

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Letras
Programa de Pós-Graduação em Estudos Linguísticos

Monique Vieira Miranda

**TRANSITIVIDADE NO VOCABULÁRIO ACADÊMICO
EM PORTUGUÊS BRASILEIRO:
uma análise baseada em artigos científicos**

Belo Horizonte
2021

Monique Vieira Miranda

**TRANSITIVIDADE NO VOCABULÁRIO ACADÊMICO
EM PORTUGUÊS BRASILEIRO:
uma análise baseada em artigos científicos**

Versão final

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Linguística Teórica e Descritiva.

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: Estudos da Língua em Uso

Orientadora: Profa. Dra. Ana Larissa Adorno Marciotto Oliveira

Belo Horizonte

2021

M672t

Miranda, Monique Vieira.

Transitividade no vocabulário acadêmico em português brasileiro [manuscrito] : uma análise baseada em artigos científicos / Monique Vieira Miranda. – 2021.

204 f., enc. : il., grafs., tabs., color.

Orientadora: Ana Larissa Adorno Marciotto Oliveira.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudos da Língua em Uso.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras

Bibliografia: f. 190-197.

Apêndices: f. 198-204.

1.Linguística de corpus – Teses. 2. Funcionalismo (Linguística) – Teses. 3. Língua portuguesa – Verbos – Teses. 4. Redação acadêmica – Teses. I. Oliveira, Ana Larissa Adorno Marciotto. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD : 418



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

**TRANSITIVIDADE NO VOCABULÁRIO ACADÊMICO EM PORTUGUÊS BRASILEIRO: uma análise
baseada em artigos científicos**

MONIQUE VIEIRA MIRANDA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDO LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 23 de fevereiro de 2021, pela banca constituída pelos membros:

Prof(a). Ana Larissa Adorno Marciotto Oliveira - Orientadora
UFMG

Prof(a). Gustavo Ximenes Cunha
UFMG

Prof(a). Aparecida de Araújo Oliveira
UFV

Prof(a). Maria Medianeira de Souza
UFPE

Prof(a). Claudio Márcio do Carmo
UFSJ

Belo Horizonte, 23 de fevereiro de 2021.



Documento assinado eletronicamente por **Gustavo Ximenes Cunha, Professor do Magistério Superior**, em 23/02/2021, às 19:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ana Larissa A Marciotto Oliveira, Professora do Magistério Superior**, em 23/02/2021, às 19:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cláudio Márcio do Carmo, Usuário Externo**, em 23/02/2021, às 19:28, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Aparecida de Araújo Oliveira, Usuário Externo**, em 23/02/2021, às 19:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Medianeira de Souza, Usuário Externo**, em 24/02/2021, às 14:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0531803** e o código CRC **377844BF**.

Para Maria.

Sem você este trabalho teria sido terminado na metade tempo, mas sem nem um terço da diversão.

AGRADECIMENTOS

À professora Ana Larissa, pela orientação e ensinamentos ao longo de todos esses anos.

À professora Cida, pelo começo e perguntas respondidas.

Ao CNPq, pelo apoio financeiro durante três quartos desta pesquisa.

Ao professor Giacomo Figueredo, pelas sugestões durante o Exame de Qualificação.

Aos colegas co-orientandos, pela companhia, discussões, figurinhas e nova forma de escrita.

Aos meus pais e avó, pelo cuidado que se estende.

Ao Luiz Max, pelo auxílio dentro e fora deste trabalho, e por todos os quatorze anos.

À Maria, por existir.

Resumo

A presente pesquisa teve como objetivo analisar o sistema da TRANSITIVIDADE instanciado por meio dos verbos pertencentes ao vocabulário acadêmico mais frequentes em artigos científicos de diferentes áreas do conhecimento, adotando como base a proposta de divisão vocabular de Nation (2001) e Webb e Nation (2017) e os pressupostos teóricos da Linguística Sistêmico-Funcional (LSF), principalmente Halliday e Matthiessen (2014). Além disso, empregaram-se variadas ferramentas da Linguística de *Corpus*, tais como #LancsBox (BREZINA *et al.*, 2018) e KitConc (MOREIRA FILHO, 2008). O presente trabalho objetivou compilar um *corpus* representativo do gênero, de modo a construir as duas partes centrais da pesquisa: na qual identifica-se os verbos pertencentes ao vocabulário acadêmico, resultando na criação de uma lista de verbos acadêmicos, frequentes em artigos científicos; e outra na qual analisa-se o sistema da TRANSITIVIDADE inerente aos lemas verbais mais frequentes nessa lista, sob a perspectiva da LSF. Tendo tais propósitos em vista, o *corpus* compilado de 12,3 milhões de palavras de textos de todas as Grandes Áreas do CNPq permitiu uma variedade de amostras linguísticas suficientes para identificar as diferentes categorias do vocabulário (geral, acadêmico, técnico e de baixa frequência), especialmente aquele alvo deste trabalho, o acadêmico. Para tanto, uma série de critérios internos e externos (GARDNER e DAVIES, 2014; COXHEAD, 2000) foram aplicados ao *corpus* a fim de identificar os verbos acadêmicos em artigos científicos em Português Brasileiro. No âmbito da LSF, o sistema da TRANSITIVIDADE refere-se à forma de o usuário representar e organizar suas ideias através da linguagem, relacionado à organização sentencial básica, cujos elementos são processos (verbos), participantes e circunstâncias. A análise sob essa perspectiva permitiu observar que as escolhas dos tipos de processos, associados à configuração da sentença e dos participantes integrantes, refletem diretamente as características do gênero acadêmico, e das Grandes Áreas em que ocorrem, tais quais a impessoalidade na redação, o foco nas atividades/procedimentos de pesquisa e a objetividade ao reportar análises e resultados. Desse modo, a combinação das duas partes da pesquisa, acrescida dos produtos finais – o *corpus* compilado disponibilizado *online*, a lista de verbos acadêmicos e a descrição do sistema da TRANSITIVIDADE – contribuem para uma compreensão mais aprofundada sobre a escrita acadêmica, avançando estudos teóricos dessas áreas com base em amostras reais da língua em uso.

Palavras-chave: Vocabulário acadêmico. Linguística sistêmico-funcional. Sistema da TRANSITIVIDADE.

Abstract

The present research aimed to analyze the TRANSITIVITY system in the most frequent academic verbs in scientific research articles from different disciplines, accordingly to Nation's (2001) and Webb and Nation's (2017) proposal of vocabulary division and to Systemic-Functional Linguistics (LSF), mainly Halliday and Matthiessen (2014). In addition, various Corpus Linguistics tools were employed, such as #LancsBox (BREZINA *et al.*, 2018) and KitConc (MOREIRA FILHO, 2008). The present work aimed to compile a corpus representative of the genre, in order to build the two central parts of the research: one in which verbs belonging to the academic vocabulary are identified, resulting in the creation of a list of academic verbs, frequent in scientific articles; and another in which the TRANSITIVITY system, inherent to the verbal lemas identified in this list is analyzed, from the perspective of LSF. Based on these goals, we compiled a corpus with approximately 12.3 million words of scientific research articles from all the CNPq's Major Areas which provided a variety of linguistic samples sufficient to identify the different vocabulary categories (general, academic, technical and low frequency), especially academic vocabulary. For this purpose, a series of internal and external criteria (GARDNER; DAVIES, 2014; COXHEAD, 2000) were applied to the corpus in order to identify academic verbs in scientific articles in Brazilian Portuguese. Within the scope of LSF, the TRANSITIVITY system refers to the way the user represents and organizes his ideas through language, related to basic sentence organization, whose elements are processes (verbs), participants and circumstances. The analysis from this perspective allowed us to observe that the choices of the types of processes, associated with the configuration of the sentence and of its participants, directly reflect the characteristics of the academic genre, and of the Great Areas in which they occur, such as impersonality in writing, the focus on research activities / procedures and objectivity in reporting analysis and results. Thus, the combination of the two parts of the research, plus the final products - the compiled corpus, which will be made available online, the list of academic verbs and the description of the TRANSITIVITY system - contribute to a deeper understanding on academic writing, advancing in theoretical studies from both areas based on real samples of the language in use.

Keywords: Academic vocabular. Sitemic-Funcional Linguistics. TRANSITIVITY system.

