

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

JOÃO EDUARDO KOZAN SILVA FORNAZARI

**RALEM: UMA FERRAMENTA PARA AUXILIAR NA IDENTIFICAÇÃO DE
PALAVRAS SIGNIFICATIVAS EM CONVERSAS EM FERRAMENTAS ON-LINE
COLABORATIVAS ATRAVÉS DE ANÁLISE LÉXICO-MORFOLÓGICA**

PONTA GROSSA

2021

JOÃO EDUARDO KOZAN SILVA FORNAZARI

**RALEM: UMA FERRAMENTA PARA AUXILIAR NA IDENTIFICAÇÃO DE
PALAVRAS SIGNIFICATIVAS EM CONVERSAS EM FERRAMENTAS ON-LINE
COLABORATIVAS ATRAVÉS DE ANÁLISE LÉXICO-MORFOLÓGICA**

**RALeM: A tool to aid on identifying meaningful words in conversations on
online collaborative tools through lexical-morphological analysis.**

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Ciência da Computação, do Departamento de Informática da Universidade Tecnológica Federal do Paraná.

Orientadora: Profa. Dra. Monica Hoetdke Pietruchinski
Coorientador: Prof. Dr. Gleifer Vaz Alves

PONTA GROSSA

2021

JOÃO EDUARDO KOZAN SILVA FORNAZARI

**RALEM: UMA FERRAMENTA PARA AUXILIAR NA IDENTIFICAÇÃO DE
PALAVRAS SIGNIFICATIVAS EM CONVERSAS EM FERRAMENTAS ON-LINE
COLABORATIVAS ATRAVÉS DE ANÁLISE LÉXICO-MORFOLÓGICA**

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Ciência da Computação, do Departamento de Informática da Universidade Tecnológica Federal do Paraná.

Data de aprovação: 25/Agosto/2021

Mônica Hoeldtke Pietruchinski
Doutorado
Universidade Tecnológica Federal do Paraná

Gleifer Vaz Alves
Doutorado
Universidade Tecnológica Federal do Paraná

Helyane Bronoski Borges
Doutorado
Universidade Tecnológica Federal do Paraná

André Pinz Borges
Doutorado
Universidade Tecnológica Federal do Paraná

PONTA GROSSA

2021

Dedico este trabalho a Nosso Senhor
Jesus Cristo e à sua Mãe, Maria
Santíssima; também à minha família, que
foi compreensiva comigo; e à Brendha
Maria, que me impulsionou para que eu
me dedicasse ao máximo a este trabalho.

RESUMO

Avaliar a colaboração entre indivíduos usando ferramentas colaborativas como salas de bate-papo requer a necessidade de analisar o significado das palavras nas frases emitidas, utilizando para isso processamento de linguagem natural. Um mapeamento sistemático foi realizado com a finalidade de identificar técnicas e ferramentas para realizar a análise léxico-morfológica das conversas. Deste mapeamento encontrou-se a LIWC (*Language of Inquiry and Word Count*), que foi incorporada ao aplicativo de *chat* Discord através do desenvolvimento de um software de tipo "bot", o RALeM (Robô de Análise Léxico-Morfológica), que permite o monitoramento das mensagens trocadas no chat, contabilizando as palavras identificadas para contribuir no processo de contagem de palavras de colaboração significativas.

Palavras-chave: Análise léxica; Análise morfológica; Processamento de linguagem natural; Ferramentas colaborativas; Sala de bate-papo.

ABSTRACT

Evaluating collaboration between people using collaborative tools like chat rooms requires the necessity to analyze words' meaning on emitted sentences, using natural language processing to do so. A systematic mapping was made aiming to identify techniques and tools to perform lexical-morphological analysis of the conversations. From this mapping LIWC (Language of Inquiry and Word Count) was found, and it was embedded to the chat application Discord through the development of a "bot"-like software, RALeM (Lexical-Morphological Analysis Bot), which allows monitoring the messages exchanged on the chat, counting identified words to contribute on meaningful collaborative words counting process.

Keywords: Lexical analysis; Morphological analysis; Natural language processing; Collaborative tools; Chat rooms.

LISTA DE ILUSTRAÇÕES

Figura 1	Processamento de Linguagem Natural: da sentença até sua forma lógica.	16
Figura 2	Captura de tela do aplicativo de voz Discord.	19
Figura 3	Categorias listadas no arquivo LIWC2007.dic.	23
Figura 4	Categorias listadas no arquivo LIWC2007.dic.	23
Figura 5	Ilustração do modo de funcionamento do ArqMAEC.	25
Figura 6	Captura de tela da seção de desenvolvedores do Discord.	36
Figura 7	Captura de tela da seção de desenvolvedores do Discord.	37
Figura 8	Tela de adicionar bot ao servidor.	37
Figura 9	Tela de adicionar bot ao servidor, com o botão Autorizar.	38
Figura 10	Interação entre Main.java e DictionaryLIWC.java.	36
Figura 11	Tela de adicionar bot ao servidor, com o botão Autorizar.	40
Figura 12	Resposta do bot à frase 'bom dia' emitida pelo usuário.	40
Figura 13	Resposta do bot à frase 'bom dia' emitida pelo usuário.	41
Figura 14	Resposta do bot ao comando -nao.	41
Figura 15	Parte da resposta do bot ao comando -sim.	42
Figura 16	Final da resposta do bot ao comando -sim.	43
Figura 17	Últimas interações do bot na sequência de operações --nova-categoria, inserindo uma categoria nova.	43
Figura 18	Resposta do bot ao comando --cancelar.	43
Figura 19	Fluxo do bot na sequência de operações --nova-categoria, onde se insere categorias já conhecidas pelo dicionário do bot.	43
Figura 20	Parte da resposta do bot ao comando --ver-categorias.	44
Figura 21	Interação usuário-bot ao adicionar categorias já existentes a uma palavra.	45
Figura 22	Interação usuário-bot ao adicionar categoria nova a uma palavra.	45
Figura 23	Interação usuário-bot ao deletar uma palavra não cadastrada em tempo de execução.	46
Figura 24	Interação usuário-bot ao deletar uma palavra cadastrada em tempo de execução.	46
Figura 25	Interação usuário-bot ao editar palavra não cadastrada em tempo de execução.	46
Figura 26	Interação ao editar palavra cadastrada em tempo de execução.	47
Figura 27	Resultado dos comandos --ocultar-bot-cat e --mostrar-bot-cat.	47
Figura 28	Interações do fluxo lógico do comando --acompanhar-cat.	48
Figura 29	Resultado do comando --ajuda.	48
Figura 30	Esquema ilustrado da interação Discord-RALeM.	49
Figura 31	Interação-teste entre membros do servidor-teste do bot.	50
Figura 32	Interação usuário-bot ao utilizar o comando --salvar.	50

Figura 33 Interação usuário-bot ao utilizar o comando --parar-acompanhar.....	51
Figura 34 Resultados da pontuação da interação-teste.	51
Figura 35 ArqMAEC com RALeM.....	53
Figura 36 Ilustração das contribuições do bot ao LIWC.	54

LISTA DE ACRÔNIMOS

ArqMAEC	Arquitetura de Monitoramento, Avaliação e Estimulação da Colaboração
EaD	Educação à Distância
RALeM	Robô de Análise Léxico-Morfológica
SciELO	Scientific Electronic Library Online

LISTA DE SIGLAS

al	Agente Individual
aG	Agente de Grupo
AMT	<i>Amazon Mechanical Turk</i>
ANEW	<i>Affective Norms for English Words</i>
CE-PLN	Comissão Especial de Processamento de Linguagem Natural
DAL	<i>Dictionary of Affect in Language</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
LEW	<i>List of Emotional Words</i>
LIWC	<i>Language of Inquiry and Word Count</i>
MRC	<i>Medical Research Council</i>
MSOL	<i>Macquarie Sentiment Orientation Lexicon</i>
PANAS	<i>Positive Affect Negative Affect Scale (</i>
PLN	Processamento de Linguagem Natural
SASA	<i>SailAil Sentiment Analyzer</i>
SBC	Sociedade Brasileira de Computação
STIL	<i>Symposium in Information and Human Language Technology</i>
TICs	Tecnologias da Informação e Comunicação
VADER	<i>Valence Aware Dictionary for Sentiment Reasoning</i>
VISL	<i>Visual Interactive Syntax Learning</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	MOTIVAÇÃO	12
1.2	OBJETIVOS	14
2	REFERENCIAL TEÓRICO	15
2.1	PROCESSAMENTO DE LINGUAGEM NATURAL	15
2.1.1	Análise Léxica	16
2.1.2	Análise Morfológica	17
2.1.3	Análise Sintática	17
2.2	FERRAMENTAS DE INFORMAÇÃO E COMUNICAÇÃO	17
2.2.1	Ferramentas Colaborativas	18
2.2.2	Ferramentas De Análise Léxico-Morfológica Aplicadas No Trabalho	19
2.3	LIWC	22
2.4	BOT	24
2.5	ARQMAEC	24
3	MAPEAMENTO SISTEMÁTICO DA LITERATURA	27
3.1	PERGUNTA	27
3.2	CRITÉRIOS DE INCLUSÃO	27
3.3	CRITÉRIOS DE EXCLUSÃO	28
3.4	BASES DE DADOS	28
3.5	PALAVRAS-CHAVE	29
3.6	RESULTADOS	30
3.7	AMEAÇAS À VALIDADE	34
3.8	CONSIDERAÇÕES FINAIS DO MAPEAMENTO	34
4	RESULTADOS	35
4.1.	PROJETO DO <i>BOT</i>	35
4.2	INSTALAÇÃO DO BOT NO AMBIENTE COLABORATIVO	36
4.3.	COMANDOS	39
4.4.	INTERAÇÃO	40
4.5.	TESTES	49
5	CONTRIBUIÇÃO E LIMITAÇÕES	53
6	CONCLUSÃO	55
7	TRABALHOS FUTUROS	57
	REFERÊNCIAS	58
	APÊNDICE A – QUADRO DA RELAÇÃO ENTRE ARTIGOS ENCONTRADOS NO MAPEAMENTO SISTEMÁTICO E FERRAMENTAS DE ANÁLISE LÉXICA	65
	APÊNDICE B - FLUXOGRAMAS LÓGICOS DOS COMANDOS DO RALEM	67

1 INTRODUÇÃO

Neste capítulo, apresenta-se o contexto que motivou o desenvolvimento deste trabalho, seu objetivo geral e seus objetivos específicos.

1.1 Motivação

A Educação à Distância (EaD), definida legalmente como “modalidade educacional na qual a mediação didático-pedagógica nos processos de ensino e aprendizagem ocorre com a utilização de meios e tecnologias de informação e comunicação, com estudantes e professores desenvolvendo atividades educativas em lugares ou tempos diversos” (BRASIL, 2005), surgiu amparada pela tecnologia e trouxe a democratização do conhecimento, tornando possível obter graduações e conhecimento sistematizado sem a necessidade de se deslocar até a instituição de ensino, influenciando até a educação presencial; e, como nota Borba (2020), o EaD “tornou-se, durante boa parte de 2020, a única forma possível de se pensar a educação”.

Porém, existem obstáculos que a EaD precisa superar para que seu impacto nos alunos seja equiparável ao da educação presencial, e um deles é a distância do professor. Essa distância “não é somente geográfica, mas vai além, configurando-se em uma distância transacional, ‘pedagógica’, a ser gerida por professores, alunos, monitores/tutores” (BEHAR, 2009) – isto é, uma distância entre aquilo que o professor ensina e aquilo que o aluno compreende. Diminuíam-se a distância pedagógica pela presença do professor em sala de aula, que conseguia perceber o comportamento dos alunos e obter reações imediatas sobre aquilo de que falou, prosseguindo com o tema, tirando dúvidas ou chamando a atenção de seus alunos.

Para tanto, surge o advento das Tecnologias da Informação e Comunicação (TICs) que visam “contribuir para ‘diminuir’ essa ‘distância pedagógica’, assegurando formas de comunicação e interação entre os ‘atores’ envolvidos no processo de construção de conhecimento pela EaD” (BEHAR, 2009), utilizando abordagens de diversas áreas para assessorar o professor e os alunos. Os *softwares* que são desenvolvidos nesse âmbito como ferramentas para aplicar uma abordagem desejada, “que tem como finalidade favorecer os processos de ensino/aprendizagem,

ou seja, ajudar o aluno a obter ou construir determinado conteúdo” (OLIVEIRA, 2001) são os chamados *softwares* educacionais.

A área de *games* chamou a atenção dos pesquisadores envolvidos no aprimoramento pedagógico da EaD por meio de *softwares*. Zirawaga (2017) observa que os jogos podem ser utilizados para auxiliar no processo de aprendizado, o que ocasionou o nascimento de uma nova área na pedagogia da EaD: gamificação.

A gamificação consiste em “um processo de aperfeiçoamento de um serviço usando elementos propícios para experiências típicas de jogos visando auxiliar o usuário a ter experiências significativas” (HAMARI, 2012). Realiza-se, então, a gamificação dos *softwares* educacionais, gerando uma nova categoria de *software* no âmbito EaD: *Software* Educacional Gamificado, que, conforme definido por Pietruchinski (2016), “é o *software* educacional que possui elementos de jogos com a finalidade de tornar o software mais atrativo aos aprendizes que o utilizam.”

Dentre os elementos que os jogos possuem, ressalta-se nesta pesquisa a colaboração entre jogadores, presente em jogos *multiplayer* colaborativos, onde não há competição entre os jogadores, mas sim colaboração mútua entre eles, desenvolvendo habilidades de comunicação, resolução de problemas, gestão de tempo e trabalho em equipe, como explica Bui (2018).

Trazendo este aspecto para os *softwares* educacionais gamificados, surge o *Software* Educacional Gamificado Colaborativo, *software* desenvolvido “para ensinar alguma habilidade ou conhecimento e que o faz através de atividades em equipes onde o objetivo principal seja construir o conhecimento juntos, compartilhando saberes e competências” (PIETRUCHINSKI, 2016), utilizando de elementos de jogos para incentivar a colaboração entre os alunos.

Com o objetivo de oferecer aos *softwares* educacionais colaborativos gamificados uma forma para monitorar, avaliar e estimular a colaboração entre os usuários encontrou-se a ArqMAEC (Arquitetura de Monitoramento, Avaliação e Estimulação da Colaboração) (PIETRUCHINSKI, 2016). Esta arquitetura, porém, não prevê como essa avaliação será realizada, e é esta a proposta do presente trabalho: uma ferramenta que realize parte do processo de avaliação da colaboração nestes *softwares*, servindo-se de análise léxico-morfológica para identificar palavras que podem indicar colaboração significativa.

A ferramenta proposta e desenvolvida neste trabalho chama-se RALeM (Robô de Análise Léxico-Morfológica), um *bot* inserido no aplicativo de *chat Discord*

(DISCORD, 2020) que realiza esta análise utilizando e aprimorando o dicionário gramático-psicológico LIWC (*Language of Inquiry and Word Count*) por meio de comandos que podem ser enviados nas mensagens enviadas no *Discord*. O RALeM utiliza o léxico fornecido pelo LIWC para acompanhar a conversa entre indivíduos no Discord, podendo assim realizar parte da avaliação da colaboração ocorrendo no *chat*.

É possível acrescentar novas palavras e novas categorias personalizadas ao dicionário, editar e deletar palavras cadastradas em tempo de execução, atualizar as palavras do dicionário acrescentando a elas as categorias recentemente criadas. Todas as alterações são registradas em uma nova versão definitiva do LIWC que pode ser reutilizada futuramente.

1.2 Objetivos

O objetivo geral deste trabalho é desenvolver uma ferramenta para auxiliar na identificação de palavras significativas em conversas on-line colaborativas através de análise léxico-morfológica. Os objetivos específicos são:

- Identificar o estado-da-arte das técnicas de análise léxico-morfológica, especialmente em Português;
- Escolher uma ferramenta para realizar análise léxico-morfológica;
- Acrescentar a funcionalidade de inserção, edição e remoção de novas categorias e novas palavras ao dicionário da ferramenta;
- Acrescentar à ferramenta a capacidade de contagem de palavras significativas emitidas por usuários do ambiente selecionado;
- Gerar arquivo externo com os dados coletados na ferramenta de análise léxico-morfológica;
- Testar a ferramenta no ambiente competitivo selecionado em um contexto pré-determinado e desconhecido pelos usuários do teste.

2 REFERENCIAL TEÓRICO

Neste capítulo, todo o arcabouço teórico que sustenta o desenvolvimento da ferramenta de análise léxico-morfológico será explicado, começando por processamento de linguagem natural, explicando suas etapas; depois, elucida-se o que são ferramentas colaborativas e quais delas foram aplicadas no presente trabalho; a seguir, explica-se o que são *bots* e onde são utilizados; apresenta-se o LIWC e como é estruturado; por fim, explana-se sobre a ArqMAEC.

2.1 Processamento de linguagem natural

Processamento de Linguagem Natural (PLN) é “um campo de pesquisa que surgiu da junção da inteligência artificial e da linguística” (DA SILVA, 2012) e pode ser definido grosso modo como “qualquer tipo de manipulação computacional da linguagem natural” (BIRD, 2009). Entende-se linguagem natural como “linguagem utilizada para comunicação cotidiana por seres humanos, como inglês, hindu ou português” (BIRD, 2009).

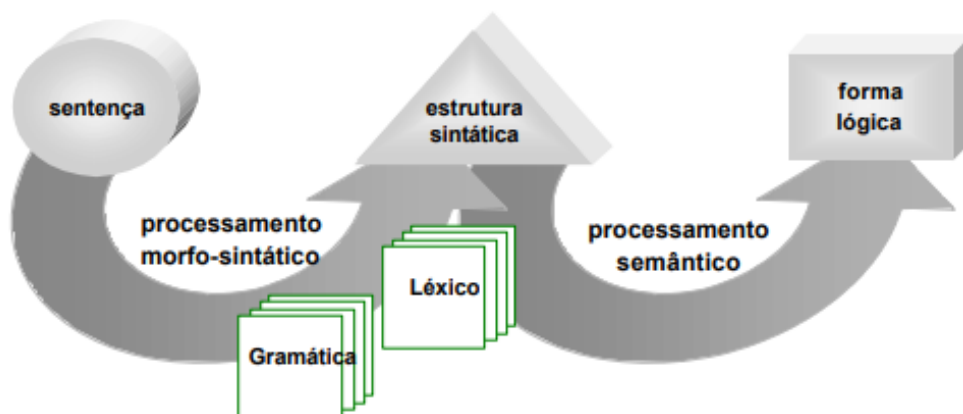
Gonzalez (2003) separa o Processamento de Linguagem Natural em cinco níveis:

1. Fonético e fonológico: relação entre as palavras e o som que produzem;
2. Morfológico: classificação da palavra conforme sua forma, elaborada a partir de significados simples;
3. Sintático: relação das palavras entre si formando frases, e das frases entre si formando sentenças;
4. Semântico: relação das palavras com seus significados e de sua combinação que origina o significado da sentença;
5. Pragmático: relação entre as frases e/ou sentenças com o contexto na qual são empregadas.

