Consegui, realmente, me imaginar na pele da Malu” pode ser visto na Figura 22.

Figura 22 – Exemplo de anotação de metáfora



Fonte: Elaborada pelo autor.

No exemplo, a expressão “na pele da Malu” foi anotada literalmente com o conceito `be-located-at-91`, que indica a presença física em um local, neste caso, “a pele da Malu”. Pode-se observar que o significado real de “imaginar-se na pele da Malu” pode ser considerado como sendo “agir da mesma forma que Malu agiria ao enfrentar as mesmas situações em que Malu se encontra”, porém, representar a metáfora dessa forma inviabiliza o processo de anotação, pois, como dito anteriormente, está sujeito a maiores subjetividades do anotador, prejudicando, assim, a estabilidade da anotação.

É importante salientar também que a troca de uma expressão (“na pele da Malu”) por outra, como “no lugar da Malu”, não resolve satisfatoriamente o problema da anotação, visto que ainda seria preciso anotar uma expressão com seu significado além da superfície textual: “no lugar da Malu” não tem sentido de um lugar pertencente à Malu ou um local em que Malu também está. Por conta dessas complicações, decidiu-se manter a anotação literal de expressões metafóricas.

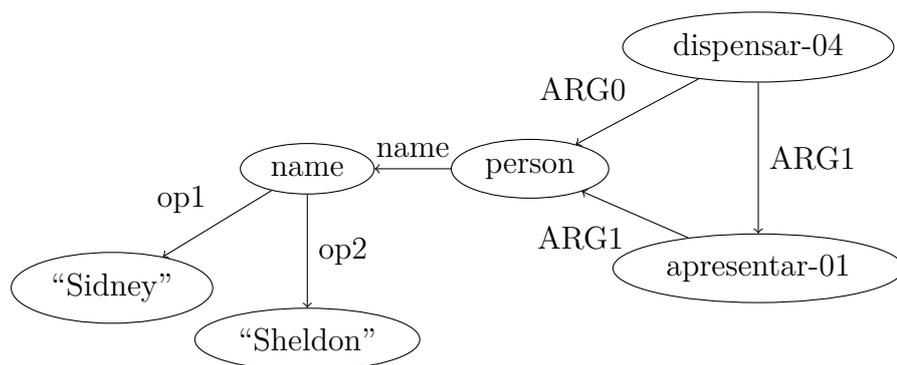
Todos os casos em que a anotação foi realizada literalmente foram marcados no corpúsculo final de forma a permitir futuros trabalhos focados especificamente nesse tipo de

³ É importante salientar que a identificação de metáforas propriamente dita é uma tarefa bastante complicada e precisaria ser feita com certa base linguística. Como esse não é o foco do trabalho, os anotadores ficaram livres para utilizar essa diretriz quando julgassem necessário.

fenômeno. Dentre as sentenças efetivamente anotadas, apenas 7 (1,73%) foram analisadas dessa forma.

Outra dificuldade encontrada na anotação foi a presença de marcadores discursivos (como “bem”, “enfim”, “mesmo” e outros). Esses marcadores foram ignorados durante a anotação, visto que não possuem um significado essencial. Um exemplo pode ser visto na [Figura 23](#), em que o marcador discursivo “bem”, da sentença “Bem, Sidney Sheldon dispensa apresentações”, foi ignorado.

Figura 23 – Exemplo de anotação com marcador discursivo



Fonte: Elaborada pelo autor.

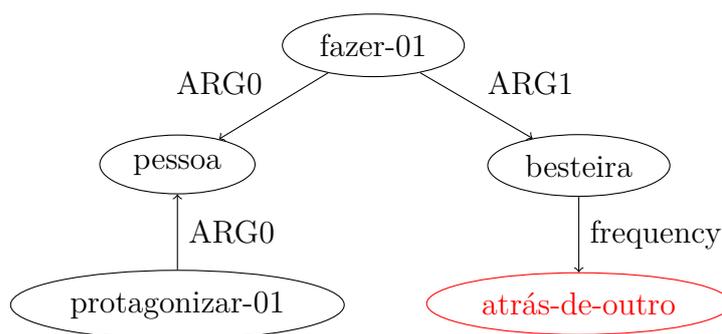
No total, 12 (2,97%) sentenças foram anotadas seguindo essa diretriz e também marcadas no córpus para facilitar futuras pesquisas no assunto.

A presença de expressões multpalavras nos textos também deve ser indicada como uma dificuldade de anotação. Apesar de as diretrizes originais elaboradas por [Banarescu et al. \(2013\)](#) indicarem que essas expressões deveriam ser substituídas por equivalentes plenos, por exemplo “dar um abraço” e “abraçar”, nem sempre isso é possível, como em “pagar mico” e “falar sério”. Nesses casos, considerou-se a expressão como um único conceito com seu significado próprio, sendo anotado com seus elementos constitutivos unidos por hífen. Um exemplo pode ser visto na [Figura 24](#), em que a representação da expressão “atrás da outra” na sentença “A protagonista faz uma besteira atrás da outra” foi anotada com o conceito único *atrás-de-outro* em conjunto com a relação *frequency*.

Dentre os fenômenos aqui discutidos, esse é o mais frequente, sendo visto em um total de 30 sentenças, isto é, 7,43% do córpus anotado.

Uma limitação encontrada nas diretrizes originais do AMR diz respeito à representação de intervalos temporais. A representação possui um conceito específico *date-interval* para representação de intervalos de tempo, porém esse conceito deve estar atrelado a entidades do tipo *date-entity*, que representam especificamente datas. Isso se mostrou uma limitação para lidar com casos de intervalos de tempo mais genéricos como o da sentença “Fala sério amiga é um livro onde a autora relata a história dela com suas amigas da infância até a adolescência”. Para solucionar esse problema, optou-se por

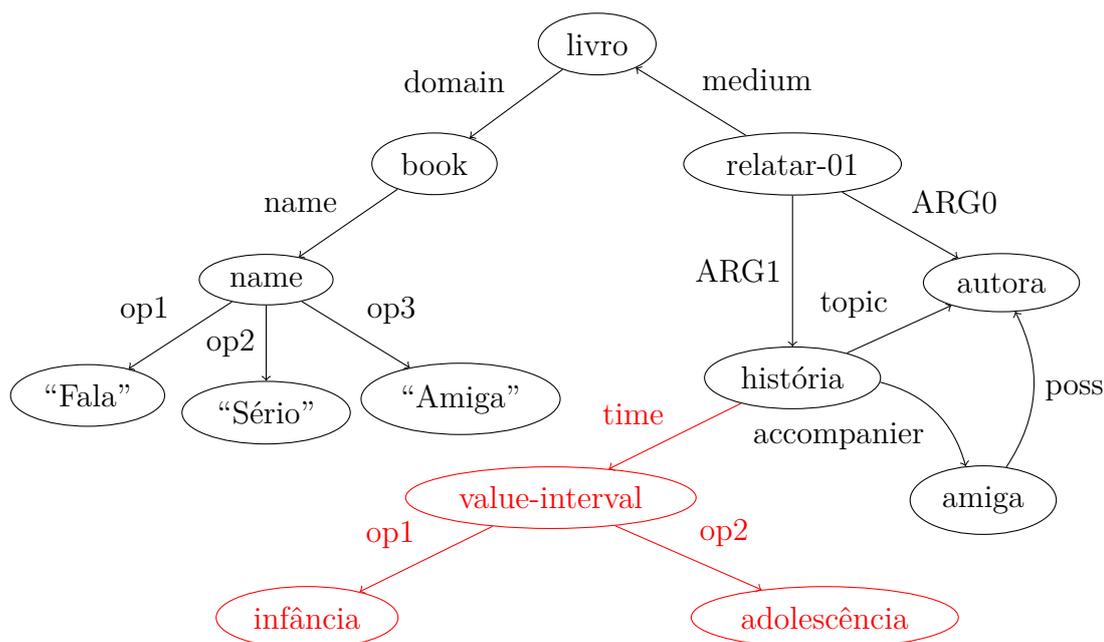
Figura 24 – Exemplo de anotação com expressão multipalavra unida por hífen



Fonte: Elaborada pelo autor.

utilizar o conceito já existente `value-interval` mais genérico, já que seus argumentos não precisam ser necessariamente datas. O grafo AMR para a sentença em questão pode ser visto na [Figura 25](#). Essa é a única sentença do cópulus que possui essa peculiaridade.

Figura 25 – Exemplo de anotação com intervalo de tempo



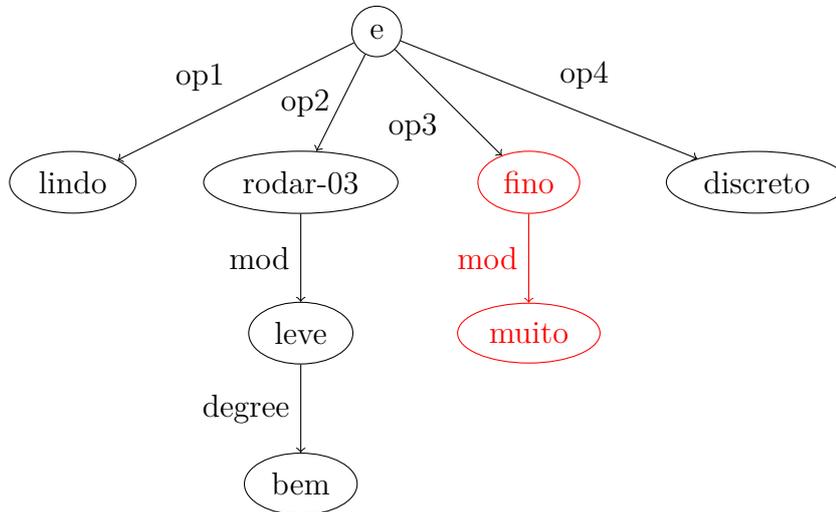
Fonte: Elaborada pelo autor.

A última observação realizada durante o processo de anotação diz respeito ao tratamento dos diminutivos sintéticos (formados a partir de sufixos, como *-inho* e *-inha*), característica comum na língua portuguesa que ainda não foi abordada pelos trabalhos anteriores. Com isso, definiu-se seguir a classificação de [Alves \(2006\)](#) dos diminutivos entre dois níveis de análise linguística: a Semântica e a Pragmática.

Os diminutivos semânticos, são aqueles com função de exprimir tamanho ou quantidades reduzidos, bem como uma intensificação do lexema ao qual se relaciona. Essa classe de diminutivo é anotada explicitamente na representação AMR, por meio de re-

lações como *mod* (modificador) ou *degree* (grau de intensidade). Um exemplo pode ser visto no grafo da Figura 26 correspondente à sentença “Lindo, roda bem leve, fininho e discreto”, em que o termo “fininho” foi anotado como correspondente a “muito fino”.

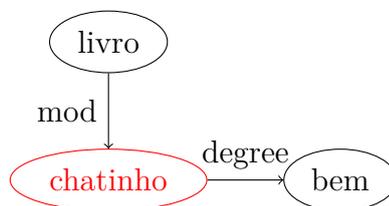
Figura 26 – Exemplo de anotação com diminutivo semântico



Fonte: Elaborada pelo autor.

Alves (2006) também aponta outros dois tipos de uso do diminutivo em português, ambos com valor pragmático, exprimindo afetividade, avaliação ou crítica do autor com relação a entidades ou situações. Dessa forma, esse tipo de fenômeno não foi anotado, mantendo-se o conceito juntamente com o sufixo correspondente ao diminutivo. Um exemplo pode ser visto na Figura 27, contendo a anotação da sentença “Livro bem chatinho”, sendo a palavra “chatinho” considerado como pragmático por expressar uma avaliação subjetiva do usuário relativamente ao livro e não a descrição de um atributo físico do objeto.

Figura 27 – Exemplo de anotação com diminutivo pragmático



Fonte: Elaborada pelo autor.

4.3.3 Comparação entre textos opinativos e jornalísticos

Além do levantamento das dificuldades, foi feita uma comparação dos resultados de anotação dos textos opinativos com textos jornalísticos com o objetivo de entender como os diferentes fenômenos se apresentam em textos de gêneros distintos. Também

tem-se o fato de que os trabalhos de sumarização semântica têm a tradição de lidar com textos exclusivamente jornalísticos, havendo uma falta de descrições da anotação AMR sobre opiniões. Para tanto, utilizaram-se as 783 sentenças anotadas por [Cabezudo e Pardo \(2019\)](#) obtidas de diferentes seções do jornal Folha de São Paulo.

Primeiramente, uma comparação das 15 relações mais frequentes entre os diferentes gêneros foi feita, podendo ser vista na [Tabela 13](#). Como mostra a relação `name`, os textos jornalísticos tendem a possuir mais entidades nomeadas (7,65%) quando comparados aos opinativos (3,96%). Isso pode se dar devido ao fato de o comentário de determinado produto ser escrito diretamente em sua página, diminuindo a necessidade de citar explicitamente sobre qual entidade se escreve.

Tabela 13 – Relações mais frequentes nos corpúscos jornalístico e opinativo

Opinativo			Jornalístico		
Relação	Frequência	%	Relação	Frequência	%
ARG1	652	20,64	ARG1	660	20,03
op	624	19,75	op	582	17,66
ARG0	485	15,35	ARG0	462	14,02
mod	314	9,94	name	252	7,65
ARG2	208	6,58	mod	240	7,28
name	125	3,96	ARG2	176	5,34
domain	96	3,04	polarity	159	4,83
polarity	80	2,53	domain	126	3,82
time	67	2,12	time	87	2,64
poss	56	1,77	quant	85	2,58
topic	56	1,77	location	64	1,94
snt	55	1,74	manner	45	1,37
quant	44	1,39	topic	45	1,37
degree	39	1,23	poss	44	1,34
ARG3	38	1,20	mode	32	0,97

Fonte: Dados da pesquisa.

O corpúscos de textos opinativos, por sua vez, possui maiores quantidades da relação `snt` (*sentence*), utilizada em conjunto com o conceito `multi-sentence` nos casos em que uma instância de anotação é constituída por mais de uma sentença. Isso se dá por conta de problemas de segmentação das sentenças do corpúscos original, o OpiSums-PT, sendo a segmentação correta feita durante o processo de anotação, seguindo as diretrizes da própria AMR. Um exemplo é a instância “Muito boa. Tudo de Positivo eu relatei nos Prós e de Negativo eu relatei nos Contras... Achei aqui por R\$1000,00 a menos q na loja física”, representada como sendo uma única sentença no corpúscos, porém, segmentada em três durante a anotação.

Com respeito à relação `degree`, os textos opinativos tendem a ter maiores quanti-

dades, já que essa relação está associada a intensificadores e comparações. Além disso, ao se levar em consideração o conceito associado a essa relação, **have-degree-91**, percebe-se mais um aumento na frequência desses fenômenos: de 38 para 59. As estatísticas para os dez **framesets** mais frequentes nos córpus são apresentadas na [Tabela 14](#). Apenas **framesets** são apresentados devido ao fato de eles representarem eventos, isto é, o foco principal da AMR, como discutido anteriormente na [Subseção 2.3.1](#). Além disso, ao olhar para esse tipo específico de conceito, é possível capturar melhor os casos de reificação (transformação de uma relação em um conceito).

Tabela 14 – *Framesets* mais frequentes no córpus jornalístico e opinativo

Opinativo			Jornalístico		
<i>Frameset</i>	Frequência	%	<i>Frameset</i>	Frequência	%
cause-01	44	5.27	ter-01	36	3,84
ler-01	42	5.03	possible-01	26	2,77
ter-01	35	4.19	contrast-01	25	2,67
gostar-01	33	3.95	dizer-01	22	2,35
contrast-01	27	3.23	fazer-01	21	2,24
escrever-01	25	2.99	haver-01	16	1,71
have-degree-91	21	2.51	querer-01	15	1,60
have-rel-role-91	21	2.51	acontecer-01	15	1,60
possible-01	17	2.04	saber-01	13	1,39
fazer-01	15	1.80	cause-01	12	1,28

Fonte: Dados da pesquisa.

Como pode ser visto na tabela de *framesets* mais frequentes, o conceito **ler-01** é bastante recorrente, visto que os produtos do córpus OpiSums-PT são, em sua maioria, livros. Da mesma forma, o conceito **have-rel-role-91**, utilizado para indicar graus de parentesco e relacionamento, também é frequente, já que existem alguns livros que abordam as relações entre personagens: “Fala sério, mãe!”, “Fala sério, pai!” e “Fala sério, amiga!”.

É interessante apontar que o conceito **gostar-01** também se mostra bastante frequente no córpus opinativo. Isso se deve, provavelmente, à natureza desse tipo de texto, em que o autor expõe suas opiniões e gostos sobre os produtos. Apesar disso, é importante ressaltar que a maioria dos comentários sobre produtos eletrônicos possuem os trechos “O que eu (não) gostei” como uma estrutura comum, provavelmente originária da página de onde os comentários foram extraídos (por exemplo em um formulário).

Uma maior frequência do conceito **cause-01** pode indicar que os usuários explicam as motivações para as opiniões expressas no texto.

No que concerne ao *frameset* **escrever-01**, sua frequência alta se dá devido ao fato de que a AMR, como descrito anteriormente, foca na representação dos eventos. Dessa

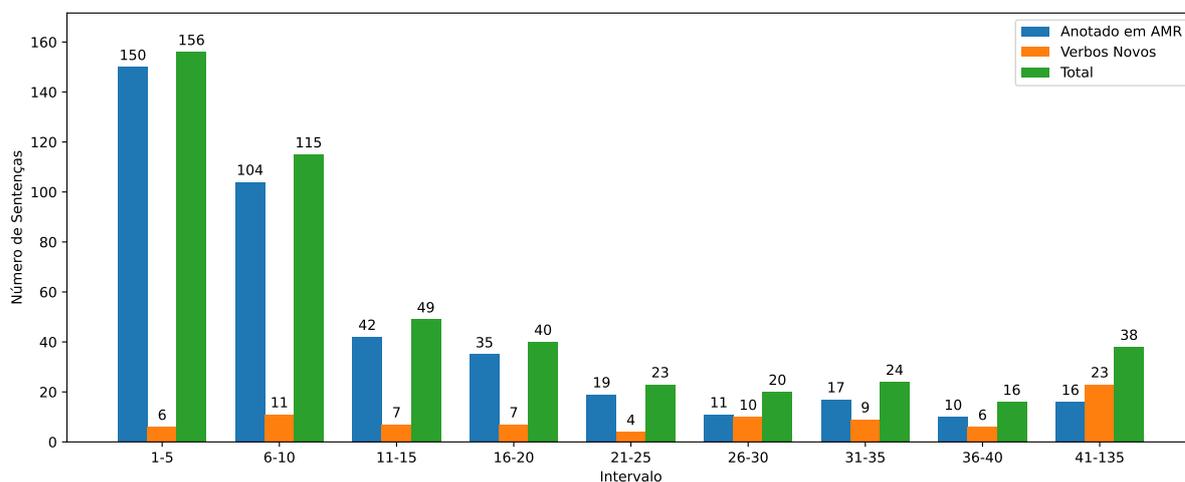
forma, trechos como “o livro de Sidney Sheldon” são anotados de forma a corresponder com “o livro escrito por Sidney Sheldon”, sendo necessária a inserção do conceito *escrever-01*.

É importante notar que os textos opinativos tendem a ter mais *framesets* utilizados em contextos com algum grau de sentimento associado (*gostar-01*, *have-degree-91*), contrário aos textos jornalísticos que possuem mais conceitos objetivos e descritivos (*ter-01*, *dizer-01*, *fazer-01*, *acontecer-01*).

4.3.4 Comparação entre diferentes tamanhos de sentenças

Para permitir uma melhor análise do comportamento da língua e da representação utilizada nos textos opinativos, os dados foram separados de acordo com o tamanho (em *tokens*) das sentenças, tendo cada grupo um intervalo de 5 *tokens*, de modo a evitar distorções nas análises devido a sentenças de tamanhos muito diferentes que podem apresentar fenômenos de tipos distintos.. O número de sentenças (anotadas e com verbos novos) em cada grupo pode ser visualizado na [Figura 28](#). Visto que os grupos de sentenças com mais de 40 tokens eram muito pequenos, eles foram todos condensados em um único grupo de sentenças longas (de 41 a 135 *tokens*).

Figura 28 – Número de sentenças em cada grupo de acordo com o número de *tokens*



Fonte: Dados da pesquisa.

No gráfico, “Anotado em AMR” representa o número de sentenças efetivamente anotadas (aquelas com um grafo AMR associado), “Verbos Novos” apresenta o número de sentenças com verbos inexistentes no Verbo Brasil (podendo ser ou não anotadas, como indicado anteriormente). Por fim, “Total” indica o número total de sentenças selecionadas para compor o grupo, independentemente de terem sido anotadas ou não.

Como pode ser visto no gráfico, sentenças maiores tendem a possuir mais verbos novos a serem incluídos no recurso, resultando em uma maior taxa de sentenças não anotadas.

As frequências, em porcentagem, das 15 relações mais frequentes para cada grupo são apresentadas na [Tabela 15](#).

Tabela 15 – Frequências (%) das 15 relações mais frequentes em textos de opinião de acordo com o tamanho das sentenças (em *tokens*)

Intervalo	[1, 5]	[6, 10]	[11, 15]	[16, 20]	[21, 25]	[26, 30]	[31, 35]	[36, 40]	[41, 136]
Relação									
:ARG1	13,73	19,03	23,65	23,41	17,51	24,21	20,00	22,4	22,20
:op	26,80	19,64	13,11	18,78	26,94	15,26	18,38	20,4	19,14
:ARG0	10,78	12,96	17,66	18,05	14,14	20,00	13,78	12,8	18,13
:mod	9,80	12,75	13,39	6,83	9,09	8,95	10,27	11,6	7,13
:ARG2	7,84	6,48	7,12	7,56	3,70	10,53	5,41	6,0	6,11
:name	9,48	4,66	2,85	3,41	4,71	1,05	3,24	2,8	2,85
:domain	3,27	5,47	3,99	2,44	3,37	1,05	2,43	1,6	2,04
:polarity	4,25	2,83	3,13	2,68	1,35	1,58	2,16	2,4	2,04
:time	1,31	2,02	2,85	1,95	1,35	3,68	0,81	2,0	3,26
:poss	0,65	0,61	0,85	2,44	1,01	1,05	1,89	2,8	3,87
:topic	0,33	1,21	2,56	1,46	2,36	0,53	2,70	2,4	2,04
:snt	1,31	0,81	1,71	0,73	1,35	1,05	4,32	2,8	1,83
:quant	0,33	1,62	0,85	1,22	2,02	2,11	2,16	2,4	0,61
:degree	3,59	1,21	0,28	0,24	1,01	0,53	2,43	0,8	1,02
:ARG3	1,31	1,82	1,14	2,20	0,67	1,05	1,08	1,2	0,20

Fonte: Dados da pesquisa.

Da tabela é possível perceber que sentenças menores possuem maiores frequências de entidades nomeadas (representado pela relação *name*). Isso se dá pelo fato de muitas sentenças pequenas, possivelmente títulos dos comentários, conterem apenas o nome do produto. Por sua vez, sentenças maiores possuem mais situações em que é necessário utilizar o conceito de *multi-sentence*, como pode ser visto a partir da relação *snt*, o que é natural, visto que são mais propensas a apresentarem problemas de segmentação.

Uma observação interessante que pode ser realizada é a de que sentenças menores possuem mais relações do tipo *mod* e *domain*, indicando que os usuários tendem a usar mais adjetivos e cópulas nesse tipo de sentença, deixando suas explicações mais elaboradas para sentenças maiores. Isso pode ser confirmado observando as estatísticas dos *framesets* mais utilizados, apresentadas na [Tabela 16](#), em especial o conceito *cause-01*, que possui uma frequência mais alta em sentenças maiores, indicando a tendência dos textos maiores apresentarem justificativas para as opiniões expressas.

Na tabela também se pode observar que sentenças maiores possuem maiores quantidades de conceitos do tipo *contrast-01*, utilizados para representar orações adversativas. Isso pode acontecer por essas sentenças apresentarem discursos mais complexos, contendo comparações, por exemplo. O desenvolvimento de um discurso mais complexo pode ser

Tabela 16 – Frequências (%) dos 10 *framesets* mais frequentes em textos de opinião de acordo com o tamanho das sentenças (em *tokens*)

Intervalo	[1, 5]	[6, 10]	[11, 15]	[16, 20]	[21, 25]	[26, 30]	[31, 35]	[36, 40]	[41, 136]
Conceito									
cause-01	3,75	4,10	5,83	4,65	4,92	3,51	5,62	7,81	6,92
ler-01	6,25	8,20	9,71	5,43	4,92	3,51	2,25	1,56	1,54
ter-01	1,25	4,92	4,85	8,53	3,28	1,75	2,25	3,12	3,85
gostar-01	3,75	9,02	3,88	4,65	4,92	0,00	2,25	0,00	3,08
contrast-01	0,00	1,64	4,85	3,88	3,28	7,02	1,12	7,81	2,31
escrever-01	0,00	4,92	3,88	2,33	8,20	3,51	2,25	1,56	1,54
have-degree-91	3,75	5,74	1,94	2,33	1,64	3,51	0,00	3,12	0,77
have-rel-role-91	7,50	1,64	2,91	1,55	1,64	1,75	2,25	0,00	3,08
possible-01	0,00	1,64	1,94	1,55	1,64	0,00	4,49	4,69	2,31
fazer-01	3,75	0,00	2,91	0,78	1,64	0,00	2,25	4,69	1,54

Fonte: Elaborada pelo autor.

visto também a partir da relação **time** na [Tabela 15](#), visto que essa relação indica informações adicionais sobre o assunto (tempo).

4.4 Comparação entre comentários e resumos

Para obter um melhor entendimento sobre o processo de criação de resumos – quais conceitos e relações são preservados na transposição dos comentários aos resumos –, pode-se analisar a distribuição dos conceitos e relações nos dois cenários. A comparação foi realizada sobre todas as sentenças do OpiSums-PT anotadas manualmente com os resumos extrativos presentes no *cópus*: caso uma sentença tenha sido selecionada para compor algum sumário, os conceitos e relações do seu grafo correspondente passam a compor a distribuição referente aos resumos.

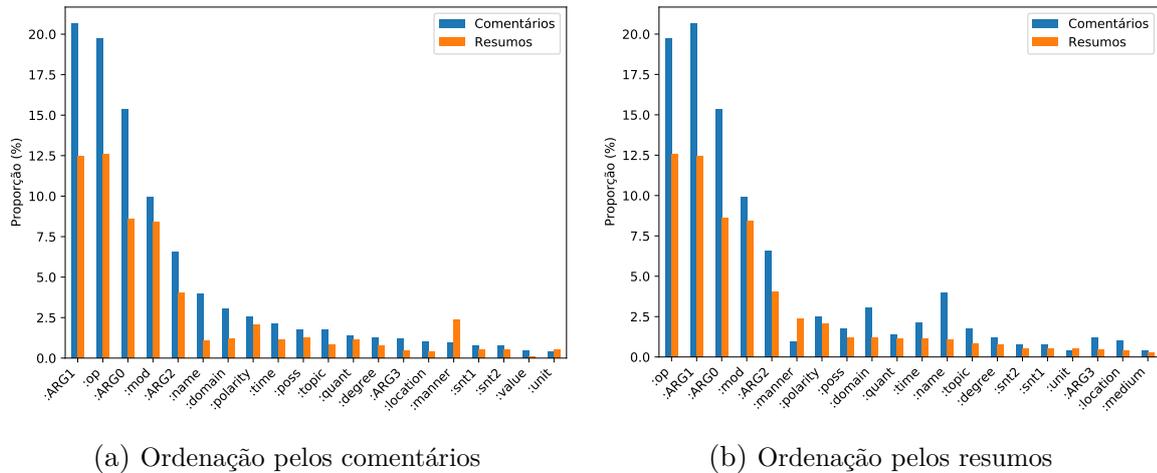
Podem-se ver as distribuições das 20 relações mais frequentes na [Figura 29](#). As duas imagens apresentam uma ordenação diferente, de acordo com os comentários e com os resumos, para garantir a visualização dos casos extremos em que um elemento seja bastante comum nos comentários, mas muito pouco frequente nos sumários e vice-versa.