LISTA DE EQUAÇÕES

Equação 1 – Modelos para o cálculo da Lei de Zipf	31
Equação 2 – Fórmula para o cálculo da frequência relativa.....	138

LISTA DE FIGURAS

Figura 1 – Distribuição da frequência absoluta das palavras pela posição no ranking, escala linear	30
Figura 2 – Distribuição da frequência absoluta das palavras pela posição no ranking, eixo X em escala log	31
Figura 3 - Estratificação	47
Figura 4 - O contínuo de instanciação	48
Figura 5 - Metafunção	51
Figura 6 – Escala de níveis e movimentos descritivos	57
Figura 7 – TIPOS DE PROCESSOS do sistema da TRANSITIVIDADE	62
Figura 8 - Tipos de participantes em orações	64
Figura 9 – Relação entre processo, participantes e circunstâncias no modelo da ergatividade	107
Figura 10 – Nomenclatura dos textos do <i>corpus</i>	121
Figura 11 - Lista lematizada dos itens lexicais do <i>corpus</i> da pesquisa no programa #LancsBox	125
Figura 12 - Ferramenta Wizard no software #LancsBox	140

LISTA DE GRÁFICOS

Gráfico 1 - Contagem dos tipos de processos por Grande Área do conhecimento	145
Gráfico 2 - Tempo verbal em orações materiais.....	147
Gráfico 3 - Tempo verbal em orações materiais por Grande Área.....	148
Gráfico 4 - Tipos de construções em orações materiais	151
Gráfico 5 - Tipos de construções em orações materiais por Grande Área	154
Gráfico 6 – Tipos de Ator em orações materiais por Grande Área	155
Gráfico 7 - Tipos de Ator explícito em orações materiais por Grande Área.....	156
Gráfico 8 – Tipos construções em orações relacionais.....	166
Gráfico 9 – Tempo verbal em orações relacionais	166
Gráfico 10 - Tempo verbal nas orações relacionais por Grande Área.....	168
Gráfico 11 - Tempo verbal e tipos de construções em orações mentais	170
Gráfico 12 - Tempo verbal em orações existenciais.....	176
Gráfico 13 - Tempo verbal em orações relacionais por Grande Área	177

LISTA DE QUADROS

Quadro 1 – Vantagens e desvantagens de <i>types</i> , lemas, lexemas e famílias de palavras	26
Quadro 2 – Exemplos de palavras em diferentes níveis de frequência no presente corpus, acompanhadas por sua posição no ranking de frequência de lemas	28
Quadro 3 - Subdivisão do vocabulário técnico.....	37
Quadro 4 – As dimensões da linguagem, seus princípios e ordens de organização.....	45
Quadro 5 – Exemplos da configuração do contexto situacional em textos	53
Quadro 6 – Variáveis de registro relacionadas às metafunções e seus sistemas léxico-gramaticais.....	55
Quadro 7 – Exemplificação dos sistemas das metafunções	56
Quadro 8 – Exemplificação do complexo oracional sob diferentes perspectivas metafuncionais	58
Quadro 9 - Exemplos de processos, participantes e circunstâncias.....	60
Quadro 10 - Exemplo de Participante, Processo e Circunstância, a última mais obrigatória. .	61
Quadro 11 – Orações materiais transitivas e intransitivas.....	66
Quadro 12 – Orações materiais (in)transitivas criativas e transformativas	67
Quadro 13 – Exemplos de processos e participantes em orações mentais	71
Quadro 14 - Classificação dos tipos de processos atributivos.....	79
Quadro 15 - Subtipos de orações relacionais identificativas	81
Quadro 16 - Exemplos de oração relacional do subtipo <i>demonstrativo</i>	82
Quadro 17 - Exemplos de orações circunstanciais identificativas	88
Quadro 18 - Exemplos de verbos e orações verbais.....	94
Quadro 19 – Processos, significados e participantes envolvidos	99
Quadro 20 – Tipos e exemplos de circunstâncias.....	102
Quadro 21 – Modelos complementares do sistema da transitividade	105
Quadro 22 – Orações classificadas de acordo com o modelo da transitividade.....	105
Quadro 23 - Orações classificadas de acordo com o modelo da ergatividade	107
Quadro 24 – Tipos de construções ergativas	109
Quadro 25 – Exemplos de classificação de acordo com a hierarquia da divisão de áreas proposta pelo CNPq	113

LISTA DE TABELAS

Tabela 1 - Exemplos da contabilização de <i>tokens</i> por diferentes softwares.....	25
Tabela 2 – Relação entre os níveis de frequência de <i>types</i> e a porcentagem de <i>tokens</i> no <i>corpus</i>	29
Tabela 3 – Composição do <i>corpus</i> da pesquisa.....	116
Tabela 4 – Médias de artigos por periódicos e de <i>tokens</i> por artigos no <i>corpus</i>	120
Tabela 5 - Lista de verbos acadêmicos	129
Tabela 6 - Critérios para distinguir os diferentes tipos de processos	141
Tabela 7 – Distribuição dos tipos de processos por verbos analisados	143

LISTA DE ABREVIATURAS E SIGLAS

ACL – Academic Collocations List

AVL – Academic Vocabulary List

AWL – Academic Word List

BNC – British National Corpus

COCA – Corpus of Contemporary American English

ESP – English for Specific Purposes

EWL – Essential Word List

GSL – General Service List

LSF – Linguística Sistêmico-Funcional

NGSL – New-General Word List

PB – Português do Brasil

PE – Português Europeu

UWL – University Word List

Sumário

1	CONSIDERAÇÕES INICIAIS	19
2	VOCABULÁRIO ACADÊMICO	24
2.1	Conceitos básicos para este estudo.....	24
2.2	Tipos de vocabulários.....	27
2.3	Palavras de alta frequência e de baixa frequência	32
2.4	Palavras técnicas.....	36
2.5	Palavras acadêmicas	39
3	LINGÜÍSTICA SISTÊMICO-FUNCIONAL	44
3.1	Conceitos básicos da linguística sistêmico-funcional	44
3.2	Oração enquanto representação: a metafunção experiencial.....	59
3.3	O sistema da TRANSITIVIDADE: processos em foco	61
3.3.1	Orações materiais	65
3.3.2	Orações mentais	70
3.3.3	Orações relacionais	77
3.3.4	Orações verbais	92
3.3.5	Orações comportamentais	95
3.3.6	Orações existenciais	97
3.3.7	Circunstâncias	99
3.4	O modelo da ergatividade: breve introdução	104
4	METODOLOGIA DE PESQUISA	111
4.1	Composição, procedimentos de coleta e organização do <i>corpus</i>	112
4.2	Procedimentos de análise dos dados	121
4.2.1	Listas de palavras	121
4.2.2	Relato de procedimentos	124
4.3	Lista de verbos acadêmicos do PB	128

4.4	Parâmetros para distinção dos tipos de processos	139
5	ANÁLISE DOS TIPOS DE PROCESSOS ENTRE OS LEMAS FREQUENTES	143
5.1	Análise do sistema da transitividade	143
5.1.1	Orações materiais	145
5.1.2	Orações relacionais.....	157
5.1.3	Orações mentais.....	169
5.1.4	Orações existenciais	174
5.1.5	Orações verbais.....	180
6	CONSIDERAÇÕES FINAIS	184
	REFERÊNCIAS	190
	ANEXO 1 – <i>Script</i> para remoção de itens não-verbais e marcações morfossintáticas	198
	ANEXO 2 – <i>Script</i> para calcular Razão	199
	ANEXO 3 – <i>Script</i> para calcular Alcance	201
	ANEXO 4 – <i>Script</i> para calcular Dispersão	203
	ANEXO 5 – <i>Script</i> para calcular Medida de disciplina e Frequência mínima.....	204

1 CONSIDERAÇÕES INICIAIS

Na literatura sobre vocabulário acadêmico, há duas grandes vertentes de pesquisas. A primeira apoia a visão de que existe um vocabulário acadêmico central e comum às variadas áreas de conhecimento em uma língua, sendo, de tal modo, relevante o desenvolvimento de listas de palavras acadêmicas para aprendizes independentemente da disciplina estudada (XUE e NATION, 1984; COXHEAD, 2000; NATION, 2001; GARDNER e DAVIES, 2014). A segunda vertente questiona a existência desse tipo de vocabulário, argumentando que parâmetros tais quais frequência, alcance, significado e colocações de uma palavra mudam ao longo das diferentes áreas de conhecimento, devido a variações no uso dentro de cada disciplina (MARTÍNEZ, 2001; HYLAND e TSE, 2007; DURRANT, 2016). Nessa perspectiva, o uso de ferramentas como lista de palavras é apenas ideal quando desenvolvidas em/para cada área específica de conhecimento.

Recentemente, estudos como de Hyland (2016) e de Dang, Coxhead e Webb (2017) começam a convergir na direção de que tais vertentes devem ser estudadas como um contínuo, não uma dicotomia. Ou seja, a construção de listas de palavras deve levar em consideração as circunstâncias tanto das habilidades pré-existentes dos aprendizes quanto do contexto de aprendizagem, isto é, não somente a disciplina, mas também o conhecimento e envolvimento prévio do estudante com a área de conhecimento e com a língua em questão.