O primeiro nível do Processamento de Linguagem Natural só se faz presente quando o contexto é reconhecimento de fala – “a habilidade de uma máquina reconhecer quais palavras são ditas” (LEVIS, 2012) – e conseqüentemente não se analisa este nível de PLN em mensagens de texto, partindo diretamente para o nível

morfológico. Para processar a linguagem natural em mensagens de texto, primeiro obtêm-se a sentença que está na mensagem; depois, utiliza-se a gramática e o léxico do idioma da sentença para realizar o processamento morfossintático, identificando cada palavra da sentença – como verbos, substantivos e pronomes; por fim, realiza-se o processamento semântico para obter o significado da sentença, independentemente de seu contexto. A Figura 1 exemplifica o processo aqui descrito.

Figura 1 Processamento de Linguagem Natural: da sentença até sua forma lógica.



Fonte: GONZALEZ, 2003.

2.1.1 Análise Léxica

A análise léxica é uma das atividades executadas no processamento morfossintático descrito na Figura 2; pode ser definida como “o processo de decompor um texto em palavras, frases e outros elementos significativos” (THANAKI, 2017). A principal função de uma análise léxica é *tokenizar* as palavras presentes no texto de entrada do processamento. *Tokens* são “os elementos significativos que são gerados usando técnicas de análise lexical” (DA SILVA, 2019).

Exemplo: quando se tem de entrada a frase “Eu quero jogar futebol amanhã”, a análise léxica retorna um vetor com os *tokens*: [“Eu”, “quero”, “jogar”, “futebol”, “amanhã”]. Com este vetor de *tokens*, o processamento morfossintático pode prosseguir para a análise morfológica.

2.1.2 Análise Morfológica

A análise morfológica é o processo “responsável por categorizar de forma isolada cada palavra de acordo com que rege as regras de classificação de cada idioma” (DA SILVA, 2019), identificando em qual categoria morfológica a palavra *tokenizada* se encaixa – como verbo, substantivo ou adjetivo, no caso do português.

Com o mesmo exemplo utilizado na subseção anterior, a entrada seria cada um dos *tokens* (“Eu”, “quero”, “jogar”, “futebol”, “amanhã”) e a saída seria as categorias gramaticais de cada *token*: “Eu” é um pronome e substantivo masculino, “quero” é um verbo na primeira pessoa do singular, “jogar” é um verbo no infinitivo, “futebol” é um substantivo masculino e “amanhã” é um substantivo masculino e um advérbio.

2.1.3 Análise Sintática

Thanaki (2017) define a análise sintática como a análise que determina o significado lógico de uma sentença, seja em parte ou em sua totalidade. O processamento morfossintático, aqui, eleva-se do nível das palavras (análise léxico-morfológica) para o nível das frases. Utiliza-se regras gramaticais para validar a estrutura da sentença recebida.

Ainda trabalhando com a frase “Eu quero jogar futebol amanhã”, a entrada da análise sintática é esta sentença; a saída é a verificação de sua estrutura lógica, buscando o sujeito e o predicado.

- Existe sujeito e predicado na frase? Sim.
- Sujeito: “Eu”
- Predicado: “quero jogar futebol amanhã”.

2.2 Ferramentas de informação e comunicação

Apresenta-se aqui, no início, o que são ferramentas colaborativas, quais delas são relevantes para este trabalho e é discorrido sobre um aplicativo que é um exemplo de uma das ferramentas; finalizando a subseção, exhibe-se as ferramentas de análise léxica encontradas e qual delas foi aplicada no trabalho.

2.2.1 Ferramentas Colaborativas

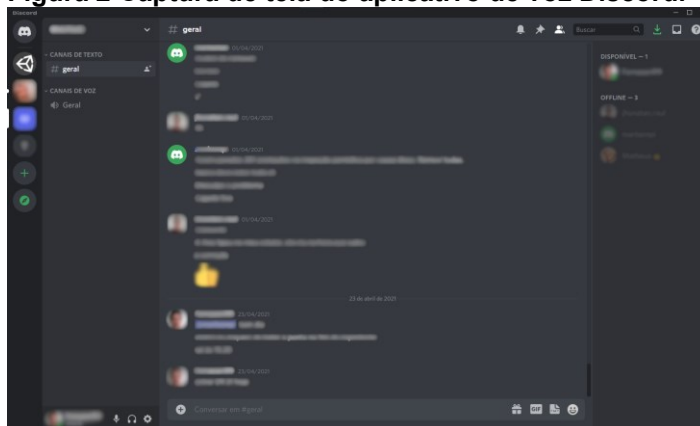
Nos *softwares* educacionais colaborativos gamificados, as ferramentas colaborativas são o meio pelo qual os alunos interagem entre si, gerando colaboração. Rabello (2010) classifica as ferramentas colaborativas em síncronas ou assíncronas, onde as síncronas permitem que os usuários da ferramenta se comuniquem em tempo real no mesmo ambiente e as assíncronas as mensagens são enviadas pelo emissor e são lidas, respondidas e/ou questionadas quando o remetente desejar – não necessariamente em tempo real. Para o presente trabalho, cita-se duas ferramentas colaborativas: o *chat* e os fóruns.

O *chat* (ou sala de bate-papo) é uma ferramenta colaborativa síncrona, utilizada amplamente no cotidiano por meio de aplicativos de dispositivos móveis – *WhatsApp, Telegram, Skype, Messenger, Discord* – e está presente também em ambientes virtuais de aprendizagem – como o *Moodle* ou *Google Classroom*. Lachi (2002) contrasta que no contexto educacional, apesar do *chat* ser um ambiente mais descontraído e proporcionar maior vazão às ideias de seus usuários, torna-se difícil para um professor avaliar a interação realizada nesses ambientes e identificar quantas contribuições significativas para a disciplina foram realizadas de fato.

Para o presente trabalho, o *Discord* é o *software* que chama mais atenção, por ser um ambiente *gamificado* pela presença e interação dos *bots*, que podem atribuir pontos para interações entre usuários que satisfazem critérios para obter conquistas. Em determinados contextos, apenas usuários com privilégios (como moderador ou dono do servidor) podem interagir com o *bot*, ordenando que execute comandos – semelhante ao ambiente colaborativo proposto na ArqMAEC (PIETRUCHINSKI, 2016).

O *Discord* também é gratuito; pode ser utilizado em qualquer sistema operacional, seja *desktop* ou *mobile*; e permite a inserção de *plugins*, como os *bots*. A interface padrão do *Discord* é exibida na Figura 2 a seguir.

Figura 2 Captura de tela do aplicativo de voz Discord.



Fonte: Autoria própria.

Pietruchinski (2016) apresenta fóruns da seguinte forma:

É uma ferramenta em que normalmente se colocam questões para serem discutidas ou respondidas pelos participantes. Nesta área um dos interlocutores apresenta a questão e os demais tem a oportunidade de responder, apresentar sugestões, dúvidas ou ainda contradizer a resposta (PIETRUCHINSKI, 2016, p. 33)

Reddit, *Describia* e *Quora* são *websites online* que utilizam fóruns como meio de interação entre seus usuários. No contexto da educação, o ambiente virtual de aprendizagem *Moodle* também permite a criação de fóruns de discussão nas disciplinas do professor. As pesquisas realizadas por Kenski (2009), Roig (2013) e Eastman (2002) apontam que o fórum é um bom ambiente para o diálogo textual e discussão de ideias por ser um ambiente menos descontraído, mas com acessibilidade semelhante ao *chat*.

2.2.2 Ferramentas de Análise Léxico-Morfológica Aplicadas no Trabalho

Muitos trabalhos encontrados no Mapeamento Sistemático (descrito na Seção 3) utilizam ferramentas de análise léxica preexistentes para parte do processamento morfossintático, *tokenizando* as palavras conforme convenção dos desenvolvedores. A abrangência da identificação morfológica varia para cada ferramenta também: WordNet (FELLBAUM, 2005) conta com 152.059 palavras somente do idioma inglês conforme relatado por Paim (2017); existem adaptações do WordNet para outros idiomas, como o WordNet-Br para o português (WORDNETBR, 2010), porém contém

menos palavras catalogadas. Já o ANEW (BRADLEY, 1999) contém 1034 vocábulos apenas, noticiado por Paim (2017).

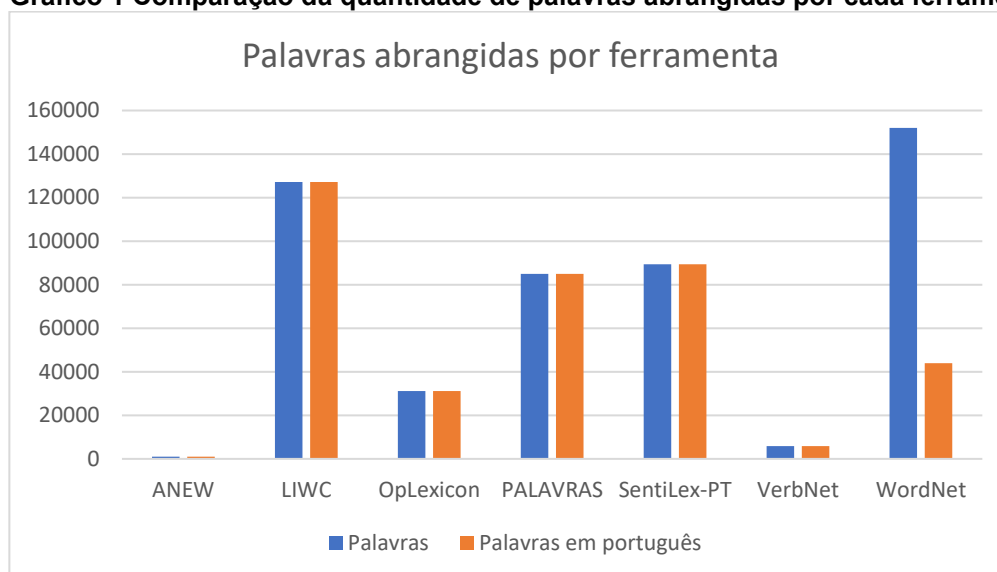
Das ferramentas encontradas no mapeamento sistemático (detalhado na Seção 3), as que possuem abordagem em português são PALAVRAS, ANEW (ANEW-Br, adaptado por Kristensen (2011)), WordNet-Br, SentiLex-PT, OpLexicon, VerbNet (versão adaptada por Scarton (2012)) e LIWC (adaptação de Filho (2013)) e as mais utilizadas são WordNet (quatro artigos), ANEW (três artigos) e WordNetAffect (três artigos), este sendo uma adaptação do WordNet para realizar também análise semântica. Como o critério de maior frequência não contempla muitas ferramentas para comparação, acrescenta-se as ferramentas que possuem abordagens para o idioma português, a saber: SentiLex-PT, VerbNet, OpLexicon, LIWC e PALAVRAS. Serão comparadas quanto à quantidade de palavras abrangidas e sua popularidade.

A Tabela 1 exibe a quantidade de palavras em sua abordagem original e em português das sete ferramentas selecionadas. Os dados foram retirados dos trabalhos de Paim (2016), Balage Filho (2013) e Bick (2000). O Gráfico 1 mostra uma comparação visual entre a quantidade de palavras de cada ferramenta.

Tabela 1 Quantidade de palavras abrangidas por cada ferramenta.

Ferramentas de análise léxica	Palavras	Palavras em português
ANEW	1034	1046
LIWC	127149	127162
OpLexicon	31144	31144
PALAVRAS	85000	85000
SentiLex-PT	89361	89361
VerbNet	5800	5800
WordNet	152059	44000

Fonte: Autoria própria.

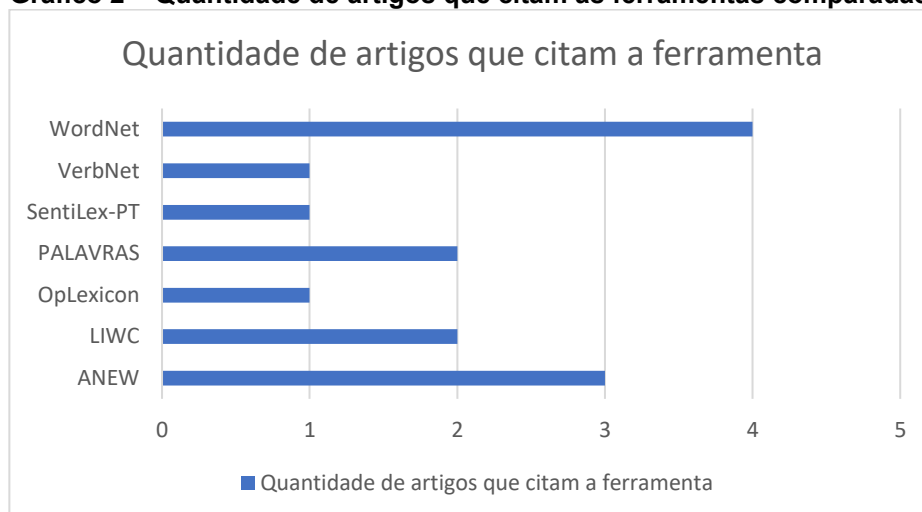
Gráfico 1 Comparação da quantidade de palavras abrangidas por cada ferramenta.

Fonte: Autoria própria.

Algumas considerações sobre as ferramentas:

- *SentiLex-PT* e *PALAVRAS* são desenvolvidas especificamente para o português;
- As adaptações de *LIWC* e *ANEW* para o português acrescentaram palavras únicas do português para maior abrangência do idioma;
- A versão brasileira de *VerbNet* utiliza de tradução semiautomática dos verbetes em inglês para aproveitar seus recursos no português;
- A versão brasileira de *WordNet* é uma versão menor do léxico original, não uma tradução;
- *OpLexicon* usa como base o léxico *OpinionLexicon* (HU, 2004).

O Gráfico 2 exhibe visualmente a quantidade de artigos que citam as ferramentas especificadas aqui. No Quadro 1 unifica-se as estatísticas apontadas aqui sobre cada ferramenta, mostra a média da quantidade de palavras de cada ferramenta e a média de popularidade. Com base nas médias, atribui-se uma pontuação para cada ferramenta. A ferramenta acima da média de popularidade ou na média de popularidade ganha um ponto, enquanto a abaixo da média não ganha ponto (P2); a ferramenta com palavras acima da média de palavras ganha um ponto, e a que estiver abaixo não ganha ponto (P1). Como critério de desempate, usa-se a quantidade de palavras.

Gráfico 2 – Quantidade de artigos que citam as ferramentas comparadas.

Fonte: Autoria própria.

Tabela 2 Estatísticas das ferramentas comparadas nesta seção.

Ferramentas	Palavras em PT-BR	Artigos	P1	P2	TOTAL
ANEW	1046	3	0	1	1
LIWC	127162	2	1	1	2
OpLexicon	31144	1	0	0	0
PALAVRAS	85000	2	1	1	2
SentiLex-PT	89361	1	1	0	1
VerbNet	5800	1	0	0	0
WordNet	44000	4	0	1	1
Média	54787	2	-	-	-

Fonte: Autoria própria.

Com base na quantidade de palavras e popularidade, obteve-se duas ferramentas candidatas para aplicação: LIWC e PALAVRAS. Usando o critério de desempate, a LIWC é a ferramenta vencedora e, portanto, afirma-se que LIWC é a melhor ferramenta para realizar análise léxica conforme critérios estabelecidos nesta subseção.

2.3 LIWC

A ferramenta *Linguistic Inquiry and Word Count* (LIWC) foi desenvolvida por Pennebaker (2001) com o intuito de “providenciar um método eficiente e eficaz para estudar os variados componentes estruturais, cognitivos e emocionais presentes em falas individuais escritas ou faladas” (PENNEBAKER, 2015), fornecendo categorias

léxicas para as palavras – como ‘verbo’, ‘pronome’ – e também psicológicas – como ‘emoção positiva’, ‘emoção negativa’.

Sua versão mais recente é de 2015 e pode ser encontrada online (LIWC, 2015), permitindo gratuitamente o teste de textos escritos de imediato ou de arquivos armazenados no computador; porém, é necessário pagar uma taxa para baixar o dicionário utilizado pela ferramenta. A versão em português mais recente disponível é de 2007 e é gratuita, podendo ser obtida pela internet (LIWC, 2013).

O dicionário do LIWC é um arquivo *.dic* onde primeiro constam as categorias e os números das respectivas categorias, depois começa a listagem das palavras juntamente com as categorias nas quais ela se encaixa (referidas pelos respectivos números). Usa-se o caractere de porcentagem (%) para separar as categorias das palavras no documento. As figuras 3 e 4 exibem essa organização.

Figura 3 Categorias listadas no arquivo LIWC2007.dic.

```
%
1      funct
2      pronoun
3      ppron
4      i
5      we
6      you
7      shehe
8      they
9      ipron
10     article
```

Fonte: Autoria própria.

Figura 4 Categorias listadas no arquivo LIWC2007.dic.

```
360    death
462    assent
463    nonfl
464    filler
%
a      1      2      3      6      7      9      10     17
121    131    138    252    463
aba    146
abafa  125    127    129
abafad* 125    127    129
abafada 125    127    129
abafadas 125    127    129
```

Fonte: Autoria própria.

Por conta da estrutura e formato simples do dicionário, é possível acrescentar palavras e categorias nele por meio de edição de texto, seja automática ou manual. Essa facilidade também permite que o dicionário seja utilizado por *softwares* externos.

2.4 Bot

De Paula (2019) apresenta *bots* como “programas de computador criados para automatizar tarefas e procedimentos repetitivos em ambiente digital”, auxiliando o ser humano na execução destas tarefas ou executando-as independente da intervenção humana. Os *bots* se tornaram populares em *telemarketing* e em ambientes de bate-papo onde há interação humana, respondendo perguntas frequentes e auxiliando o usuário a conseguir o que quer sem necessidade de outro humano respondendo suas perguntas, conforme relata Lebeuf (2018).