Como é possível ver dos gráficos, as cinco relações mais comuns são as mesmas (:ARG1, :op, :ARG0, :mod e :ARG2) com variação na ordem das duas primeiras. Também é evidente que, em sua maioria, as proporções são menores nos resumos do que nos comentários, indicando que há uma variedade alta nas escolhas das relações semânticas dos sumários, evitando que os valores fiquem muito concentrados em poucos elementos.

No que diz respeito a relações específicas, é notável a relação **:manner**, representando o modo ou método como uma ação foi realizada, que possui uma maior proporção nos resumos (2,4%) se comparado aos comentários (1%), um aumento de 140%, apontando uma certa importância que esse tipo de relação tenha no processo de sumarização.

Outro destaque é a relação **:name**, que possui uma discrepância, sendo a sexta

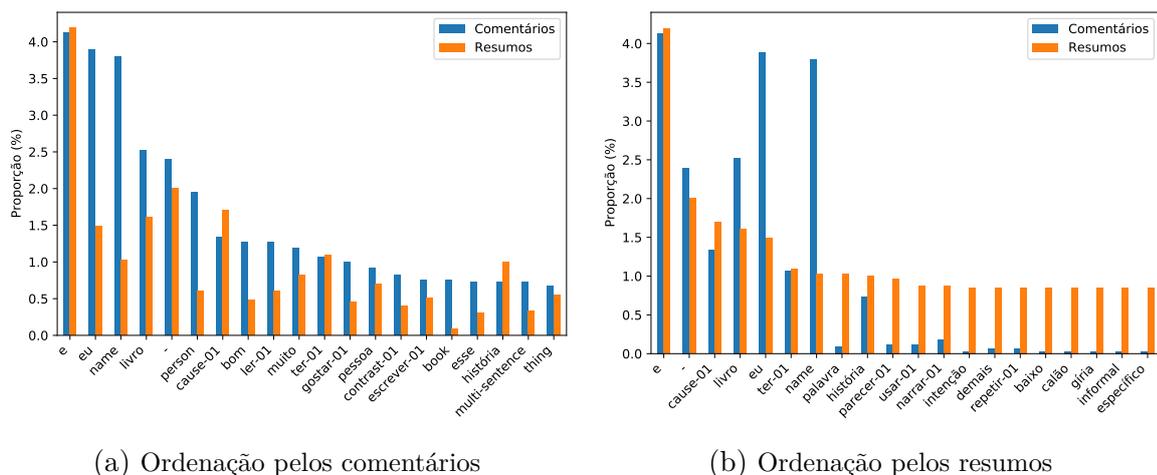
Figura 29 – Comparação da distribuição das relações entre comentários e resumos



relação mais comum nos comentários e passando à 12ª posição nos resumos, com uma queda de 72,5% (de 4% para 1,1%). Esse fenômeno indica que os sumários possuem uma menor incidência de sentenças contendo entidades nomeadas. No caso dos domínios contemplados no *córpus*, as entidades mais comuns são nomes de produtos (eletrônicos e livros) e pessoas (principalmente autores e personagens dos livros). Seria necessária, porém, uma análise mais minuciosa para determinar se existem entidades em específico que são mantidas ou descartadas na criação dos resumos.

Seguidamente, na [Figura 30](#), são apresentadas as proporções dos 20 conceitos mais frequentes nos resumos e comentários.

Figura 30 – Comparação da distribuição dos conceitos entre comentários e resumos



Numa visão geral, pode-se perceber que os conceitos menos frequentes nos resumos, do oitavo (*palavra*) ao vigésimo (*específico*), possuem uma proporção maior do que nos comentários, indicando que são relevantes, apesar de pouco frequentes no *córpus* original. Ao mesmo tempo, não existem muitos conceitos, dentre os 20 mais comuns nas opiniões,

que possuem uma proporção tão baixa (menos de 0.5%) nos sumários, apontando que a frequência pode ser, ainda assim, um dos indicativos de relevância.

O maior destaque em que ocorre uma queda na proporção de um conceito é **book** (saindo de 0,76% para 0,1%), utilizado para representar entidades nomeadas de livros. Outro rótulo utilizado no contexto de entidades nomeadas é **person**, que também possui uma queda de frequência nos resumos (passando de 1,9% para 0,6%). É possível afirmar, portanto, que as entidades referentes aos livros são, proporcionalmente, mais suprimidas nos resumos do que entidades sobre pessoas (em especial autores ou personagens), visto que houve uma queda de 86,8% da ocorrência dos conceitos **book**, enquanto os do tipo **person** caíram em apenas 68,4%.

Das observações feitas a partir dessas estatísticas, é plausível a possibilidade de se criar regras de sumarização sobre os grafos AMR, porém uma análise mais detalhada dos fenômenos se faz necessária para definir tais regras, portanto essa abordagem não foi explorada mais a fundo neste projeto.

MÉTODOS DE SUMARIZAÇÃO AUTOMÁTICA

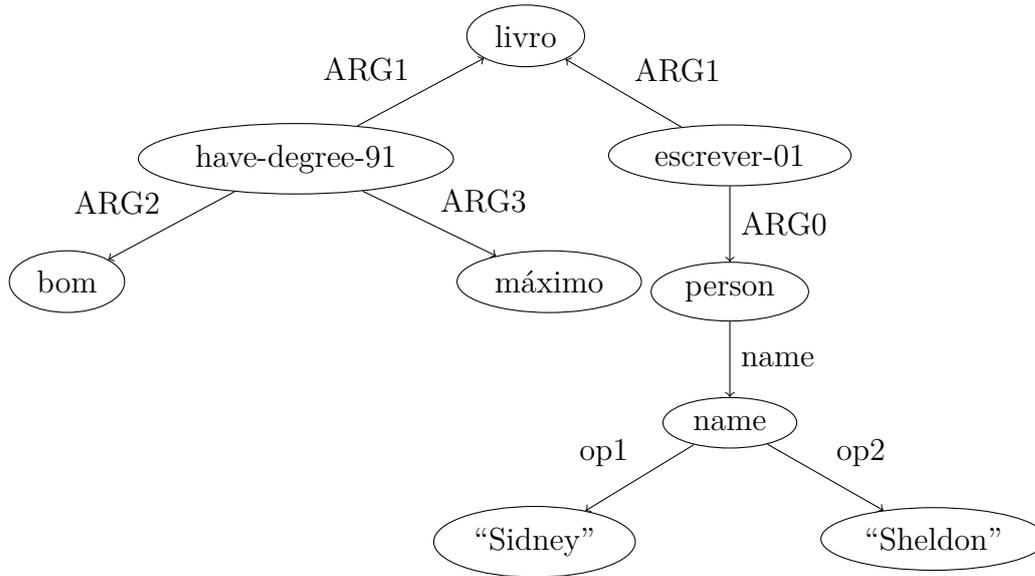
Dois grandes desafios foram investigados neste projeto: sumarização de opiniões e sumarização baseada em representações semânticas explícitas (especificamente a AMR). Como não há trabalhos específicos de sumarização de textos opinativos que utilizam representações semânticas, algumas técnicas desenvolvidas originalmente com base em textos jornalísticos foram transpostas para o contexto das opiniões.

5.1 Pré-processamento

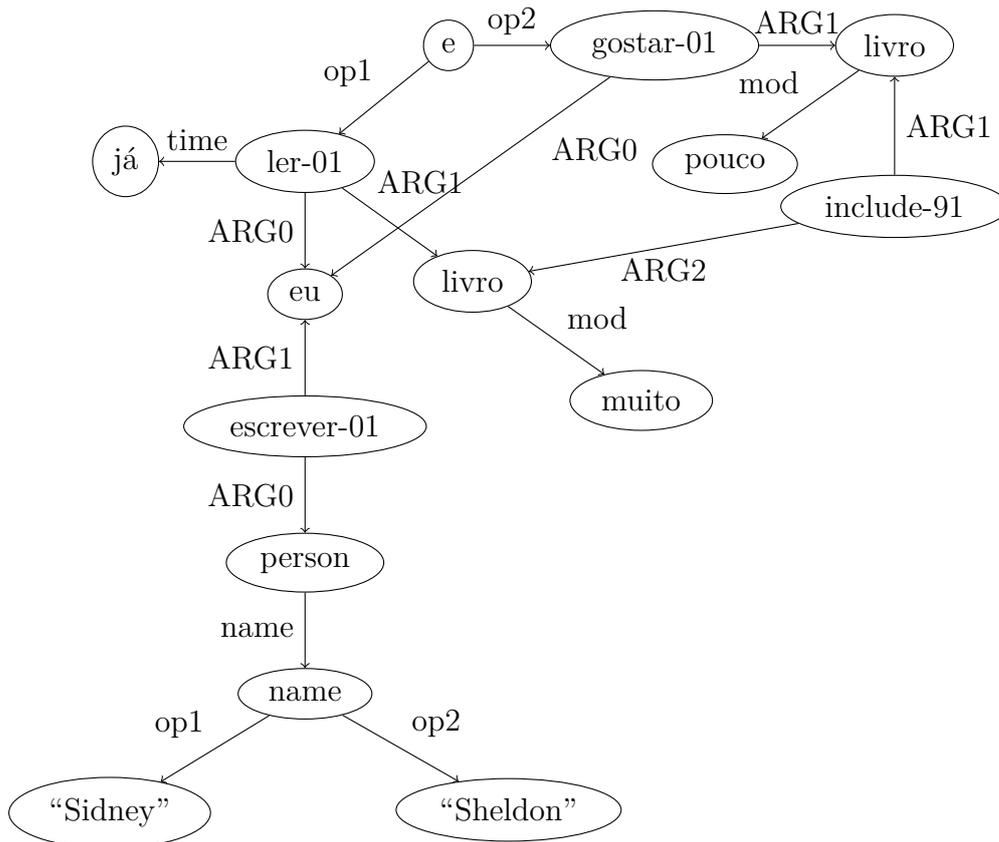
Todos os métodos de sumarização semântica explorados focam na seleção de um subgrafo resumido a partir de um único grafo AMR representando o documento (ou documentos) de entrada. Como a AMR é uma representação sentencial, é necessário, portanto, realizar-se um pré-processamento desses grafos, combinando-os em uma representação única da entrada. Para ilustrar esse processo, serão utilizados como base para os exemplos os dois grafos apresentados na [Figura 31](#).

A criação do grafo multi-sentencial é feita de forma similar ao trabalho de [Liu et al. \(2015\)](#), mesclando-se os nós que representam os mesmos conceitos (possuem o mesmo rótulo). Esse processo assume que rótulos iguais se referem a um mesmo conceito, dispensando o uso de métodos de desambiguação de sentido; as arestas são mantidas intactas durante o processo. Em seguida, as raízes dos grafos originais são todas conectadas a uma nova raiz do tipo **multi-sentence**. Apesar de aparentemente não ser preciso, foi mantido o procedimento adotado por [Dohare, Gupta e Karnick \(2018\)](#) de se escolher o maior componente conexo como a representação final do documento, a fim de garantir a conectividade do grafo. Essa etapa é necessária, porém, apenas caso ocorram erros durante a mesclagem dos nós ou por alguma má formação dos grafos de entrada.

Figura 31 – Grafos AMR de sentenças a serem mescladas em um único grafo multi-sentencial



(a) "O melhor livro de Sidney Sheldon"



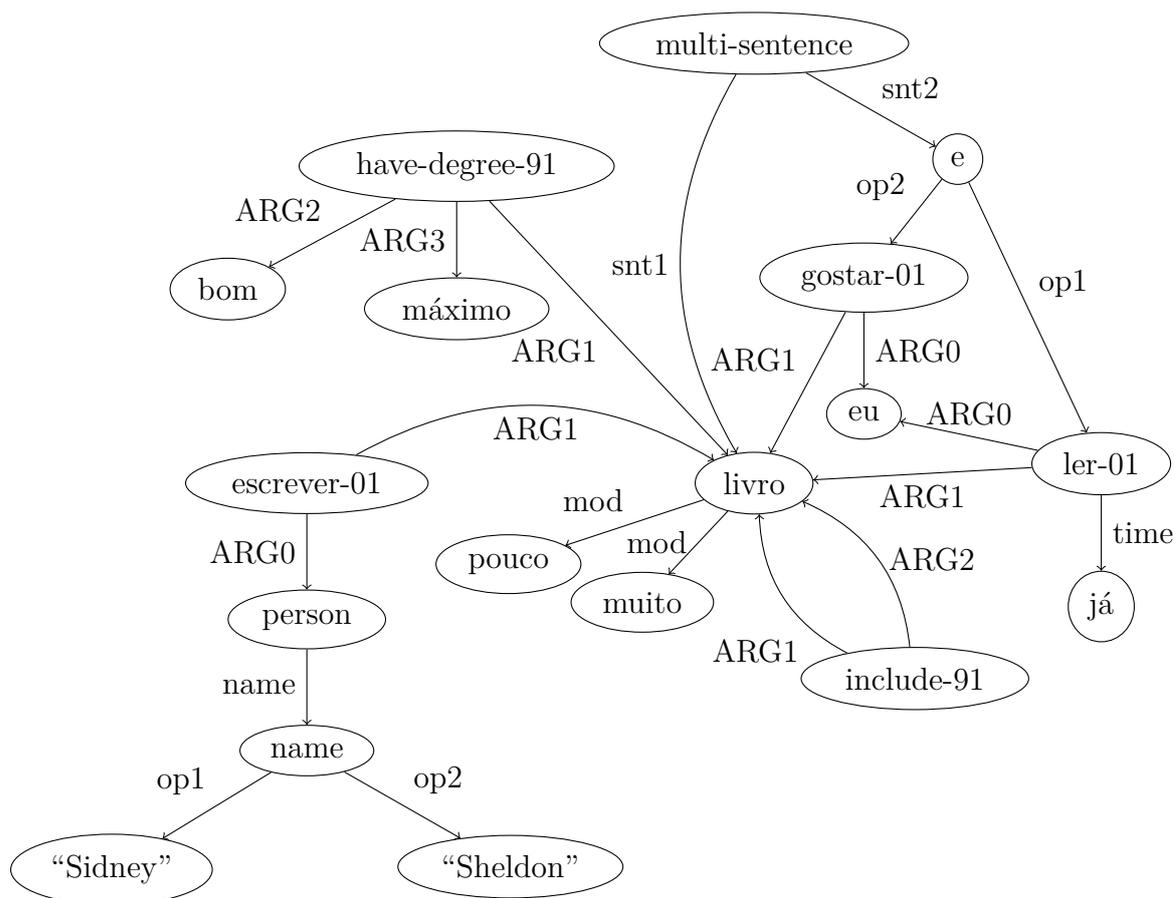
(b) "Já li muitos livros de Sidney Sheldon e gostei de poucos"

Fonte: Elaborada pelo autor.

Como entidades nomeadas e datas são representadas em AMR por múltiplos nós, essas informações foram primeiro colapsadas em um único nó antes de se prosseguir com a união, de modo a permitir que dois vértices sejam mesclados apenas se se referirem

a uma mesma entidade. Ao final do processo, esses nós são descolapsados. No exemplo apresentado, a entidade nomeada “Sidney Sheldon” seria representada por um único nó “NE:person.name."Sidney"."Sheldon"”, encapsulando o tipo de entidade e seu nome, permitindo a mesclagem correta dos grafos. O resultado final da mesclagem, com os nós já descolapsados pode ser visto na [Figura 32](#).

Figura 32 – Grafo AMR multi-sentencial final



Fonte: Elaborada pelo autor.

A partir do grafo AMR multi-sentencial é possível prosseguir com a sumarização, conforme os métodos apresentados a seguir.

5.2 Métodos de Sumarização Semântica

Os métodos de sumarização semântica investigados neste trabalho podem ser organizados em três frentes principais: baseados em Programação Linear Inteira (ILP, do inglês *Integer Linear Programming*) (LIU *et al.*, 2015; LIAO; LEBANOFF; LIU, 2018), baseados em regras (DOHARE; GUPTA; KARNICK, 2018) e baseados em Aprendizado de Máquina (AM). Foram implementados os métodos originais da literatura, assim como algumas variações e outros métodos novos derivados.

5.2.1 Sumarização baseada em Programação Linear Inteira

Os métodos de sumarização de [Liu *et al.* \(2015\)](#) e [Liao, Lebanoff e Liu \(2018\)](#) focam no uso de Programação Linear Inteira, um método de otimização combinatória, para selecionar as informações (conceitos e relações) mais relevantes de um grafo AMR.

Um problema de otimização tem como base três aspectos principais: um conjunto de variáveis, uma função objetivo sobre essas variáveis a ser otimizada (maximizada ou minimizada) e um conjunto de restrições que devem ser satisfeitas a fim de garantir que uma determinada solução (atribuição de valores às variáveis) é válida ([SIERKSMA; ZWOLS, 2015](#)). Para problemas lineares, tanto a função objetivo quanto as restrições são expressões lineares. Já no caso mais específico de problemas lineares inteiros, os valores atribuídos às variáveis são sempre inteiros.

Em termos formais, um problema linear inteiro pode ser caracterizado pela função objetivo linear genérica apresentada na [Equação 5.1](#), com n variáveis x_j e constantes c_j . A solução está sujeita a um conjunto de m restrições (inequações lineares) com a forma geral apresentada na [Equação 5.2](#), sendo a_{ij} e b_i constantes. As variáveis também precisam ser, necessariamente, inteiras positivas, como é indicado pela [Equação 5.3](#) (restrição de integralidade).

$$\text{Otimizar } \sum_{j=1}^n c_j x_j \quad (5.1)$$

$$\text{sujeito a } \sum_{j=1}^n a_{ij} x_j \geq b_i, \forall i \in \{1, \dots, m\} \quad (5.2)$$

$$x_j \in \mathbb{Z}_+, \forall j \in \{1, \dots, n\} \quad (5.3)$$

Para o processo de sumarização, [Liu *et al.* \(2015\)](#) desenvolveram uma forma de representar os grafos AMR, seus conceitos e suas relações através de vetores de atributos binários. A lista completa de atributos utilizados para representar os nós de um grafo pode ser vista no [Quadro 2](#).

Como pode ser visto no quadro, a maioria dos atributos é numérico e são discretizados seguindo a [Equação 5.4](#). Um atributo x é mapeado para N outros atributos binários y_i de acordo com alguns pontos de corte k_i pré-definidos. Não há, porém, nenhuma indicação dos autores de qual foi o critério utilizado para se determinar esses valores.

$$x \mapsto y_i = \begin{cases} 1, & \text{se } x > k_i \\ 0, & \text{caso contrário} \end{cases}, i = 1, \dots, N \quad (5.4)$$

Mais adiante, [Liao, Lebanoff e Liu \(2018\)](#) incorporaram mais um atributo referente aos conceitos: o valor de TF-IDF das palavras alinhadas ao nó. Por não haver nenhuma

Quadro 2 – Atributos utilizados por Liu *et al.* (2015) e Liao, Lebanoff e Liu (2018) para representar conceitos AMR

Atributo	Pontos de corte	Observações
Conceito		Representação <i>one-hot</i>
Frequência do conceito	0, 1, 2, 5, 10	Frequência absoluta de ocorrência do conceito sobre todas as sentenças de entrada
Profundidade mínima do conceito	1, 2, 3, 4, 5	Menor distância do conceito à raiz da sentença, sobre todas as sentenças de entrada
Profundidade média do conceito	1, 2, 3, 4, 5	Distância média dos nós com o conceito à raiz de suas respectivas sentenças
Posição mais inicial do conceito nos textos	5, 6, 7, 10, 15	Entre as sentenças contendo o conceito, a menor posição em seus respectivos documentos
Posição média do conceito nos textos	5, 6, 7, 10, 15	Posição média das sentenças contendo o conceito dentro de seus respectivos documentos
Maior tamanho de alinhamento	0, 1, 2, 5, 10	Tamanho máximo de tokens consecutivos alinhados ao conceito
Tamanho médio de alinhamento	0, 1, 2, 5, 10	Tamanho médio de tokens consecutivos alinhados ao conceito
Entidade nomeada		Atributo binário indicando se o nó representa uma entidade nomeada
Data		Atributo binário indicando se o nó representa uma entidade de data
<i>Bias</i>		Sempre 1

Fonte: Adaptada de Liu *et al.* (2015).

indicação dos pontos de corte utilizados originalmente para esse atributo em específico, neste projeto escolheu-se utilizar como base os valores dos quartis sobre os dados de treinamento. Os valores dos quartis obtidos foram: $Q_1 = 0$; $Q_2 = 0$; $Q_3 = 3$. Também foi utilizado, como limite inferior, o valor que dista do primeiro quartil em 150% do intervalo interquartil (MORETTIN; BUSSAB, 2010), dessa forma: $Q_1 - 1,5(Q_3 - Q_1) = -4,5$. Similarmente o limite superior obtido é: $Q_3 + 1,5(Q_3 - Q_1) = 7,5$. Ao final, os valores de corte para esse atributo em específico são: -4,5, 0, 3 e 7,5.

Similarmente aos nós, as relações também são representadas por atributos específicos, que podem ser vistos no Quadro 3. Os atributos dizem respeito a cada par de nós relacionados por uma ou mais arestas no grafo e não para cada relação em si (arestas).

A partir da definição dos atributos a serem utilizados, tem-se a determinação da, assim chamada, representação local $\phi(x)$ de um elemento x (um nó ou um par de nós)

Quadro 3 – Atributos utilizados por Liu *et al.* (2015) e Liao, Lebanoff e Liu (2018) para representar relações entre pares de nós AMR

Atributo	Pontos de corte	Observações
Frequência da relação mais comum entre o par de nós	0,5, 0,66, 0,75	Frequência relativa.
Frequência da segunda relação mais comum entre o par de nós	0,5, 0,66, 0,75	
Número de arestas entre o par de nós	0, 1, 2, 5, 10	
Posição mais inicial da relação nos textos	5, 6, 7, 10, 15	Entre as sentenças contendo os dois nós relacionados, a menos posição em seus respectivos documentos
Posição média da relação nos textos	5, 6, 7, 10, 15	Posição média das sentenças contendo os dois nós relacionados dentro de seus respectivos documentos
Atributos dos nós em cada ponta da relação		Todos os atributos dos dois nós relacionados, conforme apresentados anteriormente
Viés		Sempre 1

Fonte: Adaptada de Liu *et al.* (2015).

segundo a Equação 5.5, sendo n_i cada um dos atributos apresentados no Quadro 2 e e_j cada um dos atributos do Quadro 3, de modo que I e J indicam o número de atributos referentes aos conceitos e às relações, respectivamente.

$$\phi(x) = [n_1, n_2, \dots, n_I, e_1, e_2, \dots, e_J] \quad (5.5)$$

A partir da representação local de cada elemento do grafo, é possível construir uma representação global $\Phi(G)$ de um grafo G completo, isso é feito por meio da Equação 5.6.

$$\Phi(G) = \sum_{v \in V} \phi(v) + \sum_{e \in V \times V} \phi(e) \quad (5.6)$$

Essas representações são utilizadas no processo de sumarização por meio de ILP proposto por Liu *et al.* (2015). Com isso, a função objetivo da sumarização é a apresentada na Equação 5.7, sendo \mathbf{w} um vetor de constantes (pesos) para cada atributo da representação global do grafo de entrada G . Portanto, o processo de sumarização é encontrar um subgrafo que maximiza a função objetivo segundo os pesos dados.

A formalização dessa função, no contexto de ILP, é dada pela última parte da igualdade, que inclui as variáveis binárias x_i e y_{ij} que compõem a solução. A variável x_i

indica a seleção de um determinado nó v_i dentre o total de N nós do grafo, já y_{ij} representa a seleção das relações entre certo par de nós dentre os possíveis ($E = V \times V$).

$$\text{score}(G) = \mathbf{w} \cdot \Phi(G) = \sum_{i=1}^N x_i \mathbf{w} \cdot \phi(v_i) + \sum_{(i,j) \in E} y_{ij} \mathbf{w} \cdot \phi(e_{ij}) \quad (5.7)$$

Para garantir a qualidade da sumarização, os pesos da função objetivo passam por um processo de otimização próprio, de modo a guiar a seleção do subgrafo. Por se tratar de uma forma de aprendizado supervisionado, esse passo deve ser feito em um outro corpus específico de treinamento para evitar vieses nos resultados.

A otimização dos pesos é feita a partir do algoritmo AdaGrad (DUCHI; HAZAN; SINGER, 2011), apresentado na Equação 5.8. No AdaGrad, os novos pesos \mathbf{w}_{t+1} são dependentes dos pesos anteriores \mathbf{w}_t e também de um termo de atualização baseado na diferença $\Delta(G, G^*)$, calculada segundo uma função de erro pré-determinada sobre o grafo de entrada (G) e seu respectivo grafo de referência G^* (o resumo). A velocidade da atualização é controlada por um parâmetro $\eta \in (0, 1]$ e ε é um valor positivo para se evitar a divisão por 0.