Até o momento, todos os estudos mencionados são voltados à pesquisa em língua inglesa. Trabalhos similares em língua portuguesa são escassos (SANTOS, 2006; MIRANDA, 2016), não havendo ainda pesquisa que se proponha a aprofundar na elaboração de um vocabulário acadêmico com procedimentos mais rigorosos de classificação (tais quais os de GARDNER e DAVIES, 2014) ou que adentre nas questões levantadas pela segunda vertente, anteriormente mencionada, e verifique se os termos acadêmicos são de fato centrais às diferentes áreas do conhecimento, ou se são específicos de cada disciplina.

É diante desse panorama teórico que surge a presente tese, buscando suprir alguns questionamentos teóricos de ambas as vertentes de pesquisa voltados para o estudo do Português Brasileiro. De tal modo, esta tese tem como objetivo criar uma lista de verbos acadêmicos empregados em um *corpus* de artigos científicos representante de todas as áreas do conhecimento, e analisar os verbos acadêmicos mais frequentes segundo os pressupostos teóricos da Linguística Sistêmicos-Funcional (LSF), em especial Halliday e Matthiessen (2014), especificamente o sistema de TRANSITIVIDADE.

Cabe mencionar que, nesta pesquisa, a definição do termo “vocabulário acadêmico” corresponde às palavras que têm alta frequência, amplo alcance dentro do *corpus* estudado e distribuição similar e constante em todas as áreas de conhecimento analisadas (dentre outros parâmetros detalhados no Capítulo 4 deste trabalho). Assim, ao analisar como o sistema da TRANSITIVIDADE é instanciado pelos verbos acadêmicos mais frequentes nas áreas do conhecimento, pretende-se verificar se seus usos são similares – corroborando e justificando a existência de um vocabulário acadêmico central – ou se são particulares em cada disciplina – corroborando e justificando a existência de listas de vocabulário específicos e/ou técnicos em cada área.

Para solucionar tais questionamentos teóricos, optou-se por estudar um *corpus* de artigos científicos, visto que tais textos são fundamentais na comunidade acadêmica como meio de divulgar pesquisas e trabalhos realizados nas diferentes áreas de conhecimento. Assim, foi compilado um *corpus* formado por artigos científicos das oito Grandes Áreas de conhecimento do CNPq¹: Ciências Agrárias, Ciências Biológicas, Ciências Exatas e da Terra, Ciências da Saúde, Engenharias, Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes. A composição total do *corpus* é aproximadamente 12,3 milhões de palavras distribuídas ao longo de várias Subáreas, tendo sido elaborado de modo a permitir a identificação dos grupos verbais mais frequentes no gênero, sem que a compilação ficasse tendenciosa para uma determinada área.

Nesse sentido, pretendeu-se criar um *corpus* representativo do gênero, possibilitando construir as duas partes centrais desta pesquisa: a identificação dos verbos pertencentes ao vocabulário acadêmico, resultando na criação de uma lista de verbos acadêmicos, relevantes dentro do gênero artigos científico; e a análise do sistema da TRANSITIVIDADE inerente aos lemas verbais mais frequentes, sob a perspectiva da LSF. Além de atender às necessidades desta pesquisa, o *corpus* compilado também serviu de base para o projeto *Corpus Acadêmico do Português Brasileiro (CAPB)*², que será disponibilizado em sua versão ampliada para o público após a conclusão deste trabalho, a fim de contribuir para estudos futuros sobre os diferentes gêneros acadêmicos, e facilitando outras pesquisas linguísticas baseadas no uso do PB.

¹ Árvore do conhecimento. Disponível em: <<http://lattes.cnpq.br/web/dgp/arvore-do-conhecimento>>. Acesso em 06 abril 2018.

² O CAPB é um projeto interinstitucional entre a Universidade Federal de Minas Gerais e a Universidade Federal de Viçosa, que objetiva coletar diferentes textos de gêneros acadêmicos das variadas áreas de conhecimento, com a finalidade de facilitar futuras pesquisas linguísticas na área. Disponível em: <<https://sites.google.com/view/corpusacademico/>>. Acesso em 10 jan. 2019.

Tendo em vista o objetivo de lidar com itens lexicais acadêmicos, a variedade de áreas de conhecimento contempladas nesta pesquisa permite identificar as diferentes categorias do vocabulário (geral, acadêmico, técnico ou de baixa frequência, cf. Capítulo 2), pois quanto mais específico o *corpus* acadêmico, maior é a frequência do vocabulário técnico (SUTARSYAH *et al.*, 1994) e, conseqüentemente, quanto mais variadas as áreas de conhecimento analisadas, maior destaque tem o vocabulário acadêmico em si, alvo de análise desta tese.

Conforme mencionado anteriormente, vários estudos sobre vocabulário foram realizados em língua inglesa nas últimas décadas (MARTIN, 1976; BAKER, 1988, XUE e NATION, 1984; SUTARSYAH *et al.*, 1994; NATION, 2001; MARTÍNEZ, 2001; MARTÍNEZ *et al.*, 2009, TOWNSEND *et al.*, 2012, GARDNER e DAVIES, 2014; DURRANT, 2016). Entretanto, em PB, como discute o trabalho de Dellai (2016), as pesquisas são menos abundantes, geralmente enfocando o ensino de leitura no qual o vocabulário acadêmico é apenas uma pequena parte do processo. Ainda assim, relacionados ao estudo desse tipo de vocabulário, destacam-se a dissertação de Santos (2006), que se propôs elaborar uma lista do vocabulário acadêmico a partir de teses e dissertações de diferentes áreas, e a dissertação de Miranda (2016), autora desta pesquisa, que buscou elaborar uma lista de lemas verbais acadêmicos com base em um *corpus* de artigos científicos, analisando qualitativamente os processos verbais mais frequentes, na perspectiva da LSF.

Desse modo, a presente tese diferencia-se de Santos (2006) tanto no gênero a ser analisado, quanto em sua metodologia. O autor lematizou manualmente apenas substantivos e adjetivos, mantendo verbos em suas formas originais (*type*), assim como advérbios e palavras gramaticais. Nesta pesquisa serão listados os lemas verbais acadêmicos, fundamentais para a análise qualitativa do sistema da TRANSITIVIDADE. Além disso, diferentemente do que foi proposto em Miranda (2016), este trabalho utilizou um *corpus* maior e mais representativo do gênero artigo acadêmico e das oito Grandes Áreas de conhecimento. Também foi empregada uma metodologia de identificação dos itens relevantes para o vocabulário acadêmico distinta daquela utilizada anteriormente.

Cabe mencionar também que o escopo da análise desta pesquisa é mais abrangente, abarcando todo o sistema da TRANSITIVIDADE, na perspectiva de Halliday e Matthiessen (2014), com seus seis diferentes tipos de processos, não apenas os verbais isoladamente, como ocorreu em Miranda (2016).

No âmbito da LSF, o sistema da TRANSITIVIDADE corresponde a uma forma de o usuário representar e organizar suas ideias através da linguagem. Tal sistema relaciona-se à organização sentencial básica, cujos elementos são processos (verbos), participantes e circunstâncias.

Centrais à oração, os processos codificam ações e eventos, estabelecem relações e são realizados por sintagmas verbais, dividindo-se em materiais (como o verbo *fazer*), mentais (*sentir*), relacionais (*ser*, com função de verbo de ligação), verbais (*dizer*), comportamentais (*respirar*) e existenciais (*haver e existir*).

Com relação à análise do uso de processos em artigos científicos, vários estudos vêm sendo realizados sob essa perspectiva (VIVAN, 2010; BÁRBARA e MACÊDO, 2011; PENHA, 2012; MENDES, 2012; LIMA, 2012; 2014; PORTELA, 2013; MORAIS, 2014; 2015; FRUTOSO, 2015; PORTELA *et al.* 2014; PORTELA e BRAUER, 2017). Entretanto, tais trabalhos tratam majoritariamente de como processos verbais ocorrem em uma ou duas áreas de conhecimento. Por sua vez, estudos que abrangem o sistema da TRANSITIVIDADE como um todo – os diferentes tipos de processos – são mais escassos, além de tomarem como base outros gêneros, como o editorial (SOUZA, 2006), literatura infantil (HERBELE, 2018) e notícias de jornais (OLIVEIRA, 2017). Assim, a presente pesquisa se difere das anteriores na área, pois busca analisar os usos dos processos (dentro do sistema da TRANSITIVIDADE) acadêmicos mais relevantes em um *corpus* representativo do gênero estudado como um todo, e também das diferentes Grandes Áreas de conhecimento.

Destarte, a partir desse objetivo geral, pretende-se alcançar outros objetivos mais específicos que dele derivam, como:

- a) Identificar os verbos mais produtivos no *corpus* de artigos acadêmicos, elaborando uma lista de verbos acadêmicos;
- b) Afunilar a pesquisa dos verbos acadêmicos para análise do sistema da TRANSITIVIDADE por meio de amostras de linhas de concordância dos verbos acadêmicos mais frequentes;
- c) Estudar possíveis padrões de uso que configurem, ou não, particularidades de uma determinada área;
- d) Integrar o *corpus* compilado ao projeto CAPB, disponibilizando-o *online* para uso público.