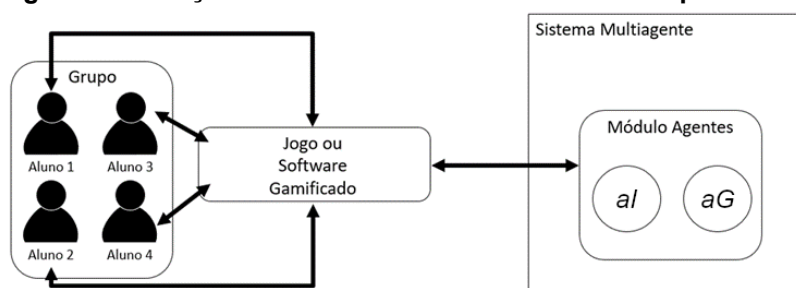
2.5 ARQMAEC

A ArqMAEC foi projetada para “automatizar o processo de monitoramento, avaliação e estimulação da colaboração entre aprendizes, através de agentes inteligentes integrados a um *software* educacional colaborativo gamificado” (PIETRUCHINSKI, 2016). Para que haja colaboração entre os aprendizes, a arquitetura estipula o uso de salas de bate-papo (ou *chat*) ou fóruns de discussão – ferramentas que permitem a comunicação síncrona ou assíncrona dos alunos e, assim, permitem a mútua colaboração, podendo ser chamadas de ferramentas colaborativas.

Monitoram-se as palavras emitidas pelos alunos em ferramentas colaborativas para medir a colaboração exercida por eles. Para analisar essa colaboração, Pietruchinski (2016) propõe a utilização da Teoria da Interdependência Social (JOHNSON, 2014), que supõe que “a interdependência social existe quando indivíduos compartilham objetivos comuns e o sucesso de cada um é afetado pelas ações dos demais” (PIETRUCHINSKI, 2016), necessitando de colaboração mútua entre indivíduos para alcançar determinado objetivo.

A forma como os alunos e os agentes se encaixam na arquitetura é ilustrada na Figura 5 exibida a seguir.

Figura 5 Ilustração do modo de funcionamento do ArqMAEC



Fonte: Autoria própria.

O sistema multiagente tem acesso à interação dos alunos, que é realizada dentro do *software* educacional gamificado por meio de mensagens emitidas em *chat* ou fóruns. Analisando essas mensagens, pode-se estimular os alunos que não estiverem apresentando determinadas características colaborativas e atribuir pontos para aqueles que estão colaborando conforme o esperado. Para realizar isso, os agentes estão organizados da seguinte forma:

O Agente Individual (*al*) tem a função de acompanhar as atividades do jogador: suas postagens, suas respostas às atividades propostas, enfim suas ações dentro do Software Gamificado (SG) e estimular o jogador que está sendo monitorado. O Agente de Grupo (*aG*), acompanha o grupo, registra suas pontuações coletivas, estimula as participações dos diferentes elementos da equipe ou propõe ao SG as ações para estimular a colaboração de todo o grupo. Cada um destes agentes é responsável pela detecção e estimulação da colaboração em seu nível particular. (PIETRUCHINSKI, 2016, p. 114)

É possível determinar qualitativamente a colaboração entre os alunos com as medidas e métricas estipuladas pela ArqMAEC para cada ferramenta colaborativa. Uma das medidas para *chat* ou fórum é calcular a quantidade de palavras na mensagem e, destas, a quantidade de palavras significativas. Calculando a proporção das palavras significativas com relação ao total de palavras da mensagem, tem-se uma das métricas para avaliar a mensagem emitida pelo aluno. Outra métrica estipulada por Pietruchinski (2016) é calcular a proporção do total de palavras de Incentivo à Colaboração com relação ao total de palavras significativas da mensagem.

Para que a avaliação e estímulo sejam postos em prática, é necessário um método ou ferramenta que identifique e classifique palavras nas mensagens emitidas pelos alunos no chat ou fórum da disciplina, obtendo, desta forma, as informações

necessárias para aplicar as medidas e, por conseguinte, as métricas; porém, a ArqMAEC não estabeleceu um método específico para executar essas atividades.

3 MAPEAMENTO SISTEMÁTICO DA LITERATURA

Para realizar a análise léxica das palavras emitidas pelos alunos no *software* educacional gamificado, deve-se ter conhecimento sobre o estado-da-arte das técnicas/ferramentas de análise léxica. Decidiu-se pelo mapeamento sistemático da literatura pertinente à análise léxica para garantir qualidade e abrangência nos artigos encontrados.

O mapeamento sistemático é “um método para construir esquemas de classificação e estruturar uma determinada área de interesse” (FASTFORMAT, 2020) e seu objetivo é determinar “a cobertura de uma determinada área de pesquisa ou tópico” (FASTFORMAT, 2020).

3.1 Pergunta

A pergunta que este mapeamento busca responder é: Quais são as técnicas de análise léxica em processamento de linguagem natural existentes para conversas em ambientes de aprendizagem?

3.2 Critérios de inclusão

Para incluir trabalhos na base de artigos do mapeamento sistemático, estabeleceu-se os seguintes critérios:

1. A técnica é utilizada em processamento de linguagem natural (PLN);
2. A técnica é aplicada no idioma português ou inglês;
3. A técnica foi testada em mensagens de salas de bate-papo ou fóruns de discussão;
4. O trabalho utiliza uma ferramenta de PLN;
5. A técnica foi utilizada em ambiente de ensino/aprendizagem.

Sobre o segundo critério, como o objetivo do presente trabalho é realizar análise léxica em palavras no português brasileiro, é preferível que os artigos utilizem técnicas sobre o português brasileiro também, mas caso não seja encontrada

nenhuma ferramenta adequada para a análise léxica em português, analisa-se as ferramentas/técnicas em inglês. Os artigos que indicarem pelo título e resumo/*abstract* que satisfazem ao menos um dos critérios de inclusão serão incluídos na base de artigos do mapeamento sistemático.

3.3 Critérios de exclusão

Os critérios de exclusão foram elaborados com base em perguntas de Revisão Sistemática específicas para Ciência da Computação (KOFOD-PETERSEN, 2014) e conforme o objetivo do presente trabalho. Eles estão listados a seguir.

- C1. A técnica não é reproduzível;
- C2. O artigo não explica detalhadamente como a técnica funciona;
- C3. O artigo não está devidamente publicado (não possui ISSN ou DOI);
- C4. O artigo é um resumo de outro artigo já encontrado;
- C5. A técnica é aplicada em idioma que não é de raiz europeia;
- C6. As métricas de desempenho da técnica não são justificadas;
- C7. As decisões de elaboração da técnica não são justificadas;
- C8. O artigo não utiliza análise léxica em seu trabalho.

Sobre o quinto critério, idiomas que não são de origem europeia possuem uma estrutura radicalmente diferente do português, tornando sua adaptação para o português mais custosa.

3.4 Bases de dados

As bases de dados consultadas para esta pesquisa foram:

1. Google Scholar;
2. Institute of Electrical and Electronic Engineers (IEEE) Xplore;
3. Symposium in Information and Human Language Technology (STIL);
4. Scientific Electronic Library Online (SciELO);
5. Springer;

6. Scopus.

Google Scholar possui acesso a várias outras bases de dados; IEEE Xplore é uma base de dados cujos artigos e periódicos são específicos da área de Tecnologia da Informação e afins; Springer e Scopus são bases com vasta quantidade de artigos das mais diversas áreas; SciELO é a base de dados que possui artigos especificamente em português; STIL é um evento bienal realizado pela Sociedade Brasileira de Computação (SBC) e pela Comissão Especial de Processamento de Linguagem Natural (CE-PLN) que possui artigos exclusivamente tratando de Processamento de Linguagem Natural (PLN).

3.5 Palavras-chave

As palavras-chave selecionadas para a realização da pesquisa nas bases de dados são: análise léxica, linguagem natural, mensagens de texto, análise sintática, ambiente de aprendizagem, ambiente educacional. Suas traduções para o inglês foram definidas como: *lexical analysis, natural language, text messages, syntactic analysis, learning environment, educational environment*.

Nas bases de dados que permitem operadores lógicos AND e OR, as palavras-chave foram combinadas das seguintes formas:

6. Análise léxica E linguagem natural E mensagens de texto E ambiente educacional;
7. Análise sintática E linguagem natural E mensagens de texto E ambiente educacional.

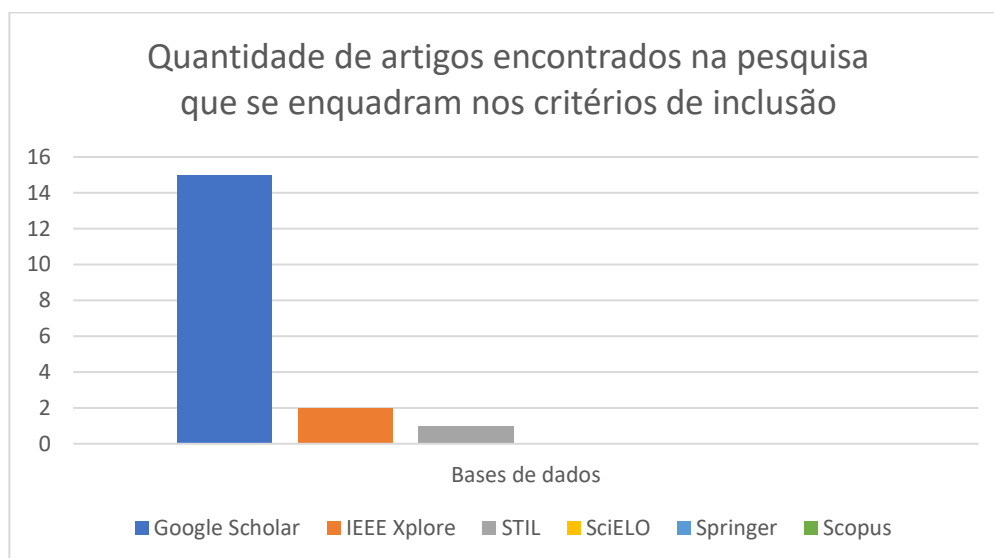
Executa-se a pesquisa com a primeira combinação, adquire-se os artigos adequados e repete-se a pesquisa, mas com a segunda combinação, adquirindo os artigos adequados novamente. Em bases cujas ferramentas de pesquisa não permitem operadores lógicos, troca-se o operador lógico “E” por vírgulas nas duas combinações de palavras-chave antes de realizar a pesquisa. No *website* da STIL, não há ferramentas de busca para filtrar os artigos pertencentes ao evento; portanto,

a pesquisa foi realizada manualmente, verificando por leitura exaustiva dos títulos e resumos presentes.

3.6 Resultados

Após a busca por artigos que se enquadravam no mínimo em um dos critérios de inclusão pelo título e resumo/*abstract*, foram obtidos quinze artigos pelo Google Scholar – dez artigos usando a primeira combinação de palavras-chave e cinco artigos usando a segunda combinação; dois artigos pelo *IEEE Xplore* com ambas as combinações de palavras-chave fornecendo os mesmos resultados; um artigo encontrado nos registros da *STIL*. As bases de dados Springer, Scopus e SciELO não forneceram nenhum resultado. Os resultados da pesquisa podem ser visualizados no Gráfico 3 a seguir.

Gráfico 3 Quantidade de artigos encontrados na pesquisa que se enquadram nos critérios de inclusão.



Fonte: Autoria própria.

Dado isso, prosseguiu-se a análise dos artigos lendo seu conteúdo por completo para verificar se satisfaziam um dos critérios de exclusão. Os artigos e os critérios de exclusão que foram satisfeitos por cada artigo estão visualmente representados no Quadro 1. Com isso, descobriu-se que a maioria dos trabalhos encontrados usavam ferramentas de análise léxica, cuja técnica utilizada não é descrita no trabalho porque a análise léxica é apenas um meio para atingir o objetivo dos respectivos trabalhos.

Quadro 1 Artigos e critérios de exclusão satisfeitos.

Artigos	C1	C2	C3	C4	C5	C6	C7	C8
(ALMEIDA, 2017)								
(BARD, 2017)								
(CARDOSO, 2003)								
(COLACE, 2015)								
(CÔRTEZ, 2016)								
(DA SILVA, 2012)								
(DE ALMEIDA, 1994)								
(HEREDIA, 2002)								
(JOSÉ, 2015)								
(MARTINS, 2013)								
(NEJME, 2013)								
(NETO, 2014)								
(PAIM, 2016)								
(PEREIRA, 2013)								
(REIS, 2015)								
(RIGO, 2013)								
(SANTOS, 2016)								
(SILVA, 2019)								
(TEIXEIRA, 2003)								
(WANG, 2007)								

Fonte: Autoria própria.

Todavia, é possível obter qual das ferramentas encontradas é a mais citada ou utilizada para análise léxica em ambientes de aprendizagem – principalmente em português – e, portanto, esse critério de exclusão será desconsiderado neste mapeamento sistemático.

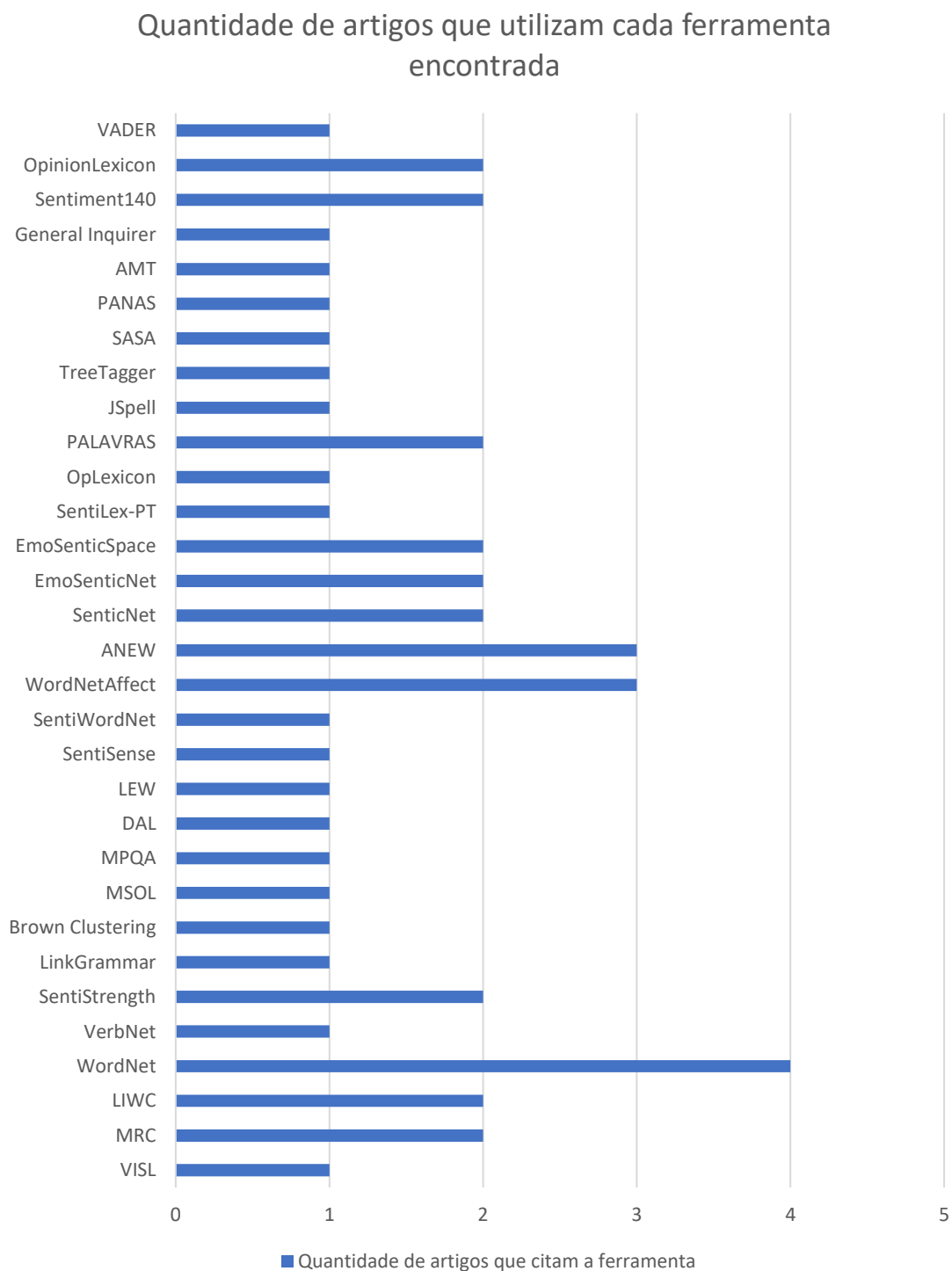
Considerando os pontos levantados e os dados apontados na tabela, os artigos de Bard (2017), Cardoso (2003), Heredia (2002), José (2015), Nejme (2013), Neto (2014), Rigo (2013), Santos (2016), Silva (2019) e Teixeira (2003) foram removidos da base.