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{\eta}{\sqrt{\Delta(G, G^*) \cdot \Delta(G, G^*) + \varepsilon}} \Delta(G, G^*) \quad (5.8)$$

Em seu trabalho, Liu *et al.* (2015) propuseram três diferentes funções de erro para o cálculo da diferença necessária para a atualização. A função mais simples é a *Perceptron Loss* indicada na Equação 5.9, em que Φ é a representação global do grafo apresentada anteriormente, G^* é o grafo de referência do resumo e G' representa o melhor grafo que se obtém com os pesos atuais \mathbf{w}_t , como indicado pela Equação 5.10. A busca pelo grafo G' é feita utilizando o próprio método de ILP sem os pesos atualizados, isto é, utilizando-se \mathbf{w}_t e não \mathbf{w}_{t+1} .

$$\Delta(G, G^*) = \Phi(G') - \Phi(G^*) \quad (5.9)$$

$$G' = \underset{G}{\operatorname{argmax}} \{ \mathbf{w}_t \cdot \Phi(G) \} \quad (5.10)$$

Uma segunda função de erro apresentada pelos autores é a *Hinge Loss*, que possui uma estrutura similar à Equação 5.9, porém com uma diferença na busca pelo grafo G' , realizada segundo a Equação 5.11, que leva em consideração um custo associado ao grafo escolhido. Esse custo é determinado pela Equação 5.12, de forma que cada elemento selecionado em G que não esteja também presente em G^* diminua a pontuação do referido grafo, tornando esses elementos mais difíceis de serem selecionados.

$$G' = \operatorname{argmax}_G \{ \mathbf{w} \cdot \Phi(G) - \operatorname{cost}(G, G^*) \} \quad (5.11)$$

$$\operatorname{cost}(G, G^*) = |G \setminus (G \cap G^*)| \quad (5.12)$$

A última função de erro aplicada por Liu *et al.* (2015) é chamada *Ramp Loss* e é descrita pela Equação 5.13. Nessa função, é utilizado o mesmo conceito de custo da *Hinge Loss*, porém a representação do grafo de referência é substituída por uma segunda busca por um grafo G'' caracterizada pela Equação 5.14. Nessa busca, o custo associado ao grafo tende, na verdade, a recompensar a escolha dos elementos em comum entre os grafos.

$$\Delta(G, G^*) = \Phi(G') - \Phi(G'') \quad (5.13)$$

$$G'' = \operatorname{argmax}_G \{ \mathbf{w} \cdot \Phi(G) + \operatorname{cost}(G, G^*) \} \quad (5.14)$$

Após a definição da função objetivo a ser otimizada (Equação 5.7), é necessário determinar as restrições de validade das soluções. O primeiro conjunto de restrições, apresentado na Equação 5.15, garante que, ao selecionar as arestas entre um par de nós (y_{ij}), os nós relacionados (x_i e x_j) devem ser escolhidos também.

$$x_i - y_{ij} \geq 0, \quad x_j - y_{ij} \geq 0, \quad \forall (i, j) \in E \quad (5.15)$$

Caso haja arestas em direções opostas entre o par de nós (y_{ij} e y_{ji}), deve-se escolher apenas uma direção (caso o par seja escolhido).

$$y_{ij} + y_{ji} \leq 1, \quad \forall (i, j) \in E \quad (5.16)$$

Também foi definido um conjunto de restrições para garantir a conectividade do subgrafo selecionado. Essas restrições se baseiam em um conjunto de variáveis (inteiras e não binárias) de fluxo f_{ij} existente para cada par de nós no grafo (independente se existirem conexões entre eles). O fluxo parte da raiz do grafo (x_0) em direção a todos os outros nós selecionados (Equação 5.17), enviando-se até N unidades de fluxo, com N sendo o número de nós no grafo. Cada nó selecionado (x_j) consome uma unidade de fluxo (Equação 5.18), isso é verificado ao se comparar o fluxo entrante no nó (f_{ij}) com o que sai dele (f_{jk}). Por último, a Equação 5.19 garante que o fluxo seja transmitido apenas entre os pares de nós selecionados.

$$\sum_{i=0}^N f_{0i} - \sum_{i=1}^N x_i = 0 \quad (5.17)$$

$$\sum_{i=0}^N f_{ij} - \sum_{k=1}^N f_{jk} - x_j = 0, \quad \forall j \in [1, N] \quad (5.18)$$

$$N \cdot y_{ij} - f_{ij} \geq 0, \quad \forall i \leq N \quad \forall j \in [1, N] \quad (5.19)$$

O método também exige que o subgrafo selecionado tenha o formato de árvore, ignorando reentrâncias, característica presente em grafos AMR.

$$\sum_{i=0}^N y_{ij} \leq 1, \quad \forall j \leq N \quad (5.20)$$

Com a função objetivo e as restrições definidas, é possível utilizar um resolvidor de ILP para encontrar o subgrafo que melhor resume a entrada. Para mais detalhes sobre o funcionamento geral desse tipo de resolvidor indica-se o estudo do algoritmo de *branch-and-bound* em [Sierksma e Zwols \(2015, p.283\)](#).

Em uma continuação do trabalho de [Liu *et al.* \(2015\)](#), [Liao, Lebanoff e Liu \(2018\)](#) incorporaram um passo anterior à otimização no processo de sumarização: o agrupamento. Os autores utilizam o método de *Spectral Clustering* para pré-selecionar as sentenças a serem utilizadas como entrada para o ILP.

No método de *Spectral Clustering*, n elementos x_i (sentenças) são representados como nós em um grafo cujas arestas são ponderadas de acordo com uma medida de similaridade, resultando em uma matriz de adjacência $W \in \mathbb{R}^{n \times n}$. A partir de W , bem como da matriz de graus D , diagonal, em que D_i representa o número de nós conectados ao nó x_i , é possível construir a matriz Laplaciana $L = D - W$. Dessa matriz, como indicado por [von Luxburg \(2007\)](#), é possível extrair os k autovetores $\mathbf{u}_j \in \mathbb{R}^n$ com menores autovalores associados, sendo k o número de grupos desejado.

Os autovetores calculados são utilizados para criar a matriz $U \in \mathbb{R}^{n \times k}$, sendo \mathbf{u}_j uma coluna de U , de forma que sua i -ésima linha é denominada y_i . Essa matriz representa um mapeamento de representação dos elementos x_i para pontos $y_i \in \mathbb{R}^k$, que condensam um grau de informação mais facilmente identificável para o agrupamento dos dados. Dessa forma, os pontos y_i podem, então, ser agrupados utilizando-se um algoritmo clássico de *k-means*.

Visto que a técnica de *Spectral Clustering* se baseia em uma determinada medida de similaridade entre os elementos, [Liao, Lebanoff e Liu \(2018\)](#) exploram 4 medidas diferentes: *Longest Common Subsequence* (LCS), *Vector Space Models* (VSM), *Smatch* ([CAI; KNIGHT, 2013](#)) e *Concept Coverage*. Apesar de ter obtido os melhores resultados na

sumarização das arestas (e segunda melhor pontuação na predição de nós) utilizando a medida de VSM, não existem informações divulgadas suficientes para reproduzir, com razoável confiança, esses experimentos. Portanto, foram exploradas neste trabalho apenas as outras três maneiras de se calcular a similaridade entre sentenças.

A primeira forma de medição de similaridade entre sentenças é a LCS, definida como a quantidade de caracteres em comum entre duas sentenças. Por se tratar originalmente de um método de correspondência de *strings*, os espaços em branco também são considerados no cálculo. O valor de similaridade entre “O menino quer brincar” e “O menino quer o biboquê”, por exemplo, é de 16.

A medida Smatch desenvolvida por Cai e Knight (2013), por sua vez, é focada na avaliação do grau de semelhança entre dois grafos AMR através do cálculo de precisão e cobertura sobre triplas lógicas derivadas do grafo. Essas triplas possuem dois formatos básicos: *relação(variável, conceito)*, para a instanciação dos conceitos, ou *relação(variável, variável)*, representando as arestas do grafo. Um exemplo de mapeamento para forma lógica da sentença “O menino quer brincar” pode ser visto na Figura 33, em que cada conceito é instanciado para suas respectivas variáveis através das relações *instance*. Elas são, então, usadas como referências na representação das relações semânticas (*ARG0* e *ARG1*).

Figura 33 – Forma lógica correspondente a um grafo AMR referente à sentença “O menino quer brincar”

$$\begin{array}{l}
 (q / \text{querer-01} \\
 :ARG0 (m / \text{menino}) \\
 :ARG1 (b / \text{brincar-01} \\
 :ARG0 m))
 \end{array}
 \rightarrow
 \begin{array}{l}
 \text{instance}(q, \text{querer-01}) \wedge \\
 \text{instance}(m, \text{menino}) \wedge \\
 \text{instance}(b, \text{brincar-01}) \wedge \\
 ARG0(q, m) \wedge \\
 ARG1(q, b) \wedge \\
 ARG0(b, m)
 \end{array}$$

Fonte: Adaptada de Cai e Knight (2013).

Para se comparar as triplas da sentença anterior com as da sentença “O menino quer o bilboquê”, apresentada na Figura 34, é necessário encontrar um mapeamento entre as variáveis utilizadas, pois elas são arbitrárias e podem variar entre as anotações.

Figura 34 – Forma lógica correspondente à sentença “O menino quer o bilboquê”

$$\begin{array}{l}
 (x / \text{querer-01} \\
 :ARG0 (y / \text{menino}) \\
 :ARG1 (z / \text{bilboquê}))
 \end{array}
 \rightarrow
 \begin{array}{l}
 \text{instance}(x, \text{querer-01}) \wedge \\
 \text{instance}(y, \text{menino}) \wedge \\
 \text{instance}(z, \text{bilboquê}) \wedge \\
 ARG0(x, y) \wedge \\
 ARG1(x, z)
 \end{array}$$

Fonte: Adaptada de Cai e Knight (2013).

A medida de Smatch entre os grafos é definida como sendo a maior medida F (em

termos de triplas lógicas) obtida dentre todos os alinhamentos possíveis. Isso pode ser verificado na Tabela 17, de forma que o melhor mapeamento ($x=q$, $y=m$, $z=b$) resulta em uma pontuação Smatch de 0,73.

Tabela 17 – Medida F obtida por cada alinhamento de variáveis possível

	Correspondências	Precisão	Cobertura	Medida F
$x=q$, $y=m$, $z=b$	4	4/5	4/6	0,73
$x=q$, $y=b$, $z=m$	1	1/5	1/6	0,18
$x=m$, $y=q$, $z=b$	0	0/5	0/6	0,00
$x=m$, $y=b$, $z=q$	0	0/5	0/6	0,00
$x=b$, $y=q$, $z=m$	0	0/5	0/6	0,00
$x=b$, $y=m$, $z=q$	2	2/5	2/6	0,36
Medida Smatch				0,73

Fonte: Adaptada de Cai e Knight (2013).

A última medida de similaridade utilizada por Liao, Lebanoff e Liu (2018) é a de cobertura de conceitos. Dessa forma, a partir do exemplo anterior, temos que, dentre os 3 conceitos existentes (*querer-01*, *menino*, *brincar-01*), 2 estão presentes em ambos os grafos (*querer-01* e *menino*). Portanto, a pontuação de cobertura é $\frac{2}{3} \approx 0,67$.

5.2.2 Sumarização baseada em regras

Dohare, Gupta e Karnick (2018) focam na criação de um resumo através da seleção de um subgrafo AMR por meio de regras. Primeiramente, cada nó (conceito) do grafo tem sua importância mensurada através da medida TF-IDF, apresentada na Equação 5.21, na qual t representa um termo em um documento d , sendo D um corpus maior com grande quantidade de textos. A pontuação depende de dois fatores: *Term Frequency* – baseado na contagem $c(t, d)$ (quantas vezes t aparece em d) – e *Inverse Document Frequency*, baseado na quantidade de documentos em D que possuem t . A partir dessa formulação é possível atribuir um grau de importância maior àqueles termos que são usados com maior frequência, porém que não são comuns a todos os textos.

$$\text{TF-IDF}(t; d, D) = \frac{c(t, d)}{\sum_{t' \in d} c(t', d)} \cdot \log \frac{|D|}{|\{d' \in D : t \in d'\}|} \quad (5.21)$$

Após o ranqueamento, os 10 conceitos mais importantes são selecionados e, para cada par de nós, a primeira sentença em que eles ocorrem simultaneamente, de acordo com a ordem original do corpus, é escolhida para a extração das relações mais relevantes, resultando em um máximo de 45 possíveis sentenças (o que pode não ocorrer, caso determinado par nunca apareça em uma mesma frase). Das sentenças separadas, o caminho mais importante p , entre os conceitos selecionados, é escolhido segundo a Equação 5.22,

sendo d a profundidade (distância à raiz) do nó u . Ou seja, é escolhido o caminho mais próximo da raiz.

$$p = \underset{p'}{\operatorname{argmin}}\{\min d(u) : \forall u \in p'\} \quad (5.22)$$

Feita a seleção dos caminhos mais importantes, eles são expandidos utilizando-se triplas de relações obtidas por meio da ferramenta OpenIE (BANKO *et al.*, 2008), adaptada por Oliveira e Claro (2019) para a língua portuguesa sob o nome DptOIE¹. Os autores indicam que seus resultados foram superiores aos obtidos por outras ferramentas anteriores do mesmo tipo. Um exemplo de saída desta ferramenta é apresentada na Quadro 4, em que três triplas diferentes foram extraídas de uma única sentença, sendo algumas triplas mais complexas, trazendo uma informação mais completa. Dessa maneira, o caminho expandido é escolhido como sendo aquele que compreende a tripla mais longa (com mais *tokens*), sendo que a tripla deve contemplar, necessariamente, todo o caminho original (antes da expansão). Caso não exista uma tripla correspondente, a sentença inteira é selecionada.

Quadro 4 – Exemplo de triplas extraídas pela ferramenta DptOIE

Texto	Entidade 1	Relação	Entidade 2
	ele	mostra	de forma clara, simples e divertida
ele mostra de forma clara, simples e divertida os pensamentos, conflitos e dilemas de um adolescente aparentemente revoltado com o mundo, todas as suas qualidades e defeitos.	ele	mostra	de forma clara, simples e divertida os pensamentos, conflitos e dilemas de um adolescente aparentemente revoltado com o mundo
	ele	mostra	os pensamentos, conflitos e dilemas de um adolescente aparentemente revoltado com o mundo

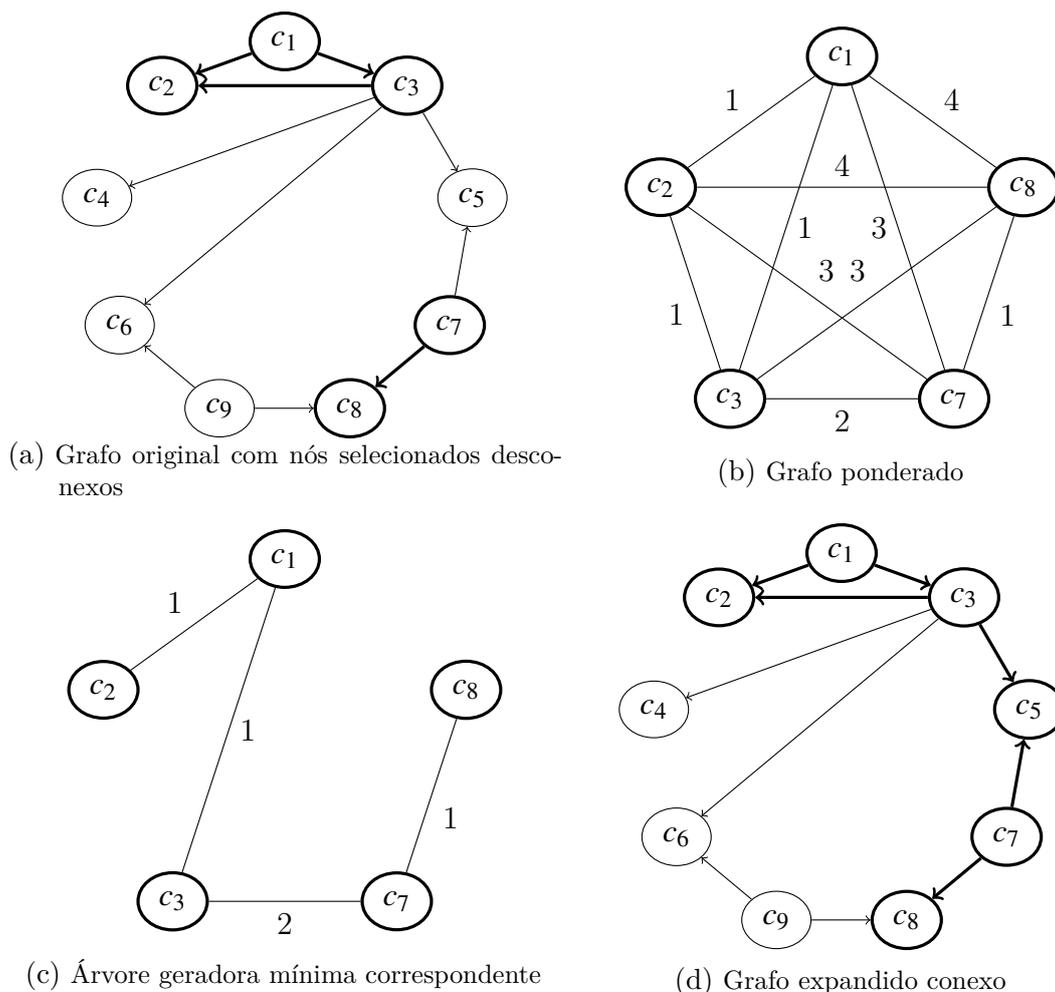
Fonte: Elaborada pelo autor.

Existem casos não tratados no método original de Dohare, Gupta e Karnick (2018) em que a inclusão dos conceitos e arestas não resulta em um único grafo conexo, pré-condição para se ter uma representação AMR. Dessa maneira, é importante garantir essa conectividade por meio de mais uma expansão, o que é realizado por meio de uma Árvore Geradora Mínima. Um exemplo é apresentado na Figura 35, em que o grafo da Figura 35a possui os nós c_1 , c_2 , c_3 , c_7 e c_8 selecionados para compor o resumo, o que resultaria em um grafo desconexo. Assim, constrói-se um grafo ponderado totalmente conexo entre os

¹ Disponível em: <<http://formas.ufba.br/page/downloads>>. Acesso em: 27 abr. 2021.

nós relevantes (Figura 35b) de forma que o peso das arestas corresponde ao tamanho do caminho mínimo entre os dois vértices correspondentes no grafo original (desconsiderando-se as direções); por exemplo, o peso entre c_2 e c_8 é 4, pois o caminho mínimo entre eles é $c_2 \rightarrow c_3 \rightarrow c_5 \rightarrow c_7 \rightarrow c_8$, de tamanho 4.

Figura 35 – Expansão de um grafo usando sua Árvore Geradora Mínima para garantir conectividade



Fonte: Elaborada pelo autor.

A partir do grafo ponderado, pode-se extrair sua Árvore Geradora Mínima (Figura 35c), que irá indicar o menor número de nós que precisamos incluir no caminho original de forma a torná-lo conexo. Nesse exemplo, selecionou-se a aresta entre c_3 e c_7 , com peso 2 correspondente ao caminho $c_3 \rightarrow c_5 \rightarrow c_7$, portanto o nó c_5 também deve ser incluso para garantir a conectividade do grafo de resumo, como mostrado na Figura 35d.

Com os grafos de cada sentença expandidos e conexos, pode-se, então criar o grafo final do resumo mesclando-se cada grafo obtido pelos processos anteriores.

As variações do trabalho de Dohare, Gupta e Karnick (2018) exploradas neste projeto dizem respeito à forma de ranqueamento dos conceitos para a sumarização. A

primeira alteração visa a substituir o uso da medida de TF-IDF pela pontuação TF, visto que é possível que a inclusão do termo IDF ignore expressões que sejam importantes e que, por esse motivo, estejam presentes em diversos documentos sobre um mesmo domínio ou produto. Um exemplo é o termo “bom” – adjetivo que pode ser utilizado para indicar o sentimento do usuário com relação ao produto, o que, por sua vez, é uma informação bastante relevante –, que está presente em 23.440 de um total de 132.373 documentos no córpus B2W-Reviews01 (REAL; OSHIRO; MAFRA, 2019). Uma comparação pode ser feita com o termo “mas”, conjunção que pode não ser tão importante quanto o adjetivo, presente em 20.721 documentos no mesmo córpus. Dessa forma, o termo “bom” seria prejudicado pelo uso da medida TF-IDF, por estar presente em diversos comentários, o que é natural, já que é um termo comum e importante para exprimir sentimento.

Outra forma de pontuação investigada foca na combinação linear de diferentes medidas de centralidade (cálculo da importância dos nós) referente aos grafos, baseado em alguns atributos que Leskovec, Milic-Frayling e Grobelnik (2005) utilizaram para a classificação de nós em um grafo semântico usando técnicas de Aprendizado de Máquina. As medidas utilizadas foram:

Grau de incidência Uma medida simples de importância é o grau de incidência do nó u , definido como a quantidade de arestas conectadas a ele: $d(u) = |I_u|$, sendo I_u o número de vizinhos que possuem conexões apontando para u (em um grafo direcionado).

Grau médio dos vizinhos Também leva-se em consideração um certo nível de informação de como os nós se conectam através do grau médio dos vizinhos de um dado nó u , como indicado na Equação 5.23.

$$\bar{d}(u) = \frac{1}{|N_u|} \sum_{v \in N_u} d(v) \quad (5.23)$$

Centralidade de Grau Essa medida é definida como sendo a fração de vértices no grafo que são conectados ao nó u dado, como indicado na Equação 5.24. Esse valor é, em sua essência, o grau de incidência – apresentado anteriormente – relativo ao tamanho do grafo em número de vértices ($|V|$).

$$c_d(u) = \frac{|I_u|}{|V|} \quad (5.24)$$

Centralidade de Autovetor A medida de Centralidade de Grau assume que cada nó contribui igualmente na importância de seus vizinhos, ignorando que a relevância de um vértice pode aumentar quando conectado a outros nós também importantes (NEWMAN, 2010). Para lidar com esse problema, a Centralidade de Autovetor leva em consideração a importância dos vizinhos de uma forma recursiva, como é mostrado na Equação 5.25, sendo A a matriz de adjacência contendo os pesos das arestas presentes no grafo.

$$c_a(u) = \sum_{v \in I_u} A_{vu} \cdot c_a(v) \quad (5.25)$$

Visto que um grafo AMR não possui originalmente pesos, substituiu-se, para o cálculo da centralidade, quaisquer relações paralelas entre dois nós por uma aresta única com peso correspondente ao número original de arestas. Em um grafo, por exemplo, com dois nós u e v conectados por três relações diferentes r_1 , r_2 e r_3 saindo de v em direção a u , o cálculo da centralidade levará em consideração uma única aresta de peso 3 entre os dois vértices ($A_{vu} = 3$).

Pagerank Esse algoritmo desenvolvido por Brin e Page (1998) também define a importância de um nó recursivamente, conforme a Equação 5.26. O valor de Pagerank atribuído a um nó u depende da importância de cada nó v incidente a u (I_u) ponderado pela quantidade de arestas saindo de v ($|O_v|$), de forma a impedir que um nó importante conectado a muitos nós “distribua” essa relevância indiscriminadamente no grafo. A contribuição desses valores também é controlada por um fator d . Por se tratar de um método recorrente, a pontuação de cada nó é inicializada com um valor fixo e a mesma fórmula é aplicada diversas vezes a fim de se atingir uma estabilidade nos valores, isso é, o valor de Pagerank final.

$$PR(u) = (1 - d) + d \sum_{v \in I_u} \frac{PR(v)}{|O_v|} \quad (5.26)$$

HITS Similarmente ao Pagerank, o algoritmo HITS, desenvolvido por Kleinberg (1999), atua de uma maneira iterativa, associando duas características diferentes aos nós: *hub* e *authority*. Os nós *hub* são aqueles que apontam para uma grande quantidade de nós *authority*. Esses, por sua vez, são aqueles que são referenciados por vários *hubs* em uma relação de reforço mútuo. Dessa forma, as pontuações de *hub* ($h(u)$) e *authority* ($a(u)$) de um vértice u são definidas conforme as Equações 5.27 e 5.28, respectivamente, sendo O_u o conjunto de nós apontados por u e I_u o conjunto de nós incidentes a u .

$$h(u) = \sum_{v \in O_u} a(v) \quad (5.27)$$

$$a(u) = \sum_{v \in I_u} h(v) \quad (5.28)$$

Também por ser um algoritmo iterativo, os valores são todos iniciados uniformemente ao longo do grafo, sendo aplicadas as fórmulas alternadamente até se atingir o equilíbrio.

Closeness Centrality Distanciando-se da definição de importância por meio do número de conexões de um nó, a medida de *Closeness* foca na distância média de um nó

a todos os outros do grafo (NEWMAN, 2010, p. 181), como é mostrado na Equação 5.29, sendo $d(v,u)$ o tamanho do menor caminho entre v e u . Como o uso da distância média resultaria em valores altos para os nós mais distantes dos outros e valores baixos àqueles mais centrais, o valor inverso é utilizado. A distância de um nó a si mesmo também é ignorada, desconsiderando, assim, laços no grafo.

$$c_c(u) = \frac{|V| - 1}{\sum_{v \in V: v \neq u} d(v,u)} \quad (5.29)$$

Betweenness Centrality O valor de *Betweenness* de um vértice u é calculado como sendo a fração de caminhos mínimos entre todos os pares de nós (s,t) do grafo que passam por u , como mostrado na Equação 5.30, em que $\sigma(s,t)$ representa o número de caminhos mínimos entre s e t , dentre os quais $\sigma(s,t|u)$ passam obrigatoriamente por u . Essa medida indica os nós que, se removidos do grafo, causariam uma maior quebra no fluxo de comunicação entre os nós (NEWMAN, 2010, p. 186).

$$c_b(u) = \sum_{s,t \in V} \frac{\sigma(s,t|u)}{\sigma(s,t)} \quad (5.30)$$

Local Clustering Coefficient Similarmente à medida de *betweenness*, esse valor aponta os nós que possuem maior controle do fluxo de informação, porém com foco apenas em sua própria vizinhança, apontando áreas do grafo mais densamente conectadas. Para tanto, define-se esse coeficiente como sendo o número de vizinhos do nó u conectados entre si (triângulos passando por u) com relação ao total de conexões possíveis (NEWMAN, 2010, p.202). Similarmente à Centralidade de Autovetor, construiu-se um grafo ponderado, mas com os pesos normalizados entre 0 e 1, pré-condição para se garantir que o coeficiente seja bem calculado. No caso de um grafo ponderado, o número de triângulos passando por u é substituído pela soma das médias geométricas dos pesos (força) das arestas que compõem cada triângulo, como apresentado na Equação 5.31 (FAGIOLO, 2007).

$$CC(u) = \frac{\frac{1}{2} \sum_{v \neq u} \sum_{w \neq (u,v)} [(A_{uv} + A_{vu})(A_{uw} + A_{wu})(A_{vw} + A_{wv})]^{\frac{1}{3}}}{d^{\text{tot}}(u)(d^{\text{tot}}(u) - 1) - 2d^{\leftrightarrow}(u)} \quad (5.31)$$

Na equação, u possui como vizinhos os nós v e w com pesos armazenados na matriz de adjacência A . O grau total do nó u é definido como o número de arestas conectadas a u (independentemente da direção): $d^{\text{tot}} = |I_u| + |O_u|$. Já $d^{\leftrightarrow}(u)$ representa o número de arestas recíprocas entre u e seus vizinhos, isto é, o número de vértices v que possuem arestas partindo e chegando a u ($u \rightarrow v$ e $v \rightarrow u$).

Profundidade do nó A profundidade de um vértice u é definida como sendo o tamanho do caminho mínimo de u à raiz do grafo, ignorando-se a direção das arestas.