Esses objetivos visam responder às seguintes perguntas de pesquisa:

- a) É possível elaborar uma lista de verbos acadêmicos, representativa das diferentes Grandes Áreas do gênero em PB?
- b) A partir da análise dos usos desses verbos, sob a perspectiva do sistema da TRANSITIVIDADE da LSF, é possível identificar padrões comuns às Grandes Áreas e, portanto, ao gênero acadêmico, ou as particularidades de cada Grande Área estão em maior evidência?

De tal forma, os resultados dessa pesquisa, provenientes dos objetivos e das respostas às perguntas acima, pretendem contribuir para avançar os estudos teóricos na área de tipos de vocabulário, em especial o acadêmico, voltados para o PB, assim como para avançar as pesquisas sobre o sistema da TRANSITIVIDADE em artigos acadêmicos de diferentes áreas do conhecimento. Além disso, o próprio *corpus* disponibilizado após a conclusão do trabalho poderá contribuir para futuras pesquisas sobre o gênero, assim como seu ensino enquanto língua materna ou adicional/estrangeira.

Com base no panorama inicial descrito nesta seção, as duas próximas introduzem o referencial teórico que norteia esta pesquisa – a divisão vocabular da língua (Capítulo 2) e o sistema de TRANSITIVIDADE (Capítulo 3) –, seguidos pela descrição dos procedimentos metodológicos adotados para coleta do *corpus* e elaboração da lista de verbos acadêmicos, assim como a lista em si (Capítulo 4). Na sequência, apresentam-se as análises sobre os usos dos processos mais relevantes em artigos acadêmicos nas diferentes áreas do conhecimento (Capítulo 5). Por último, as Considerações Finais retomam o caminho percorrido para responder os questionamentos levantados neste capítulo.

2 VOCABULÁRIO ACADÊMICO

A presente pesquisa realiza uma análise do sistema da transitividade em um *corpus* de artigos acadêmicos. Para atingir tal objetivo, restringiu-se o escopo do objeto de estudo, selecionando-se apenas os verbos relevantes para o gênero, segundo a perspectiva da divisão do vocabulário de uma língua (NATION, 2001). Desse modo, o presente capítulo discute os diferentes tipos de vocabulário, pormenorizando a revisão teórica a respeito do vocabulário acadêmico, que foi o alvo desta pesquisa. Embora não se tenha pretendido listar um número abrangente dos verbos acadêmicos, já que esta tese se restringe ao gênero artigo acadêmico, acredita-se que, aliado aos procedimentos metodológicos detalhados no Capítulo 3, é possível descrever os grupos verbais relevantes para o gênero em diferentes áreas do conhecimento.

Em vista disso, as próximas seções dedicam-se, primeiramente, a esclarecer os termos utilizados neste e nos demais capítulos; em seguida, a discorrer sobre a divisão do vocabulário em palavras de alta e baixa frequência, técnicas e acadêmicas, dando especial atenção àquelas que nomeiam este capítulo.

2.1 Conceitos básicos para este estudo

Antes de iniciar a discussão sobre os diferentes tipos de vocabulário, é necessário estabelecer alguns conceitos que nortearam esta pesquisa – alguns nem sempre de definição tranquila na literatura. Assim, a seguir, introduz-se os termos empregados tanto nesta revisão teórica, quanto na construção do *corpus* e na análise dos dados, a partir do exemplo no trecho (1), a seguir.

- (1) Como a serapilheira forma uma parte substancial na cadeia trófica do solo, qualquer processo que modifique a sua taxa de aporte e qualidade, como mudanças na composição de espécies arbóreas, causa alterações na temperatura, na umidade, na incidência de luminosidade e nas condições físico-químicas do solo e, potencialmente, influencia diretamente riqueza e densidades de espécies da comunidade macrofauna (A_FL_2017_0_AA_084)³

³ Trecho retirado do *corpus* desta pesquisa, assim como todos sob sigla similar, explicada em detalhes na seção 4.2.

O trecho acima possui 59 *tokens*, 44 *types*, 41 lemas, 41 lexemas e 41 famílias de palavras. Isso significa dizer que os 59 *tokens* se referem às ocorrências únicas de cada palavra do texto, a cada vez que uma sequência de letras ou números é separada por um espaço em branco (SAUTCHUK, 2010, p. 5; BREZINA, 2018, p. 39). Tal conceito de palavra é também denominado “palavra ortográfica” por Trask (2004, p. 218).

Tendo isso em mente, cabe observar que pode haver um problema ao lidarmos com um mesmo *corpus* em mais de um *software* concordanciador: diferentes programas podem fornecer números de *tokens* diferentes para o mesmo *corpus*, de acordo com os critérios próprios utilizados para contabilizar palavras. Alguns, como #LancsBox, contam tanto sequências de letras quanto números; outros, como WordSmith Tools, contam letras, números e pontuação associadas a ele (por exemplo, “2018”, “2017-2018” e “2018,” são considerados três *tokens* distintos). Programas como Sketch Engine e CQPweb também contam a pontuação do texto. Já outros *softwares*, como Kitconc e AntConc, consideram apenas letras. Além desses exemplos, variações no número de *tokens* podem ocorrer em função da presença de clíticos, palavras hifenizadas, entre outros.

A Tabela 1 abaixo mostra a diferença no número de *tokens* do *corpus* desta pesquisa fornecidos por diferentes *softwares*, a fim de demonstrar a importância de sempre explicitar o *software* utilizado para análise, visto que tal informação pode impactar quantitativamente os dados. Assim, cabe registrar que o *software* utilizado para determinar o número de palavras deste *corpus* foi o KitConc devido tanto aos seus critérios (contabiliza apenas letras), quanto às suas demais ferramentas disponíveis para análise linguística.

Tabela 1 - Exemplos da contabilização de *tokens* por diferentes softwares

<i>Software</i>	<i>Tokens</i> totais	<i>Types</i>
#LancsBox	12.540.589	271.096
Wordsmith Tools	12.606.924	239.204
AntConc	12.331.643	174.174
KitConc	12.320.167	174.856

Fonte: Elaboração própria.

Retomando o trecho (1), conforme mencionado, há 44 *types*, que são as formas únicas de uma palavra no *corpus*. Isto é, se *tokens* lidam com as ocorrências totais das sequências de

palavras ou números, os *types* referem-se a uma mesma sequência. Assim, no exemplo, a palavra “na” ocorreu 5 vezes (*tokens*), entretanto, é considerada apenas 1 *type*.

Diferente de *tokens* e *types*, que se baseiam na forma de uma palavra, para identificação de lemas e lexemas é necessário analisar linguisticamente o texto. **Lemas** são determinados a partir de análise morfológica, agrupando todas as formas flexionadas de um mesmo radical que pertencem a uma única classe de palavras (KUČERA; FRANCIS, 1967, p. 1 apud BREZINA, 2018, p. 40). Assim, são agrupadas formas plural e singular de uma palavra, ou flexões verbais, ou superlativas de um adjetivo. Em (41), “na” e “nas” pertencem ao mesmo lema.

Por sua vez, **lexemas** são lemas com um significado específico atribuído a eles, a fim de distinguir palavras polissêmicas, conforme, por exemplo, as diferentes entradas de um mesmo lema no dicionário (CRYSTAL, 2008, p. 276). No exemplo (1), “taxa” com significado de ‘índice, proporção’ é um lexema distinto de “taxa” enquanto ‘imposto’.

Englobando os conceitos de lema e lexema, a **família de palavras** consiste na forma básica da palavra, acrescida de todas as suas formas flexionadas e derivadas, mantendo aproximado o sentido ao da forma base (BAUER e NATION, 1993). Assim, os lemas “alterações” e “alterar” pertenceriam à mesma família de palavras.

Nesse sentido, ao levar em consideração cada conceito descrito anteriormente em uma análise linguística de *corpus*, há certas vantagens e desvantagens que foram resumidas no Quadro 1, a seguir.