No *corpus* obtido com o mapeamento, as ferramentas de análise léxica utilizadas ou citadas foram:

1. *Visual Interactive Syntax Learning (VISL) Complex Parsing (VISL2, 2011)*;
2. *Medical Research Council (MRC) Psycholinguistic Database (COLTHEART, 1981)*;
3. *Language of Inquiry and Word Count (LIWC) (PENNEBAKER et al., 2001)*;
4. *WordNet (FELLBAUM, 2005)*;
5. *VerbNet (KIPPER-SCHULER, 2005)*;
6. *SentiStrength (THELWALL, 2010)*;
7. *SentiWordNet (ESULI, 2006)*;
8. *WordNetAffect (STRAPPARAVA, 2004)*;
9. *Affective Norms for English Words (ANEW) (BRADLEY, 1999)*;
10. *SenticNet (CAMBRIA, 2010)*;
11. *EmoSenticNet (PORIA, 2014)*;
12. *EmoSenticSpace (PORIA, 2014)*;
13. *SentiLex-PT (SILVA, 2012)*;
14. *OpLexicon (SOUZA, 2011)*;
15. *PALAVRAS (BICK, 2000)*;
16. *JSpell (DE ALMEIDA, 1994)*;
17. *TreeTagger (SCHMID, 1995)*;
18. *SailAil Sentiment Analyzer (SASA) (WANG, 2012)*;
19. *Positive Affect Negative Affect Scale (PANAS) (WATSON, 1985)*;
20. *Amazon Mechanical Turk (AMT) service (TURK, 2012)*;
21. *General Inquirer (STONE, 1966)*;
22. *Sentiment140Lexicon (MOHAMMAD, 2013)*;
23. *OpinionLexicon / SentimentLexicon (HU, 2004)*;
24. *Valence Aware Dictionary for Sentiment Reasoning (VADER) (GILBERT, 2014)*;
25. *Link Grammar (SLEATOR, 1995)*;
26. *Brown Clustering (BROWN, 1992)*;
27. *Macquarie Sentiment Orientation Lexicon (MSOL) (MOHAMMAD, 2009)*;
28. *MPQA Subjectivity Lexicon (WIEBE, 2005)*;
29. *Dictionary of Affect in Language (DAL) (WHISSELL, 2009)*;
30. *SentiSense (DE ALBORNOZ, 2012)*;
31. *List of Emotional Words (LEW) (GO, 2009)*;

Os artigos nos quais as ferramentas apareceram são listados no Quadro 5 localizado no Apêndice A, e a quantidade de artigos que utilizam cada ferramenta é exibida graficamente no Gráfico 4.

Gráfico 4 Quantidade de artigos que utilizam cada ferramenta encontrada.



Fonte: Autoria própria.

3.7 Ameaças à validade

Alguns pontos que podem ameaçar a validade deste mapeamento:

- Caso não houvesse a reavaliação dos critérios de exclusão, apenas dois artigos (DE ALMEIDA, 1994) (ALMEIDA, 2017) permaneceriam na base de dados;
- A forma como as ferramentas realizam análise léxica não são exibidas nos artigos;
- O rigor nos critérios de exclusão pode ser além do necessário para o mapeamento;
- O rigor nos critérios de inclusão pode ser aquém do necessário para o mapeamento.

3.8 Considerações finais do mapeamento

Foram encontradas 31 (trinta e uma) ferramentas de análise léxica que também são utilizadas em ambientes educacionais. A mais citada é a *WordNet*, presente em quatro artigos; ANEW e *WordNetAffect* são as segundas mais citadas, presentes em três artigos, o que colabora com a popularidade de *WordNet*, visto que *WordNetAffect* utiliza-se de *WordNet*; 9 (nove) ferramentas são citadas apenas duas vezes; e existem 18 (dezoito) ferramentas citadas apenas uma vez.

4 RESULTADOS

Neste capítulo é apresentada a ferramenta desenvolvida neste trabalho, o RALeM (Robô de Análise Léxico-Morfológica), um *software* do tipo “*bot*”. Este *bot* foi desenvolvido em Java (ORACLE, 2019), utilizando JDK 16, o ambiente de desenvolvimento integrado IntelliJ (JETBRAINS, 2020) e uma API (*Application Programming Interface*) para a interação Java-Discord, a JDA (Java-Discord API). O IntelliJ permite a utilização de APIs no desenvolvimento de projetos Java, portanto foi importado a JDA para os arquivos utilizados na compilação do projeto conforme instrui Keener (2021).

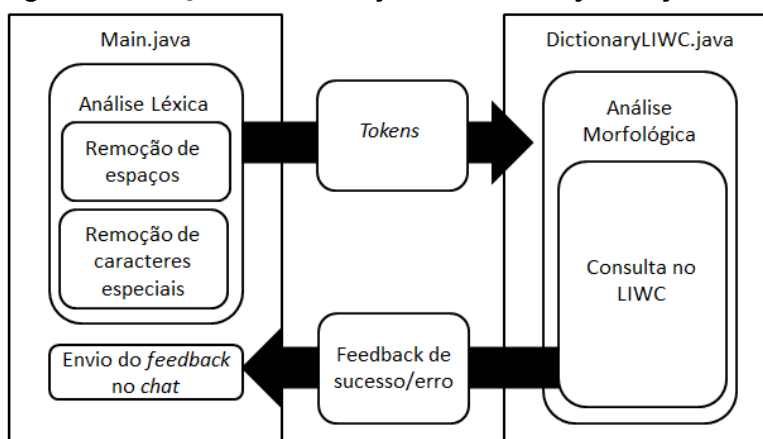
4.1. Projeto do *Bot*

Os arquivos que compõem o projeto Java do *bot* RALeM são:

- *LIWC2007.dic*, o dicionário em si;
- *VERSION.txt*, que armazena a versão atual do dicionário;
- *DictionaryLIWC.java*, arquivo Java que possui as classes *Category*, *DictionaryLIWC*, *ReturnWordPts* e *Word*, sendo responsável pela análise morfológica das palavras, consultando o dicionário;
- *Main.java*, classe principal Java do projeto, que inicializa o bot, instancia as classes dentro de *DictionaryLIWC.java*, carrega as informações do dicionário dentro de *DictionaryLIWC.java* e realiza a análise léxica, enviando os *tokens* para a análise morfológica.

A forma como *Main.java* e *DictionaryLIWC.java* interagem entre si e trabalham é ilustrada na Figura 10 a seguir.

Figura 6 Interação entre Main.java e DictionaryLIWC.java.



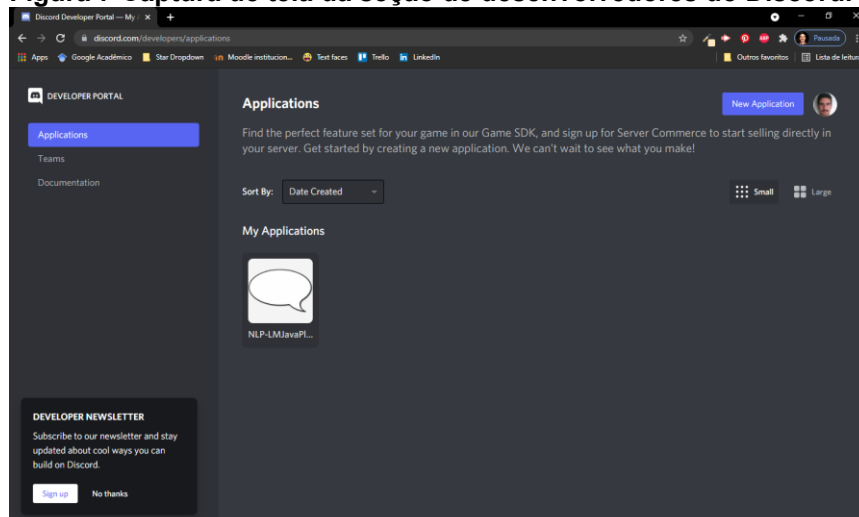
Fonte: Autoria própria.

4.2 Instalação do Bot no ambiente colaborativo

Por conta da popularidade, facilidade de uso, gratuidade, fácil instalação e uso em Windows, Linux e Mac e por ser aberto a *plugins*, foi escolhido o aplicativo de voz *Discord* (DISCORD, 2020) para a aplicação do *bot* RALeM. Os *bots* inseridos no *Discord* podem ser desenvolvidos em qualquer linguagem, porém cada linguagem possui uma API (*Application Programming Interface*) específica para comunicação com o *Discord*.

O portal on-line do *Discord* possui uma seção para que desenvolvedores possam enviar seus *bots* para teste no aplicativo, cadastrando-o no portal (exibido na Figura 7). O *bot* recebe um *token* único e que não deve ser compartilhado – pois o *token* permite que qualquer outro desenvolvedor possa utilizar o *bot* como desejar. O *token* é utilizado para conectar o *bot* ao servidor desejado.

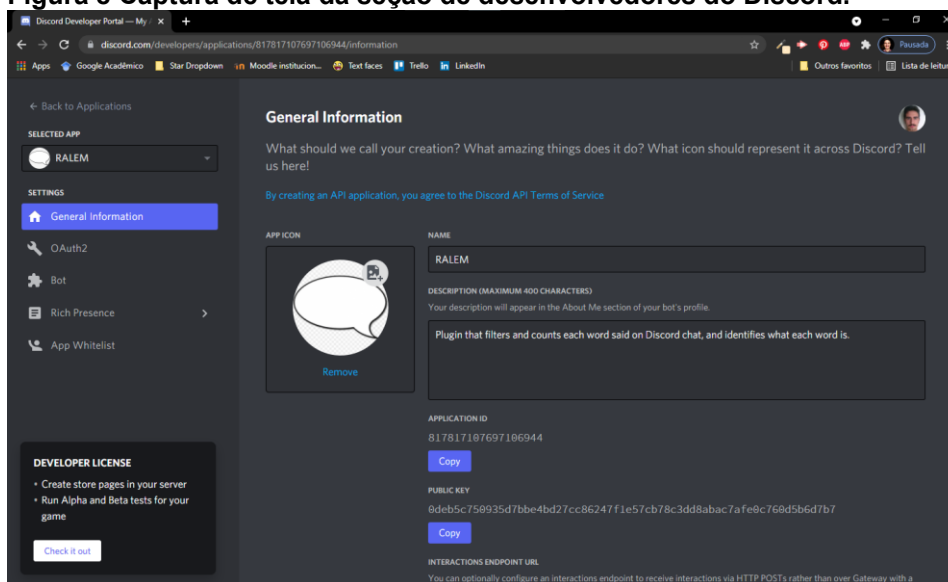
Figura 7 Captura de tela da seção de desenvolvedores do Discord.



Fonte: Autoria própria.

No portal do *bot* (Figura 7), você pode editar o nome do *bot*, colocar *link* de Termos de Serviço, inserir uma descrição do *bot*, acrescentar permissões de usuário a ele, dentre outros.

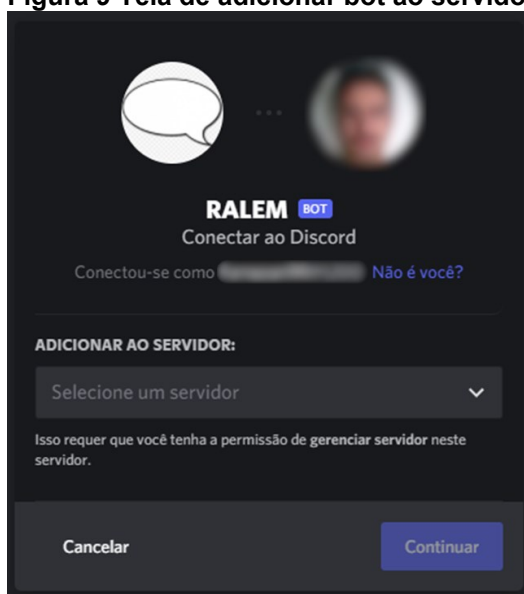
Figura 8 Captura de tela da seção de desenvolvedores do Discord.



Fonte: Autoria própria.

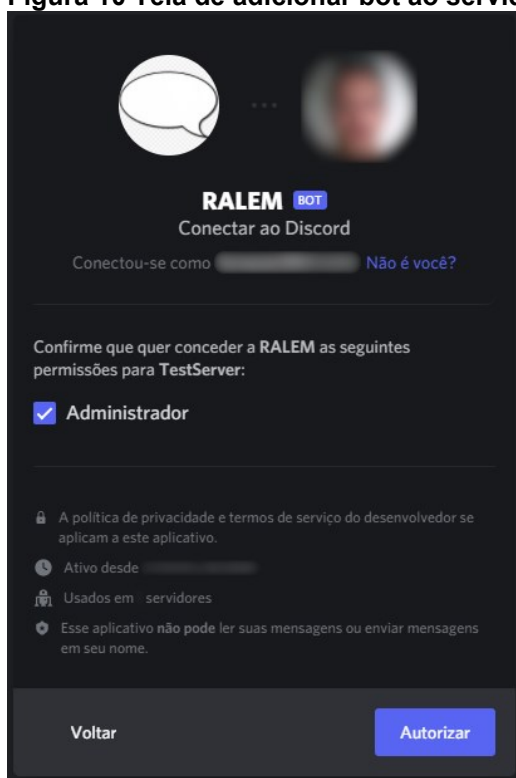
Para adicionar o bot a um servidor, precisa-se copiar o ID de cliente do bot e inseri-lo em uma URL específica. Aparece, então, a janela exibida na Figura 8. Basta selecionar o servidor, clicar em Continuar e depois em Autorizar (Figura 9).

Figura 9 Tela de adicionar bot ao servidor.



Fonte: Autoria própria.

Figura 10 Tela de adicionar bot ao servidor, com o botão Autorizar.



Fonte: Autoria própria.

Após isso, aparece uma janela de confirmação e pode-se fechar o navegador. Ao executar o *software* do bot, é possível ver o bot 'online' na lista de usuários do servidor que foi escolhido.

4.3. Comandos

Para personalização do dicionário e interação com o *bot*, foram desenvolvidos comandos que são reconhecidos ao serem enviados pelo usuário no *chat*. A lista de comandos e sua descrição está na tabela a seguir.

Tabela 3 Lista de comandos do RALeM.

Comando	Descrição
--ajuda	Mostra os comandos e suas descrições para o usuário no servidor.
--cancelar	Cancela qualquer operação iniciada por qualquer comando.
--acompanhar-cat	Acompanha palavras de determinada categoria, calculando quantas palavras em 100 palavras emitidas pertenciam à categoria escolhida.
--parar-acompanhar	Suspende o acompanhamento da categoria escolhida e retorna ao fluxo padrão do <i>bot</i> .
--ver-categorias	Exibe uma lista com as categorias existentes no dicionário.
--nova-categoria	Inicia o processo de cadastro de nova categoria.
--adic-cat-palavra	Inicia o processo de adição de categorias já existentes a uma palavra já existente.
--deletar-palavra	Deleta uma palavra (apenas as criadas pelo usuário antes de salvar e reiniciar o <i>bot</i>)
--editar-palavra	Inicia o processo de edição da grafia de uma palavra criada pelo usuário.
--ocultar-bot-cat	Ocultar a análise que o <i>bot</i> faz para cada palavra.
--mostrar-bot-cat	Exibe a análise que o <i>bot</i> faz para cada palavra.
--salvar	Salva todas as alterações em uma nova versão do dicionário LIWC.
--sim	Quando o <i>bot</i> não reconheceu uma palavra e esse comando é enviado, o <i>bot</i> começa o procedimento de adicionar a nova palavra ao dicionário.
--nao	Quando o <i>bot</i> não reconheceu uma palavra e esse comando é enviado, ele envia uma mensagem amigável informando que nenhuma operação será feita.

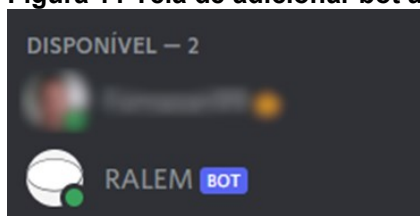
Fonte: Autoria própria.

Os comandos podem ser enviados no próprio *chat* do servidor, permitindo, assim, que um usuário sem conhecimento de programação e sem experiência prévia com interfaces de linha de comando possa editar e criar sua própria versão do dicionário com menos dificuldades.

4.4. INTERAÇÃO

Assim que o *bot* é executado, ele carrega todas as categorias e todas as palavras presentes no dicionário. Poucos segundos depois da execução do *software*, pode-se ver o *bot* online no servidor, como mostrado na Figura 11 a seguir.

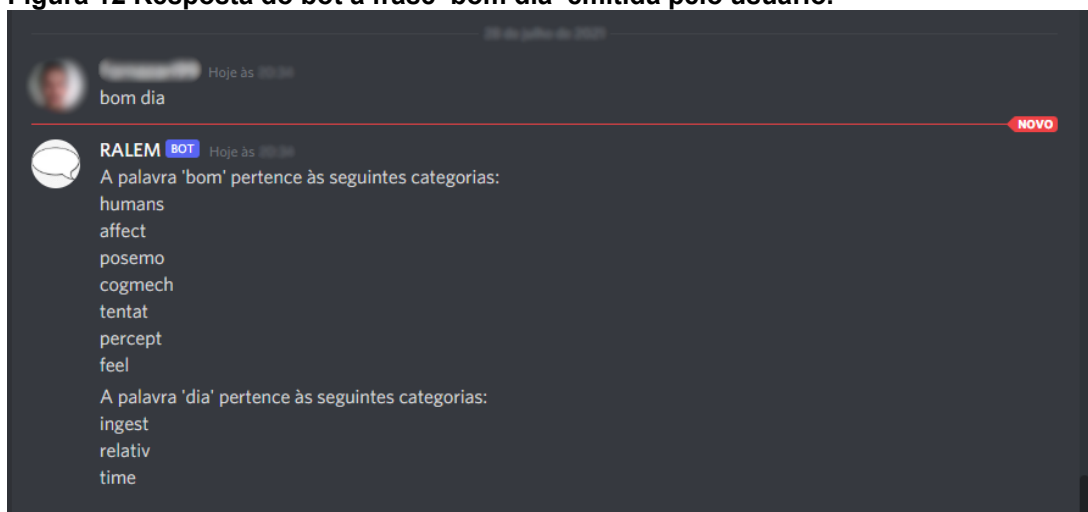
Figura 11 Tela de adicionar bot ao servidor, com o botão Autorizar.



Fonte: Autoria própria.

Pode-se interagir com o *bot* por meio do *chat* do servidor. Ao enviar uma palavra que o *bot* “conhece”, ele exibe as categorias que a palavra é pertencente e, ao enviar uma palavra desconhecida, ele pergunta se o usuário quer inserir aquela palavra nova no dicionário.

Figura 12 Resposta do bot à frase ‘bom dia’ emitida pelo usuário.



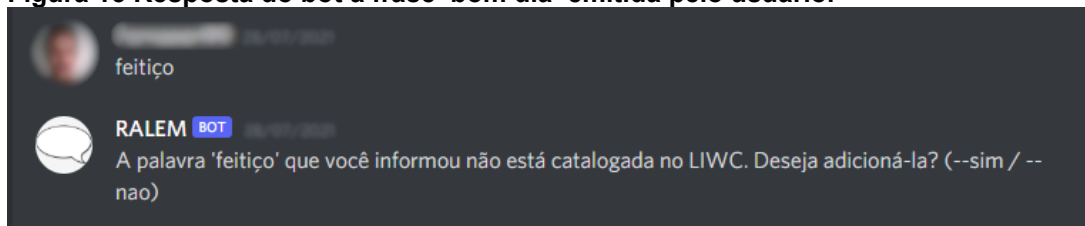
Fonte: Autoria própria.