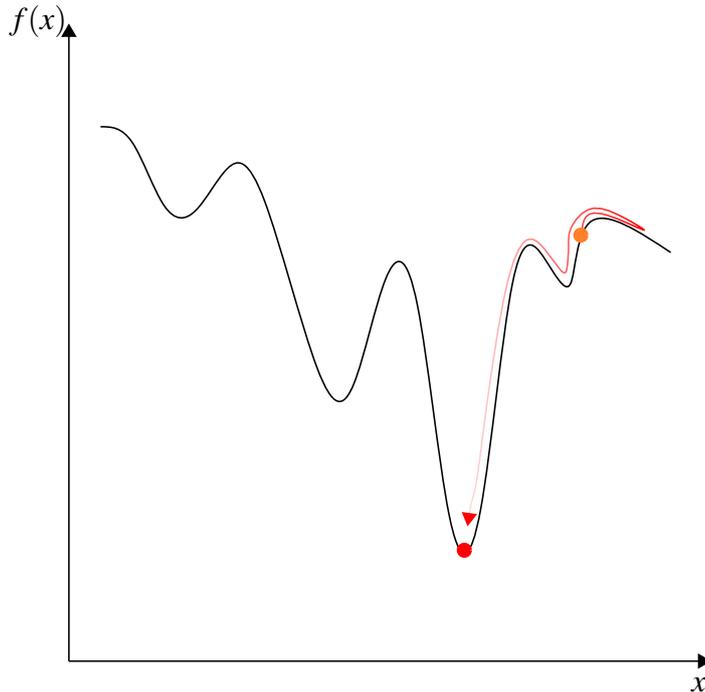
Além das medidas referentes à estrutura do grafo, foram mantidos, para a combinação linear, os valores de TF e TF-IDF dos conceitos. Também foram investigados alguns atributos específicos da área de Análise de Sentimento, visto que este projeto tem como foco os textos opinativos:

Sentimento do conceito A partir do conceito presente no grafo, pode-se utilizar um léxico, como o OpLexicon (SOUZA *et al.*, 2011), para determinar o grau de sentimento de cada vértice do grafo. Os sentimento positivo, negativo ou neutro foram associados a três atributos diferentes, por meio de uma codificação *one-hot*, em que as três posições contêm valores binários, um para cada sentimento. O conceito “pífio”, por exemplo, possui uma polaridade negativa segundo o léxico, resultando em três atributos de valores 0, 1 e 0 (positivo, negativo e neutro, respectivamente). Já “destemido”, por possuir um sentimento positivo, tem como valores associados: 1, 0 e 0.

Aspectos Também adicionou-se um atributo booleano indicando se o nó representa um aspecto do produto ou não, como por exemplo, a bateria ou a câmera de um *smartphone*. Esses aspectos são obtidos pela anotação manual presente no corpus OpiSums-PT utilizado (CONDORI *et al.*, 2015).

Definidos os atributos a serem utilizados para o cálculo da importância dos vértices, os pesos para a combinação desses valores devem ser otimizados de modo a atribuir pontuações mais altas aos nós com maiores chances de compor os resumos. Essa otimização pode ser feita utilizando-se diferentes algoritmos e abordagens. Uma dessas abordagens é através do método de *Simulated Annealing* (KIRKPATRICK; GELATT; VECCHI, 1983 apud LAARHOVEN; AARTS, 1987), também já explorado na otimização tanto de estruturas quanto de pesos para Redes Neurais (LUDERMIR; YAMAZAKI; ZANCHETTIN, 2006).

Simulated Annealing é uma metaheurística probabilística de busca local baseada no conceito de temperatura. O sistema inicia com uma temperatura alta permitindo que a busca resulte em soluções piores do que a anterior, e, conforme as iterações se passam, essa temperatura abaixa, forçando a busca a seguir o gradiente da função a ser otimizada (sempre encontrar uma solução melhor). Esse processo é ilustrado na Figura 36, na qual se quer minimizar uma função $f(x)$, com solução inicial representada pelo círculo laranja. Conforme as iterações de busca são executadas, novas soluções são encontradas (representadas pela seta), porém, devido à temperatura alta do sistema, essas soluções têm uma chance de serem piores do que as anteriores (geram valores mais altos em $f(x)$), porém, em temperaturas mais baixas, essa tendência não é tão presente, levando a solução a seguir o gradiente de $f(x)$ em direção a um mínimo (ou máximo, nos casos de maximização) local, representado pelo círculo vermelho. Essa técnica de busca foi criada para tentar escapar dos mínimos (ou máximos) locais da função, aumentando as chances de encontrar soluções melhores.

Figura 36 – Ilustração de funcionamento do método de *Simulated Annealing*

Fonte: Elaborada pelo autor.

Formalmente, o *Simulated Annealing* é baseado na distribuição de Boltzmann, descrita pela [Equação 5.32](#) (KIRKPATRICK; GELATT; VECCHI, 1983), em que a probabilidade do sistema se encontrar em um estado com nível de energia E é dependente da temperatura T , sendo $Z(T)$ um termo de normalização dependente dessa temperatura e k_B constante. Conforme a temperatura do sistema diminui, as probabilidades da distribuição se concentram nos estados com menor energia (E baixo).

$$P(\mathbf{E} = E) = \frac{1}{Z(T)} \cdot \exp\left(-\frac{E}{k_B T}\right) \quad (5.32)$$

No caso da metaheurística de busca, a energia do sistema é definida como $E = f(\boldsymbol{\omega}) - f(\mathbf{w})$, sendo $\boldsymbol{\omega}$ a melhor dentre as novas soluções obtidas a partir de \mathbf{w} . A função objetivo $f(\mathbf{w})$ é a que se busca otimizar durante a busca.

Neste trabalho, a função objetivo é definida como sendo a cobertura dos conceitos selecionados pelo processo de sumarização com relação aos conceitos presentes em um resumo de referência criado manualmente. Com os pesos \mathbf{w} , é possível calcular uma pontuação s para cada conceito representado por um vetor $\mathbf{c} \in \mathbb{R}^N$ de atributos: $s(\mathbf{c}|\mathbf{w}) = \mathbf{w} \cdot \mathbf{c}$. Essas pontuações são utilizadas, substituindo o valor de TF-IDF original, para seleção dos nós importantes no grafo seguindo as regras apresentadas anteriormente, gerando o resumo automático S com referência R . Portanto a função é definida de acordo com a

Equação 5.33, sendo $\sigma(S, R)$ o número de conceitos em comum a S e R e $|R|$ o número de conceitos em R .

$$f(S, R) = \frac{\sigma(S, R)}{|R|} \quad (5.33)$$

A geração de novas soluções \mathbf{w}' a partir de uma solução \mathbf{w} dada é realizada seguindo a Equação 5.34, sendo α um vetor de números aleatórios provenientes de uma distribuição uniforme no intervalo de $[-1.0, +1.0]$ (LUDERMIR; YAMAZAKI; ZANCHETTIN, 2006).

$$\mathbf{w}' = \mathbf{w} + \alpha \quad (5.34)$$

O controle de resfriamento do sistema utilizado é o mesmo aplicado por Ludermir, Yamazaki e Zanchettin (2006), apresentado na Equação 5.35. Na fórmula, a temperatura T_i na iteração i é definida de acordo com um fator de resfriamento r . Esse resfriamento é realizado a cada I_T iterações, até um máximo de I_{\max} iterações.

$$T_i = \begin{cases} rT_{i-1} & \text{se } i = kI_T, k = 1, 2, \dots, \frac{I_{\max}}{I_T} \\ T_{i-1} & \text{caso contrário} \end{cases} \quad (5.35)$$

A partir dessas definições, pode-se construir o Algoritmo 1 utilizado para a otimização por *Simulated Annealing*.

5.2.3 Sumarização baseada em Aprendizado de Máquina

A seleção dos conceitos relevantes no grafo também pode ser realizada pelo viés do Aprendizado de Máquina, interpretada como sendo um problema de classificação binária (relevante ou não). Para tanto, os atributos utilizados anteriormente (medidas de centralidade, profundidade, sentimento, aspecto, etc.) podem ser aproveitados para se representar os nós e, assim, permitir o treinamento dos algoritmos de AM escolhidos para investigação neste trabalho: Árvore de Decisão, *Random Forest*, SVM e *Multilayer Perceptron* (MLP). O trabalho original de Leskovec, Milic-Frayling e Grobelnik (2005) utiliza, para a classificação, apenas um modelo de SVM, porém neste projeto foram incluídos os métodos baseados em árvores – mais simples e interpretáveis – bem como a rede neural MLP, um método mais comum recentemente.

Árvores de Decisão (em inglês *Decision Trees*) se baseiam no particionamento do espaço de entrada por múltiplos passos de decisão sobre os atributos disponíveis, como pode ser visto no exemplo da Figura 37, sendo x_i os atributos com pontos de corte θ_j . Esses passos são definidos de maneira a melhor separar as classes do problema, o que pode ser mensurado por medidas como entropia cruzada ou o índice Gini (BISHOP, 2006, p. 666).

Algoritmo 1 – *Simulated Annealing***Entrada:**

- $D \in \mathbb{R}^{|V| \times N}$: Dados de treinamento ($|V|$ nós com N atributos cada)
- $\mathbf{w}_0 \in \mathbb{R}^N$: Pesos iniciais
- $I_{\max} \in \mathbb{Z}_+^*$: Número máximo de iterações
- $K \in \mathbb{Z}_+^*$: Número de soluções novas a buscar em uma iteração
- $T_0 \in \mathbb{R}_+^*$: Temperatura inicial
- $r \in (0, 1.0)$: Fator de resfriamento
- $I_T \in \mathbb{Z}_+^*$: Frequência de resfriamento (em iterações)

Saída:

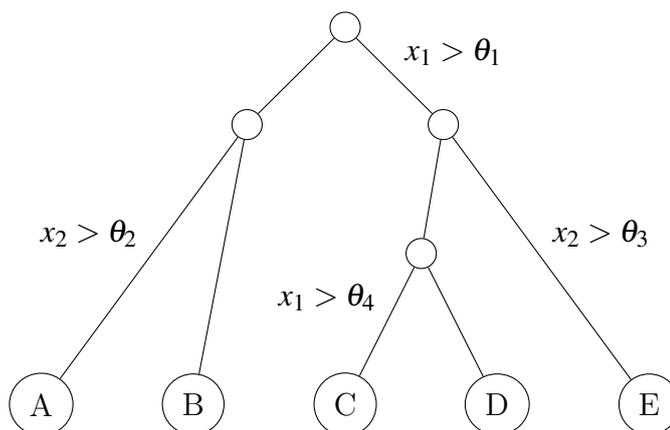
$\mathbf{w} \in \mathbb{R}^N$: Pesos otimizados

- 1: **procedimento** SIMULATED__ANNEALING($D, \mathbf{w}_0, I_{\max}, K, T_0, r, I_T$)
- 2: $T_i \leftarrow T_0$ ▷ Temperatura atual
- 3: $\mathbf{w} \leftarrow \mathbf{w}_0$ ▷ Solução atual
- 4: **para** $i = 0$ até $I_{\max} - 1$ **faça**
- 5: **se** $(i + 1) \bmod I_T = 0$ **então** ▷ Resfriamento (Equação 5.35)
- 6: $T_i \leftarrow rT_{i-1}$
- 7: **senão**
- 8: $T_i \leftarrow T_{i-1}$
- 9: **fim se**
- 10: $W' \leftarrow \{\mathbf{w}'_k : k = 1, \dots, K\}$ ▷ K possíveis soluções novas (Equação 5.34)
- 11: $\omega \leftarrow \underset{\mathbf{w}' \in W'}{\operatorname{argmax}}\{f(\mathbf{w}'|D)\}$ ▷ Encontrar melhor solução nova
- 12: **se** $f(\omega|D) > f(\mathbf{w}|D)$ **então** ▷ Sempre ir a uma solução melhor
- 13: $\mathbf{w} \leftarrow \omega$
- 14: **senão**
- 15: $p \leftarrow \exp\left(\frac{f(\omega|D) - f(\mathbf{w}|D)}{T_i}\right)$ ▷ Probabilidade de ir a uma solução pior
- 16: **se** $p > \beta$ **então** ▷ $\beta \in [0, 1]$ aleatório
- 17: $\mathbf{w} \leftarrow \omega$
- 18: **fim se**
- 19: **fim se**
- 20: **fim para**
- 21: **return** \mathbf{w}
- 22: **fim procedimento**

Fonte: Adaptada de [Ludermir, Yamazaki e Zanchettin \(2006\)](#).

Outro método investigado é o de *Random Forest*, definido como uma combinação de diversas Árvore de Decisão ([DENISKO; HOFFMAN, 2018](#)). Essa técnica se baseia em construir um novo conjunto de dados X_B a partir do conjunto de treinamento X , retirando-se aleatoriamente N atributos dos elementos de X . Dessa forma, um classificador de Árvore de Decisão é treinado sobre X_B , sem olhar para todos os atributos presentes no conjunto de dados. Esse processo é repetido L vezes, de forma que cada classificador treinado perceba mais nuances nos atributos, prevenindo também o *overfitting*. Ao final, a classificação de um novo item é dada pela combinação dos resultados de todas as L árvores treinadas.

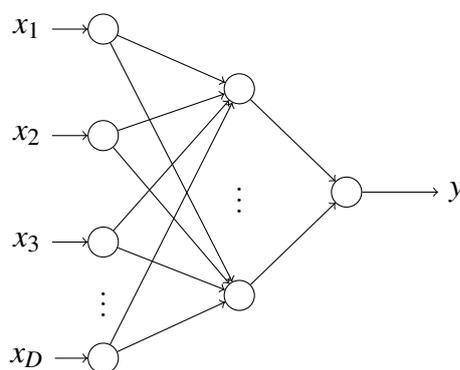
Figura 37 – Exemplo de Árvore de Decisão



Fonte: Adaptada de Bishop (2006, p. 664).

Uma outra técnica de AM explorada é a de SVM (do inglês *Support Vector Machine*), que se baseia na maximização de uma margem referente à superfície de decisão do modelo (hiperplano que separa as classes) (BISHOP, 2006, p. 326). A margem é definida como sendo a distância da superfície de decisão à instância mais próxima. Para casos não linearmente separáveis, é feito, ainda, um mapeamento dos dados para um espaço de maior dimensão, para que haja, então, uma superfície linear que separe os dados.

O último método investigado é o de uma Rede Neural *Multilayer Perceptron*, que pode ser entendido como uma sucessão de transformações sobre a entrada (BISHOP, 2006), como exemplificado na Figura 38, sendo cada círculo uma representação de um neurônio artificial, descrito pela Equação 5.36.

Figura 38 – Exemplo de *Multilayer Perceptron*

Fonte: Elaborada pelo autor.

$$a_j = h(\mathbf{w}_j \phi(\mathbf{x}) + b_j) \quad (5.36)$$

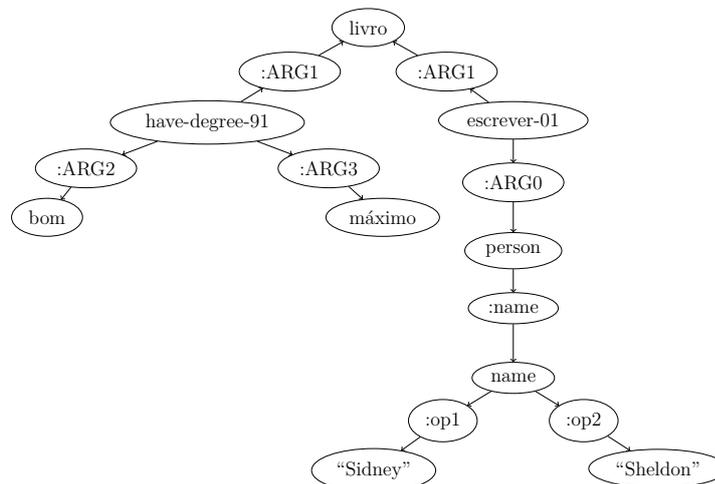
Na equação, a saída de um neurônio j (a_j) é dependente de uma combinação linear,

com pesos \mathbf{w}_j , sobre a entrada $\mathbf{x} \in \mathbb{R}^D$ e com um viés independente b_j . O valor obtido através da combinação é, então, modificado por uma função $h(\cdot)$ não linear, chamada função de ativação. A função $\phi(\cdot)$ sobre a entrada, por sua vez, pode ser a identidade ($\phi(\mathbf{x}) = \mathbf{x}$), como é o caso dos neurônios da primeira camada (mais a esquerda); ou pode também assumir a forma da própria equação do neurônio, originando as conexões entre as diferentes camadas, ou seja, a saída a_j dos neurônios (que, por sua vez, são transformações da entrada) são combinadas entre si nas próximas camadas da rede.

Após a seleção dos nós mais importantes através da classificação utilizando os algoritmos de AM, o processo de sumarização continua conforme as regras apresentadas na [Subseção 5.2.2](#): são selecionados os caminhos mais importantes entre os nós que são, posteriormente, expandidos utilizando-se a ferramenta DptOIE ([OLIVEIRA; CLARO, 2019](#)).

Como alternativa às regras de seleção das arestas e expansão, também investigou-se uma abordagem puramente baseada em AM, tratando-se as relações como elementos a serem classificados. Para isso, construiu-se o grafo de Levi correspondente ao AMR, substituindo-se cada aresta por um nó correspondente com arestas conectando-o às duas pontas da relação. Um exemplo é o grafo de Levi apresentado na [Figura 39](#), correspondente ao AMR da [Figura 31a](#). Na figura, os vértices que representam as relações são rotulados com dois pontos seguido do nome da relação (como `:ARG1` ou `:op2`).

Figura 39 – Exemplo de grafo de Levi



Fonte: Elaborada pelo autor.

A partir do grafo de Levi, é possível aplicar o método de classificação de vértices, a fim de selecionar tanto nós quanto arestas relevantes. A partir dessa seleção, o grafo do sumário foi expandido de forma a incluir todas as relações relevantes que possuem ao menos uma de suas pontas também selecionadas, incluindo o outro conceito na relação (caso não tenha sido selecionado), por exemplo, se `:ARG1` e `livro` forem selecionados,

`escrever-01` será incluído também apesar de não ter sido inicialmente classificado como relevante. Isso garante que as relações escolhidas possuem as duas extremidades no grafo do resumo. Como passo final, escolhe-se o maior componente conexo como representação do sumário, a fim de garantir a conectividade do grafo AMR.

5.3 Medidas de avaliação

Além da medida *Smatch* (CAI; KNIGHT, 2013), apresentada ao final da Subseção 5.2.1, utilizaram-se para a avaliação da qualidade dos sumários as medidas SEMA (ANCHIÊTA; CABEZUDO; PARDO, 2019) e ROUGE-1 (LIN, 2004).

A medida SEMA, desenvolvida por Anchiêta, Cabezudo e Pardo (2019) como uma alternativa à *Smatch*, tem como foco comparar dois grafos AMR de uma maneira determinística seguindo uma estratégia de Busca por Largura. Como exemplo, serão utilizados os grafos AMR das Figuras 33 e 34. Um dos grafos, por exemplo o da Figura 33, é mantido como referência (a ser consultado), enquanto a busca percorre o outro grafo.

A busca se inicia na raiz (\mathbf{x}), conceito também existente no grafo de referência (\mathbf{q}). A seguir, as relações vizinhas a esse conceito são avaliadas: (`querer-01 :ARG0 menino`) e (`querer-01 :ARG1 bilboquê`), dentre as quais apenas a primeira está presente na referência. Feitas as correspondências das relações, os conceitos vizinhos são avaliados, porém apenas aqueles que estiverem presentes em uma relação válida previamente selecionada, ou seja, apenas `menino`. Com isso, a contagem de triplas corretas apresentada pela SEMA é de 3 (`instance(x, querer-01), instance(y, menino), ARG0(x, y)`), enquanto que a reportada pela *Smatch* é de 4 (`instance(x, querer-01), instance(y, menino), ARG0(x, y), ARG1(x, z)`). Por fim, o valor definitivo da *Smatch* é definido como o F-Score obtido através das triplas relevantes: com precisão $\frac{3}{5} = 0,6$ e cobertura $\frac{3}{6} = 0,5$, resultando numa medida de aproximadamente 0,55.

Outra medida utilizada é a ROUGE-1, apresentada por Lin (2004), amplamente utilizada na área de Sumarização Automática para avaliar resumos textuais. Essa pontuação é definida como sendo a cobertura de palavras no resumo automático (s), comparado a um resumo de referência (r), como é mostrado na Equação 5.37, sendo $\sigma(s, r)$ o número de palavras em comum entre s e r , e $|r|$, o número de palavras em r .

$$\text{ROUGE} - 1(s, r) = \frac{\sigma(s, r)}{|r|} \quad (5.37)$$

Essa medida aponta a taxa de informação presente na referência (informações relevantes) que é retida no resumo automático, independentemente se houver informações extras ou não, isto é, se o resultado automático cobre a maioria das ideias principais do texto original.

5.4 Experimentos

Todos os experimentos foram realizados utilizando-se a ferramenta SemOpinionS², implementada no âmbito deste projeto. Toda a manipulação e análise dos grafos foram feitas por meio da biblioteca NetworkX³ (HAGBERG; SCHULT; SWART, 2008). Os algoritmos de AM utilizados foram implementados pela ferramenta Scikit-learn⁴ (PEDREGOSA *et al.*, 2011).

A avaliação das técnicas foi feita sobre uma parte do corpus OpiSums-PT de Condori *et al.* (2015). Como o processo de anotação – discutido no Capítulo 4 – resultou em dois produtos com todos os seus comentários anotados (Iphone 5 e O Apanhador no Campo de Centeio), esses foram os dados selecionados para os experimentos. Com o objetivo de identificar o impacto do uso de uma anotação automática, dois outros produtos similares (um *smartphone* e um livro) também foram escolhidos para análise: Galaxy SIII e O Outro Lado da Meia Noite, tendo suas sentenças anotadas em AMR por um *parser* automático (ANCHIÊTA; PARDO, 2018a).

Cada um dos métodos investigados nos experimentos – e descritos anteriormente na Seção 5.2 – possuem suas particularidades de execução e hiperparâmetros. Com isso, cada configuração de execução é descrita a seguir.

LiuEtAl2015 Método de Liu *et al.* (2015). Para o treinamento usando AdaGrad, similarmente aos autores originais, $\eta = 1,0$; $\epsilon = 1,0$. Foram comparadas duas funções de erro: *Perceptron* e *Ramp*.

LiaoEtAl2018 Método de Liao, Lebanoff e Liu (2018). $\eta = 1,0$; $\epsilon = 1,0$. Apenas a função *Ramp* foi avaliada. Para o *Spectral Clustering*, comparou-se as medidas de similaridade: LCS, *Smatch* e Cobertura de Conceitos.

DohareEtAl2018 Método de Dohare, Gupta e Karnick (2018). O número de nós relevantes selecionados inicialmente, antes das expansões, é 10. Para o cálculo das frequências de documento (utilizadas no TF-IDF), utilizou-se dois corpus diferentes: um genérico e um específico ao domínio do produto. Como representante do domínio de produtos eletrônicos, utilizou-se o corpus B2W-Reviews01⁵ (REAL; OSHIRO; MAFRA, 2019). Do domínio de livros, por sua vez, o corpus ReLi (FREITAS *et al.*, 2014) foi usado, com a eliminação dos dois produtos utilizados para a avaliação (“O Apanhador no Campo de Centeio” e “O Outro Lado da Meia Noite”). Para a criação do corpus genérico, foi feita uma combinação dos dois corpus específicos.

DohareEtAl2018_TF Variação de DohareEtAl2018, substituindo-se a medida

² Disponível em: <<https://github.com/Superar/SemOpinionS>>. Acesso em: 30 mar. 2021.

³ Disponível em: <<https://networkx.org/>>. Acesso em: 30 mar. 2021.

⁴ Disponível em: <<https://scikit-learn.org/>>. Acesso em: 30 mar. 2021.

⁵ Disponível em: <<https://github.com/b2wdigital/b2w-reviews01>>. Acesso em: 31 mar. 2021.

TF-IDF pela TF. O número de nós relevantes selecionados inicialmente também é 10.

score_optimization Variação de DohareEtAl2018, substituindo-se a medida TF-IDF por uma pontuação levando em conta diversos atributos da estrutura do grafo AMR, sentimentos associados aos conceitos e aspectos dos produtos. Devido à natureza estocástica da otimização por *Simulated Annealing*, esse método foi executado diversas vezes para se ter resultados mais robustos.

O número de vértices selecionados inicialmente foi mantido em 10. Para a otimização: $I_{\max} = 1000$; $K = 5$; $T_0 = 1,0$; $r = 0,9$; $I_T = 10$, seguindo o trabalho de [Ludermir, Yamazaki e Zanchettin \(2006\)](#).

machine_learning Seleção de nós relevantes por meio de AM. Foram avaliados quatro modelos diferentes com os hiperparâmetros padrões indicado pela ferramenta:

- *Árvore de decisão*
 - Critério de divisão dos nós: índice Gini.
- *Random Forest*
 - Número de árvores: 100.
- SVM
 - *Kernel*: RBF.
- MLP
 - Número de unidades ocultas: 100.
 - Função de ativação: ReLU.
 - Taxa de aprendizado: 10^{-4} .

Existe a possibilidade dos hiperparâmetros não serem os melhores para a tarefa, porém, para um trabalho preliminar de estudo da viabilidade do método, acredita-se serem suficientes.

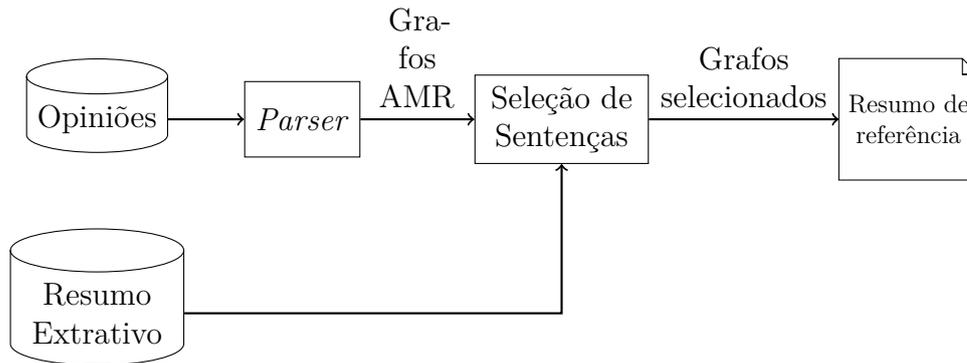
machine_learning_levi Seleção de vértices e arestas relevantes através de AM. Foram mantidos os hiperparâmetros do modelo `machine_learning`.

5.4.1 Treinamento

A maioria das técnicas investigadas precisam de um treinamento, ou otimização, de parâmetros. Por esse motivo, foi necessária a criação de um *cópus* de treinamento. Devido à falta de anotação manual de sentenças opinativas em AMR, criou-se um *cópus* artificial

através do *parser* de Anchiêta e Pardo (2018a) contendo as sentenças (provenientes do córpus OpiSums-PT) dos comentários sobre os produtos que não foram utilizados durante a avaliação. A criação dos resumos de referência nesse córpus foi feita por meio dos resumos extrativos, isto é, selecionando algumas sentenças dos comentários de entrada. Esse processo é ilustrado na Figura 40.

Figura 40 – Processo de criação automática do córpus de treinamento



Fonte: Elaborada pelo autor.

O córpus de treinamento construído possui 65 instâncias (comentários e seu resumo correspondente), contendo um total de 3290 sentenças de comentários anotadas, sendo 388 selecionadas para compor os resumos.

5.4.2 Geração textual

Visto que uma das medidas de avaliação utilizadas, a ROUGE, se baseia na comparação de tokens entre dois textos, é preciso, a partir do grafo AMR do resumo, seu texto correlato. Devido ao fato de não haver, ainda, bons geradores textuais AMR para o português – e da medida ROUGE-1 não levar em consideração a ordem dos *tokens* no texto – a geração do resumo é realizada de forma bastante rudimentar.