Quadro 1 – Vantagens e desvantagens de *types*, lemas, lexemas e famílias de palavras

Definição de palavra	Vantagens	Desvantagens
<i>Type</i>	Categoria de baixa-inferência	Não há distinção gramatical ou semântica entre as formas.
Lema	Distinção entre formas com diferentes funções gramaticais	Envolve processamento automatizado para etiquetamento morfossintático e lematização, sujeitos a erros
Lexema	Categoria mais específica em que o significado é levado em consideração	Categoria de alta-inferência, sujeita a erros, ainda não disponível de modo completamente automático

Família de palavras	Agrupamento de palavras com o mesmo radical, próximas semanticamente	Não há distinção gramatical entre as formas.
---------------------	--	--

Fonte: Adaptado de Brezina (2018, p. 41)

2.2 Tipos de vocabulários

Uma das motivações para pesquisas na área de tipos de vocabulário e ensino de línguas, segundo Coxhead (2013, p. 113), é a pergunta “De qual vocabulário precisam os alunos de *English for Specific Purposes* (ESP)?”. Embora a resposta não seja sempre objetiva, suscita outras perguntas a respeito do propósito do aluno, de sua área de estudo ou de seu tempo de aprendizagem – preocupações que transcendem o ensino de língua estrangeira. Consequentemente, para responder tal questão deve-se considerar que certas palavras possuem significado técnico dentro de uma determinada área de conhecimento (como *massa* em Física) e outras que costumam ocorrer em um conjunto limitado de áreas (como *esterno* em áreas de Saúde e Biológicas).

Tais termos, ensinados a partir da perspectiva ESP, possuem diferentes denominações tais quais: vocabulário 1) para propósitos específicos; 2) especializado; 3) técnico; 4) subtécnico e 5) semitécnico. A variação se dá apenas em questão de nomenclatura, pois todos se referem aos termos específicos de uma área profissional ou de estudos.

Ainda segundo a autora (COXHEAD, 2013, p. 116), o ensino do vocabulário específico é importante para a aprendizagem de uma língua, pois professores e aprendizes precisam ter certeza de que o conteúdo administrado no tempo limitado de sala de aula será proveitoso para seus propósitos. Nesse âmbito teórico, parte-se do princípio de que o conhecimento de uma área esteja diretamente relacionado à linguagem daquela disciplina. O bom domínio da língua utilizada naquele meio é um indicador do pertencimento dos aprendizes àquele grupo, pois é através da língua que se demonstrará o conhecimento sobre a área (WOODWARD-KRON, 2008, p. 246).

Pressupõe-se, portanto, que exista um vocabulário especializado em determinadas áreas e que, em contrapartida, exista também um vocabulário comum aos diferentes gêneros e usos da língua. Consequentemente, argumenta-se a favor da existência de um vocabulário central de uso geral, abrangendo as diferentes áreas da vida e do trabalho, denominada “linguagem básica”

(BASTURKMEN, 2006, p. 16). Nessa abordagem, segundo Quirk *et al.* (1972) e Nation (2001; 2008), acredita-se que o aprendiz deve primeiro dominar a linguagem básica, de uso geral cotidiano, para depois estudá-la com propósitos específicos.

Partindo dessa premissa, Coxhead e Nation (2001) categorizam o vocabulário em quatro grupos de palavras: as **palavras de alta frequência** (linguagem básica) e as **palavras de baixa frequência**, que são as mais utilizadas em geral; e os termos **vocabulário técnico** e **vocabulário acadêmico**, referindo-se àquelas mais frequentes em contextos discursivos específicos. Conforme Webb e Nation (2017), tal classificação é importante por definir a quais palavras maior atenção deve ser dada durante o ensino.

Quanto às categorias, as palavras de alta frequência são comumente encontradas nas linguagens escrita e falada, por essa razão, não compreendê-las pode comprometer o entendimento e dificultar a comunicação. Um exemplo é o par *conhecimento* e *compreensão* que pode ser considerados sinônimo, entretanto, o primeiro está entre os mais frequentes da língua portuguesa (entre as 1.000 primeiras⁴, especificamente na posição 311°), de forma que sua aprendizagem se torna prioridade sobre a outra palavra, cujas ocorrências a colocam na 2.132° posição do *ranking* do *corpus* PtTenTen11 (representante de vários gêneros escritos da Língua Portuguesa, em um total aproximado de 4,6 bilhões de palavras).

Por sua vez, tomando como base o *corpus* da presente tese, o Quadro 2 exemplifica amostras aleatórias⁵ de diferentes palavras em variadas posições do *ranking*⁶ deste *corpus*. O gráfico procura mostrar que quanto mais frequente, maior a chance de familiaridade do usuário com a as palavras. Por outro lado, as palavras mais próximas da coluna direita são menos frequentes e, em função disso, menos usadas e pouco prováveis de serem conhecidas em toda sua extensão semântica, inclusive pelo usuário nativo da língua.

Quadro 2 – Exemplos de palavras em diferentes níveis de frequência no presente corpus, acompanhadas por sua posição no ranking de frequência de lemas

Entre as 1.000 primeiras	Entre 1.000 e 2.000	Entre 2.000 e 5.000	Entre 5.000 e 10.000	Entre 10.000 e 20.000
médio (105)	espaço (1.110)	simulação (2.495)	reformulação (7.486)	patógeno (11.698)

⁴ O cálculo das posições no *ranking* de frequência foi baseado na lista lematizada do PtTenTen11.

⁵ As amostras aleatórias foram escolhidas com auxílio deste site

<https://www.4devs.com.br/gerador_de_numeros_aleatorios>, pareando números e itens lexicais.

⁶ O cálculo das posições no *ranking* de frequência foi baseado na lista lematizada do *corpus* desta pesquisa.

final (417)	cobertura (1.202)	progresso (3.466)	cateter (8.673)	diafragmático (15.779)
decisão (622)	alternativa (1.430)	hidráulico (4.676)	estocástico (9.154)	imputar (18.161)

Fonte: Elaboração própria.

A fim de melhor compreender a relevância de utilizar o *ranking* de frequência de *types*, assim como a sua subdivisão em níveis, a Tabela 2 a seguir apresenta a relação entre os níveis de frequência e a porcentagem de *tokens* em relação ao total de ocorrências do *corpus*.

Tabela 2 – Relação entre os níveis de frequência de *types* e a porcentagem de *tokens* no *corpus*

Nível de frequência	Posição	<i>Tokens</i>	Porcentagem referente às ocorrências totais ⁷	Número de ocorrências do último lema do nível
1° 1.000	1 – 1.000	8.534.968	68,1	1.424
2° 1.000	1.001 – 2.000	992.094	7,9	703
3° 1000	2.001 – 3.000	546.278	4,4	424
5° 1.000	4.001 – 5.000	250.699	2,0	217
10° 1.000	9.001 – 10.000	87.920	0,7	81
20° 1.000	19.001 – 20.000	28.101	0,2	27

Fonte: Elaboração própria.

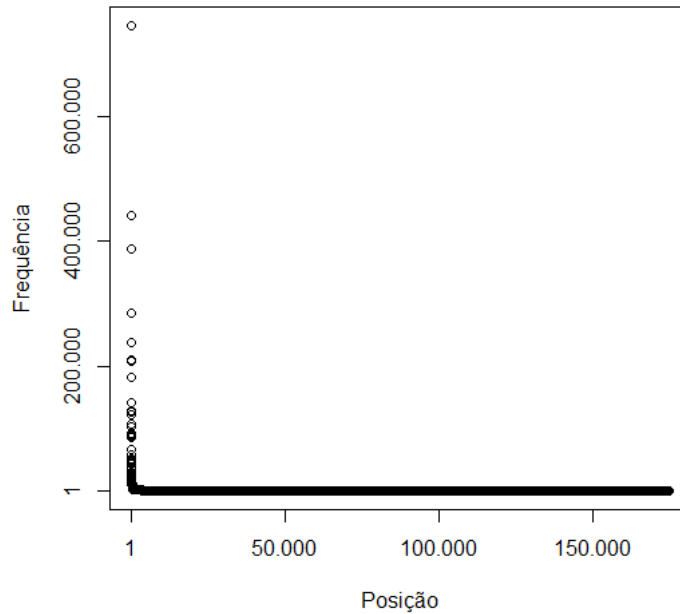
A diminuição drástica observada nos valores percentuais da Tabela 2 pode ser explicada pela Lei de Zipf (1949), nomeada em homenagem ao psicolinguista George Kingsley Zipf, na qual se determina que, ao começar pela palavra mais frequente de uma lista de palavras, a segunda mais frequente terá apenas metade da frequência do primeiro item. Já a terceira mais frequente terá um terço da primeira e assim progressivamente. Isso implica que a quantidade de evidência que um *corpus* pode fornecer diminui com rapidez. Além dessa característica, a lei de Zipf estabelece a existência de um número muito pequeno de palavras muito frequentes e, concomitantemente, um número muito alto de palavras de baixa frequência. Ou seja, embora haja um número elevado de palavras que se repetem, geralmente palavras gramaticais como

⁷ Cálculo baseado na lista lematizada de frequência fornecida pelo *software* #LancsBox, utilizando o etiquetador e os parâmetros morfossintáticos TreeTagger.

artigos, conjunções etc., a maior parte das palavras ocorrerá apenas uma ou duas vezes em um mesmo texto (WEBB e NATION, 2017).