Na figura 12, o usuário enviou a frase “bom dia”. RALeM executa os seguintes passos:

1. obtém a cadeia de caracteres enviada pelo usuário;
2. remove os espaços da cadeia de caracteres;
3. remove os caracteres especiais, como traço, vírgula, ponto-e-vírgula, exclamação, interrogação, dois-pontos e ponto;

4. busca no dicionário do LIWC as palavras obtidas após esta filtragem, transformando todos os caracteres da palavra em minúsculas;
5. como a palavra foi encontrada, o bot manda no servidor a mensagem “A palavra <palavra> pertence às seguintes categorias:” e manda a sequência de categorias da palavra.

Figura 13 Resposta do bot à frase ‘bom dia’ emitida pelo usuário.



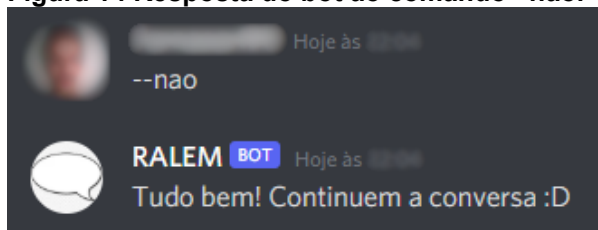
Fonte: Autoria própria.

Na figura 13, o usuário enviou a palavra “feitiço”. Os passos executados pelo RALeM são:

1. obtém a cadeia de caracteres enviada pelo usuário;
2. remove os espaços da cadeia de caracteres;
3. remove os caracteres especiais, como traço, vírgula, ponto-e-vírgula, exclamação, interrogação, dois-pontos e ponto;
4. busca no dicionário do LIWC as palavras obtidas após esta filtragem, transformando todos os caracteres da palavra em minúsculas;
5. como a palavra não existe no dicionário, o *bot* sugere ao usuário que cadastre a palavra nele, com a mensagem “A palavra ‘<palavra>’ que você informou não está catalogada no LIWC. Deseja adicioná-la?” e sugere os comandos *--sim* e *--nao*.

Ao inserir o comando *--nao*, o *bot* apenas emite a mensagem “Tudo bem! Continuem a conversa :D”, como exibido na Figura 14.

Figura 14 Resposta do bot ao comando *--nao*.



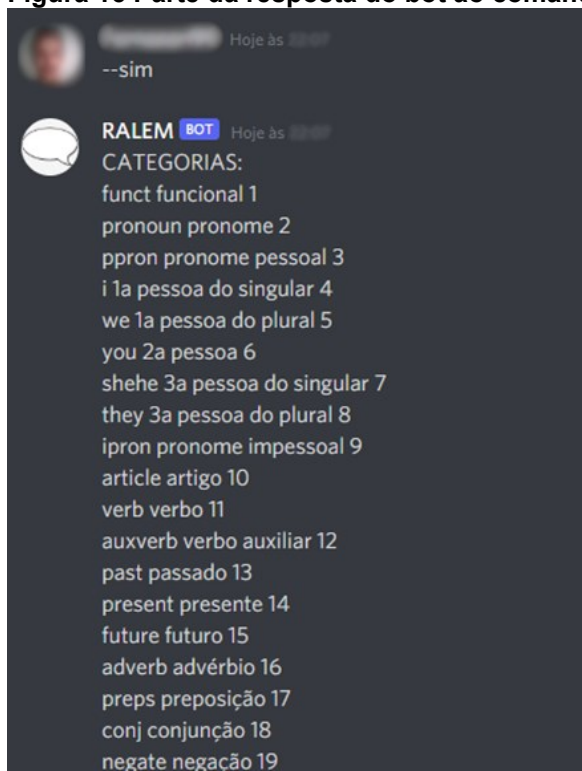
Fonte: Autoria própria.

Nas figuras 15 e 16, o *bot* exibe as categorias existentes para o desenvolvedor, e ao final pede que o usuário verifique as categorias para que encontre a categoria na qual a palavra nova se encaixa. Esse fluxo é exibido na Figura 19. Caso a palavra nova não se encaixe em nenhuma das categorias, o usuário deve inserir o comando *--nova-categoria*.

Ao inserir o comando *--cancelar* aqui, a operação é desfeita. O *bot* retorna a mensagem “Operação cancelada.”, como exibido na Figura 18, para confirmar o cancelamento ao usuário. Este comportamento é padrão para qualquer momento onde o comando *--cancelar* for inserido.

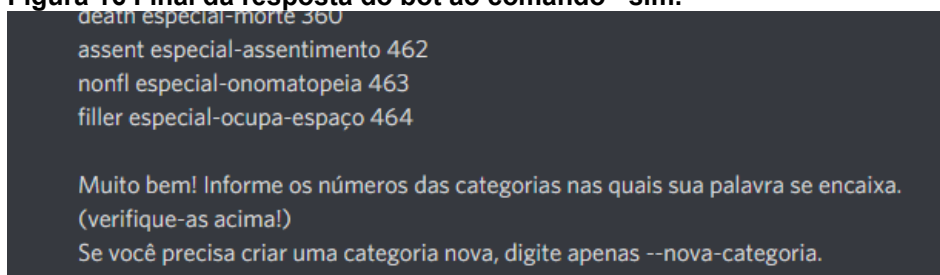
Se o usuário inserir o comando *--nova-categoria*, o *bot* solicita ao usuário o nome da categoria (Figura 17). A numeração é feita considerando o número da última categoria, tornando o número da nova categoria seu sucessor numérico.

Figura 15 Parte da resposta do bot ao comando *--sim*.



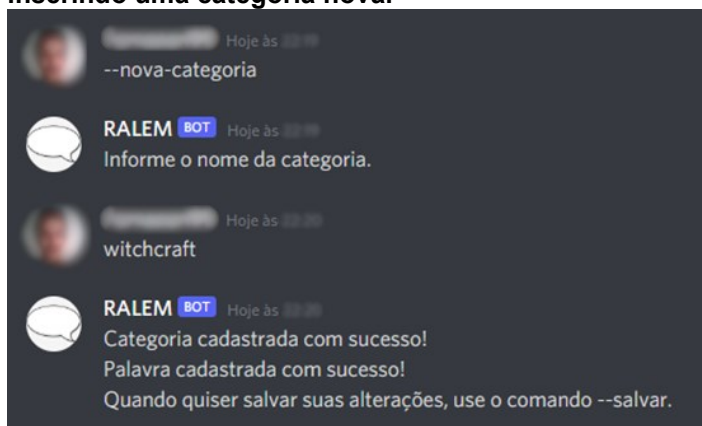
Fonte: Autoria própria.

Figura 16 Final da resposta do bot ao comando --sim.



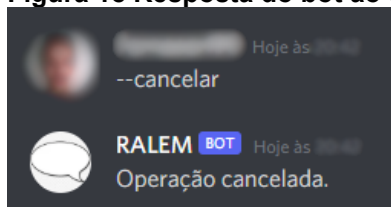
Fonte: Autoria própria.

Figura 17 Últimas interações do bot na sequência de operações --nova-categoria, inserindo uma categoria nova.



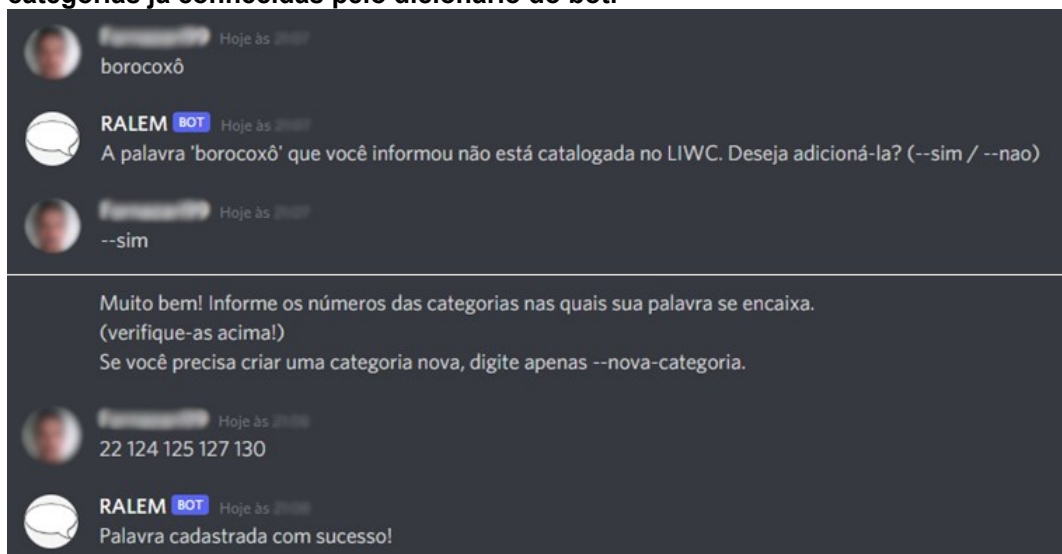
Fonte: Autoria própria.

Figura 18 Resposta do bot ao comando --cancelar.



Fonte: Autoria própria.

Figura 19 Fluxo do bot na sequência de operações --nova-categoria, onde se insere categorias já conhecidas pelo dicionário do bot.

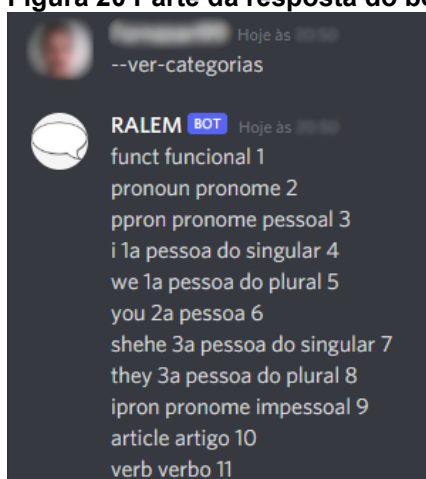


Fonte: Autoria própria.

Após o término do cadastro da categoria e da palavra, o *bot* retorna uma mensagem de confirmação de que a categoria e a palavra foram cadastradas com sucesso, e informa que, para salvar definitivamente as alterações realizadas, é necessário inserir o comando `--salvar`.

Se o usuário inserir o número de uma ou mais categorias já existentes, RALeM retorna uma mensagem confirmando que a palavra foi cadastrada com sucesso. A partir daqui, já é possível obter o resultado da análise léxica que o *bot* realiza, como mostrado anteriormente na Figura 12. Também é possível apenas cadastrar uma nova categoria sem inserir palavras desconhecidas, apenas inserindo o comando `--nova-categoria`.

Figura 20 Parte da resposta do bot ao comando `--ver-categorias`.

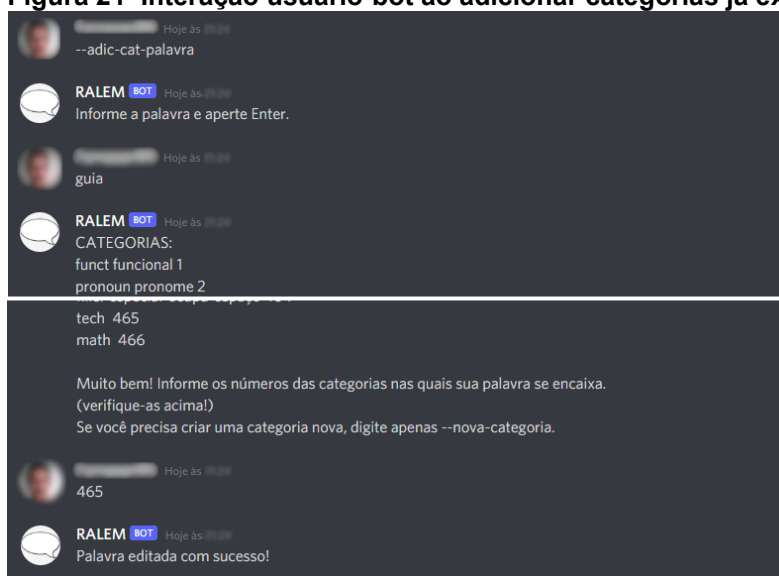


Fonte: Autoria própria.

Inserindo o comando `--ver-categorias`, o *bot* exibe todas as categorias – as padrão e as já cadastradas. A lista atualiza-se com as categorias cadastradas em tempo de execução.

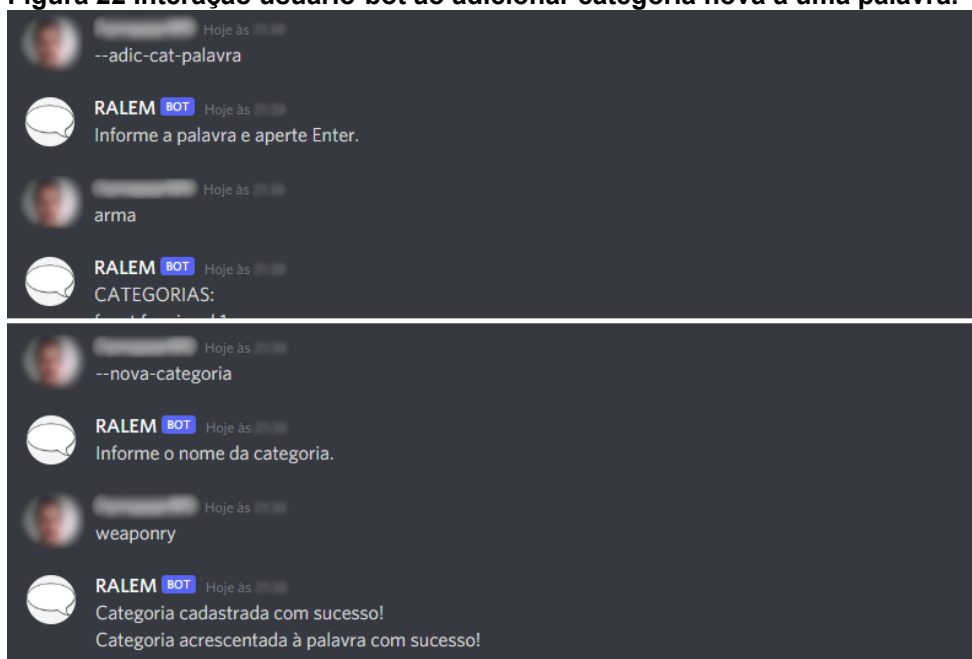
Para adicionar categorias a uma palavra existente no dicionário, cadastrada pelo usuário ou não, usa-se o comando `--adic-cat-palavra`. Após o comando, o *bot* solicita ao usuário que informe a palavra que deseja adicionar uma categoria, depois mostra a lista de categorias existentes (Figura 21) como nas figuras 15 e 16. Também é possível acrescentar uma categoria nova durante este processo ao inserir o comando `--nova-categoria` (Figura 22), com processo semelhante ao da Figura 17. Ao final, o *bot* retorna uma mensagem de confirmação.

Figura 21 Interação usuário-bot ao adicionar categorias já existentes a uma palavra.



Fonte: Autoria própria.

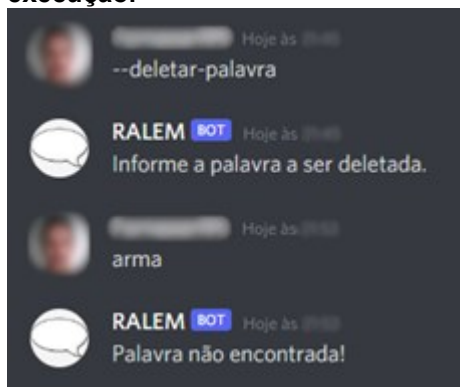
Figura 22 Interação usuário-bot ao adicionar categoria nova a uma palavra.



Fonte: Autoria própria.

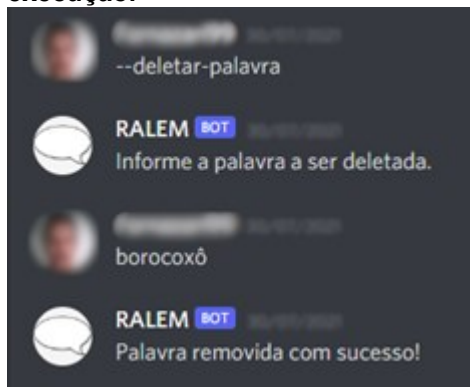
Inserindo o comando `--deletar-palavra`, o *bot* solicita ao usuário que informe a palavra que deseja deletar. Após o envio da palavra, o *bot* confirma a remoção da palavra (Figura 24). Se a palavra não for encontrada, o *bot* informa o erro e o usuário deve inserir o comando `--deletar-palavra` e tentar novamente (Figura 23). O *bot* só permite a remoção de palavras que foram cadastradas em tempo de execução.

Figura 23 Interação usuário-bot ao deletar uma palavra não cadastrada em tempo de execução.



Fonte: Autoria própria.

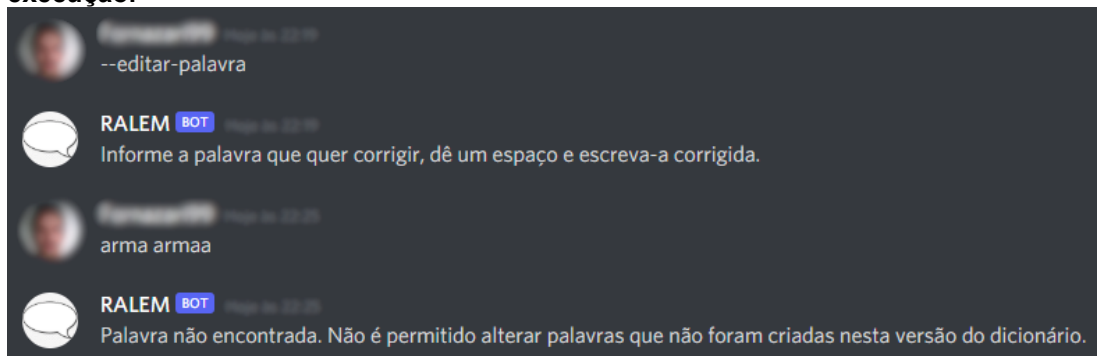
Figura 24 Interação usuário-bot ao deletar uma palavra cadastrada em tempo de execução.



Fonte: Autoria própria.