Como o grafo do resumo é construído, em todos os métodos aqui avaliados, a partir da seleção de conteúdo proveniente de um grafo maior construído com as sentenças dos comentários (como discutido na Seção 5.1), é possível se aproveitar de informações das sentenças originais para a geração textual. Desse modo, obteve-se o alinhamento, por meio da ferramenta de Pourdamghani *et al.* (2014) adaptada para o Português⁶, entre os conceitos AMR e as palavras nas sentenças originais e, posteriormente, as palavras alinhadas aos conceitos do grafo de resumo foram utilizadas para se construir o texto final (em uma espécie de *bag of words*). Por se tratar apenas de uma sequência de palavras, sem ordem específica, o texto do resumo pode não fazer necessariamente sentido, porém

⁶ As adaptações foram pequenas, focando-se em modificação na lista de *stopwords* e nos processos de etiquetagem morfosintática e tokenização do texto. O alinhado está disponível em: <<https://github.com/msobrevillac/TB-AMRAligner>>. Acesso em: 29 abr. 2021.

não afetará a medida ROUGE-1. Essa mesma abordagem foi também utilizada por [Liu et al. \(2015\)](#) em seu trabalho.

Exemplos de grafos de comentários obtidos através do pré-processamento descrito, seus respectivos resumos e os textos (*bag of words*) correspondentes serão apresentados com mais detalhe no [Apêndice A](#).

5.5 Resultados

Os resultados serão apresentados para cada configuração descrita na [Seção 5.4](#), visto que existem diferentes análises a serem feitas sobre cada método. Ao final, uma comparação entre os métodos será feita, tendo-se como base o melhor resultado dentre as variações de um mesmo modelo. Devido ao fato de cada produto possuir 5 resumos criados manualmente, as tabelas apresentam a média e desvio padrão das medidas obtidas a partir da comparação do resumo automático com cada referência.

Para dar maior robustez às conclusões obtidas, utilizou-se o teste de significância de Kolmogorov-Smirnov por ser um teste não-paramétrico com poucas suposições acerca da distribuição dos dados ([CORDER; FOREMAN, 2011](#), p.80–84)⁷. Como três medidas de avaliação diferentes foram utilizadas, será pontuado sobre quais valores o teste foi realizado. O teste foi calculado comparando-se, par a par, todos os 20 resultados obtidos para cada método sendo avaliado (5 resumos de referência para cada um de 4 produtos).

Os resultados obtidos ao se utilizar a técnica [LiuEtAl2015](#) são apresentados na [Tabela 18](#).

Dos dados da tabela, é possível corroborar as observações obtidas originalmente por [Liu et al. \(2015\)](#) de que a função *Ramp* é melhor adaptada para a tarefa de sumarização, visto que os valores obtidos utilizando essa função são maiores, sendo que o resultado do teste estatístico sobre a medida SEMA ($D = 0,9; p \ll 0,001$) mostra que a diferença é significativa. A única exceção é a pontuação *Smatch* para o produto Galaxy SIII, o que pode ser devido ao otimismo intrínseco da medida, o que não ocorre com a SEMA.

Ao se comparar os diferentes tipos de produtos entre si (*smartphones* e livros), pode-se ver que, ao usar a função *Perceptron*, as medidas de *Smatch* ($D = 0,8; p \ll 0,001$) e ROUGE-1 ($D = 0,5; p = 0,17$) são consistentemente maiores para os telefones. Isso provavelmente se deve ao fato de os resumos automáticos produzidos para esse tipo de produto terem sido maiores – 11 e 13 conceitos – se comparados aos dos livros – 5 e 7 conceitos –, aumentando a cobertura tanto de conceitos (para a *Smatch*) quanto de palavras (para a

⁷ O teste possui dois valores a serem reportados: D , indicando o grau de divergência dos dados, e o p-valor, probabilidade das distribuições serem a mesma, isto é, quanto menor o p-valor, mais significativa é a diferença entre os resultados. É usual definir um limiar para o valor-p, comumente definido em 0,05.

Tabela 18 – Resultado dos experimentos com a configuração LiuEtAl2015

Produto		<i>Smatch</i>	SEMA	ROUGE-1
Função de erro: <i>Perceptron</i>				
Iphone 5	Média	0,20	0,04	0,1624
	Desvio Padrão	0,0505	0,0205	0,0679
O Apanhador no Campo de Centeio	Média	0,12	0,09	0,1236
	Desvio Padrão	0,0303	0,0638	0,0130
Galaxy SIII	Média	0,26	0,12	0,2440
	Desvio Padrão	0,0370	0,0114	0,0553
O Outro Lado da Meia Noite	Média	0,15	0,08	0,1723
	Desvio Padrão	0,0207	0,0507	0,0479
Função de erro: <i>Ramp</i>				
Iphone 5	Média	0,30	0,27	0,3414
	Desvio Padrão	0,0992	0,0896	0,1186
O Apanhador no Campo de Centeio	Média	0,34	0,30	0,4252
	Desvio Padrão	0,0358	0,0396	0,0345
Galaxy SIII	Média	0,23	0,20	0,2572
	Desvio Padrão	0,0422	0,0438	0,0418
O Outro Lado da Meia Noite	Média	0,21	0,17	0,2516
	Desvio Padrão	0,0190	0,0207	0,0257

Fonte: Dados da pesquisa.

ROUGE-1). É importante, porém, salientar que o teste estatístico mostrou que apenas as diferenças a respeito da medida *Smatch* são significantes ($p \leq 0,05$), podendo ser necessários mais dados para verificar se a distinção em relação à medida ROUGE-1 se torna mais evidente. É interessante observar que o mesmo não ocorre ao se utilizar a função *Ramp*, pois os resumos produzidos são muito maiores (135, 160, 252 e 245 conceitos, segundo a ordem apresentada na tabela).

Ainda dos resultados com a função *Ramp*, é possível observar que os resultados dos produtos anotados manualmente são superiores aos do *parser* ($D = 0,7; p = 0,01$ para a medida SEMA), mesmo com a otimização dos pesos tendo sido realizada sobre grafos obtidos automaticamente. Isso indica que o uso dessa função é capaz de sobrepujar especificidades do *parser*, generalizando os pesos, também, para dados com anotações manuais.

O fato de os resumos sobre os comentários anotados manualmente serem menores e, ainda assim, possuírem melhores pontuações também é notável. Isso indica que o método possui uma seleção de conteúdo mais precisa com dados devidamente anotados.

A segunda configuração investigada – LiaoEtAl2018 – possui como variações as

diferentes medidas de similaridades usadas durante o passo de *Spectral Clustering*. Os resultados são apresentados na [Tabela 19](#).

Tabela 19 – Resultado dos experimentos com a configuração LiaoEtAl2018

Produto		<i>Smatch</i>	SEMA	ROUGE-1
Medida de similaridade: LCS				
Iphone 5	Média	0,41	0,29	0,4231
	Desvio Padrão	0,0868	0,1584	0,1418
O Apanhador no Campo de Centeio	Média	0,38	0,28	0,4325
	Desvio Padrão	0,0976	0,1139	0,1400
Galaxy SIII	Média	0,39	0,25	0,4026
	Desvio Padrão	0,1010	0,1494	0,1220
O Outro Lado da Meia Noite	Média	0,24	0,10	0,2638
	Desvio Padrão	0,0455	0,0712	0,0795
Medida de similaridade: <i>Smatch</i>				
Iphone 5	Média	0,33	0,28	0,3875
	Desvio Padrão	0,1050	0,1210	0,1155
O Apanhador no Campo de Centeio	Média	0,33	0,20	0,3801
	Desvio Padrão	0,0507	0,0800	0,0714
Galaxy SIII	Média	0,27	0,15	0,2613
	Desvio Padrão	0,0495	0,0736	0,0679
O Outro Lado da Meia Noite	Média	0,21	0,09	0,2138
	Desvio Padrão	0,0409	0,0513	0,0529
Medida de similaridade: Cobertura de Conceitos				
Iphone 5	Média	0,35	0,32	0,3959
	Desvio Padrão	0,1101	0,1061	0,1345
O Apanhador no Campo de Centeio	Média	0,33	0,22	0,3722
	Desvio Padrão	0,0308	0,0493	0,0536
Galaxy SIII	Média	0,32	0,23	0,3475
	Desvio Padrão	0,0864	0,1019	0,0983
O Outro Lado da Meia Noite	Média	0,18	0,04	0,2170
	Desvio Padrão	0,0311	0,0522	0,0377

Fonte: Dados da pesquisa.

Da tabela, pode-se verificar que a medida LCS aparenta ser melhor do que as outras, em termos de *Smatch* e ROUGE-1. A medida SEMA também é, em sua maioria, melhor quando o *Spectral Clustering* é realizado com a LCS. O teste estatístico, porém, não aponta diferenças significativas entre os métodos com relação a nenhuma medida de avaliação. Esse resultado difere do obtido por [Liao, Lebanoff e Liu \(2018\)](#), visto que os autores indicam que a medida de Cobertura de Conceitos se mostrou melhor. Isso pode

ter ocorrido por diferenças nas análises, pois a pontuação *Smatch* leva em consideração triplas (relação e seus argumentos) completas, enquanto os autores originais focaram na comparação de nós e arestas separadamente.

Outra observação interessante a ser pontuada é a de que os valores de Desvio Padrão obtidos pelo método [LiaoEtAl2018](#) são maiores se comparados à sua versão mais simples ([LiuEtAl2015](#)). Isso provavelmente se dá pela seleção de sentenças para a sumarização: depois de agrupadas, um determinado número de sentenças é selecionado aleatoriamente de cada grupo, o que resulta em resumos menores, já que não foram utilizadas todas as sentenças possíveis como na técnica anterior. Os resumos menores impactam nas medidas, visto que existem mais chances de as informações selecionadas pertencerem a uma referência e não a outras. Apesar disso, o agrupamento parece auxiliar no processo de sumarização pois, mesmo com resumos menores, as medidas se mostraram melhores e, portanto, aparenta-se haver uma seleção do conteúdo mais adequada. Apesar disso, o resultado do teste estatístico com relação à pontuação *Smatch* não indica uma diferença significativa ($D = 0,4; p = 0,08$) entre a técnica utilizando a medida LCS e o método de [LiuEtAl2015](#) com a função *Ramp*.

Comparando-se os produtos entre si, pode-se perceber que o livro “O Outro Lado da Meia Noite” parece ser de mais difícil sumarização, pois as medidas obtidas de seus resultados são sistematicamente menores, o que também pôde ser visto na [Tabela 18](#).

A próxima configuração avaliada é a [DohareEtAl2018](#), com variações nos córpis utilizados para o cálculo das frequências de documento (IDF). Os resultados podem ser vistos na [Tabela 20](#).

A primeira observação a ser feita é a de que o uso do córpis genérico se mostrou melhor do que os córpis específicos. Isso é mais evidente ao analisar os valores obtidos para os livros, uma vez que o B2W-Reviews01 é maior do que o ReLi e, no córpis genérico, acaba por sobrepujá-lo. Esse resultado indica que, para o cálculo do IDF, é mais importante ter uma grande quantidade de documentos, ainda que de domínios diferentes. Essa observação deve ser, porém, tomada com cautela, visto que o teste estatístico aponta pouca significância na diferença entre os resultados para os livros ($D = 0,79; p = 0,3$).

Das medidas de ROUGE-1 obtidas, pode-se notar que a anotação manual das sentenças em AMR ajudou a produzir melhores resultados, na média. Isso se dá, provavelmente, no momento do alinhamento dos conceitos AMR com as palavras do texto correspondentes. Como os grafos obtidos a partir do *parser* possuem normalmente erros, conceitos faltantes ou conceitos e relações nulos (`null_tag` e `null_edge`, respectivamente), o alinhamento é prejudicado, afetando tanto o cálculo das frequências TF-IDF quanto a geração textual. Contudo, o teste estatístico indica que, levando em consideração a distribuição dos valores para os experimentos, essas diferenças são pouco significativas entre si ($D = 0; p = 0,80$), sendo necessária uma maior quantidade de dados e testes para compro-

Tabela 20 – Resultado dos experimentos com a configuração DohareEtA12018

Produto		<i>Smatch</i>	SEMA	ROUGE-1
Córpus para IDF: genérico				
Iphone 5	Média	0,40	0,31	0,4223
	Desvio Padrão	0,0735	0,1557	0,1065
O Apanhador no Campo de Centeio	Média	0,36	0,23	0,4229
	Desvio Padrão	0,0699	0,0978	0,0721
Galaxy SIII	Média	0,33	0,21	0,2898
	Desvio Padrão	0,1419	0,2049	0,1813
O Outro Lado da Meia Noite	Média	0,24	0,03	0,2630
	Desvio Padrão	0,0167	0,0148	0,0192
Córpus para IDF: específico				
Iphone 5	Média	0,40	0,31	0,4223
	Desvio Padrão	0,0731	0,1557	0,1065
O Apanhador no Campo de Centeio	Média	0,29	0,14	0,3196
	Desvio Padrão	0,1751	0,2423	0,1445
Galaxy SIII	Média	0,33	0,21	0,2898
	Desvio Padrão	0,1387	0,2049	0,1813
O Outro Lado da Meia Noite	Média	0,30	0,10	0,2462
	Desvio Padrão	0,0757	0,0865	0,0955

Fonte: Dados da pesquisa.

var essa visão.

Os resultados utilizando a variação dessa configuração, utilizando apenas a medida TF (DohareEtA12018_TF) são apresentados na Tabela 21. Da tabela, pode-se observar que a remoção do termo IDF prejudicou, em geral, os resultados, refutando a hipótese inicial de que, para uma sumarização multidocumento esse termo seria irrelevante (ou até mesmo prejudicaria a criação de bons resumos). No entanto, o teste estatístico sobre a medida *Smatch* não aponta uma significância forte nesses valores ($D = 0,35$; $p = 0,17$).

Curiosamente, o livro “O Apanhador no Campo de Centeio” obteve resultados melhores ao se utilizar a medida TF puramente, sem levar em consideração as frequências de documentos. Para realizar uma análise mais detalhada, os 10 conceitos mais bem ranqueados de acordo com cada método são apresentados na Tabela 22, bem como o número de resumos de referência que contêm o referido conceito.

Como pode ser visto na tabela, o método utilizando apenas a medida TF seleciona conceitos presentes em mais resumos de referência (3,4 em média) do que por meio da medida TF-IDF (2,7 em média). As expansões subsequentes no grafo são realizadas a partir dessa lista inicial. Portanto, ao serem inseridos conceitos menos relevantes, os

Tabela 21 – Resultado dos experimentos com a configuração DohareEtA12018_TF

Produto		<i>Smatch</i>	SEMA	ROUGE-1
Iphone 5	Média	0,24	0,17	0,2831
	Desvio Padrão	0,1228	0,1346	0,1304
O Apanhador no Campo de Centeio	Média	0,44	0,35	0,5082
	Desvio Padrão	0,0503	0,0789	0,0679
Galaxy SIII	Média	0,27	0,14	0,2384
	Desvio Padrão	0,0792	0,1240	0,0943
O Outro Lado da Meia Noite	Média	0,20	0,07	0,2063
	Desvio Padrão	0,0332	0,0391	0,0402

Fonte: Dados da pesquisa.

Tabela 22 – Conceitos selecionados nos experimentos DohareEtA12018 e DohareEtA12018_TF

Experimento: DohareEtA12018			Experimento: DohareEtA12018_TF		
Conceito	Pontuação	Referências	Conceito	Pontuação	Referências
livro	73,85	5	ele	32	2
percentage-entity	73,85	0	eu	25	5
historia	36,30	0	livro	19	5
história	36,30	5	percentage-entity	19	0
ler-01	35,87	5	e	15	5
personagem	26,89	1	mais	9	2
ele	23,52	2	cause-01	9	5
clássico	22,38	2	ler-01	7	5
literatura	21,57	2	historia	7	0
cause-01	19,93	5	história	7	5

Fonte: Dados da pesquisa.

processos seguintes também são prejudicados, afetando assim o resumo final como um todo. É importante salientar, novamente, que essas observações são feitas apenas para um produto em específico e que, em geral, usar a TF-IDF se mostrou melhor.

O próximo experimento realizado é o `score_optimization`, que incorpora ao processo de sumarização diversas medidas referentes à estrutura do grafo de entrada. Visto que o método é aleatório, ele foi executado 10 vezes, sendo a média dos resultados apresentada na Tabela 23. É importante citar que o valor de desvio padrão na tabela diz respeito à media do desvio padrão, sobre os 5 resumos de referência, entre todas as execuções – e não o desvio padrão das médias para cada referência.

A tabela mostra que, na média, o método baseado em *Simulated Annealing* é capaz de obter resultados competitivos com os outros métodos da literatura, indicando, também, que as medidas referentes à estrutura do grafo incluem um certo grau de informação importante para a sumarização.

Tabela 23 – Resultado dos experimentos com a configuração `score_optimization`

Produto		<i>Smatch</i>	SEMA	ROUGE-1
Iphone 5	Média	0,27	0,18	0,2559
	Desvio Padrão	0,0725	0,1147	0,0812
O Apanhador no Campo de Centeio	Média	0,34	0,25	0,3235
	Desvio Padrão	0,0943	0,1303	0,0797
Galaxy SIII	Média	0,43	0,35	0,4840
	Desvio Padrão	0,0944	0,1298	0,1351
O Outro Lado da Meia Noite	Média	0,26	0,08	0,2227
	Desvio Padrão	0,0389	0,0578	0,0532

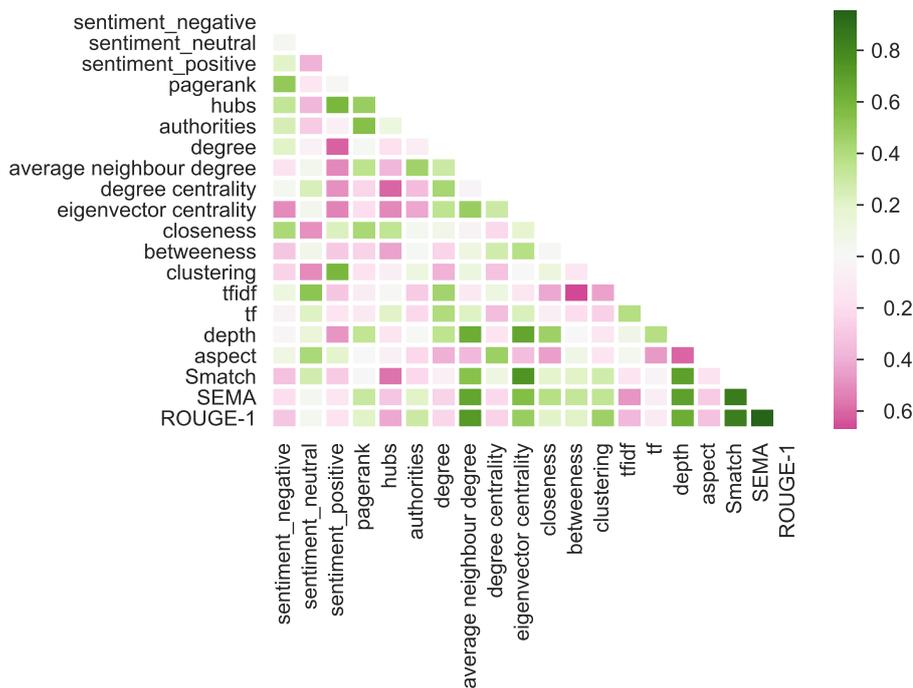
Fonte: Dados da pesquisa.

Para avaliar a importância efetiva dos atributos utilizados, a correlação do valor absoluto dos pesos otimizados, entre si e também com as medidas de avaliação, é apresentada na [Figura 41](#). Foi utilizado o módulo pois valores negativos podem indicar uma correlação forte com atributos que não necessariamente sejam importantes; por exemplo um atributo cujos pesos tenha uma média negativa e correlação positiva com alguma medida de avaliação não será importante, pois os valores que aumentam a qualidade da saída são aqueles que são acima da média (negativa), ou seja, aqueles valores mais próximos de 0 (com pouca influência na pontuação). Ao se utilizar o módulo dos pesos, a medida de correlação passa a ser uma indicação da importância dos nós, sendo uma correlação positiva uma importância maior (maior influência na pontuação do nó).

Na figura, quanto mais escura (rosa ou verde) a célula referente a dois atributos, mais correlatos (negativamente ou positivamente) eles são. A importância foi calculada utilizando-se todas as 10 execuções do método e o valor das medidas de avaliação utilizadas para cada execução foi a média sobre cada produto.

Do gráfico, é possível perceber que as medidas de avaliação são mais correlatas aos valores absolutos dos pesos dos atributos de `average neighbour degree` (grau médio dos vizinhos), `depth` (profundidade do nó) e `eigenvector centrality` (centralidade de autovetor), ou seja, esses são os atributos que, quando impactam a pontuação dos nós, mais aumentam a qualidade dos sumários. Já com os atributos `hubs` (do algoritmo HITS) e `tfidf`, os valores de correlação são negativos, indicando que eles possuem uma maior tendência de serem ignorados para criar resumos melhores.

Ao analisar os atributos relativos ao sentimento do conceito (`sentiment_negative`, `sentiment_neutral` e `sentiment_positive`), pode-se perceber que não são atributos tão relevantes, visto que, quando possuem pesos maiores, prejudicam os resultados ou influenciam muito pouco (no caso de conceitos de sentimento neutro). O mesmo ocorre com a anotação de aspectos (`aspect`).

Figura 41 – Visualização dos pesos otimizados no experimento `score_optimization`

Fonte: Dados da pesquisa.

Examinando as medidas de importância relativas entre os atributos em si, pode-se inferir que aqueles que possuem valores positivos (verde) são atributos que acompanham suas importâncias específicas, isto é, se complementam no cálculo das pontuações (*average neighbour degree*, *depth* e *eigenvector centrality*, por exemplo). Já os pares de atributos com correlação negativa, são atributos “complementares”, ou seja, apenas um deles mostrou destaque na pontuação; esse é o caso dos atributos de *degree centrality* (centralidade de grau) e *hubs*, por exemplo.

O próximo experimento realizado diz respeito à configuração `machine_learning`, com quatro técnicas de AM diferentes. Os resultados podem ser vistos na [Tabela 24](#).

A partir da análise dos resultados da tabela, pode-se indicar que, dentre os métodos avaliados, o de *Random Forest* parece ser capaz de extrair mais informação e, com isso, obter melhores resultados no geral, ainda que não supere outras técnicas na sumarização dos comentários sobre livros. Apesar disso, seus valores de Desvio Padrão são de certa forma maiores do que os obtidos por outros métodos, indicando que seu resultado pode estar distante de algumas referências – mas não tantas a ponto de resultar em baixas pontuações de avaliação. Como indicado pelo teste estatístico sobre a medida SEMA, essa técnica possui resultados significativos em comparação ao MLP e ao SVM ($D = 0,45; p = 0,03$ para ambos). A diferença com relação ao algoritmo de Árvore de Decisão não é, porém, significativa ($D = 0,35; p = 0,17$).

Tabela 24 – Resultado dos experimentos com a configuração `machine_learning`

Produto		<i>Smatch</i>	SEMA	ROUGE-1
<i>Árvore de Decisão</i>				
Iphone 5	Média	0,33	0,30	0,3434
	Desvio Padrão	0,0844	0,0695	0,0885
O Apanhador no Campo de Centeio	Média	0,39	0,32	0,4353
	Desvio Padrão	0,0620	0,0853	0,0838
Galaxy SIII	Média	0,26	0,21	0,2708
	Desvio Padrão	0,0455	0,0698	0,0359
O Outro Lado da Meia Noite	Média	0,23	0,16	0,2871
	Desvio Padrão	0,0378	0,0430	0,0556
<i>Random Forest</i>				
Iphone 5	Média	0,40	0,32	0,3896
	Desvio Padrão	0,1159	0,1698	0,1010
O Apanhador no Campo de Centeio	Média	0,47	0,40	0,5100
	Desvio Padrão	0,0915	0,1243	0,1355
Galaxy SIII	Média	0,35	0,24	0,3194
	Desvio Padrão	0,1145	0,1776	0,1322
O Outro Lado da Meia Noite	Média	0,23	0,03	0,1998
	Desvio Padrão	0,0110	0,0089	0,0179
SVM				
Iphone 5	Média	0,27	0,08	0,1676
	Desvio Padrão	0,0416	0,0387	0,0463
O Apanhador no Campo de Centeio	Média	0,51	0,40	0,5519
	Desvio Padrão	0,1850	0,2392	0,2036
Galaxy SIII	Média	0,28	0,07	0,1686
	Desvio Padrão	0,0391	0,0195	0,0327
O Outro Lado da Meia Noite	Média	0,26	0,03	0,2245
	Desvio Padrão	0,0439	0,0537	0,0668
MLP				
Iphone 5	Média	0,27	0,10	0,1843
	Desvio Padrão	0,0351	0,0466	0,0334
O Apanhador no Campo de Centeio	Média	0,51	0,40	0,5519
	Desvio Padrão	0,1890	0,2392	0,2036
Galaxy SIII	Média	0,25	0,08	0,1888
	Desvio Padrão	0,0130	0,0404	0,0224
O Outro Lado da Meia Noite	Média	0,27	0,08	0,2255
	Desvio Padrão	0,0455	0,0622	0,0618

Fonte: Dados da pesquisa.

Comparando os produtos entre si, o livro “O Outro Lado da Meia Noite” parece ter, realmente, os comentários mais complexos de se sumarizar, o que também pôde ser visto nos experimentos anteriores `LiaoEtAl2018` e `score_optimization`. Além disso, os resultados obtidos pelo melhor modelo (*Random Forest*) para os dados anotados manualmente foram, mais uma vez, melhores quando comparados aos dados do *parser*, indicando que a qualidade dos grafos de entrada é importante no processo de sumarização.

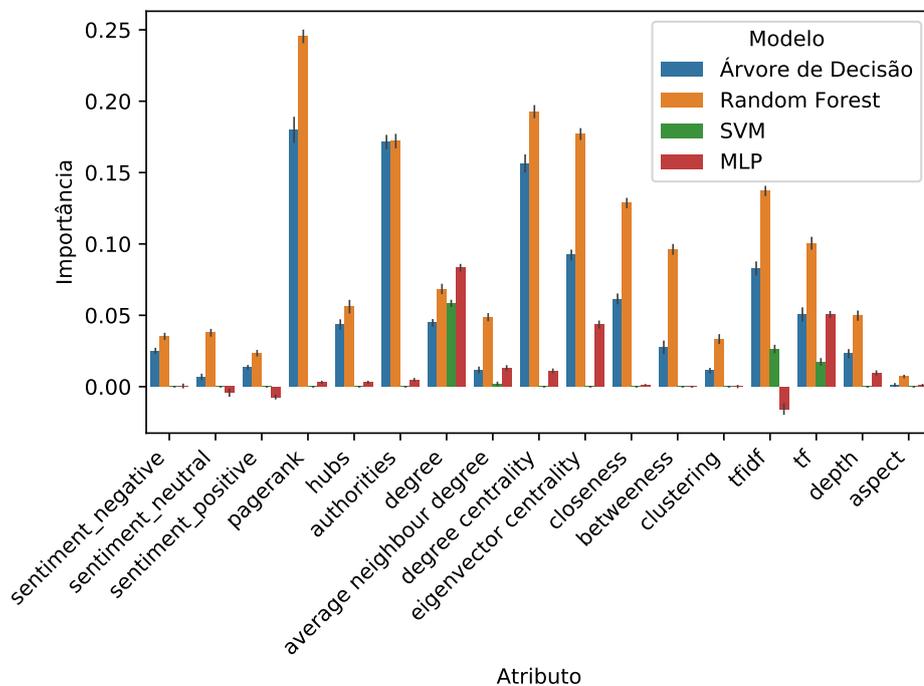
A partir dos modelos treinados, pode-se realizar também uma análise de como os atributos utilizados impactam no resultado de cada classificador, ou seja, suas respectivas

importâncias. Isso é realizado através do método indicado por Breiman (2001), avaliando-se o quanto a inclusão de ruído em determinado atributo impacta negativamente os resultados do classificador já treinado. Dessa forma, executa-se a classificação nos dados D , a fim de gerar a pontuação de referência s , e depois, para cada atributo j , cria-se uma versão ruidosa \tilde{D}_j desses dados fazendo-se uma permutação aleatória de todos os valores de j sobre D ; para atributos mais importantes, essa permutação impactará negativamente os resultados. Esse procedimento é repetido K vezes a fim de garantir uma boa aproximação do grau médio de importância, calculado como indicado na Equação 5.38, em que i_j é a importância do atributo j e $s_{k,j}$ é a medida de avaliação obtida pelo classificador na repetição k sobre \tilde{D}_j .