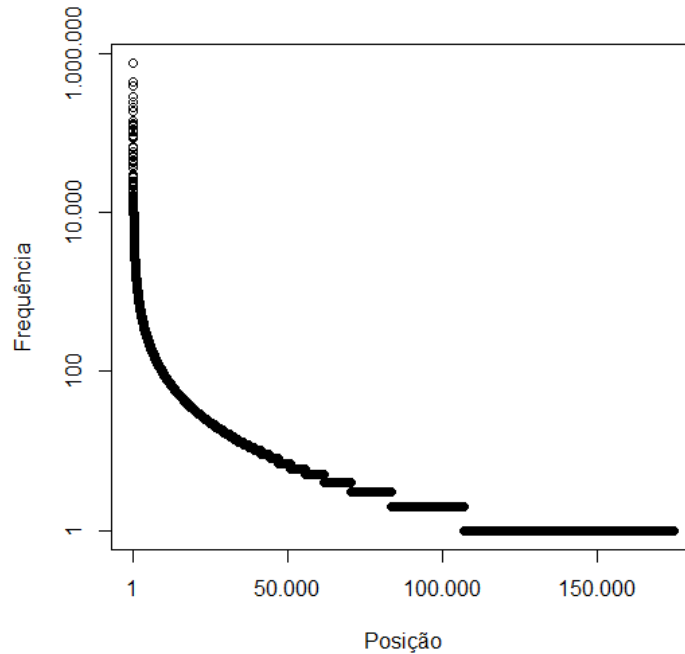
A Figura 1 e Figura 2 abaixo ilustram tais propriedades ao representar a distribuição da frequência das palavras deste *corpus* por posição em escala linear (Figura 1) e em escala logarítmica (Figura 2) – na qual cada marca do eixo X é a marca anterior multiplicada por 10 –, a fim de melhor visualizar o ponto em que a curva começa a se dispersar. Após as figuras segue a Equação 1, que resume o modelo matemático pelo qual a Lei de Zipf pode ser expressa e, logo, calculada.

Figura 1 – Distribuição da frequência absoluta das palavras pela posição no ranking, escala linear



Fonte: Elaboração própria.

Figura 2 – Distribuição da frequência absoluta das palavras pela posição no ranking, eixo X em escala log



Fonte: Elaboração própria.

Equação 1 – Modelos para o cálculo da Lei de Zipf

$$\text{Frequência absoluta da palavra} \times \text{Posição na lista de palavras} \cong \text{Constante}$$

ou

$$\text{Frequência absoluta} \cong \frac{\text{Constante}}{\text{Posição na lista de palavras}}$$

em que a *constante* é a frequência do primeiro item na lista de palavras.

Cabe observar que a lei de Zipf representa uma aproximação, pois as frequências em um *corpus* real podem diferir levemente daquelas previstas por este modelo. Entretanto, a partir dessa lei, pode-se chegar a interessantes implicações para análise linguística, tais quais: 1) a necessidade de se avaliar criticamente o tamanho da evidência da língua que há no *corpus* ao analisar determinados fenômenos, especialmente ao se trabalhar com palavras de baixa frequência, pois não há grande diferença entre os valores das palavras nesse nível, de acordo com o *corpus* compilado (WEBB e NATION, 2017); 2) a importância de se utilizar um *corpus* representativo e coerente com os fenômenos linguísticos que se deseja investigar; 3) a

relevância de se procurar outras fontes comparáveis a fim de verificar a evidência (ou falta) de uma palavra e/ou fenômeno (BREZINA, 2018, p. 46).

A partir dessas características estruturais de um *corpus*, pode-se perceber que trabalhar com a subdivisão em níveis de frequência auxilia o ensino e aprendizagem de língua estrangeira. Nesse âmbito, pesquisas (SCHIMITT; SCHIMITT; CAPHAM, 2001; WEBB e CHANG, 2012) demonstram que aprendizes de língua estrangeira tendem a aprender o vocabulário de acordo com sua frequência de uso, dominando boa parte das palavras do primeiro nível, as 1.000 primeiras, mais facilmente quando comparadas com as pertencentes ao segundo nível. Do mesmo modo, aprendizes dominam melhor as palavras do segundo nível do que as do terceiro. A partir daí, segundo Webb e Nation (2017), há pouca necessidade de priorizar o ensino de determinadas palavras se não considerar os propósitos do aprendiz.

Nas seções a seguir, discutem-se as quatro diferentes categorias do vocabulário propostas por Nation (2001), a fim de melhor definir o vocabulário acadêmico, que norteia a presente pesquisa.

2.3 Palavras de alta frequência e de baixa frequência

Palavras de alta frequência na língua tendem a ter maior valor no ensino-aprendizagem do que aquelas menos frequentes, pois são mais prováveis de serem utilizadas em uma situação comunicativa. Assim, palavras diferentes possuem valores diferentes para aprendizes sendo, por exemplo, mais útil conhecer *setor*, *profissional* e *público*⁸, que são frequentes na língua e, portanto, possuem maior probabilidade de uso, do que *embarcadouro*, *alpestre*, e *reestimativa*⁹, cujas ocorrências são escassas ao longo dos *corpora* consultados. Em vista disso, ao lado dos propósitos do aprendiz, a frequência também deve ser um dos critérios para identificação das palavras mais valiosas para esse vocabulário.

Desse modo, listas de palavras podem ser bastante úteis para o ensino, especialmente aquelas de alta frequência. Em língua inglesa, uma das mais conhecidas é *General Service List* (GSL), de West (1953), contendo aproximadamente 2.000 famílias de palavras baseadas em um *corpus* de cinco milhões de palavras coletadas de textos escritos. Estudos (COXHEAD e

⁸ Todos os exemplos estão entre as 1.000 palavras mais frequentes do *Corpus* Brasileiro (com 1 bilhão de palavras) e do PtEnTen11 (com 4,5 bilhões).

⁹ Todos os exemplos foram palavras de baixa ocorrência no *Corpus* Brasileiro e no PtEnTen11, com menos de 0,001% de ocorrências relativas a cada *corpus* (e não figuraram no *corpus* desta pesquisa).

HIRSH, 2007; HYLAND e TSE, 2007; BREZINA e GABLASOVA, 2015) demonstram que, apesar da idade da lista, a GSL continua atual cobrindo cerca de 71,5 a 89,6% do vocabulário em diferentes gêneros textuais. Em função desse amplo alcance, a lista de West pode ser considerada o ponto de partida para o ensino de vocabulário de uma língua (WEBB e NATION, 2017). Entretanto, sua relevância vem decaindo progressivamente, especialmente em função do surgimento de novas palavras – *computer* e *internet* – e queda de outras – *carriage* e *telegraph*, que impactam na atualidade da lista. Além dessa limitação, a GSL é tendenciosa para a variedade escrita da língua inglesa, assim como a baixa cobertura de palavras abaixo das 1.000 primeiras (DANG e WEBB, 2016).

Visando solucionar tais problemas, a GSL foi reformulada e recompilada sob a denominação de *New-General Service List* (NGSL), por Brezina e Gablasova (2015). A nova lista é composta por 2.492 lemas baseados em um *corpus* que inclui registros escritos e orais, embora majoritariamente o primeiro, coletados a partir de quatro outros *corpora* (LOB, BNC, BE06 e EnTenTen12), em um total de 12 bilhões de palavras. Tal lista se diferencia das semelhantes por ter sido criada a partir de uma abordagem quantitativa, que levou em consideração três critérios: frequência, dispersão e distribuição das palavras ao longo dos diferentes *corpora*. Outra característica relevante é o agrupamento de *types* em lemas – a unidade da lista

Duas vantagens dessa lista são a grande quantidade de amostras da língua em uso, consideravelmente maior que as demais, e a divisão de seus itens lexicais em vocabulário central e vocabulário atual, o que permite a professores e aprendizes observar a variação na língua ao longo do tempo (DANG e WEBB, 2016a). Apesar de todos os pontos positivos, cabe mencionar que a NGSL possui duas limitações: a tendência para o variedade grã-bretã do inglês e a limitação metodológica aos dados quantitativos, possibilitando a não-inclusão de itens que não são de alta frequência, mas ainda assim são relevantes para aprendizes da língua.

Ao lado da GSL e da NGSL, há também outras recentes, como a lista do *British National Corpus* (BNC), representativa da variedade formal escrita do inglês da Grã-Bretanha; o *Corpus of Contemporary American English* (COCA), que abrange as modalidades da língua inglesa escrita e falada. Nation (2012) utilizou ambos *corpora* para criação da BNC/COCA2000, que consiste em uma lista baseada em um *corpus* de 6 milhões de palavras de fontes orais e 4 milhões de palavras escritas, abarcando diferentes gêneros, desde filmes até ficção infantil, e as diversas variedades da língua, como o inglês da Nova Zelândia ao lado do grã-bretão e o estadunidense. Nessa lista, Nation incluiu palavras de baixa frequência que poderiam ser úteis para aprendizes de língua estrangeira, a quem se dedica a lista.