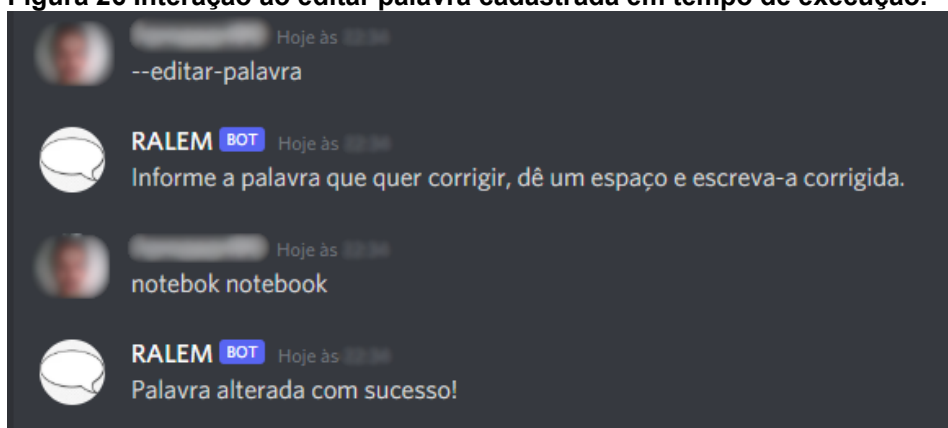
Para renomear uma palavra inserida, deve-se inserir o comando *--editar-palavra* e informar a palavra que foi inserida e a palavra reescrita, nessa ordem (Figuras 25 e 26). Então o *bot* retorna uma mensagem confirmando o sucesso da edição. O *bot* não permite a edição de palavras que não foram cadastradas em tempo de execução.

Figura 25 Interação usuário-bot ao editar palavra não cadastrada em tempo de execução.



Fonte: Autoria própria.

Figura 26 Interação ao editar palavra cadastrada em tempo de execução.

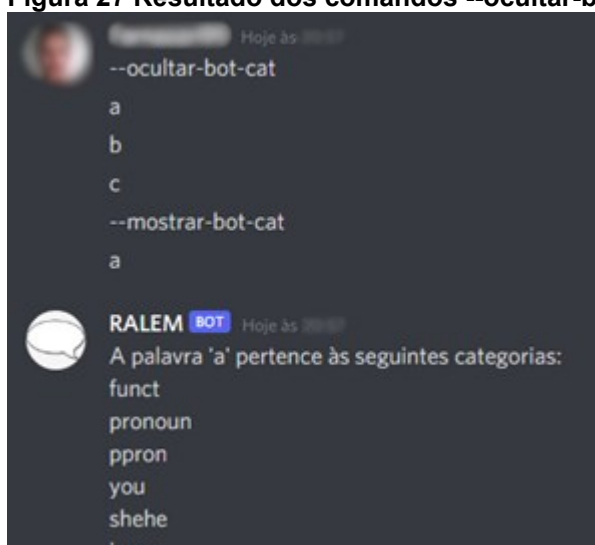


Fonte: Autoria própria.

Para acompanhar uma categoria, utiliza-se o comando `--acompanhar-cat`. Assim que o *bot* recebe o comando, ele mostra a lista de categorias existentes. O usuário deve informar apenas uma categoria e enviá-la no *chat*. Quando a categoria é enviada, o *bot* avisa que está acompanhando as palavras enviadas no *chat* (Figura 28).

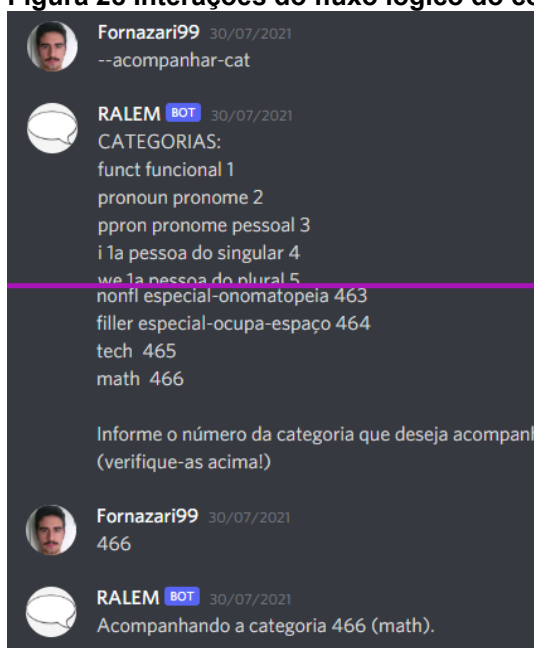
Ele continua realizando as análises de cada palavra e perguntando se o usuário deseja inserir novas – isso não afeta a pontuação final – mas, se o usuário julgar essas análises inconvenientes para o contexto, ele pode silenciá-las com o comando `--ocultar-bot-cat`. Para exibi-las novamente, usa-se o comando `--mostrar-bot-cat`. O uso dos comandos é exemplificado na Figura 27.

Figura 27 Resultado dos comandos `--ocultar-bot-cat` e `--mostrar-bot-cat`.



Fonte: Autoria própria.

Figura 28 Interações do fluxo lógico do comando --acompanhar-cat.

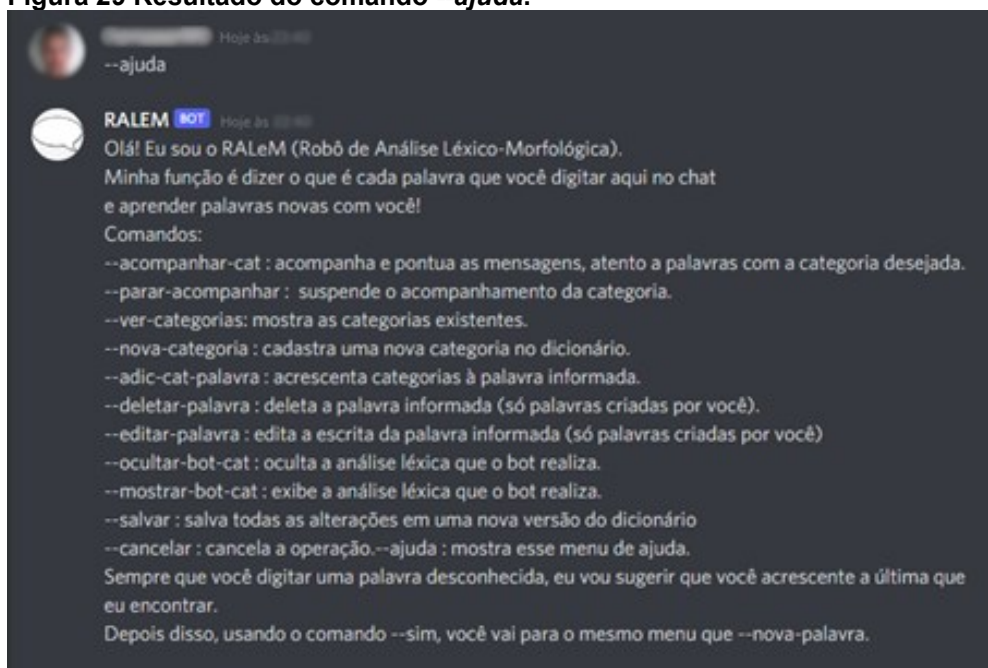


Fonte: Autoria própria.

Para que o usuário veja todos os comandos existentes e suas descrições, utiliza-se o comando `--ajuda` como mostrado na Figura 29.

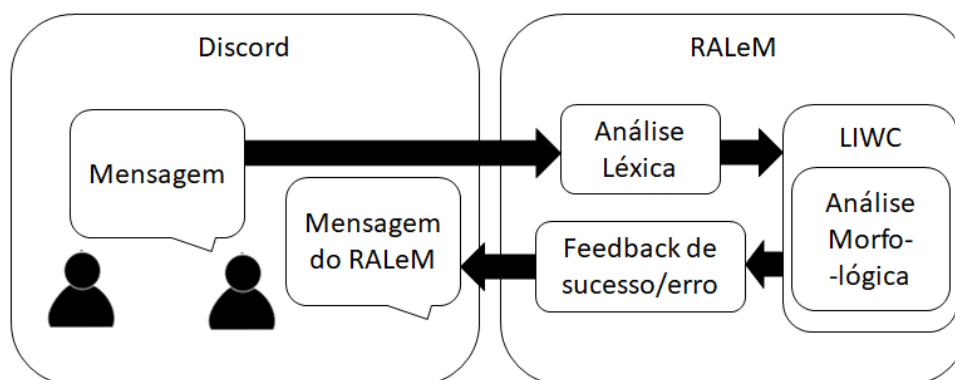
Todas as interações exibidas aqui estão ilustradas nos diagramas do Apêndice A. A síntese de como o RALeM interage com o *Discord* e o LIWC é ilustrada na Figura 30.

Figura 29 Resultado do comando --ajuda.



Fonte: Autoria própria.

Figura 30 Esquema ilustrado da interação Discord-RALeM.



Fonte: Autoria própria.

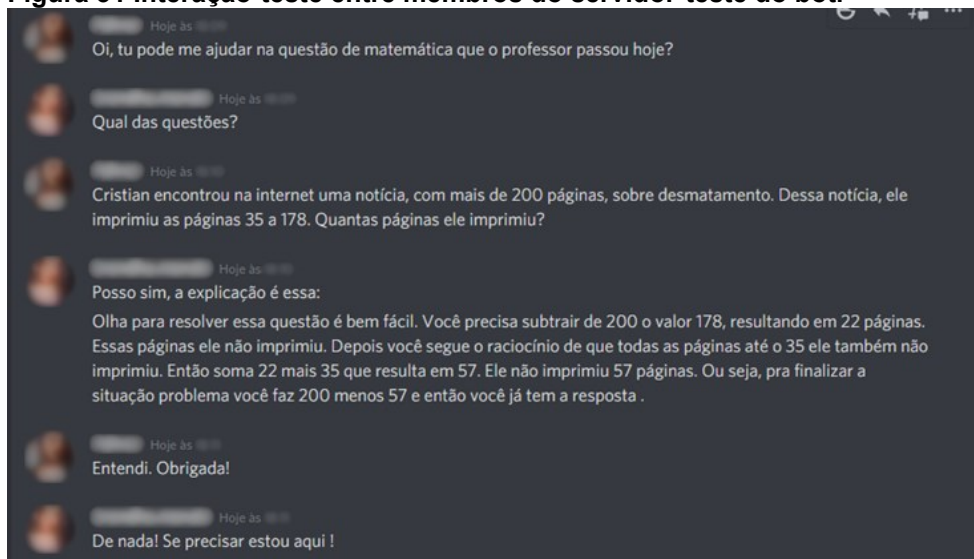
4.5. TESTES

Todos os comandos do *bot* servem para que o comando *--acompanhar-cat* seja utilizado em contextos específicos determinados pelo usuário - tal como o professor acompanharia as palavras emitidas no ambiente educacional gamificado na arquitetura proposta por Pietruchinski (2015). Foi feito, então, um teste para validar se, com essa funcionalidade, o bot consegue pontuar uma conversa entre usuários que desconhecem as palavras significativas.

No teste exibido na Figura 31, duas pessoas que desconheciam o dicionário personalizado tinham que interagir apenas conhecendo o tema – matemática. A identidade das pessoas foi borrada nas capturas de tela para preservar sua privacidade. Como uma simples conversa é uma interação entre indivíduos e permite colaboração entre eles, duas pessoas são suficientes para realizar o teste.

Enquanto elas conversavam, o *bot* acompanhava e pontuava as palavras emitidas por cada usuário, e, a cada cem palavras, gerava uma pontuação. O tema escolhido é educacional, o que permite a avaliação conforme o que foi proposto pela ArqMAEC (PIETRUCHINSKI, 2015).

Figura 31 Interação-teste entre membros do servidor-teste do bot.



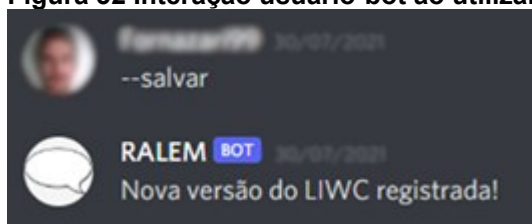
Fonte: Autoria própria.

Para interromper o acompanhamento, basta inserir o comando `--salvar` ou `--parar-acompanhar` (Figuras 32 e 33) que a pontuação é salva no arquivo *chatPoints* – semelhante ao que estipulou Pietruchinski (2016) no monitoramento do ambiente educacional gamificado – que possui em seu nome um valor gerado com base na data e hora de quando o arquivo foi instanciado, assim permitindo que cada registro de pontuação seja único e prevenindo sobrescritas acidentais.

O comando `--salvar` (Figura 32) também salva todas as alterações no dicionário feitas em tempo de execução e salva uma nova versão do dicionário. A versão nova fica registrada no arquivo *VERSION.txt*, que é usado pelo *bot* ao iniciar novamente para abrir o dicionário correto. O *bot* retorna uma mensagem de confirmação de que a operação de salvar foi bem sucedida.

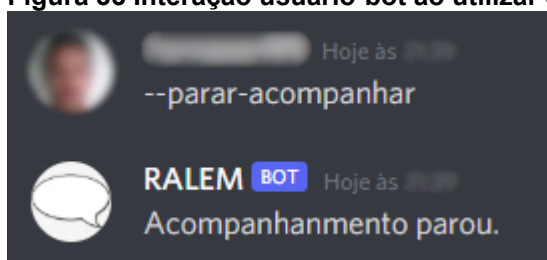
O comando `--parar-acompanhar` (Figura 33) não gera uma nova versão do dicionário; apenas salva a pontuação no arquivo *chatPoints* (Figura 34) e fecha o arquivo. Esses dois comandos salvam a pontuação gerada em um arquivo de texto, gerado automaticamente ao ligar o *bot*. Só é possível gerar outro arquivo ao reiniciar o *bot*.

Figura 32 Interação usuário-bot ao utilizar o comando `--salvar`.



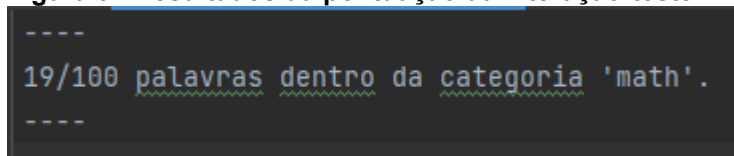
Fonte: Autoria própria.

Figura 33 Interação usuário-bot ao utilizar o comando --parar-acompanhar.



Fonte: Autoria própria.

Figura 34 Resultados da pontuação da interação-teste.



Fonte: Autoria própria.

As palavras significativas do dicionário usado no teste (com exceção dos algarismos numéricos) estão listadas no Quadro 3. Um resumo do teste está descrito no Quadro 4.

Quadro 2 Palavras significativas para o tema Matemática no teste.

Palavras significativas – Dicionário de Matemática
conta atividade tarefa lição equação x y z resultado errado certo quanto número vezes * / gráfico + = ajudar cálculo calculadora vezes tentei potência raiz quadrada cúbica cubo quadrado parábola professor matéria quantos qual resposta pergunta letra negativo positivo bhaskara fez faz entrega ajuda ajudado ajudaram ajudamos ajudem ajude multiplicar multiplica multiplique multiplicando dividir dividindo divisão divide divida divido somo somar somando soma subtração subtraio tiro subtrair menos mais questão alternativa adicionando adicione adiciona adicionar resolver resolvendo resolva resolve resolvi resolvemos resolução terminei termino termina terminamos terminou explica explico explicando explicar explique consegui conseguiu consegue entendeu entendi entendido difícil fácil chato chata complicado complicada nota notas prova trabalho exercício exercícios aula aulas ponto pontos grau graus radianos hipotenusa triângulo cateto catetos sinal operação produto a b errei acertei acertar errar acertou errou errado contas questões perguntas respostas corrigir corrigiu corrigi corrige

Fonte: Autoria própria.

Quadro 3 Resumo do teste realizado.

Detalhes do teste	
Quantidade de pessoas	2
Tema	Matemática
Palavras emitidas	134
Pontuação obtida	19 / 100

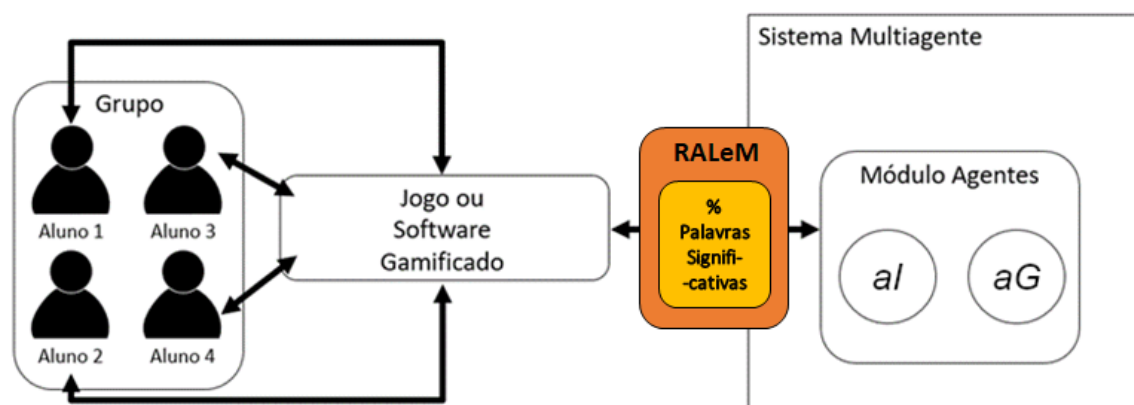
Fonte: Autoria própria.

5 CONTRIBUIÇÃO E LIMITAÇÕES

Para contribuir com o projeto da ArqMAEC (PIETRUCHINSKI, 2016), a análise léxico-sintática deveria ser realizada em um ambiente similar ao proposto na arquitetura e pontuar palavras consideradas significativas para um contexto determinado pelo professor, passando pela análise léxica, morfológica e sintática.

A análise léxica e morfológica é realizada com sucesso pelo *bot*, como mostrado nas Figuras 12 e 13, e palavras desconhecidas podem ser acrescentadas para adaptar a análise para o contexto desejado, conforme os comandos e interações exibidas nas Figuras 14 a 27. O acompanhamento e pontuação das palavras significativas é realizado com sucesso, como exibido nas Figuras 28 a 32. RALeM pode ser acrescentado ao ArqMAEC como mediador entre os agentes inteligentes e o *software* gamificado, como ilustrado na Figura 35 a seguir.