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j} \quad (5.38)$$

Os resultados da importância de cada atributo podem ser vistos no gráfico da Figura 42. A medida de avaliação utilizada foi a cobertura dos conceitos classificados, visto que é mais importante, em uma sumarização, não deixar de incluir informações relevantes do que incluir dados menos importantes.

Figura 42 – Importância dos atributos para cada técnica de AM utilizada



Fonte: Dados da pesquisa.

Do gráfico pode-se perceber que os modelos baseados em Árvores são os que dependem de mais atributos para garantir uma qualidade de classificação, focando-se principal-

mente nos valores de Pagerank, *Authorities*, Centralidade de Grau (`degree centrality`) e Centralidade de Autovalores (`eigenvector centrality`).

É possível observar que o algoritmo de SVM usa como base principal o atributo de grau (`degree`) do nó, levando em conta também os valores de TF e TF-IDF em menor escala. Com relação ao modelo de MLP, é interessante notar que o valor de importância para o atributo TF-IDF é negativo, indicando que, em média, a cobertura dos conceitos foi maior ao utilizar-se esse atributo ruidoso.

Comparando-se aos resultados observados anteriormente na análise dos atributos durante o experimento `score_optimization`, pode-se perceber que as técnicas de AM não se basearam tanto no uso dos atributos de grau médio dos vizinhos (`average neighbour degree`) e de profundidade do nó (`depth`). Também é possível ver que, similarmente à observação anterior, os atributos referentes ao sentimento dos nós ou anotação de aspecto também não possuem uma importância relevante para os algoritmos de Aprendizado de Máquina.

O último experimento realizado é o segundo a configuração `machine_learning_levi`, que trata as arestas como nós, permitindo que elas sejam também classificadas pelo modelo. Os resultados são apresentados na [Tabela 25](#).

Dos resultados da tabela é possível perceber que a introdução de novos nós – referentes às arestas – tem, de fato, um impacto negativo nos resultados da sumarização. Os valores, porém, não distam muito daqueles obtidos em outros experimentos, como, por exemplo, o algoritmo de MLP com `LiuEtAl2015` usando a função *Ramp* ($D = 0,1; p = 1,0$) e `DohareEtAl2018_TF` ($D = 0,25; p = 0,57$) indicando, portanto, que essa técnica pode ser um bom caminho de investigação, visto que a seleção automática das arestas por classificação dispensa a criação de regras para seleção de caminhos entre nós relevantes (como as desenvolvidas por [Dohare, Gupta e Karnick \(2018\)](#)). Contudo, a seleção das relações por meio de regras ainda produz melhores resultados, conforme comparação com os resultados do experimento `machine_learning` ($D = 0,4; p = 0,08$, MLP).

É possível que os maus resultados obtidos sobre os grafos de Levi deve-se ao conjunto de atributos não muito bem adequados aos nós referentes às relações. Atributos como “sentimento”, “aspecto” não se aplicam a relações AMR, apenas aos conceitos. Algumas medidas de centralidade, como a centralidade de grau ou a de autovetor também não se adequam a esse tipo de nó, visto que eles possuem apenas uma aresta entrante e uma saindo (os conceitos relacionados). Dessa maneira, com um conjunto de atributos melhor adaptado, pode ser possível obter melhores resultados.

Uma outra análise que pode ser realizada sobre os resultados obtidos dos experimentos diz respeito à taxa de compressão dos resumos criados, uma vez que a maioria dos métodos não possui esse tipo de controle, como já abordado no [Capítulo 3](#). Essa taxa

Tabela 25 – Resultado dos experimentos com a configuração `machine_learning_levi`

Produto		<i>Smatch</i>	SEMA	ROUGE-1
<i>Árvore de Decisão</i>				
Iphone 5	Média	0,27	0,25	0,3038
	Desvio Padrão	0,0770	0,0620	0,0839
O Apanhador no Campo de Centeio	Média	0,36	0,36	0,4146
	Desvio Padrão	0,0239	0,0245	0,0523
Galaxy SIII	Média	0,23	0,19	0,2538
	Desvio Padrão	0,0335	0,0493	0,0291
O Outro Lado da Meia Noite	Média	0,19	0,14	0,2462
	Desvio Padrão	0,0534	0,0709	0,0684
<i>Random Forest</i>				
Iphone 5	Média	0,29	0,28	0,3179
	Desvio Padrão	0,0907	0,0870	0,1148
O Apanhador no Campo de Centeio	Média	0,36	0,35	0,4143
	Desvio Padrão	0,0167	0,0179	0,0415
Galaxy SIII	Média	0,25	0,19	0,2622
	Desvio Padrão	0,0606	0,0934	0,0616
O Outro Lado da Meia Noite	Média	0,19	0,13	0,2390
	Desvio Padrão	0,0297	0,0409	0,0403
SVM				
Iphone 5	Média	0,31	0,29	0,3177
	Desvio Padrão	0,0834	0,0673	0,0884
O Apanhador no Campo de Centeio	Média	0,36	0,35	0,4141
	Desvio Padrão	0,0192	0,0179	0,0407
Galaxy SIII	Média	0,23	0,18	0,2551
	Desvio Padrão	0,0600	0,0879	0,0642
O Outro Lado da Meia Noite	Média	0,20	0,09	0,2316
	Desvio Padrão	0,0396	0,0653	0,0451
MLP				
Iphone 5	Média	0,29	0,24	0,2902
	Desvio Padrão	0,0817	0,1150	0,0773
O Apanhador no Campo de Centeio	Média	0,36	0,35	0,4101
	Desvio Padrão	0,0192	0,0207	0,0492
Galaxy SIII	Média	0,23	0,18	0,2480
	Desvio Padrão	0,0432	0,0750	0,0352
O Outro Lado da Meia Noite	Média	0,18	0,09	0,1978
	Desvio Padrão	0,0235	0,0363	0,0182

Fonte: Dados da pesquisa.

foi calculada sobre os grafos AMR, visto que a geração de texto utilizada ainda é muito rudimentar e não produz de fato textos completos, falhando na representação de alguns conceitos e relações. Dessa forma, foram computadas a proporção de nós e arestas que foram removidos durante o processo de sumarização, ou seja, quanto maior a taxa, menos foram os resumos criados. Esses valores podem ser vistos na [Tabela 26](#).

Da tabela pode-se perceber que o experimento LiuEtAl2015, com uso da função de erro Perceptron, gerou os resumos mais comprimidos, eliminando mais de 95% dos nós e das arestas de entrada. Também é possível observar que todos os métodos têm maior

Tabela 26 – Taxa de compressão dos diferentes métodos explorados

Método	Variação	Taxa de compressão	
		Nós	Arestas
LiuEtAl2015	Função de erro: Perceptron	95,63%	97,04%
machine_learning	SVM	82,64%	86,72%
DohareEtAl2018	Córpus específico	76,64%	80,42%
machine_learning	MLP	74,70%	79,44%
LiaoEtAl2018	Similaridade: LCS	67,73%	80,72%
DohareEtAl2018	Córpus genérico	67,31%	71,02%
machine_learning	<i>Random Forest</i>	62,88%	68,43%
score_optimization		57,47%	63,27%
DohareEtAl2018_TF		52,32%	56,99%
LiaoEtAl2018	Similaridade: <i>Smatch</i>	50,78%	70,29%
LiaoEtAl2018	Similaridade: Cobertura de conceitos	43,85%	65,72%
machine_learning	Árvore de decisão	38,67%	45,18%
machine_learning_levi	MLP	29,00%	34,22%
machine_learning_levi	SVM	26,10%	33,09%
machine_learning_levi	<i>Random Forest</i>	23,62%	27,96%
machine_learning_levi	Árvore de decisão	16,89%	20,35%
LiuEtAl2015	Função de erro: <i>Ramp</i>	8,71%	44,01%

Fonte: Dados da pesquisa.

capacidade de eliminar arestas do que nós. Isso se deve ao fato de os nós poderem ter diversas arestas entre si; dessa forma, a seleção de uma única aresta implica a inclusão de ao menos dois nós relacionados, sem a necessidade da incorporação de outras arestas.

Analisando os resultados dos experimentos `machine_learning_levi`, pode-se ver que os resumos criados são maiores do que suas contrapartes sem o uso de Grafos de Levi. Isso indica que a inserção de novos nós (referentes às relações) modificou a estrutura do grafo de tal forma que levou o modelo a incluir mais nós. Além disso, a classificação das arestas – como se fossem vértices – induziu os algoritmos a selecionarem muito mais arestas do que as regras desenvolvidas por [Dohare, Gupta e Karnick \(2018\)](#); entre os modelos de MLP, por exemplo, o uso dos Grafos de Levi resultou na seleção de aproximadamente 234,53% arestas a mais (taxa de 34,22%) do que o uso das regras (taxa de 79,44%). Isso se dá devido à restrição das regras utilizadas, que podem ignorar várias arestas durante o processo, enquanto a classificação analisa necessariamente todas as relações da entrada (podendo selecioná-las ou não).

É importante, na análise do desempenho dos métodos de sumarização, levar em conta suas taxas de compressão além das medidas de avaliação, visto que resumos muito pequenos (como os produzidos pelo método de [Liu et al. \(2015\)](#)) podem não carregar muita informação relevante. Com isso, pode-se comparar as técnicas desenvolvidas com a

taxa de compressão dos resumos de referência criados manualmente: 80,46% para vértices e 85,93% para arestas. Dessa forma, os métodos avaliados que produzem resumos de tamanhos mais razoáveis são, principalmente, o de [Dohare, Gupta e Karnick \(2018\)](#) – baseado em regras – e o uso de AM, por meio do algoritmo de MLP.

Para se ter uma visão geral da qualidade dos modelos investigados, a [Tabela 27](#) apresenta os melhores e piores resultados (entre os obtidos para cada produto avaliado) de cada método. Por brevidade, escolheu-se omitir a medida *Smatch* – visto que a SEMA pode ser usada na análise dos grafos e a pontuação ROUGE-1 dos resumos textuais – para esta avaliação.

Da tabela pode-se ver que o método baseado em AM, com o uso do algoritmo de MLP, obteve os melhores resultados, juntamente com os de SVM e *Random Forest*. Como já discutido anteriormente, ao se analisar produto a produto, os experimentos com esse método resultaram em valores maiores apenas para um único produto, tendo pontuações mais baixas entre os outros. Há, porém, espaço para melhorias nesses métodos, que podem acabar por contornar esse problema.

As técnicas que se utilizam dos Grafos de Levi para a seleção de conceitos e relações relevantes, apesar de apresentarem altos resultados, produzem resumos muito grandes.

Tabela 27 – Melhores e piores resultados de cada modelo de sumarização semântica

Método	Variação	SEMA		ROUGE-1	
		Melhor	Pior	Melhor	Pior
machine_learning	MLP	0,40	0,08	0,5519	0,1843
machine_learning	SVM	0,40	0,03	0,5519	0,1676
machine_learning	<i>Random Forest</i>	0,40	0,03	0,5100	0,1998
machine_learning_levi	Árvore de decisão	0,36	0,14	0,4146	0,2462
machine_learning_levi	<i>Random Forest</i>	0,35	0,13	0,4143	0,2390
machine_learning_levi	MLP	0,35	0,09	0,4101	0,1978
machine_learning_levi	SVM	0,35	0,09	0,4141	0,2316
score_optimization		0,35	0,08	0,4840	0,2227
DohareEtAl2018_TF		0,35	0,07	0,5082	0,2063
machine_learning	Árvore de decisão	0,32	0,16	0,4353	0,2708
LiaoEtAl2018	Similaridade: Cobertura de conceitos	0,32	0,04	0,3959	0,2170
DohareEtAl2018	Córpus específico	0,31	0,10	0,4223	0,2462
DohareEtAl2018	Córpus genérico	0,31	0,03	0,4229	0,2630
LiuEtAl2015	Função de erro: <i>Ramp</i>	0,30	0,17	0,4252	0,2516
LiaoEtAl2018	Similaridade: LCS	0,29	0,10	0,4325	0,2638
LiaoEtAl2018	Similaridade: <i>Smatch</i>	0,28	0,09	0,3875	0,2138
LiuEtAl2015	Função de erro: <i>Perceptron</i>	0,12	0,04	0,2440	0,1236

Fonte: Dados da pesquisa.

Os experimentos `score_optimization` e `DohareEtAl2018_TF` estão ranqueados em um grau intermediário tanto na qualidade dos resumos quanto no seus tamanhos. Com isso, se mostram promissores para melhorias.

Portanto, pode-se escolher como melhor método, levando em consideração a qualidade e tamanho médio dos sumários, o uso de Aprendizado de Máquina – principalmente por meio do algoritmo de *Multilayer Perceptron* – visto que apresentou o melhor resultado (SEMA) em geral e também gera resumos de tamanho semelhante às referências. Apesar disso, é importante salientar que existem melhorias capazes de serem feitas a fim de tornar os resultados mais robustos entre diferentes produtos, como um incremento nos atributos utilizados e também nas regras de criação dos resumos. Tais mudanças também podem beneficiar outros métodos explorados neste projeto. Mais detalhes de possíveis trabalhos futuros serão discutidos mais à frente no [Capítulo 6](#).

5.5.1 Sumarização de opiniões

Para fins de comparação, serão apresentados os resultados de sumarização obtidos por meio de técnicas específicas de opiniões. Foram escolhidos os trabalhos, já apresentados no [Capítulo 3](#), de [Ganesan, Zhai e Han \(2010\)](#) (`Opinosis`) – por também lidar com grafos, ainda que não semânticos – e de [Condori e Pardo \(2017\)](#) (`Opizer-A`), visto que é um sistema focado na língua portuguesa. Foram utilizadas as implementações da ferramenta `Opizer`⁸.

É importante apontar que os dois métodos selecionados são abstrativos, portanto a avaliação é realizada com relação aos resumos abstrativos de referência, e apenas em nível textual (medida ROUGE), já que não são baseados em AMR. Dessa maneira, também deve ser notado que a comparação com os métodos de sumarização semântica apresentados anteriormente não pode ser integral, devido à falta de uma etapa de geração textual satisfatória (como abordado na [Subseção 5.4.2](#)). Assim sendo, deve-se manter em mente que esses resultados são reportados apenas para se ter um visão geral do atual Estado da Arte na área de Sumarização de Opiniões em língua portuguesa. Os resultados obtidos podem ser vistos na [Tabela 28](#).

Como é possível observar na tabela, o método de [Ganesan, Zhai e Han \(2010\)](#) apresentou melhores resultados do que o `Opizer-A` para a maioria dos produtos, com exceção do livro *O Outro Lado da Meia Noite* que, como já discutido anteriormente, tende a apresentar avaliações piores nos experimentos. O valor de ROUGE obtido pelo `Opizer-A` sobre os comentários desse livro são, por sua vez, maiores, indicando que seu processo pode ser capaz de contornar algumas dificuldades postas nos comentários (ou referências) desse produto.

⁸ Disponível em: <https://github.com/roquelopez/opizer>. Acesso em: 29 abr. 2021

Tabela 28 – Resultado dos experimentos com métodos de sumarização de opiniões

Produto	ROUGE-1	
Opizer-A		
Iphone 5	Média	0,2320
	Desvio Padrão	0,0138
O Apanhador no Campo de Centeio	Média	0,1793
	Desvio Padrão	0,0239
Galaxy SIII	Média	0,2683
	Desvio Padrão	0,0326
O Outro Lado da Meia Noite	Média	0,2183
	Desvio Padrão	0,0379
Opinosis		
Iphone 5	Média	0,3163
	Desvio Padrão	0,0385
O Apanhador no Campo de Centeio	Média	0,3300
	Desvio Padrão	0,0517
Galaxy SIII	Média	0,2369
	Desvio Padrão	0,0251
O Outro Lado da Meia Noite	Média	0,1517
	Desvio Padrão	0,0254

Fonte: Dados da pesquisa.

É importante apontar que o método de geração textual das duas técnicas são diferentes: enquanto o *Opinosis* se utiliza quase exclusivamente de palavras do texto original, o *Opizer-A* tem como base o uso de modelos com lacunas, podendo inserir um maior número de palavras novas ao texto, o que pode impactar negativamente a medida ROUGE, visto que ela utiliza de comparação direta de *tokens*, ignorando paráfrases e construções linguísticas diversas (por meio de relações de sinonímia, por exemplo). Essa inadequação da medida ROUGE para sumarização abstrativa é apontada por vários autores na literatura (GENEST; LAPALME, 2011; LI, 2015; NG; ABRECHT, 2015; KHAN *et al.*, 2018; HUANG; WU; WANG, 2020). Portanto, por ter uma geração textual mais próxima da extrativa, essa medida tende a beneficiar a técnica usada por Ganesan, Zhai e Han (2010); o mesmo ocorre com os métodos de sumarização semântica já apresentados. Todas essas observações foram também realizadas por Condori e Pardo (2017).

Em uma comparação mais direta com os métodos baseados em semântica, apresentados na Tabela 27, é possível ver que tanto o *Opizer-A* quanto o *Opinosis* se mostram piores do que a maioria das técnicas avaliadas, sendo melhores, apenas, do que a variação da abordagem de Liu *et al.* (2015) utilizando-se a função de erro *Perceptron*.

Pelos motivos já expostos a respeito da geração textual, não se pode tirar uma conclusão mais assertiva sobre o sistema **Opizer-A**. Porém, devido ao fato do **Opinosis** retirar a maior parte de seu texto dos comentários originais, é possível concluir com mais assertividade que o uso da representação semântica auxiliou na criação de resumos mais próximos do ideal definido pelas referências.

CONSIDERAÇÕES FINAIS

Como discutido na [Seção 1.2](#), o objetivo principal desse trabalho foi realizar uma investigação da sumarização de opiniões sob uma abordagem por meio de semântica profunda, motivado pela proposta do *framework* CLSA de [Cambria et al. \(2015\)](#). Para alcançar tal objetivo, foi realizada a anotação de parte do *corpus* OpiSums-PT ([CONDORI et al., 2015](#)), com a representação semântica AMR, e foram feitos experimentos com diferentes abordagens de sumarização semântica. A partir desses processos foi possível chegar a algumas conclusões, além do surgimento de algumas perspectivas para pesquisas futuras.

6.1 Conclusões

A anotação realizada resultou, como apresentado em detalhes na [Seção 4.3](#), em 404 (26,90% do *corpus*) sentenças anotadas, além de uma lista com diversos verbos novos a serem incluídos no recurso do Verbo Brasil ([DURAN; ALUÍSIO, 2015](#)).

A análise dos resultados da anotação mostrou um alto valor de concordância, comparado a trabalhos anteriores. Além disso, foi feita uma avaliação de como os diferentes conceitos e relações se comportam na representação AMR sobre textos de dois gêneros diferentes: jornalísticos e opinativos. Com isso, observou-se que os textos jornalísticos possuem mais entidades nomeadas. O *corpus* de opiniões, por sua vez, possui mais problemas de separação de sentenças, bem como mais relações do tipo *degree* (e seu conceito relacionado *have-degree-91*), indicando maior presença de intensificadores e comparações.

Textos opinativos também apresentam maiores frequências de conceitos que ocorrem em contextos de subjetividade, como *gostar-01*. Comparativamente, o *corpus* jornalístico apresenta conceitos mais factuais, como *acontecer-01*.

Ao se analisar a distribuição dos fenômenos sobre sentenças opinativas de diferentes tamanhos, foi possível observar que sentenças maiores apresentam justificativas às opiniões

(por meio do conceito *cause-01*) e também algumas outras informações contextuais de tempo e local.

Para a sumarização automática, foram realizados um total de 17 experimentos, entre todos os métodos avaliados e suas variações. Os métodos todos se mostraram competitivos, com valores máximos de SEMA (ANCHIÊTA; CABEZUDO; PARDO, 2019) por volta dos 30%, chegando a um máximo de 40% utilizando-se o método de Aprendizado de Máquina (sobre os comentários do livro “O Apanhador no Campo de Centeio”). Dentre os produtos avaliados, as opiniões sobre o produto “O Outro Lado da Meia Noite” parecem ser de mais difícil sumarização, por ter os menores valores de avaliação na maioria dos experimentos feitos.

Também fez-se uma análise dos atributos – a respeito da estrutura do grafo – utilizados nos métodos de pontuação de conceitos e de AM. Pôde-se levantar que os atributos mais importantes foram o de grau médio dos vizinhos do nó, sua profundidade e sua centralidade de autovetor. Complementarmente, os valores de *hub* (do algoritmo HITS (KLEINBERG, 1999)) e TF-IDF (utilizado originalmente por Dohare, Gupta e Karnick (2018)) se mostraram mais ineficientes. Apesar disso, o método de Dohare, Gupta e Karnick (2018) ainda obteve resultados satisfatórios, indicando que as regras de expansão do subgrafo relevante (por meio da ferramenta OpenIE) é eficiente na criação dos resumos.

Outra avaliação realizada diz respeito à taxa de compressão de cada método, mostrando que os métodos que mais se aproximam à taxa de compressão dos resumos de referência são o de Dohare, Gupta e Karnick (2018) e também o baseado em AM (com o algoritmo de MLP). Além disso, nas extremidades temos o método de Liu *et al.* (2015), com cada função de erro, *Perceptron* ou *Ramp*, gerando os menores e maiores resumos, respectivamente (em termos de número de conceitos).

Comparando-se os resultados dos métodos entre si, levando em consideração as taxas de compressão obtidas, pôde-se apontar o uso de Aprendizado de Máquina como sendo o melhor método entre os avaliados, com foco principal no algoritmo de *Multilayer Perceptron*, levando em consideração tanto a qualidade quanto o tamanho dos resumos produzidos. É necessário, porém, ter-se em mente que muitos testes estatísticos mostraram pouca diferença significativa entre os diversos algoritmos de AM explorados. Também, o fato de não ter sido feito um ajuste dos hiperparâmetros pode ter beneficiado alguns métodos e prejudicado outros.

A fim de avaliar, dentro da área de Sumarização de Opiniões, o impacto de se utilizar métodos baseados em representações semânticas explícitas (nomeadamente, a AMR), foram analisados as ferramentas de Ganesan, Zhai e Han (2010) e Condori e Pardo (2017). Apesar de certas ressalvas quanto ao uso da medida ROUGE como avaliação de resumos abstrativos, foi observado que o uso de semântica explícita levou a melhores resultados.

Por fim, a partir desses resultados, pode-se fazer uma avaliação de cada hipótese levantada ao começo da pesquisa, apresentadas no [Seção 1.2 \(Capítulo 1\)](#).

Referente à anotação do corpus, pode-se corroborar a hipótese de que existem fenômenos encontrados durante o processo de anotação que necessitam de um cuidado maior fora das diretrizes originais da representação. Porém, é importante apontar que a intuição inicial de que esses seriam fenômenos específicos do gênero opinativo não se confirma, visto que os fenômenos encontrados (expressões multipalavras, intervalos de tempo ou marcadores discursivos) podem ocorrer também nos gêneros sobre os quais a AMR foi originalmente desenvolvida (jornalístico e científico). Discutivelmente, pode-se levantar a possibilidade de que as expressões metafóricas sejam mais frequentes no gênero opinativo – ou menos comuns em textos jornalísticos e científicos –, porém, é necessária uma análise linguística mais detalhada desses textos para se poder afirmar algo nesse sentido.

Com relação à ideia de que existem fenômenos da língua portuguesa que ainda não possuíam diretrizes de anotação bem definidas, pode-se destacar o uso do diminutivo sintético, dividindo-se entre dois tipos: semânticos e pragmáticos ([ALVES, 2006](#)). Visto que esse fenômeno não foi contemplado nos trabalhos de anotação anteriores para a língua portuguesa, pode-se considerar essa hipótese confirmada.

No que se refere à hipótese de que existem conceitos e relações específicas que são mais importantes para a criação dos resumos, foi possível encontrar, através de uma análise estatística do corpus, alguns padrões, como a eliminação de entidades nomeadas (principalmente livros) e o destaque de relações do tipo *:manner*. Não foi possível delinear, entretanto, um conjunto de regras de sumarização a partir dessa análise. Dessa forma, essa hipótese pode ser considerada parcialmente corroborada, sendo necessário um detalhamento mais profundo de análise linguística para dar mais força a esses resultados.

Quanto à sumarização, a partir da comparação com métodos de sumarização de opiniões – principalmente o de [Ganesan, Zhai e Han \(2010\)](#) – foi possível comprovar a hipótese principal de que o uso de métodos baseados em semântica explícita produzem melhores resultados de acordo com as medidas utilizadas.

A hipótese de que a inclusão de informações provenientes da Análise de Sentimentos – anotação de sentimentos e aspectos – poderia gerar resultados melhores não foi corroborada pelas análises dos atributos realizadas, visto que esses atributos pouco influenciaram na qualidade dos resumos (analisando-se os métodos de otimização de pontuação por *Simulated Annealing* e classificação por AM).

6.1.1 Contribuições

De forma sucinta, as contribuições deste trabalho são:

- Exploração de métodos de sumarização de opiniões com base em AMR;
- Um corpus de opiniões parcialmente anotado em AMR;
- Uma lista de verbos faltantes no recurso Verbo Brasil, permitindo futuras expansões;
- Descrição de casos de anotação mais difícil;
- Análise comparativa de como fenômenos da língua se comportam em AMR em diferentes gêneros textuais;
- Implementação de ferramenta de sumarização de opiniões com base em semântica (SemOpinionS), incluindo diferentes técnicas, tanto originais da literatura quanto novas;
- Avaliação dos métodos de sumarização de opiniões (baseadas em semântica ou não).

6.2 Trabalhos Futuros

Ao longo do trabalho e de sua avaliação surgem, naturalmente, ideias e possibilidades de trabalhos futuros. Primeiramente, no que diz respeito à anotação do corpus em AMR, é importante a continuação desse trabalho sobre as sentenças restantes do OpiSums-PT (tanto dos comentários quanto dos resumos). Além disso, é necessária a expansão do repositório Verbo Brasil, com os predicadores, ou sentidos, faltantes; dessa forma a anotação de certas sentenças poderá ser viabilizada. A partir das dificuldades de anotação, aparecem algumas necessidades mais específicas, como uma melhor formalização linguístico-computacional da semântica de metáforas e expressões multipalavras.