Tendo em vista tais informações, cabe mencionar que as listas anteriores de itens lexicais mais frequentes da língua são organizadas em famílias de palavras (com exceção da NGSL). Entretanto, Dang e Webb (2016b) demonstram em seu estudo que, se utilizados lemas na formação das listas, elas cobririam maior proporção de textos do que qualquer uma das listas originais. Essa pesquisa resultou na elaboração da *Essential Word List* (EWL), que representa um vocabulário prático e específico dos 800 lemas mais frequentes em língua inglesa escrita e falada, fornecendo material lexical fundamental para iniciantes (WEBB e NATION, 2017).

Quanto à sua aprendizagem, as palavras de alta frequência geralmente são incorporadas ao vocabulário do aprendiz ao se deparar com elas em uso no dia-a-dia, porém, seu ensino deliberado pode ser necessário, visto que nem sempre é possível para o aprendiz estar imerso na cultura e na língua estudadas. Portanto, para Webb e Nation (2017) o ensino dessas palavras deve ser recorrente em sala de aula, pois o tempo dedicado à aprendizagem desse grupo reduzido de palavras se justifica em função de sua frequência, cobertura e alcance ao longo dos diferentes propósitos pelos quais a língua é aprendida (NATION, 2001, p. 16).

Ainda cabe discutir dois *corpora* semelhantes aos utilizados para criar as listas mencionadas anteriormente, porém, em língua portuguesa: o *Corpus Brasileiro* e o *PtTenTen11*, que abarcam amostras de textos de diferentes gêneros. O primeiro, de acordo com Berber-Sardinha (2014), conta com textos representantes da língua escrita como jornais, acadêmicos, páginas da internet (Wikipédia, por exemplo), manuais, entre outros; assim como da língua falada, tais quais audiodescrição de entrevistas narrações futebolísticas e debates e discursos presidenciais, ambas modalidades totalizando um *corpus* de aproximadamente 1 bilhão de palavras. Já o *PtTenTen11* corresponde a um *corpus* em língua portuguesa (englobando as variações PB e Português Europeu, PE) compilado a partir de textos variados coletados da Web e que pertence à Família TenTen – um projeto que visa coletar *corpora* de diferentes línguas e conta, atualmente, com cerca de 30 idiomas disponíveis para uso público.

O *PtTenTen 11* possui aproximadamente 4,6 bilhões de *tokens* gerais e, apesar de abranger duas variantes de língua portuguesa, pode-se escolher entre uma das duas. Na versão PB, o *corpus* possui cerca de 3,6 bilhões de *tokens*, englobando diferentes gêneros disponíveis na Web, desde blogs até artigos científicos.

Como ambos *corpora* dedicaram uma parte de sua compilação a textos acadêmicos, é pertinente mencionar que o *Corpus Brasileiro* aloca uma parte consideravelmente maior aos textos acadêmicos do que aos demais gêneros – mais de 50% do *corpus* é composta por gêneros científicos, como artigos, teses e dissertações. Por outro lado, a parte acadêmica do *PtTenTen11* não pode ser mensurada da mesma forma, pois a compilação se deu por meio da busca de texto

automática. Entretanto, alguns parâmetros foram determinados quanto à não duplicação de material linguístico, assim como à busca em diferenciados e relativamente equilibrados domínios de textos que oscilam entre diferentes graus de formalidade (KILGARRIFF *et al.*, 2014).

Em vista dessas características, este capítulo trabalha com ambos *corpora* como referência ao lidar com amostras da língua geral, não-acadêmica, a fim de ilustrar o PB em diferentes gêneros, mesmo sendo tendencioso para o acadêmico no *Corpus Brasileiro*.

Retomando a descrição dos tipos de vocabulário, por sua vez, as palavras de baixa frequência correspondem a todas aquelas que não são consideradas de alta frequência, de modo que há milhares de palavras nessa categoria, conforme ilustrado na Figura 1 e na Figura 2, representantes da Lei de Zipf, que englobam palavras como *subsistir* (dentro do décimo primeiro conjunto de mil palavras¹⁰) e também *conjetura* (dentro do conjunto do trigésimo quinto milhar). Segundo Goulden, Nation e Read (1990), falantes nativos reconhecem tipicamente entre 15 e 20 mil famílias de palavras, de forma que conseguiriam classificar o último exemplo como de baixa frequência, tendo em vista as respectivas posições na lista.

Schmitt e Schmitt (2014) dividem as palavras de baixa frequência em dois grupos: de média frequência e de baixa frequência. Nessa classificação, as palavras de alta frequência correspondem às primeiras 3.000 famílias de palavras, as de média frequência cobrem de 4.000 a 9.000 e as de baixa frequência referem-se às restantes.

Para os autores, o conhecimento do vocabulário de média frequência é suficiente para compreender interações tanto escritas quanto faladas, assim, a distinção de palavras até esse nível pode ser útil para o ensino (NATION, 2006). Apesar disso, cabe lembrar que a fronteira entre os grupos (de alta, média e baixa frequência) é arbitrária, de modo que qualquer palavra entre as categorias médias ou baixas pode ser candidata a uma das de alta frequência em função de sua posição, do seu alcance em diferentes textos ou do tipo de *corpus* no qual se baseia a lista.

No caso do *corpus* desta pesquisa, formado por artigos científicos de diferentes áreas do conhecimento, espera-se que palavras de média e baixa frequência na língua (tendo como referência o *corpus* PtTenTen11) sejam de alta frequência neste. Assim, considerando um *subcorpus* da pesquisa, como Ciências da Saúde, espera-se que palavras de baixa frequência no *corpus* de referência – por exemplo, *regressão* e *prontuário* – tenham frequência relativa maior no *subcorpus* desta tese. Isso ocorre porque os itens lexicais mencionados podem ser

¹⁰ Tomando como referência o corpus PtTenTen11 em ambos casos exemplificados.

considerados de baixa frequência na língua (correspondendo a 0,038% e 0,029% dos *tokens* totais do *corpus* e ocupando as posições 10.880° e 12.320°, respectivamente), porém são técnicos na área de Saúde, em função de sua frequência e posição (0,32% e 0,34% dos *tokens* totais e posições 386° e 411°, respectivamente), entre outros critérios explicitados na seção 2.4.

Apesar dessa possibilidade de dupla classificação, algumas palavras de baixa frequência são apenas isso, palavras pouco empregadas na língua, tais quais, *âmago*, *tergiversar*, *pernóstico*, que dificilmente são utilizadas para expressar ideias cotidianas¹¹, cujos significados se aproximam daqueles de outras palavras mais frequentes (como *essência*, *hesitar* e *pretensioso*, respectivamente), ou cujo uso seja marcado por arcaísmos, regionalismos, estrangeirismos ou (in)formalidade.

Em vista disso, Nation (2001) sugere que as palavras de baixa frequência sejam ensinadas utilizando estratégias alternativas (diferentes das empregadas com o vocabulário de alta frequência, para qual ampla exposição do aprendiz às palavras é peça central), tais como inferir seu significado a partir do contexto ou através de seus radicais, prefixos ou sufixos, além de encorajar consultas a dicionários. Isto é, professores não devem passar boa parte da aula ensinando tal vocabulário extenso e pouco frequente, mas se concentrar em ensinar aos alunos estratégias sobre como lidar com ele.

2.4 Palavras técnicas

As palavras técnicas são aquelas relativamente pouco utilizadas na língua cotidiana, tais quais as de baixa frequência, porém, são frequentes em uma determinada área de conhecimento, de modo que são consideradas técnicas daquela área – conforme o par *regressão* e *prontuário*, discutidos na seção anterior, cujas ocorrências em um *corpus* de língua geral são baixas, mas frequentes em um *corpus* especializado.

De tal modo, compreende-se que as palavras técnicas ocupam lugar privilegiado em certas áreas, sendo, portanto, seu conhecimento é essencial para usuários inseridos naquele contexto. Cabe mencionar que, diferente das palavras de alta frequência, o vocabulário técnico é aprendido concomitantemente com o conhecimento sobre o tema e, por isso, para ele geralmente há glossários e explicações do autor do texto, entre outros recursos de aprendizagem, pois seus significados se restringem àquela área (NATION, 2001).

¹¹ Ocupam as posições 22.667°, 64.017° e 167.784°, respectivamente, no *Corpus* Brasileiro. Por sua vez, no PtTenTen11, ocupam as posições 18.162°, 38.311° e 64.752°, respectivamente.