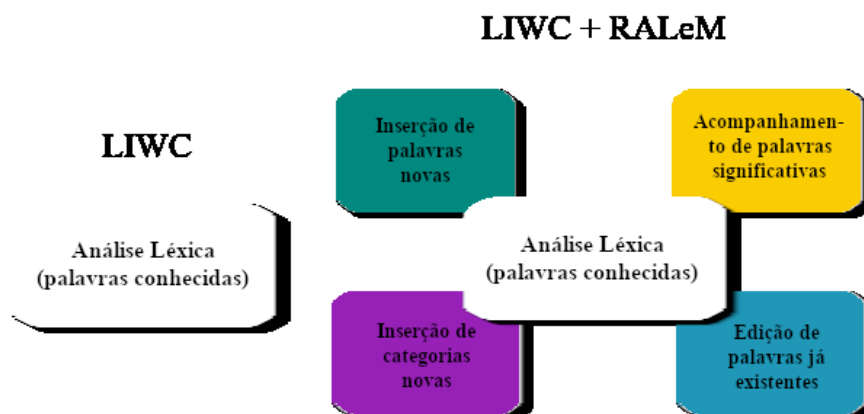
Figura 35 ArqMAEC com RALeM.



Fonte: Autoria própria (adaptado de PIETRUCHINSKI (2016))

Com o LIWC, era possível apenas a análise léxico-morfológica das palavras existentes no arquivo do dicionário, não possibilitando por si só a análise de palavras desconhecidas. Foi desenvolvido, então, o mecanismo de personalização do dicionário com o *bot* RALeM, onde o professor pode inserir as palavras ausentes e acrescentar categorias às palavras já existentes, servindo-se de comandos de texto. Estas contribuições estão ilustradas na Figura 36.

Figura 36 Ilustração das contribuições do bot ao LIWC.



Fonte: Autoria própria.

Todavia, mesmo com o *bot*, não é possível realizar análise sintática das frases emitidas no *chat*. Isso exige um arcabouço linguístico que o LIWC não oferece - o dicionário contém apenas informações referentes à forma da palavra (morfologia) e suas categorias gramaticais (léxico). O *bot* também não é capaz de avaliar, com sua estrutura atual, se a interação significativa foi positiva ou negativa, o que não permite a identificação dos critérios da Interdependência Social (JOHNSON, 2014) com precisão.

6 CONCLUSÃO

A Arquitetura de Monitoramento, Avaliação e Estimulação da Colaboração (ArqMAEC) foi projetada para auxiliar professores de EaD a monitorar alunos de suas disciplinas por meio de *softwares* educacionais colaborativos gamificados. Esse monitoramento, avaliação e estimulação devem ser realizados em mensagens emitidas pelos alunos em *chats* ou fóruns, que servem como ferramentas colaborativas para o *software*, por agentes inteligentes. Para que os agentes consigam realizar a avaliação das mensagens, é necessário processar computacionalmente a linguagem natural presente nas mensagens.

O processamento de linguagem natural (PLN) começa com uma sentença de entrada, geralmente contendo uma ou mais frases. Realiza-se o processamento morfossintático da sentença para obter sua estrutura sintática e, a partir do processamento semântico desta, retornar de saída a forma lógica da sentença, pronta para ser aplicada em algum contexto.

A análise léxica é etapa crucial do processamento morfossintático e, graças ao advento da Internet e, mais recentemente, das redes sociais e aplicativos de *chat*, possui vasto ambiente para teste e aplicação. Inúmeras ferramentas de análise léxica são utilizadas, a mais comum sendo a *WordNet*, contando com mais de cem mil léxicos. Porém, como a análise léxica é utilizada mais como meio para realizar processamento de linguagem natural do que como um fim em si mesma, as pesquisas não descrevem minuciosamente o *modus operandi* das ferramentas que utilizam.

Para encontrar a ferramenta mais adequada para análise léxica neste trabalho, utilizou-se como critérios a) popularidade (quantidade de citações da ferramenta em artigos encontrados em mapeamento sistemático) e b) quantidade de palavras em português contempladas pela ferramenta, sendo b) o critério com maior peso. A ferramenta que satisfizesse estes dois critérios foi a *Language of Inquiry and Word Count* (LIWC), com 127162 (cento e vinte e sete mil e cento e sessenta e duas) palavras em português e citada em 2 (dois) artigos.

Para poder testar o dicionário do LIWC e adaptá-lo ao contexto do projeto do ArqMAEC (PIETRUCHINSKI, 2015), precisava-se de um *software* que permitisse a inserção de *plugins* e a interação entre membros por *chat*. Escolheu-se o Discord (DISCORD, 2020) por estes aspectos e a acessibilidade, popularidade e gratuidade do *software*, a linguagem Java (ORACLE, 2019) para implementação do *bot* por

possuir uma API Java para *Discord* e a IDE IntelliJ (JETBRAINS, 2020) para desenvolvimento em Java pela disponibilidade gratuita do *software* e a possibilidade de utilizar a API JDA nela.

O *bot* permite ao usuário, por meio de comandos em mensagens no servidor, a edição e remoção de palavras cadastradas em tempo de execução, inserção de novas palavras, inserção de categorias a palavras, inserção de novas categorias, acompanhamento de palavras que estão numa determinada categoria com pontuação das palavras registradas em arquivo texto e efetivação das alterações em um novo arquivo de dicionário. Ele analisa cada palavra emitida no *chat* com base na versão mais recente do dicionário.

Com essa capacidade de acompanhamento das palavras e pontuações, satisfaz-se um dos requisitos para análise léxico-sintática proposta pelo ArqMAEC: o professor pode utilizar o *bot* para criar uma nova versão do dicionário com a categoria desejada e as palavras significativas daquela categoria, acompanhando as mensagens para avaliar a conversação entre alunos, obtendo assim, a cada cem palavras, a quantidade de palavras significativas emitidas pelos usuários.

7 TRABALHOS FUTUROS

Para complementar a contribuição do presente trabalho, precisa-se de uma ferramenta de análise sintática para complementar as funções realizadas pelo RALeM e realizar, efetivamente, a análise léxico-sintática, cobrindo todas as suas fases. É necessário conhecer o estado-da-arte das ferramentas de análise sintática, analisar com critérios objetivos qual ferramenta é a mais adequada para o projeto e combinar a análise sintática com a análise léxico-morfológica já realizada pelo *bot*.

Pode-se também ampliar as funcionalidades atuais do *bot*, permitindo a inserção de várias palavras com um único comando em determinada categoria, classificação automática de palavras desconhecidas em determinada categoria, inserção de várias categorias simultaneamente, remoção de várias palavras de determinada categoria e contagem de palavras repetidas significativas, com o objetivo de identificar possíveis sinais de fraude na pontuação; e programar o *bot* para que os comandos de modificação do dicionário e acompanhamento de palavras sejam exclusivos a um usuário com privilégios, simulando a interação professor-*bot*.

Com relação ao LIWC, pode-se elaborar uma forma de trabalhar com as categorias psicológico-afetivas do dicionário, utilizando-o para análise da qualidade das interações realizadas no *chat* e do provável estado emocional dos usuários que estão se comunicando, contribuindo para a análise da interação dos alunos conforme a Teoria da Interdependência Social (JOHNSON, 2014).

Para adequar o projeto ainda mais à arquitetura proposta por Pietruchinski (2015), além das alterações já sugeridas, pode-se alterar o arquivo de registro de pontuação do acompanhamento das palavras para o formato XML e desenvolver uma forma de gerar um arquivo novo sempre que o acompanhamento for encerrado.

REFERÊNCIAS

- ALMEIDA, A. W. **Sentiment Analysis in short messages using affective lexicon**. Dissertação (Mestrado) – Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica do Paraná. Curitiba, 2017.
- BALAGE FILHO, Pedro; PARDO, Thiago Alexandre Salgueiro; ALUÍSIO, Sandra. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. 2013.
- BARD, P.; LUIS, Renan Lopes; MORAES, Sílvia. Normalizador de Texto para Língua Portuguesa baseado em Modelo de Linguagem (A Normalizer based on Language Model for Texts in Portuguese)[In Portuguese]. In: **Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology**. 2017. p. 142-150.
- BICK, Eckhard. **The parsing system palavras: Automatic grammatical analysis of Portuguese in a constraint grammar framework**. Aarhus Universitetsforlag, 2000.
- BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural language processing with Python: analyzing text with the natural language toolkit**. " O'Reilly Media, Inc.", 2009.
- BORBA, Marcelo de Carvalho; PAULA DOS SANTOS MALHEIROS, Ana; ZULATTO, Rúbia Barcelos Amaral. **Educação a distância online**. Autêntica Editora, 2020.
- BRADLEY, Margaret M.; LANG, Peter J. **Affective norms for English words (ANEW): Instruction manual and affective ratings**. Technical report C-1, the center for research in psychophysiology, University of Florida, 1999.
- BRASIL. Decreto nº 5.622, de 19 de dezembro de 2005. Regulamenta o art. 80 da Lei no 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União**: seção 1, Brasília, DF, ano 243, p. 1-4, 20 dez. 2005.
- BROWN, Peter F. et al. Class-based n-gram models of natural language. **Computational linguistics**, v. 18, n. 4, p. 467-480, 1992.
- BUI, Thanh; NGUYEN, Hung. **Essential Feature-Cooperative Gameplay**. 2018.
- CAMBRIA, Erik et al. Senticnet: A publicly available semantic resource for opinion mining. In: **AAAI fall symposium: commonsense knowledge**. 2010.
- CARDOSO, Antônio Luiz Mattos de Souza; DE MENEZES, Crediné Silva. Processamento eficiente de consultas em linguagem natural em ambientes educacionais. In: **1º Workshop em Tecnologia da Informação e da Linguagem Humana-TIL**. 2003.

COLACE, Francesco et al. Sentiment detection in social networks and in collaborative learning environments. **Computers in Human Behavior**, v. 51, p. 1061-1067, 2015.

COLTHEART, Max. The MRC psycholinguistic database. **The Quarterly Journal of Experimental Psychology Section A**, v. 33, n. 4, p. 497-505, 1981.

CÔRTEZ, Eduardo Gabriel. **Processamento de linguagem natural para interação entre usuários e agentes pedagógicos em ambientes virtuais de aprendizagem**. 2016. 65 f. Trabalho de Conclusão de Curso (Ciência da Computação) – Universidade de Santa Cruz do Sul. Santa Cruz do Sul, 2016.

DA SILVA, Carlos Alexandre Siqueira; DE MENEZES, Crediné Silva. Etiquetagem e rotulagem de frases no ambiente de apoio a disciplinas on-line “Assistente de Reuniões”. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2012.

DE ALMEIDA, José Joao Dias; PINTO, Ulisses. JSpell – um módulo para a análise léxica genérica de linguagem natural. **Encontro da Associação Portuguesa de Linguística**, 1994.

DE ALBORNOZ, Jorge Carrillo; PLAZA, Laura; GERVÁS, Pablo. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In: **LREC**. 2012. p. 3562-3567.

DE OLIVEIRA, Celina Couto; DA COSTA, José Wilson; MOREIRA, Marcia. **Ambientes informatizados de aprendizagem: produção e avaliação de software educativo**. Papirus, 2001.

DE PAULA, Lorena Tavares; MICHALSKI, Rafael. Os bots de disseminação de informação na conjuntura das campanhas presidenciais de 2018 no Brasil. **Múltiplos Olhares em Ciência da Informação**, v. 9, n. 1, 2019.

DISCORD. **DISCORD | Seu Lugar para Papear e Ficar de Boa**. [S.l.] 2020. Disponível em: <<https://discord.com/>> Acesso em: 20 jul. 2020.

EASTMAN, Jacqueline K.; SWIFT, Cathy Owens. Enhancing collaborative learning: Discussion boards and chat rooms as project communication tools. **Business Communication Quarterly**, v. 65, n. 3, p. 29-41, 2002.

ESULI, Andrea; SEBASTIANI, Fabrizio. Sentiwordnet: A publicly available lexical resource for opinion mining. In: **LREC**. 2006. p. 417-422.

FASTFORMAT. **Como fazer citação de artigos online e sites da internet? – FastFormat**. [S.l.] 2020. Disponível em: <<https://blog.fastformat.co/como-fazer-citacao-de-artigos-online-e-sites-da-internet/>> Acesso em: 29 set. 2020.

FELLBAUM, C: WordNet and wordnets. In: **Encyclopedia of Language & Linguistics Bd. 13** (2005), S. 665–670 — ISBN 978-0-08-044854-1.

GILBERT, C. H. E.; HUTTO, Erric. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: **Eighth International Conference on Weblogs and Social Media (ICWSM-14)**. Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>. 2014. p. 82.

GO, Alec; BHAYANI, Richa; HUANG, Lei. Twitter sentiment classification using distant supervision. **CS224N project report, Stanford**, v. 1, n. 12, p. 2009, 2009.

GONZALEZ, Marco; LIMA, Vera Lúcia Strube. Recuperação de informação e processamento da linguagem natural. In: **XXIII Congresso da Sociedade Brasileira de Computação**. 2003. p. 347-395.

HAMARI, Juho; HUOTARI, Kai. Defining gamification: a service marketing perspective. In: **Proceedings of the 16th international academic MindTrek conference**. 2012. p. 17-22.

HEREDIA, Eduardo; OMAR, Nizam; MENDONÇA, Jadir Custódio. Um agente de comunicação em linguagem natural no contexto de um sistema multiagentes orientado à tutoria inteligente na WWW. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2002. p. 456-463.

HU, Mingqing; LIU, Bing. Mining and summarizing customer reviews. In: **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining**. 2004. p. 168-177.

JETBRAINS S.R.O. **IntelliJ IDEA: o Java IDE capaz e ergonômico da JetBrains**. [S.l.] 2020. Disponível em: <<https://www.jetbrains.com/pt-br/idea/>>. Acesso em 02 ago. 2021.

JOHNSON, David W.; JOHNSON, Roger T. Learning together. **Handbook of cooperative learning methods**, v. 51, p. 65, 1994.

JOHNSON, David W.; JOHNSON, Roger T.; SMITH, Karl A. Cooperative learning: Improving university instruction by basing practice on validated theory. **Journal on Excellence in University Teaching**, v. 25, n. 4, p. 1-26, 2014.

JOSÉ, Jário; PAIVA, Ranilson; BITTENCOURT, Ig Ibert. AVALIAÇÃO AUTOMÁTICA DE ATIVIDADES ESCRITAS BASEADA EM ALGORITMO GENÉTICO E PROCESSAMENTO DE LINGUAGEM NATURAL: Avaliador Ortográfico-Gramatical. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2015. p. 95.

KEENER, Austin. **DV8FromTheWorld/JDA: Java wrapper for the popular chat & VOIP service: Discord**. [S.l.] 2021. Disponível em: <<https://github.com/DV8FromTheWorld/JDA>>. Acesso em: 12 nov. 2020.

KENSKI, Vani Moreira et al. Ensinar e aprender em ambientes virtuais. **ETD-Educação Temática Digital**, v. 10, n. 2, p. 223-249, 2009.

KRISTENSEN C. H; GOMES C. F. A; JUSTO A. R; VIEIRA K. Normas brasileiras para o Affective Norms for English Words. In: **Trends Psychiatry Psychother**, 2011.

KOFOD-PETERSEN, A. How to do a structured literature review in computer science (version 0.2). **Copenhagen: Alexandra Institute**, 2014.

LEVIS, John; SUVOROV, Ruslan. Automatic speech recognition. **The encyclopedia of applied linguistics**, 2012.

LEBEUF, Carlene R. **A taxonomy of software bots: towards a deeper understanding of software bot characteristics**. 2018. Tese de Doutorado.

LIWC. **LIWC**. São Paulo, Brasil. 2021. Disponível em: <<http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>>. Acesso em: 27 jul. 2021.

LIWC. **<https://liwc.wpengine.com>**. [S.l.] 2021. Disponível em: <<https://liwc.wpengine.com>>. Acesso em: 27 jul. 2021.

MARTINS, Lucas L. et al. Integrando Recursos de Processamento de Linguagem Natural e Web-Semântica em um Ambiente Educacional, para Apoio aos Alunos. In: **Anais do III Student Workshop on Information and Human Language Technology-2013-Fortaleza, CE. Sociedade Brasileira de Computação**. 2013. p. 10-12.

MOHAMMAD, Saif; DUNNE, Cody; DORR, Bonnie. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: **Proceedings of the 2009 conference on empirical methods in natural language processing**. 2009. p. 599-608.

MOHAMMAD, Saif M.; KIRITCHENKO, Svetlana; ZHU, Xiaodan. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. **arXiv preprint arXiv:1308.6242**, 2013.

NEJME, Fatima Zahra; BOUALKNADEL, Siham; ABOUTAJDINE, Driss. Toward a noun morphological analyser of standard Amazigh. In: **2013 ACS International Conference on Computer Systems and Applications (AICCSA)**. IEEE, 2013. p. 1-1.

NETO, Mauro da Rocha Xavier et al. **Desenvolvimento do projeto HUMANI como ferramenta didática para a aprendizagem de linguagens de programação**. TISE, 2014.

ORACLE. **Java | Oracle**. [S.l.] 2019. Disponível em: <<https://www.java.com/pt-BR/>>. Acesso em 2 ago. 2021.

PAIM, ALDO MARCELO. **Inferência de personalidade a partir de textos em português brasileiro utilizando léxicos**. 2016. Tese de Doutorado. PhD thesis, Pontificia Universidade Católica do Paraná.

PENNEBAKER, James W.; FRANCIS, Martha E.; BOOTH, Roger J. Linguistic inquiry and word count: LIWC 2001. **Mahway: Lawrence Erlbaum Associates**, v. 71, n. 2001, p. 2001, 2001.

PENNEBAKER, James W. et al. **The development and psychometric properties of LIWC2015**. 2015.

PEREIRA, Felipe Ribeiro; RIGO, Sandro Jose. Utilização de processamento de linguagem natural e ontologias na análise qualitativa de frases curtas. **RENOTE-Revista Novas Tecnologias na Educação**, v. 11, n. 3, 2013.

PIETRUCHINSKI, M.H. **ArqMAEC: Um modelo arquitetural baseado em agentes para monitorar, avaliar e estimular a colaboração em ambientes educacionais gamificados**. Tese (Doutorado) – Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, Universidade Federal do Paraná. Curitiba, 2016.