A respeito dos métodos de sumarização, é interessante investigar os métodos baseados em ILP – de Liu *et al.* (2015) e Liao, Lebanoff e Liu (2018) – por uma perspectiva unicamente baseada na seleção das arestas mais importantes, visto que os aqui avaliados já possuem uma restrição que seleciona, obrigatoriamente, os conceitos relacionados, ao se escolher uma relação; além disso, os atributos dos nós estão também já codificados na representação das arestas, permitindo que um método focado nas arestas possa ser mais leve (com menos atributos) e, ainda assim, produzir resultados comparáveis aos de Liao, Lebanoff e Liu (2018). Essa abordagem pode também suprir a observação de Liao, Lebanoff e Liu (2018) de que a seleção de relações relevantes é mais difícil do que a de conceitos.

Também é interessante a avaliação dos atributos desenvolvidos por Liu *et al.* (2015) nos métodos baseados em ILP, de modo a encontrar quais realmente são importantes à sumarização. O atributo de frequência do conceito, por exemplo, tem um dos pontos de binarização em zero, o que pode não ser o ideal, visto que todos os conceitos terão, ao menos,

uma ocorrência – se estão sendo representados, eles existem no texto –, gerando apenas um tipo de valor para esse atributo em específico, não auxiliando, assim, na distinção dos nós. Dessa forma, uma melhor definição dos atributos, bem como suas combinações em uma representação única do grafo, se faz pertinente.

Ainda sobre atributos, também é importante melhorar o conjunto utilizado nos métodos de pontuação e Aprendizado de Máquina, já que eles têm como base apenas a estrutura do grafo. Como indicado pelos próprios autores do trabalho que motivou o uso desse tipo de informação, [Leskovec, Milic-Frayling e Grobelnik \(2005\)](#), as informações estruturais não são suficientes para determinar a importância dos conceitos, sendo necessário, para se obter melhores resultados, a inclusão de “atributos linguísticos”. Apesar disso, não é citado, pelos autores, exatamente quais atributos são utilizados além de etiquetas morfossintáticas e gênero¹. Dessa maneira, é necessário um trabalho mais minucioso de quais informações linguísticas podem ser incluídas nesse contexto. É importante salientar que esse conjunto de atributos melhorado também pode ter seu comportamento avaliado sobre os métodos de ILP, como o de [Liao, Lebanoff e Liu \(2018\)](#) investigado neste projeto.

Com relação às regras desenvolvidas por [Dohare, Gupta e Karnick \(2018\)](#) – e utilizadas em diversos métodos aqui avaliados – é interessante investigar outras seleções de caminhos importantes entre os conceitos importantes, visto que, em um cenário de sumarização multidocumento, é possível que a sentença mais importante com dois conceitos não seja a que ocorre “primeiro” no cópuz; o próprio conceito de “primeiro”, nesse contexto, se perde, visto que está sujeito à ordem arbitrária em que o cópuz foi concebido. A escolha dos caminhos importantes podem, por exemplo, se dar com base no grau de utilidade das sentenças, subárea da Mineração de Opiniões que mede o quão útil uma opinião é para o usuário tomar suas decisões.

Nesse trabalho, escolheu-se como método de otimização dos pesos, para a pontuação dos nós, o *Simulated Annealing*, devido à sua simplicidade de implementação. Porém, podem-se investigar outras abordagens de otimização, como Algoritmos Genéticos, tal qual realizado por [Khan et al. \(2018\)](#).

Uma grande melhora no fluxo de sumarização deste trabalho que pode ser realizada diz respeito à geração textual, de maneira a produzir o texto final do resumo a partir do grafo correspondente. Isso apresenta um desafio por si só no nível das sentenças, sendo mais complexo ainda no contexto deste projeto, visto que o grafo do resumo é uma representação semântica multi-sentencial do mesmo, sendo necessária uma etapa anterior de separação das sentenças no grafo e sua organização em um texto único e coerente. Para tanto, pode ser possível utilizar modelos de língua mais recentes, como o GPT-3 ([BROWN et al., 2020](#)) ou o BERT ([DEVLIN et al., 2019](#)), para gerar o texto de uma maneira guiada pelas informações presentes no grafo do sumário.

¹ Os autores indicam que utilizaram mais de 70 etiquetas semânticas, porém sem especificá-las.

A criação do grafo multi-sentencial também apresenta um desafio, sendo que a mesclagem dos nós de conceitos semelhantes pode não ser o suficiente para garantir uma consistência semântica na representação. Para tanto, pode-se investigar as aplicações do trabalho de O’Gorman *et al.* (2018), que desenvolve uma extensão, semanticamente motivada, da AMR para lidar com múltiplas sentenças por meio de uma nova camada de anotação.

REFERÊNCIAS

- ABEND, O.; RAPPOPORT, A. Universal conceptual cognitive annotation (UCCA). In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 51., 2013, Sófia. **Proceedings...** Sófia: Association for Computational Linguistics, 2013. v. 1, p. 228–238. Citado na página [42](#).
- AGGARWAL, C. C. **Machine Learning for Text**. Cham: Springer International Publishing, 2018. ISBN 978-3-319-73530-6. Citado na página [21](#).
- AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining Text Data**. Boston: Springer US, 2012. ISBN 978-1-4614-3222-7. Citado na página [21](#).
- AKHTAR, N.; ZUBAIR, N.; KUMAR, A.; AHMAD, T. Aspect based sentiment oriented summarization of hotel reviews. **Procedia Computer Science**, Elsevier, v. 115, p. 563–571, jan. 2017. ISSN 1877-0509. Citado na página [38](#).
- AKSOY, C.; BUGDAYCI, A.; GUR, T.; UYSAL, I.; CAN, F. Semantic argument frequency-based multi-document summarization. In: INTERNATIONAL SYMPOSIUM ON COMPUTER AND INFORMATION SCIENCES, 24., 2009, Guzelyurt. **Proceedings...** Guzelyurt: Institute of Electrical and Electronics Engineers, 2009. p. 460–464. Citado na página [51](#).
- ALVES, E. O diminutivo no português do Brasil: Funcionalidade e tipologia. **Estudos Linguísticos**, Campinas, v. 35, p. 694–701, 2006. Citado nas páginas [30](#), [73](#), [74](#) e [129](#).
- ANCHIÊTA, R. T.; CABEZUDO, M. A. S.; PARDO, T. A. S. SEMA: An extended semantic evaluation for AMR. In: GELBUKH, A. (Ed.). **Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing**. [S.l.]: Springer, 2019. No prelo. Citado nas páginas [70](#), [105](#) e [128](#).
- ANCHIÊTA, R. T.; PARDO, T. A. S. A Rule-Based AMR Parser for Portuguese. In: SIMARI, G. R.; FERMÉ, E.; SEGURA, F. G.; MELQUIADES, J. A. R. (Ed.). **Advances in Artificial Intelligence - IBERAMIA 2018**. Cham: Springer, 2018. v. 11238, p. 341–353. ISBN 978-3-030-03927-1 978-3-030-03928-8. Citado nas páginas [106](#) e [108](#).
- _____. Towards AMR-BR: A SemBank for brazilian portuguese language. In: CALZOLARI, N.; CHOUKRI, K.; CIERI, C.; DECLERCK, T.; HASIDA, K.; ISAHARA, H.; MAEGAARD, B.; MARIANI, J.; MORENO, A.; ODIJK, J.; PIPERIDIS, S.; TOKUNAGA, T.; GOGGI, S.; MAZO, H. (Ed.). **Proceedings of the 11th International Conference on Language Resources and Evaluation**. Miyazaki: European Languages Resources Association, 2018. p. 974–979. Citado nas páginas [28](#), [29](#), [42](#) e [46](#).
- ANGELIDIS, S.; LAPATA, M. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2018, Bruxelas. **Proceedings...** Bruxelas: Association for Computational Linguistics, 2018. p. 3675–3686. Citado na página [38](#).

- AVANÇO, L. V. **Sobre Normalização e Classificação de Polaridade de Textos Opinativos Na Web**. 102 p. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, São Carlos, 2015. Citado nas páginas 34 e 37.
- BANARESCU, L.; BONIAL, C.; CAI, S.; GEORGESCU, M.; GRIFFITT, K.; HERM-JAKOB, U.; KNIGHT, K.; KOEHN, P.; PALMER, M.; SCHNEIDER, N. Abstract meaning representation for sembanking. In: PAREJA-LORA, A.; LIAKATA, M.; DIPPER, S. (Ed.). **Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse**. Sófia: Association for Computational Linguistics, 2013. p. 178–186. Citado nas páginas 27, 28, 42, 66, 67 e 72.
- BANKO, M.; CAFARELLA, M. J.; SODERLAND, S.; BOARDHEAD, M.; ETZIONI, O. Open information extraction from the web. **Communications of the ACM**, v. 51, n. 12, p. 68–74, 2008. ISSN 00010782. Citado nas páginas 55 e 94.
- BEINEKE, P.; HASTIE, T.; MANNING, C.; VAITHYANATHAN, S. Exploring sentiment summarization. In: QU, Y.; SHANAHAN, J.; WIEBE, J. (Ed.). **Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text Theories and Applications**. Menlo Park: AAAI Press, 2004. v. 07, p. 1–4. Citado nas páginas 58 e 66.
- BENDER, E. M.; KOLLER, A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 58., 2020. **Proceedings...** [S.l.]: Association for Computational Linguistics, 2020. p. 5185–5198. Citado na página 25.
- BERTAGLIA, T. F. C.; NUNES, M. d. G. V. Exploring word embeddings for unsupervised textual user-generated content normalization. In: HAN, B.; RITTER, A.; DERCZYNSKI, L.; XU, W.; BALDWIN, T. (Ed.). **Proceedings of the 2nd Workshop on Noisy User-Generated Text (WNUT)**. Osaka: The COLING 2016 Organizing Committee, 2016. p. 112–120. Citado nas páginas 35 e 68.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. Nova Iorque: Springer, 2006. (Information Science and Statistics). ISBN 978-0-387-31073-2. Citado nas páginas 101 e 103.
- BLAIR-GOLDENSOHN, S.; HANNAN, K.; MCDONALD, R.; NEYLON, T.; REIS, G.; REYNAR, J. Building a Sentiment Summarizer for Local Service Reviews. In: WORKSHOP ON NLP CHALLENGES IN THE INFORMATION EXPLOSION ERA, 2008, Pequim. **Proceedings...** Nova Iorque: Association for Computing Machinery, 2008. ISSN 2329924X. Citado nas páginas 38 e 58.
- BONIAL, C.; BADARAU, B.; GRIFFITT, K.; HERM-JAKOB, U.; KNIGHT, K.; O’GORMAN, T.; PALMER, M.; SCHNEIDER, N. Abstract meaning representation of constructions: The more we include, the better the representation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 11., 2018, Miyazaki. **Proceedings...** Miyazaki: European Language Resources Association, 2018. ISBN 979-10-95546-00-9. Citado na página 46.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, out. 2001. ISSN 1573-0565. Citado na página 118.

- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. **Computer networks and ISDN systems**, Elsevier, v. 30, n. 1-7, p. 107–117, 1998. Citado nas páginas 57 e 97.
- BROWN, T.; MANN, B.; RYDER, N.; SUBBIAH, M.; KAPLAN, J. D.; DHARIWAL, P.; NEELAKANTAN, A.; SHYAM, P.; SASTRY, G.; ASKELL, A.; AGARWAL, S.; Herbert-Voss, A.; KRUEGER, G.; HENIGHAN, T.; CHILD, R.; RAMESH, A.; ZIEGLER, D.; WU, J.; WINTER, C.; HESSE, C.; CHEN, M.; SIGLER, E.; LITWIN, M.; GRAY, S.; CHESSE, B.; CLARK, J.; BERNER, C.; MCCANDLISH, S.; RADFORD, A.; SUTSKEVER, I.; AMODEI, D. Language models are few-shot learners. In: LAROCHELLE, H.; RANZATO, M.; HADSELL, R.; BALCAN, M. F.; LIN, H. (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2020. v. 33, p. 1877–1901. Citado na página 131.
- BRUM, H.; NUNES, M. d. G. V. Building a sentiment corpus of tweets in brazilian portuguese. In: International Conference on Language Resources and Evaluation, 11., 2018, Miyazaki. **Proceedings...** Miyazaki: European Language Resources Association, 2018. Citado na página 49.
- CABEZUDO, M. A. S.; PARDO, T. Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In: LINGUISTIC ANNOTATION WORKSHOP, 13., 2019, Florença. **Proceedings...** Florença: Association for Computational Linguistics, 2019. p. 236–244. Citado nas páginas 29, 42, 46, 67, 68, 69, 70 e 75.
- CAI, S.; KNIGHT, K. Smatch: An evaluation metric for semantic feature structures. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 51., 2013, Sófia. **Proceedings...** Sófia, 2013. p. 748–752. ISBN 978-1-937284-51-0. ISSN 13607456. Citado nas páginas 56, 69, 91, 92, 93 e 105.
- CAMARGO, R. T.; AGOSTINI, V.; FELIPPO, A. D.; PARDO, T. A. Manual Typification of Source Texts and Multi-document Summaries Alignments. **Procedia - Social and Behavioral Sciences**, v. 95, p. 498–506, out. 2013. ISSN 18770428. Citado na página 25.
- CAMBRIA, E. An introduction to concept-level sentiment analysis. In: CASTRO, F.; GELBUKH, A.; GONZÁLEZ, M. (Ed.). **Advances in Soft Computing and Its Applications**. Berlim: Springer, 2013. p. 478–483. ISBN 978-3-642-45110-2. ISSN 03029743. Citado nas páginas 26, 27 e 35.
- CAMBRIA, E.; PORIA, S.; BISIO, F.; BAJPAI, R.; CHATURVEDI, I. The CLSA model: A novel framework for concept-level sentiment analysis. In: GELBUKH, A. (Ed.). **Computational Linguistics and Intelligent Text Processing**. Cham: Springer, 2015. p. 3–22. ISBN 978-3-319-18116-5. ISSN 16113349. Citado nas páginas 35, 36, 37 e 127.
- CAMPBELL, R.; SUZUKI, H. **Language-Neutral Syntax : An Overview**. Redmond, 2002. 27 p. Citado na página 57.
- CAO, Z.; WEI, F.; LI, W.; LI, S. Faithful to the original: Fact aware neural abstractive summarization. In: MCILRAITH, S. A.; WEINBERGER, K. Q. (Ed.). **Proceedings of the 32nd AAAI Conference on Artificial Intelligence**. Nova Orleans: AAAI Press, 2018. p. 4784–4791. Citado nas páginas 25 e 58.

CARENINI, G.; NG, R.; PAULS, A. Multi-Document Summarization of Evaluative Text. In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 11., 2006, Trento. **Proceedings...** Trento: Association for Computational Linguistics, 2006. p. 305–312. Citado nas páginas [15](#), [59](#), [60](#), [64](#), [65](#) e [66](#).

CHIERCHIA, G. **Semântica**. Tradução de L. A. Pagani, L. Negri e R. Ilari. Campinas: Editora da UNICAMP, 2003. ISBN 85-268-0652-1. Citado na página [41](#).

CONDORI, R. E. L.; PARDO, T. A. S. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. **Expert Systems With Applications**, Elsevier, v. 78, p. 124–134, 2017. Citado nas páginas [15](#), [19](#), [24](#), [38](#), [49](#), [58](#), [63](#), [64](#), [65](#), [66](#), [123](#), [124](#) e [128](#).

CONDORI, R. E. L.; PARDO, T. A. S.; AVANÇO, L. V.; FILHO, P.; BOKAN, A.; CARDOSO, P.; DIAS, M.; NÓBREGA, F.; CABEZUDO, M.; SOUZA, J.; ZACARIAS, A.; SENO, E.; Di Felippo, A. A Qualitative Analysis of a Corpus of Opinion Summaries based on Aspects. In: LINGUISTIC ANNOTATION WORKSHOP, 9., 2015, Denver. **Proceedings...** Denver: Association for Computational Linguistics, 2015. p. 62–71. Citado nas páginas [21](#), [22](#), [26](#), [29](#), [64](#), [67](#), [68](#), [99](#), [106](#) e [127](#).

CORDER, G. W.; FOREMAN, D. I. **Nonparametric Statistics: A step-by-step approach**. [S.l.]: Wiley, 2011. ISBN 978-1-118-16582-9. Citado na página [109](#).

DAVE, K.; LAWRENCE, S.; PENNOCK, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 12., 2003, Budapeste. **Proceedings...** Nova Iorque: Association for Computing Machinery, 2003. p. 519–528. ISBN 1-58113-680-3. ISSN 12709638. Citado nas páginas [34](#), [35](#) e [37](#).

DENISKO, D.; HOFFMAN, M. M. Classification and interaction in random forests. **Proceedings of the National Academy of Sciences**, v. 115, n. 8, p. 1690–1692, fev. 2018. ISSN 0027-8424, 1091-6490. Citado na página [102](#).

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES, 2019. **Proceedings...** Minneapolis: Association for Computational Linguistics, 2019. v. 1, p. 4171–4186. Citado na página [131](#).

DOHARE, S.; GUPTA, V.; KARNICK, H. Unsupervised Semantic Abstractive Summarization. In: SHWARTZ, V.; TABASSUM, J.; VOIGT, R.; CHE, W.; MARNEFFE, M.-C. de; NISSIM, M. (Ed.). **Proceedings of ACL 2018, Student Research Workshop**. Melbourne: Association for Computational Linguistics, 2018. p. 74–83. ISBN 9781538607527. Citado nas páginas [15](#), [25](#), [28](#), [42](#), [51](#), [52](#), [53](#), [54](#), [55](#), [56](#), [65](#), [83](#), [85](#), [93](#), [94](#), [95](#), [106](#), [119](#), [121](#), [122](#), [128](#) e [131](#).

DONATELLI, L.; REGAN, M.; CROFT, W.; SCHNEIDER, N. Annotation of tense and aspect semantics for sentential AMR. In: SAVARY, A.; RAMISCH, C.; HWANG, J. D.; SCHNEIDER, N.; ANDRESEN, M.; PRADHAN, S.; PETRUCK, M. R. L. (Ed.). **Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions**

and Constructions (LAW-MWE-CxG-2018). Santa Fe: Association for Computational Linguistics, 2018. p. 96–108. Citado na página 46.

DONG, Y.; FU, Y.; WANG, L.; CHEN, Y.; DONG, Y.; LI, J. A sentiment analysis method of capsule network based on BiLSTM. **IEEE Access**, v. 8, p. 37014–37020, 2020. Citado na página 34.

DUCHI, J.; HAZAN, E.; SINGER, Y. Adaptive subgradient methods for online learning and stochastic optimization. **Journal of Machine Learning Research**, v. 12, n. 61, p. 2121–2159, 2011. Citado na página 89.

DURAN, M. S.; ALUÍSIO, S. M. Automatic Generation of a Lexical Resource to support Semantic Role Labeling in Portuguese. In: JOINT CONFERENCE ON LEXICAL AND COMPUTATIONAL SEMANTICS, 4., 2015, Denver. **Proceedings...** Denver: Association for Computational Linguistics, 2015. p. 216–221. Citado nas páginas 28, 31, 42, 69 e 127.

FAGIOLO, G. Clustering in complex directed networks. **Physical Review E**, American Physical Society, v. 76, n. 2, ago. 2007. ISSN 1539-3755, 1550-2376. Citado na página 98.

FELDMAN, R.; SANGER, J. Advanced approaches in analyzing unstructured data. In: **The Text Mining Handbook**. Cambridge: Cambridge University Press, 2006. p. 410. ISBN 978-0-511-54691-4. Citado nas páginas 21, 33 e 34.

FILLMORE, C. J. The case for case. In: BACH, E.; HARMS, R. T. (Org.). **Universals in Linguistic Theory**. Nova Iorque: Holt, Rinehart and Winston, 1968. p. 1–88. Citado na página 41.

FLANIGAN, J.; THOMSON, S.; CARBONELL, J.; DYER, C.; SMITH, N. A. A discriminative graph-based parser for the abstract meaning representation. In: Annual Meeting of the Association for Computational Linguistics, 52., 2014, Baltimore. **Proceedings...** Baltimore: Association for Computational Linguistics, 2014. v. 1, p. 1426–1436. Citado na página 52.

FREITAS, C.; MOTTA, E.; MILIDIÚ, R. L.; CÉSAR, J. Sparkling vampire... LOL! Annotating opinions in a book review corpus. In: Corpus Linguistics Symposium, 11., 2012, São Carlos. **New Language Technologies and Linguistic Research: A two-way road**. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014. p. 128–146. Citado nas páginas 67 e 106.

GANESAN, K.; ZHAI, C.; HAN, J. Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 23., 2010, Pequim. **Proceedings...** Pequim: Coling 2010 Organizing Committee, 2010. p. 340–348. Citado nas páginas 15, 38, 60, 61, 62, 65, 66, 123, 124, 128 e 129.

GATT, A.; KRAHMER, E. Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. **Journal of Artificial Intelligence Research**, v. 61, p. 65–170, jan. 2018. ISSN 1076-9757. Citado na página 23.

GEHRMANN, S.; DENG, Y.; RUSH, A. Bottom-up abstractive summarization. In: RILOFF, E.; CHIANG, D.; HOCKENMAIER, J.; TSUJII, J. (Ed.). **Proceedings of the**

2018 Conference on Empirical Methods in Natural Language Processing. Bruxelles: Association for Computational Linguistics, 2018. p. 4098–4109. Citado na página [25](#).

GENEST, P.-E.; LAPALME, G. Framework for Abstractive Summarization using Text-to-Text Generation. In: WORKSHOP ON MONOLINGUAL TEXT-TO-TEXT GENERATION, 2011, Portland. **Proceedings...** Portland: Association for Computational Linguistics, 2011. p. 64–73. Citado nas páginas [26](#), [51](#), [57](#), [58](#) e [124](#).

GREENBACKER, C. Towards a Framework for Abstractive Summarization of Multimodal Documents. In: ACL 2011 STUDENT SESSION, 2011, Portland. **Proceedings...** Portland: Association for Computational Linguistics, 2011. p. 75–80. Citado nas páginas [25](#) e [58](#).

HAGBERG, A. A.; SCHULT, D. A.; SWART, P. J. Exploring network structure, dynamics, and function using NetworkX. In: VAROQUAUX, G.; VAUGHT, T.; MILLMAN, J. (Ed.). **Proceedings of the 7th Python in Science Conference**. Pasadena, 2008. p. 11–15. Citado na página [106](#).

HARABAGIU, S. M.; LĂCĂTUȘU, F. Generating single and multi-document summaries with GISTEXTER. In: DOCUMENT UNDERSTANDING CONFERENCE, 2002, Filadélfia. **Proceedings...** Filadélfia, 2002. p. 30–38. Citado na página [51](#).

HERMANN, K. M.; KOCISKY, T.; GREFFENSTETTE, E.; ESPEHOLT, L.; KAY, W.; SULEYMAN, M.; BLUNSOM, P. Teaching machines to read and comprehend. In: CORTES, C.; LAWRENCE, N. D.; LEE, D. D.; SUGIYAMA, M.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems 28**: Annual conference on neural information processing systems. Montreal: Curran Associates, Inc., 2015. p. 1693–1701. Citado na página [55](#).

HOVY, E.; LAVID, J. Towards a ‘science’ of corpus annotation: A new methodological challenge for corpus linguistics. **International Journal of Translation**, v. 22, n. 1, p. 25, 2010. Citado nas páginas [48](#), [49](#), [67](#) e [70](#).

HU, M.; LIU, B. Mining and summarizing customer reviews. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 10., 2004, Nova Iorque. **Proceedings...** Nova Iorque: Association for Computing Machinery, 2004. p. 168–177. ISBN 1581138889. Citado nas páginas [38](#), [58](#) e [66](#).

HUANG, L.; WU, L.; WANG, L. Knowledge Graph-Augmented Abstractive Summarization with Semantic-Driven Cloze Reward. In: JURAFSKY, D.; CHAI, J.; SCHLUTER, N.; TETREAULT, J. (Ed.). **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. [S.l.]: Association for Computational Linguistics, 2020. p. 5094–5107. Citado nas páginas [24](#), [26](#) e [124](#).

HUETTNER, A.; SUBASIC, P. Fuzzy typing for document management. **ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes**, p. 26–27, 2000. Citado na página [34](#).

JACKENDOFF, R. **Semantic Structures**. Cambridge: The MIT Press, 1990. v. 18. (Current Studies in Linguistics, v. 18). Citado na página [41](#).

JURAFSKY, D.; MARTIN, J. H. Vector Semantics and Embeddings. In: **Speech and Language Processing**. 3. ed. Upper Saddle River: Prentice Hall, 2019. cap. 6. Citado na página 55.

KAMAL, A. Subjectivity classification using machine learning techniques for mining feature-opinion pairs from web opinion sources. **International Journal of Computer Science Issues**, v. 10, n. 5, p. 191–200, 2013. Citado nas páginas 34, 36 e 37.

KE, P.; JI, H.; LIU, S.; ZHU, X.; HUANG, M. SentiLARE: Sentiment-aware language representation learning with linguistic knowledge. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2020. **Proceedings...** [S.l.]: Association for Computational Linguistics, 2020. p. 6975–6988. Citado na página 35.

KHAN, A.; SALIM, N.; FARMAN, H. Clustered genetic semantic graph approach for multi-document abstractive summarization. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS ENGINEERING, 2016, Islamabad. **2016 International Conference on Intelligent Systems Engineering (ICISE)**. Islamabad: Institute of Electrical and Electronics Engineers, 2016. p. 63–70. Citado na página 25.

KHAN, A.; SALIM, N.; FARMAN, H.; KHAN, M.; JAN, B.; AHMAD, A.; AHMED, I.; PAUL, A. Abstractive text summarization based on improved semantic graph approach. **International Journal of Parallel Programming**, v. 46, n. 5, p. 992–1016, 2018. ISSN 1573-7640. Citado nas páginas 58, 124 e 131.

KHAN, A.; SALIM, N.; KUMAR, Y. J. A framework for multi-document abstractive summarization based on semantic role labelling. **Applied Soft Computing**, Elsevier, v. 30, p. 737–747, maio 2015. ISSN 1568-4946. Citado nas páginas 51, 57 e 58.

KIM, H. D.; GANESAN, K.; SONDHI, P.; ZHAI, C. **Comprehensive Review of Opinion Summarization**. Urbana, 2011. 30 p. Disponível em: <https://www.ideals.illinois.edu/bitstream/handle/2142/18702/survey_opinionSummarization.pdf>. Acesso em: 23 abr. 2021. Citado nas páginas 23, 37 e 38.

KIM, H. D.; ZHAI, C. Generating comparative summaries of contradictory opinions in text. In: CHEUNG, D. W.-L.; SONG, I.-Y.; CHU, W. W.; HU, X.; LIN, J. J. (Ed.). **Proceeding of the 18th ACM Conference on Information and Knowledge Management - CIKM '09**. Nova Iorque: Association for Computing Machinery, 2009. p. 385–394. ISBN 978-1-60558-512-3. Citado na página 39.

KINGSBURY, P.; PALMER, M. From Treebank to PropBank. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 3., 2002, Las Palmas. **Proceedings...** Las Palmas: European Language Resources Association, 2002. p. 1989–1993. ISBN 978-0-87421-656-1. Citado nas páginas 28 e 42.

KIPPER, K.; KORHONEN, A.; RYANT, N.; PALMER, M. A large-scale classification of english verbs. **Language Resources and Evaluation**, v. 42, p. 21–40, 2008. Citado na página 42.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by Simulated Annealing. **Science**, American Association for the Advancement of Science, v. 220, n. 4598, p. 671–680, maio 1983. ISSN 1095-9203. Citado nas páginas 99 e 100.

- KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. **Journal of the ACM**, Association for Computing Machinery, Nova Iorque, v. 46, n. 5, p. 604–632, set. 1999. ISSN 0004-5411. Citado nas páginas [97](#) e [128](#).
- KNIGHT, K.; BARANESCU, L.; BONIAL, C.; GEORGESCU, M.; GRIFFITT, K.; HERMJAKOB, U.; MARCU, D.; PALMER, M.; SCHNEIDER, N. Abstract meaning representation (AMR) annotation release 1.0 LDC2014T12. Linguistic Data Consortium, Filadélfia, 2014. Citado na página [53](#).
- KONSTAS, I.; IYER, S.; YATSKAR, M.; CHOI, Y.; ZETTLEMOYER, L. Neural AMR: Sequence-to-sequence models for parsing and generation. In: Annual Meeting of the Association for Computational Linguistics, 55., 2017, Vancouver. **Proceedings...** Vancouver: Association for Computational Linguistics, 2017. v. 1, p. 146–157. Citado na página [55](#).
- KU, L.-W.; LIANG, Y.-T.; CHEN, H.-H. Opinion extraction, summarization and tracking in news and blog corpora. In: NICOLOV, N.; SALVETTI, F.; LIBERMAN, M.; MARTIN, J. H. (Ed.). **Computational Approaches to Analyzing Weblogs**. Stanford, 2006. p. 100–107. ISBN 1-57735-264-5. Citado na página [39](#).
- LAARHOVEN, P. J. M.; AARTS, E. H. L. **Simulated Annealing**: Theory and applications. 1. ed. Dordrecht: Springer, 1987. v. 37. 187 p. (Mathematics and Its Applications, v. 37). ISBN 978-94-015-7744-1. Citado na página [99](#).
- LAPPIN, S. Semantics. In: MITKOV, R. (Ed.). **The Oxford Handbook of Computational Linguistics**. Nova Iorque: Oxford University Press, 2003. p. 91–111. Citado na página [27](#).
- LEE, K.; HE, L.; LEWIS, M.; ZETTLEMOYER, L. End-to-end neural coreference resolution. In: Conference on Empirical Methods in Natural Language Processing, 2017, Copenhagen. **Proceedings...** Copenhagen: Association for Computational Linguistics, 2017. p. 188–197. Citado nas páginas [54](#) e [55](#).
- LEGROSKI, M. Definindo Metáfora. **Voos Revista Polidisciplinar Eletrônica da Faculdade Guairacá**, v. 1, n. 2, 2010. ISSN 1808-9305. Citado na página [70](#).
- LESKOVEC, J.; GROBELNIK, M.; Milic-Frayling, N. Learning Semantic Graph Mapping for Document Summarization. In: ECML/PKDD Workshop ON Knowledge Discovery AND Ontologies, 2004. **Proceedings...** Pisa: Pisa University Press, 2004. p. 1–6. Citado na página [58](#).
- LESKOVEC, J.; Milic-Frayling, N.; GROBELNIK, M. **Extracting Summary Sentences Based on the Document Semantic Graph**. Redmond, 2005. Citado nas páginas [25](#), [96](#), [101](#) e [131](#).
- LEVIN, B. **English Verb Classes and Alternations**: A preliminary investigation. Chicago: University of Chicago Press, 1993. ISBN 978-0-226-47533-2. Citado na página [42](#).
- LI, W. Abstractive multi-document summarization with semantic information extraction. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2015, Lisboa. **Proceedings...** Lisboa: Association for Computational Linguistics, 2015. p. 1908–1913. Citado nas páginas [26](#), [51](#), [57](#), [58](#) e [124](#).

- LI, W.; HE, L.; ZHUGE, H. Abstractive News Summarization based on Event Semantic Link Network. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 26., 2016, Osaka. **Proceedings**. Osaka: Association for Computational Linguistics, 2016. p. 236–246. Citado nas páginas 57 e 58.
- LI, X.; BING, L.; ZHANG, W.; LAM, W. Exploiting BERT for end-to-end aspect-based sentiment analysis. In: XU, W.; RITTER, A.; BALDWIN, T.; RAHIMI, A. (Ed.). **Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)**. Hong Kong: Association for Computational Linguistics, 2019. p. 34–41. Citado na página 35.
- LIAO, K.; LEBANOFF, L.; LIU, F. Abstract Meaning Representation for Multi-Document Summarization. In: BENDER, E. M.; DERCZYNSKI, L.; ISABELLE, P. (Ed.). **Proceedings of the 27th International Conference on Computational Linguistics**. Santa Fé: Association for Computational Linguistics, 2018. p. 1178–1190. Citado nas páginas 13, 15, 28, 42, 52, 56, 57, 65, 85, 86, 87, 88, 91, 93, 106, 111, 130 e 131.
- LIN, C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries. In: MOENS, M.-F.; SZAPKOWICZ, S. (Ed.). **Text Summarization Branches Out**. Barcelona: Association for Computational Linguistics, 2004. p. 74–81. Citado nas páginas 53 e 105.
- LIU, B. Sentiment analysis and opinion mining. In: HIRST, G. (Ed.). **Synthesis Lectures on Human Language Technologies**. [S.l.]: Morgan & Claypool Publishers, 2012. v. 5, p. 167. ISBN 978-1-60845-884-4. Citado nas páginas 23, 33, 34 e 37.
- LIU, B.; HU, M.; CHENG, J. Opinion observer: analyzing and comparing opinions on the Web. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 14., 2005, Chiba. **Proceedings...** Nova Iorque: Association for Computing Machinery, 2005. p. 342–351. ISBN 1-59593-046-9. Citado nas páginas 39 e 40.
- LIU, F.; FLANIGAN, J.; THOMSON, S.; SADEH, N.; SMITH, N. A. Toward Abstractive Summarization Using Semantic Representations. In: MIHALCEA, R.; CHAI, J.; SARKAR, A. (Ed.). **Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. Denver: Association for Computational Linguistics, 2015. p. 1077–1086. Citado nas páginas 13, 15, 28, 42, 51, 52, 53, 54, 56, 65, 66, 83, 85, 86, 87, 88, 89, 90, 91, 106, 109, 121, 124, 128 e 130.
- LIU, Y.; LAPATA, M. Text summarization with pretrained encoders. In: INUI, K.; JIANG, J.; NG, V.; WAN, X. (Ed.). **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**. Hong Kong: Association for Computational Linguistics, 2019. p. 3730–3740. Citado nas páginas 25 e 58.
- LOVINGER, J.; VALOVA, I.; CLOUGH, C. Gist: General integrated summarization of text and reviews. **Soft Computing**, Springer, Berlim, v. 23, n. 5, p. 1589–1601, mar. 2019. ISSN 1433-7479. Citado na página 38.
- LU, Y.; ZHAI, C.; SUNDARESAN, N. Rated aspect summarization of short comments. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 18., 2009, Madri. **Proceedings...** Madri: Association for Computing Machinery, 2009. p. 131–140. Citado nas páginas 38 e 58.

LUDERMIR, T. B.; YAMAZAKI, A.; ZANCHETTIN, C. An optimization methodology for neural network weights and architectures. **IEEE Transactions on Neural Networks**, Institute of Electrical and Electronics Engineers, v. 17, n. 6, p. 1452–1459, nov. 2006. Citado nas páginas 99, 101, 102 e 107.

MANI, I. **Automatic Summarization**. Amsterdam: John Benjamins Publishing Company, 2001. v. 3. 298 p. (Natural Language Processing, v. 3). ISBN 90-272-4986-5. Citado nas páginas 21, 22, 23, 24, 26 e 66.

MARTINS, C. B.; RINO, L. H. M. Pruning UNL texts for Summarizing Purposes. In: NATURAL LANGUAGE PROCESSING PACIFIC RIM SYMPOSIUM, 6., 2001, Tóquio. **Proceedings...** Tóquio, 2001. p. 539–544. Citado na página 58.

_____. **Heurísticas de Poda de Sentenças Para a Sumarização Automática de Textos UNL**. São Carlos, 2002. Citado na página 51.

MATTHIESSEN, C. M. I. M.; BATEMAN, J. A. **Text Generation and Systemic-Functional Linguistics: Experiences from English and Japanese**. Londres: Pinter Pub Ltd, 1992. Citado nas páginas 28 e 43.

MCENERY, T. Corpus linguistics. In: MITKOV, R. (Ed.). **The Oxford Handbook of Computational Linguistics**. Nova Iorque: Oxford University Press, 2003. p. 448–463. Citado na página 48.

Miguelés-Abraira, N. **A Study Towards Spanish Abstract Meaning Representation**. 98 p. Dissertação (Mestrado em Análise e Processamento da Linguagem) — Universidad del País Vasco, Donostia-San Sebastián, jun. 2017. Citado na página 42.

MISHNE, G.; RIJKE, M. de. MoodViews: Tools for blog mood analysis. In: AAAI SYMPOSIUM ON COMPUTATIONAL APPROACHES TO ANALYSING WEBLOGS, 2006, Palo Alto. **Papers from the AAAI Spring Symposium**. Palo Alto: AAAI Press, 2006. SS-06-03, p. 100–107. ISBN 1-57735-264-5. Citado nas páginas 39 e 40.

MITHUN, S.; KOSSEIM, L. Summarizing blog entries versus news texts. In: ORASAN, C.; HASLER, L.; FORASCU, C. (Ed.). **Proceedings of the Workshop on Events in Emerging Text Types**. Borovets: Association for Computational Linguistics, 2009. p. 1–8. ISBN 978-954-452-011-3. Citado na página 37.

MOAWAD, I. F.; AREF, M. Semantic graph reduction approach for abstractive Text Summarization. In: INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING & SYSTEMS, 7., 2012, Cairo. **Proceedings...** Piscataway: Institute of Electrical and Electronics Engineers, 2012. p. 132–138. Citado nas páginas 57 e 58.

MORAES, S. M.; SANTOS, A. L.; REDECKER, M.; MACHADO, R. M.; MENEGUZZI, F. R. Comparing approaches to subjectivity classification: A study on Portuguese tweets. In: SILVA, J.; RIBEIRO, R.; QUARESMA, P.; ADAMI, A.; BRANCO, A. (Ed.). **Computational Processing of the Portuguese Language**. Cham: Springer, 2016. v. 9727, p. 86–94. ISBN 978-3-319-41551-2. ISSN 16113349. Citado na página 36.

MORETTIN, P. A.; BUSSAB, W. d. O. **Estatística básica**. São Paulo: Saraiva, 2010. ISBN 978-85-02-08177-2. Citado na página 87.

NENKOVA, A.; MCKEOWN, K. A Survey on Text Summarization Techniques. In: AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining Text Data**. Boston, Massachusetts, USA: Springer US, 2012. p. 43–76. ISBN 978-1-4614-3222-7. Citado na página 22.

NEWMAN, M. E. J. **Networks: An introduction**. Oxford: Oxford University Press, 2010. ISBN 978-0-19-920665-0. Citado nas páginas 96 e 98.

NG, J.-P.; ABRECHT, V. Better Summarization Evaluation with Word Embeddings for ROUGE. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2015, Lisboa. **Proceedings...** Lisboa: Association for Computational Linguistics, 2015. p. 1925–1930. Citado na página 124.

O’GORMAN, T.; REGAN, M.; GRIFFITT, K.; HERMJAKOB, U.; KNIGHT, K.; PALMER, M. AMR beyond the sentence: The multi-sentence AMR corpus. In: BENDER, E. M.; DERCZYNSKI, L.; ISABELLE, P. (Ed.). **Proceedings of the 27th International Conference on Computational Linguistics**. Santa Fé: Association for Computational Linguistics, 2018. p. 3693–3702. Citado na página 132.

OLIVEIRA, L. de; CLARO, D. B. DptOIE: A portuguese Open Information Extraction system based on dependency analysis. **Computer Speech and Language Journal**, out. 2019. No prelo. Disponível em: <<http://repositorio.ufba.br/ri/handle/ri/30719>>. Acesso em: 27 abr. 2021. Citado nas páginas 94 e 104.

PANG, B.; LEE, L. **Opinion Mining and Sentiment Analysis**. Hanover: now Publishers Inc, 2008. 135 p. ISBN 978-1-60198-150-9. Citado nas páginas 23 e 34.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment classification using machine learning techniques. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2002, Filadélfia. **Proceedings...** Filadélfia, 2002. p. 79–86. ISBN 978-1-4503-4068-7. ISSN 0003150X. Citado nas páginas 34 e 37.

PECAR, S. Towards opinion summarization of customer reviews. In: SHWARTZ, V.; TABASSUM, J.; VOIGT, R.; CHE, W.; MARNEFFE, M.-C. de; NISSIM, M. (Ed.). **Proceedings of ACL 2018, Student Research Workshop**. Melbourne: Association for Computational Linguistics, 2018. p. 1–8. Citado na página 38.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Citado na página 106.

PEREIRA, D. A. A survey of sentiment analysis in the Portuguese language. **Artificial Intelligence Review**, v. 54, n. 2, p. 1087–1115, 2021. Citado nas páginas 23 e 34.

POURDAMGHANI, N.; GAO, Y.; HERMJAKOB, U.; KNIGHT, K. Aligning English strings with Abstract Meaning Representation graphs. In: MOSCHITTI, A.; PANG, B.; DAELEMANS, W. (Ed.). **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha: Association for Computational Linguistics, 2014. p. 425–429. Citado na página 108.

- PUSTEJOVSKY, J.; STUBBS, A. **Natural Language Annotation for Machine Learning**. 1. ed. Sebastopol: O'Reilly Media, 2013. ISBN 978-1-4493-0666-3. Citado na página 49.
- RADEV, D. R.; TEUFEL, S.; SAGGION, H.; LAM, W.; BLITZER, J.; QI, H.; ÇELEBI, A.; LIU, D.; DRABEK, E. Evaluation challenges in large-scale document summarization. In: Annual Meeting of the Association for Computational Linguistics, 41., Sapporo. **Proceedings...** Sapporo: Association for Computational Linguistics, 2003. p. 375–382. Citado nas páginas 59, 60 e 62.
- REAL, L.; OSHIRO, M.; MAFRA, A. B2W-Reviews01: An open product reviews corpus. In: CONTE, T.; CASTRO, T.; SILVA, B. S. da; BARBOSA, S. D. J. (Ed.). **Proceedings of the 12th Brazilian Symposium in Information and Human Language Technology**. Salvador: Sociedade Brasileira de Computação, 2019. p. 200–208. Citado nas páginas 96 e 106.
- REIMER, U.; HAHN, U. Text condensation as knowledge base abstraction. In: CONFERENCE ON ARTIFICIAL INTELLIGENCE APPLICATIONS, 4., 1988, San Diego. **Proceedings...** San Diego: Institute of Electrical and Electronics Engineers, 1988. p. 338–344. Citado nas páginas 51 e 66.
- RUSU, D.; FORTUNA, B.; GROBELNIK, M.; MLADENIÆ, D. Semantic graphs derived from triplets with application in document summarization. **Informatica**, v. 33, n. 3, p. 357–362, 2009. ISSN 03505596. Citado na página 51.
- SANTOS, D. Corporizando algumas questões. In: TAGNIN, S. E. O.; VALE, O. A. (Ed.). **Avanços Da Linguística de Corpus No Brasil**. São Paulo: Humanitas, 2008. p. 41–66. Citado na página 47.
- SARDINHA, T. B. Corpus Linguistics: History and problematization. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, v. 16, n. 2, p. 323–367, 2000. ISSN 0102-4450. Citado na página 48.
- SCHANK, R. C. **Conceptual Information Processing**. Amsterdam: North-Holland, 1975. Citado na página 41.
- SEE, A.; LIU, P. J.; MANNING, C. D. Get to the point: Summarization with pointer-generator networks. In: BARZILAY, R.; KAN, M.-Y. (Ed.). **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Vancouver: Association for Computational Linguistics, 2017. v. 1, p. 1073–1083. Citado nas páginas 26 e 58.
- Sherry; BHATIA, P. Multilingual text summarization with UNL. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTER ENGINEERING AND APPLICATIONS, 2015, Ghaziabad. **2015 International Conference on Advances in Computer Engineering and Applications**. Ghaziabad: Institute of Electrical and Electronics Engineers, 2015. p. 740–745. ISBN 978-1-4673-6911-4. Citado na página 25.
- SIERKSMA, G.; ZWOLS, Y. **Linear and Integer Optimization: Theory and practice**. 3. ed. Boca Ratón: Chapman & Hall/CRC, 2015. (Advances in Applied Mathematics). ISBN 978-1-4987-1016-9. Citado nas páginas 86 e 91.

SILVA, R. R. **Sumarização Contrastiva de Opinião**. 151 p. Dissertação (Mestrado Em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, São Carlos, 2019. Citado na página 41.

SILVA, R. R.; PARDO, T. A. S. **Córpus 4P: Um córpus anotado de opiniões em português sobre produtos eletrônicos para fins de sumarização contrastiva de opinião**. In: JORNADA DE DESCRIÇÃO DO PORTUGUÊS, 6., 2019, Salvador. **Anais...** Salvador, 2019. p. 330–338. Citado na página 49.

SORNLERLAMVANICH, V.; POTIPITI, T.; CHAROENPORN, T. UNL Document Summarization. In: INTERNATIONAL WORKSHOP ON MULTIMEDIA ANNOTATION, 1., 2001, Tóquio. **Proceedings...** Tóquio, 2001. Citado na página 58.

SOUZA, M.; VIEIRA, R.; Busetti, D.; CHISHMAN, R.; ALVES, I. M.; UNISINOS, F. D. L. Construction of a portuguese opinion lexicon from multiple resources. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 8., 2011, Cuiabá. **Proceedings...** Cuiabá, 2011. Citado na página 99.

SPECIA, L.; RINO, L. H. **Representação Semântica: Alguns Modelos Ilustrativos**. São Carlos, 2002. 33 p. Citado nas páginas 27 e 41.

TITOV, I.; MCDONALD, R. Modeling online reviews with multi-grain topic models. In: INTERNATIONAL CONFERENCE ON World Wide Web, 17., 2008, Pequim. **Proceeding...** Nova Iorque: Association for Computing Machinery, 2008. p. 111–120. ISBN 978-1-60558-085-2. Citado na página 38.

UCHIDA, H.; ZHU, M.; SENTA, T. D. **A Gift for a Millenium**. Tóquio: United Nations University, 1999. Citado nas páginas 41 e 57.

VANDERWENDE, L.; BANKO, M.; MENEZES, A. Event-centric summary generation. In: DOCUMENT UNDERSTANDING CONFERENCE, 2004, Boston. **Proceedings...** Boston, 2004. p. 76–81. Citado nas páginas 51 e 58.

VARGAS, F. A.; PARDO, T. A. S. Aspect clustering methods for sentiment analysis. In: VILLAVICENCIO, A.; MOREIRA, V.; ABAD, A.; CASELI, H.; GAMALLO, P.; RAMISCH, C.; OLIVEIRA, H. G.; PAETZOLD, G. H. (Ed.). **Proceedings of the 13th International Conference on the Computational Processing of Portuguese (PROPOR)**. Cham: Springer, 2018a. v. 11122, p. 365–374. ISBN 978-3-319-99721-6. ISSN 16113349. Citado na página 38.

_____. Hierarchical clustering of aspects for opinion mining: A corpus study. In: FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Org.). **Linguística de Corpus : Perspectivas**. Porto Alegre: Instituto de Letras da UFRGS, 2018b. Citado na página 49.

Vieira da Silva, F. J.; ROMAN, N. T.; CARVALHO, A. M. Stock market tweets annotated with emotions. **Corpora**, v. 15, n. 3, p. 343–354, nov. 2020. ISSN 1749-5032, 1755-1676. Citado na página 49.

von Luxburg, U. A tutorial on spectral clustering. **Statistics and Computing**, v. 17, n. 4, p. 395–416, dez. 2007. ISSN 0960-3174, 1573-1375. Citado na página 91.

WANG, D.; LI, T.; ZHU, S.; DING, C. H. Q. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In: CHUA, T.-S.; LEONG, M.-K.; MYAENG, S. H.; OARD, D. W.; SEBASTIANI, F. (Ed.). **Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**. Singapore: Association for Computing Machinery, 2008. p. 307–314. Citado na página 51.

WANG, L.; LING, W. Neural network-based abstract generation for opinions and arguments. In: KNIGHT, K.; NENKOVA, A.; RAMBOW, O. (Ed.). **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. San Diego: Association for Computational Linguistics, 2016. p. 47–57. Citado na página 38.

ZHUANG, L.; JING, F.; ZHU, X.-Y. Movie review mining and summarization. In: YU, P. S.; TSOTRAS, V.; FOX, E.; LIU, B. (Ed.). **Proceedings of the 15th ACM International Conference on Information and Knowledge Management - CIKM '06**. Nova Iorque: Association for Computing Machinery, 2006. p. 43. ISBN 1-59593-433-2. Citado na página 38.

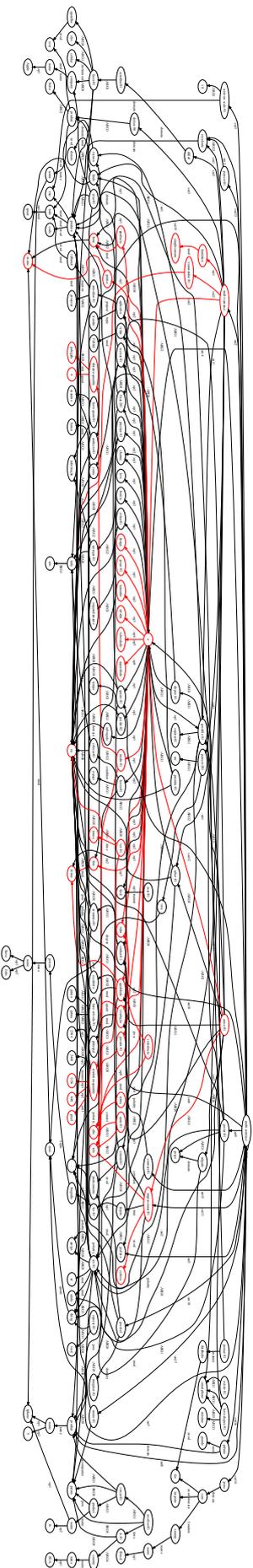
EXEMPLOS DE GRAFOS AMR

Este apêndice tem como objetivo apresentar exemplos de grafos de comentários sobre um determinado produto (provenientes da junção de várias sentenças anotadas em AMR), bem como um resumo correspondente. Os elementos do grafo (nós e arestas) que também estão presentes no resumo – isto é, o subgrafo sumarizado – estão apresentados em destaque na cor vermelha. Em seguida é apresentado o texto gerado automaticamente para o resumo, no formato *bag of words*. Todos esses procedimentos são explicados com detalhes no [Capítulo 5](#).

O primeiro exemplo é apresentado na [Figura 43](#), cujo grafo provém dos comentários (anotados manualmente) sobre o produto “iPhone 5”. O subgrafo destacado diz respeito a um resumo de referência criado por um humano. São apresentados também o texto original (do resumo de referência) e o correspondente ao grafo por meio do método de geração automática.

Em seguida, é apresentado na [Figura 44](#) um exemplo sobre o produto “Galaxy SIII”, cujos comentários foram, por sua vez, anotados automaticamente, o que é evidente devido à ocorrência de várias relações do tipo `:null_edge`.

Figura 43 – Exemplo de grafo AMR e resumo para o produto “Iphone 5”



(a) Grafo AMR

A tela com certeza é algo que impressiona, possuindo 4 polegadas, uma resolução de 1136 x 640 pixels, e ainda com um vidro de alta resistência.

Com certeza trata-se de um iOS muito avançado, e apesar de tudo e mais um pouco que qualquer pessoa deseje em um celular, não é pesado e nem muito grande, além de ser extremamente fino com apenas 7.6mm, o que facilita na hora de carregá-lo para todos os lugares.

Muito prático, com boa aparência, e ótima qualidade. DESDE O DESIGN, AO PESO, TAMANHO, NITIDEZ, VELOCIDADE, CAPACIDADE, OTIMAS FOTOS, TELA DE RETINA (NUNCA VI IGUAL).

Simplesmente fantástico! Eu recomendo!

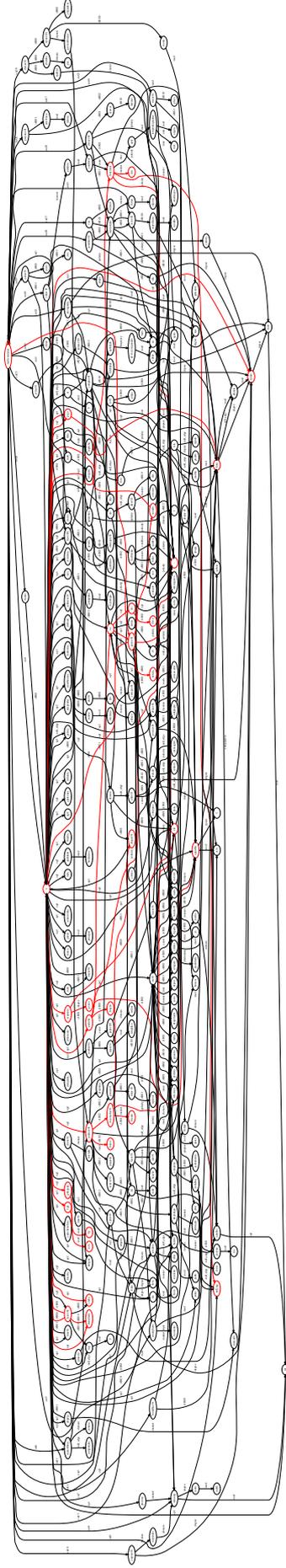
(c) Resumo de referência

impressiona tela certeza ainda e algo possuindo resolução vidro prático aparência qualidades design ao tamanho nitidez velocidade capacidade fotos ex muito bom ótimo polegadas pixels resistência alta retina peso vi que nunca no fantástico simplesmente recomendando no

(b) Bag of words correspondente ao subgrafo de resumo

Fonte: Dados da pesquisa.

Figura 44 – Exemplo de grafo AMR e resumo para o produto “Galaxy SIII”



(a) Grafo AMR

gostei núcleos notícias e jogos 1gb densidade po-
legadas muito vídeos permite melhor ram faz 0p
função aparelho ou ler livros esquenta ele pouco
mas baixo incomodar chega branca funcionando
sim brasil

(b) *Bag of words* correspondente ao subgrafo de resumo

O que gostei: Ótimo desempenho em jogos, Processador de quatro núcleos de 1,4GHz e 1GB de RAM, Display Super AMOLED 4.8 Polegadas 720p e 306 PPI de densidade de pixels que faz com que seja muito bom para ver vídeos e para jogos em HD, Função não permite que o aparelho desligue enquanto se estiver olhando para ele oque é bom para que goste de ler livros ou goste de ler notícias na internet,

O que não gostei: Ele esquenta um Pouco na parte de baixo mas não chega a incomodar, na cor branca ele parece ser muito frágil e o S Voice ainda não funciona em português.

(c) Resumo de referência

Fonte: Dados da pesquisa.