PORIA, Soujanya et al. EmoSenticSpace: A novel framework for affective common-sense reasoning. **Knowledge-Based Systems**, v. 69, p. 108-123, 2014.

REIS, Julio et al. Uma abordagem multilíngue para análise de sentimentos. In: **Anais do IV Brazilian Workshop on Social Network Analysis and Mining**. SBC, 2015.

ROIG, Anna Escofet et al. Indicadores de análisis de procesos de aprendizaje colaborativo en entornos virtuales de formación universitaria. In: **Enseñanza and Teaching**. 2012.

RABELLO, R. d. S. **Interação e autismo: uso de agentes inteligentes para detectar déficits de comunicação em ambientes síncronos**. Tese (Doutorado) – Centro Interdisciplinar de Novas Tecnologias na Educação, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2010.

RIGO, Sandro J. et al. Abordagem linguística para identificação da dimensão afetiva expressa em textos de Ambientes Virtuais de Aprendizagem—um Léxico da Emoção. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2013. p. 738.

SANTOS, Jário José; PAIVA, Ranilson; BITTENCOURT, Ig Ibert. Avaliação Léxico-Sintática de Atividades Escritas em Algoritmo Genético e Processamento de Linguagem Natural: Um Experimento no ENEM. **Revista Brasileira de Informática na Educação**, v. 24, n. 02, p. 92, 2016.

SCARTON, C. E.; ALUISIO, S. Towards a cross-linguistic VerbNet-style lexicon for Brazilian Portuguese. In: **Workshop on Creating Cross-language Resources for Disconnected Languages and Styles Workshop Programme**, p.11, 2012.

SCHMID, Helmut. TreeTagger - a language independent part-of-speech tagger. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>, 1994.

SCHULER, Karin Kipper. **VerbNet: A broad-coverage, comprehensive verb lexicon**. 2005.

SILVA, Mário J.; CARVALHO, Paula; SARMENTO, Luís. Building a sentiment lexicon for social judgement mining. In: **International Conference on Computational Processing of the Portuguese Language**. Springer, Berlin, Heidelberg, 2012. p. 218-228.

SILVA, Fabiano Moreira da. **Reconhecimento de Padrões por Compressão aplicado ao Processamento da Linguagem Natural**. 2019. 58 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Instituto Federal de Educação, Ciência e Tecnologia Goiano. Rio Verde, 2019.

SLEATOR, Daniel DK; TEMPERLEY, Davy. Parsing English with a link grammar. **arXiv preprint cmp-lg/9508004**, 1995.

SOUZA, Marlo; VIEIRA, Renata. Construction of a portuguese opinion lexicon from multiple resources. **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Brasil.**, 2011.

STONE, P. J.; DUNPHY, D. C.; SMITH, M. S.; and OGILVIE, D. M. **The General Inquirer: A Computer Approach to Content Analysis**. MIT Press, 1966.

STRAPPARAVA, Carlo et al. Wordnet affect: an affective extension of wordnet. In: **LREC**. 2004. p. 40.

TEIXEIRA, Sérgio; DE MENEZES, Crediné Silva. Facilitando o uso de Ambientes Virtuais através de Agentes de Conversação. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2003. p. 455-464.

THANAKI, Jalaj. **Python natural language processing**. Packt Publishing Ltd, 2017.

THELWALL, Mike et al. Sentiment strength detection in short informal text. **Journal of the American society for information science and technology**, v. 61, n. 12, p. 2544-2558, 2010.

TURK, Amazon Mechanical. Amazon mechanical turk. **Retrieved August**, v. 17, p. 2012, 2012.

VISL2, 2012. **Complex Parsing Interface - Visual Interactive Syntax Learning**. <<http://beta.visl.sdu.dk/visl/pt/parsing/automatic/complex.php>>. Acesso em: 29 set. 2020.

WANG, Hao et al. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: **Proceedings of the ACL 2012 system demonstrations**. 2012. p. 115-120.

WANG, Ying-Hong; WANG, Wen-Nan; HUANG, Chu-Chi. Enhanced semantic question answering system for e-learning environment. In: **21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)**. IEEE, 2007. p. 1023-1028.

WATSON, David; CLARK, Lee Anna; TELLEGEN, Auke. Development and validation of brief measures of positive and negative affect: the PANAS scales. **Journal of personality and social psychology**, v. 54, n. 6, p. 1063, 1988.

WHISSELL, Cynthia. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. **Psychological reports**, v. 105, n. 2, p. 509-521, 2009.

WIEBE, Janyce; WILSON, Theresa; CARDIE, Claire. Annotating expressions of opinions and emotions in language. **Language resources and evaluation**, v. 39, n. 2-3, p. 165-210, 2005.

WORDNETBR. **WordNet-Br 1.0 – Base de verbos**. [S.l.] 2012. Disponível em: <<http://www.nilc.icmc.usp.br/wordnetbr/>> Acesso em: 29 set. 2020.

ZIRAWAGA, Victor Samuel; OLUSANYA, Adeleye Idowu; MADUKU, Tinovimbanashe. Gaming in education: Using games as a support tool to teach history. **Journal of Education and Practice**, v. 8, n. 15, p. 55-64, 2017.

APÊNDICE A – Quadro da relação entre artigos encontrados no mapeamento sistemático e ferramentas de análise léxica

Quadro 4 Relação entre artigos do *corpus* e ferramentas utilizadas (primeira parte).

Ferramentas	ALMEIDA 2017	COLACE 2015	CORTÊS 2016	DA SILVA 2012	DE ALMEIDA 1994	MARTINS 2013	PAIM 2016	PEREIRA 2013	REIS 2015	WANG 2007
VISL										
MRC										
LIWC										
WordNet										
VerbNet										
SentiStrength										
Link Grammar										
Brown Clustering										
MSOL										
MPQA										
DAL										
LEW										
SentiSense										
SentiWordNet										

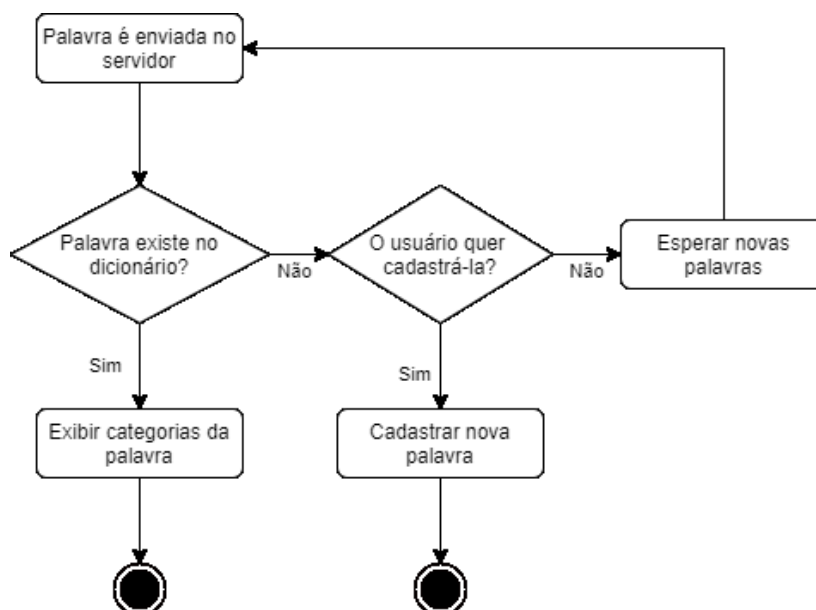
Fonte: Autoria própria.

Quadro 4 Relação entre artigos do *corpus* e ferramentas utilizadas (segunda parte).

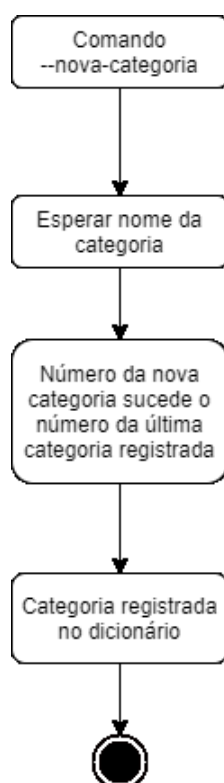
Ferramentas	ALMEIDA 2017	COLACE 2015	CORTÊS 2016	DA SILVA 2012	DE ALMEIDA 1994	MARTINS 2013	PAIM 2016	PEREIRA 2013	REIS 2015	WANG 2007
WordNetAffect										
ANEW										
SenticNet										
EmoSenticNet										
EmoSenticSpace										
SentiLex-PT										
OpLexicon										
PALAVRAS										
JSpell										
Tree Tagger										
SASA										
PANAS										
AMT										
General Inquirer										
Sentiment140										
Opinion Lexicon										
VADER										

Fonte: Autoria própria.

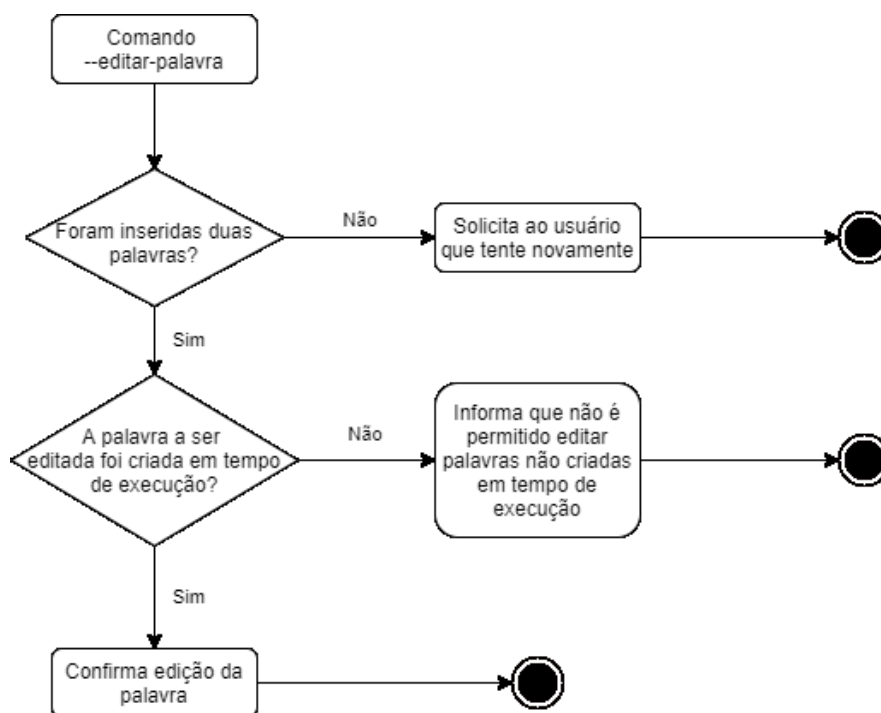
APÊNDICE B - Fluxogramas Lógicos dos Comandos do RALeM



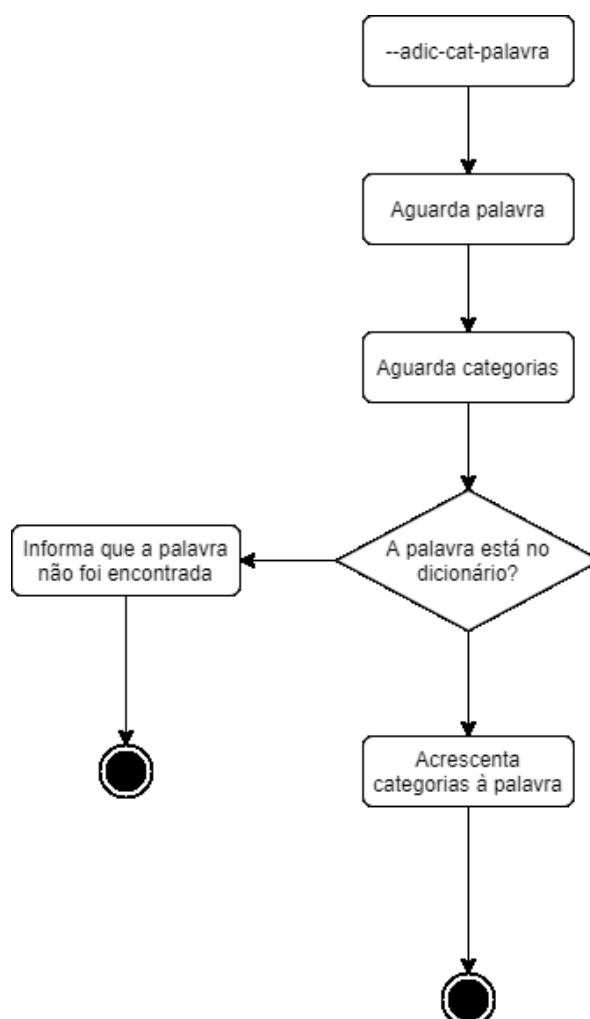
Fluxo padrão - palavra enviada no servidor



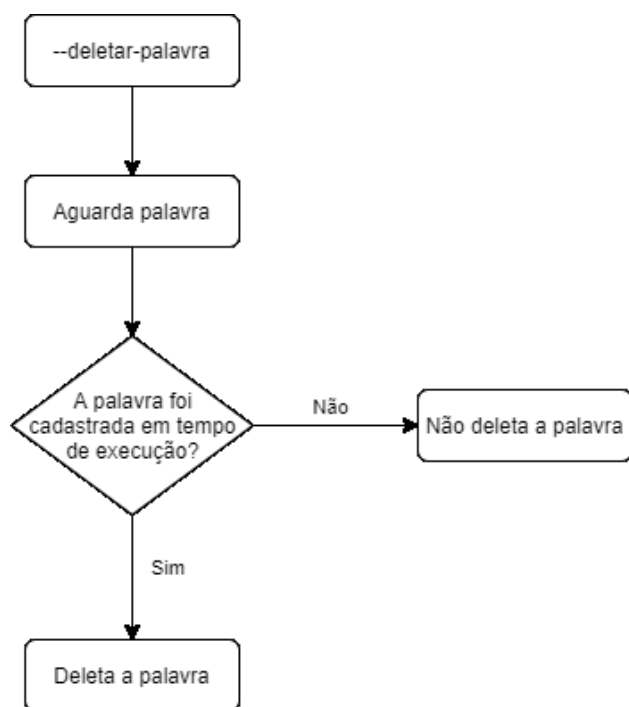
Fluxo de Nova Categoria



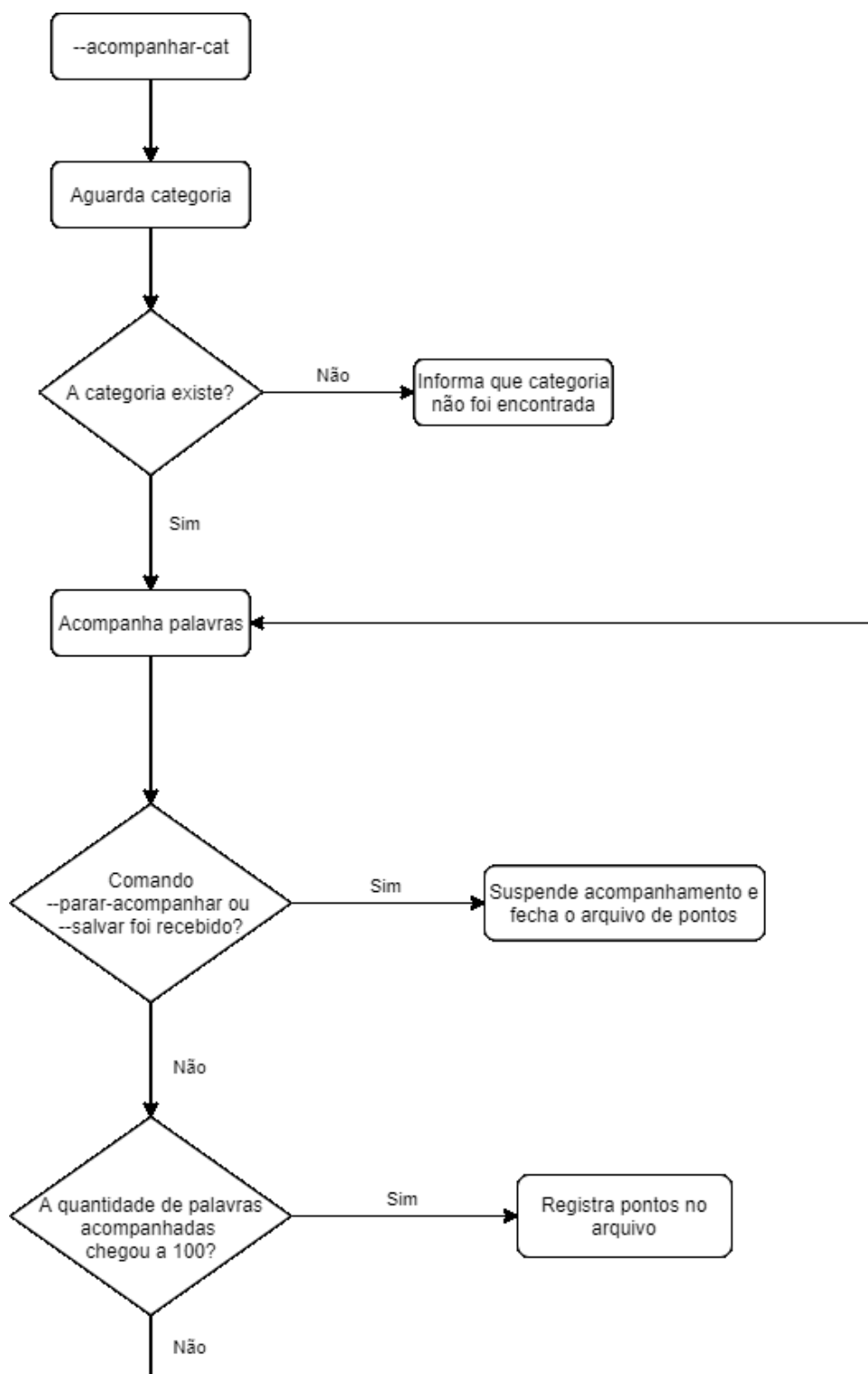
Fluxo lógico de edição de palavra.



Fluxo lógico de adicionar categorias a uma palavra.



Fluxo lógico de remoção da palavra.



Fluxo lógico de acompanhamento das palavras.