Além disso, vale ressaltar que o vocabulário técnico não é apenas composto de palavras pouco frequentes na língua, mas também daquelas pertencentes às frequências alta e média (CHUNG e NATION, 2003). Por exemplo, termos como *coração*, *sangue e pele* são palavra de alta frequência na língua portuguesa¹² e, concomitantemente, são consideradas técnicas da área de Saúde. Por outro lado, *vértebra*, *tendão* e *traqueia* são palavras de baixa frequência na língua¹³ e também podem ser classificadas como técnica nesse mesmo contexto.

Resumidamente, uma palavra técnica é aquela que pode ser reconhecida como frequente e/ou relevante dentro de uma certa área. Assim, existem níveis de tecnicidade que dependem do quão restrito o uso de uma palavra é em relação à essa área. Conforme Nation (2001), tais níveis são organizados em quatro categorias, do mais técnico (primeira) ao menos técnico (quarta), conforme apresentado no Quadro 3, a seguir.

Quadro 3 - Subdivisão do vocabulário técnico

	Exemplos
Categoria 1	
Itens que raramente ocorrem fora de determinada disciplina.	Saúde: colagênese, tricômio. Letras: lexema, morfema.
Categoria 2	
Itens usados dentro e fora de determinada disciplina, mas não com mesmo significado.	Direito: ação, recurso. Computação: executar, rodar.
Categoria 3	
Itens utilizados dentro e fora de uma disciplina, mas cuja maioria dos usos pertence àquela área, estando o significado disponível/acessível fora da disciplina.	Saúde: tórax, esterno. Engenharia elétrica: energia, potência.
Categoria 4	
O item é mais comum na disciplina do que fora dela, não havendo especialização do significado, entretanto, alguém com	Saúde: coração, cérebro. Letras: palavra, sentido. Computação: programa, arquivo.

¹² As três palavras figuram entre os 3.000 lemas mais frequentes, tanto no PtTenTen11, nas posições 844°, 1409° e 1.351°, respectivamente; quanto no *Corpus* Brasileiro, nas posições 2.056°, 1.636° e 2.618°, respectivamente.

¹³ No *corpus* PtTenTen11, configuram respectivamente as posições 19.678°, 12.256° e 29.114°, já no *Corpus* Brasileiro, estão posicionadas respectivamente em 20.338°, 22.863° e 23.370°

conhecimento da área tem maior discernimento sobre seu significado.	
---	--

Fonte: Parcialmente adaptado de Nation (2001, p. 317), com alguns exemplos retirados do *corpus* da pesquisa.

Desse modo, as palavras da Categoria 1 são técnicas, particulares àquela área do conhecimento, e podem ser encontradas através da frequência e alcance em um *corpus* representativo da(s) disciplina(s) (YANG, 1986 apud NATION, 2001, p. 317). Os indivíduos que dominam as palavras desta categoria também dominam a área de estudo, pois sua aprendizagem ocorre concomitantemente com a aprendizagem da disciplina. Já na Categoria 2 são consideradas palavras técnicas aquelas cujo significado utilizado na linguagem comum não fornece *insight* sobre o significado técnico.

Por sua vez, as Categorias 3 e 4 são menos transparentes, pois não são exclusivas da área de conhecimento. Assim, na Categoria 3, as palavras ocorrem dentro e fora da área, embora a maior parte dos usos sejam dentro da disciplina, enquanto na Categoria 4 os itens lexicais e seu significado estão prontamente disponíveis fora da área de conhecimento, entretanto, um especialista terá maior conhecimento técnico sobre aquele item do que um leigo.

Classificar palavras nessas quatro categorias depende de critérios relativos à sua frequência e significado, além de seu *alcance* ao longo de diferentes textos de variadas disciplinas. Cabe ressaltar que as Categorias 2 e 3 demonstram que apenas tais critérios não são suficientes, pois necessita-se também de uma análise semântica do significado do item lexical. Com base no exposto, Nation (2001) determina que cabe ao pesquisador estabelecer o ponto de corte para classificação de palavras técnicas, principalmente referente às Categorias 2, 3 e 4.

Destarte, as palavras técnicas, que também podem ser denominadas **terminológicas** (NATION, 2001, p. 318), requerem conhecimento científico ou especializado para sua compreensão devido ao seu significado e a baixa frequência fora daquela área de conhecimento, o que faz seu alcance ao longo das disciplinas ser bastante limitado. Consequentemente, em um *corpus* como o da presente pesquisa, especializado e simultaneamente englobando distintas disciplinas, as palavras técnicas tendem a ser de baixa/média frequência, ocorrendo em um número limitado de textos e *subcorpora*.

De acordo com a pesquisa de Chung e Nation (2003) sobre esse tipo de vocabulário, cerca de 31,2% das palavras de um texto de Anatomia e 20,6% das palavras de um texto de Linguística Aplicada foram consideradas técnicas por profissionais da área, a partir de testes elaborados pelos pesquisados. Tal resultado implica dizer que em uma página, com média de

400 palavras, 125 são técnicas na primeira disciplina, e 82 são técnicas quanto à segunda. Embora esses valores considerem a recorrência da mesma palavra, a dificuldade de compreender o texto seria expressiva para um leitor não-familiarizado com os termos.

A partir de tal pesquisa, observa-se que áreas distintas possuem volumes diferentes de palavras técnicas. Em geral, isso se dá porque algumas dessas palavras são frequentemente utilizadas com significado técnico fora de sua determinada área, de modo que esses termos podem ser considerados de alta frequência ou parte do vocabulário acadêmico, embora sejam técnicos naquela disciplina (FLOWERDEW, 1992; CHUNG e NATION, 2003).

Além da escala de níveis, Coxhead (2013) e Chung e Nation (2003) propõem outras estratégias para identificação do vocabulário técnico, tais quais o uso de dicionários especializados, a consulta a especialistas na área ou a utilização de ferramentas da linguística de *corpus*, como listas de frequência.

Em relação a essas listas, algumas vêm sendo desenvolvidas voltadas para disciplinas específicas (tais quais HSU, 2013; COXHEAD e HIRSH, 2007), embora majoritariamente voltadas à língua inglesa, na qual proliferam estudos sobre os tipos de vocabulário. Em PB, algumas pesquisas também começam a surgir, associadas à vertente de estudos da Terminologia e à construção de glossários.

Do ponto de vista do ensino acadêmico, segundo Nation (2001, p. 19), como o vocabulário técnico também é de alta frequência, os alunos ou profissionais da área devem ser ensinados a estabelecer conexões entre eles de modo que consigam distinguir entre significados de alta frequência e os técnicos. Por exemplo, é preciso saber diferenciar *parede celular* de seus itens lexicais composicionais (*parede* e *celular*). Logo, ao se requerer conhecimento específico de uma determinada área de conhecimento, professores podem fornecer estratégias que auxiliem os alunos a compreender e lembrar desse tipo de vocabulário, considerando que tais palavras só fazem sentido quando aprendidas em contexto, conforme explica o autor, como o termo *morfema* cujo significado precisa ser aprendido enquanto parte do estudo de linguística, não antes.

2.5 Palavras acadêmicas

O vocabulário acadêmico se refere àquele que se estende constantemente ao longo de textos acadêmicos de diferentes disciplinas, ao mesmo tempo em que pode ser infrequente em textos não-acadêmicos. A essas palavras também denomina-se vocabulário **especializado**

(COXHEAD, 2013) ou **subtécnico** (BAKER, 1989; NATION, 2001), pois, enquanto palavras técnicas possuem significado relevante para a compreensão de um certo tópico, as acadêmicas são empregadas de modo a auxiliar o uso das técnicas – daí provém a terminologia *subtécnica* – e são geralmente compostas por palavras formais (NATION, 2001, p. 17). Tal vocabulário não costuma ser ensinado em cursos, por ser menos visível que o técnico (COXHEAD, 2000), entretanto, não menos importante.

Com base nisso, de acordo com Webb e Nation (2017), listas de palavras acadêmicas devem ser ensinadas somente após a aprendizagem daqueles de alta frequência, em função de sua grande proporção e seu impacto sobre a compreensão do vocabulário pelo aprendiz. Em conjunto, tais listas fornecem bases empíricas para ensino com fins específicos, impactando positivamente a compreensão dos discursos acadêmicos – desde aulas à confecção de textos, já que o domínio sobre o vocabulário tem o maior impacto sobre se o que foi falado ou escrito será compreendido ou não (LAUFER e SIM, 1985 *apud* WEBB e NATION, 2017).

Vale destacar que, às vezes, algumas das palavras acadêmicas podem se relacionar à área (ou um determinado grupo de áreas) em que ocorrem, entretanto, contribuem majoritariamente para aquilo que os pesquisadores/autores fazem. Ou seja, tais palavras permitem referenciar o trabalho de outros (*assume, establish, indicate, conclude e maintain*, conforme enumera NATION, 2001), assim como permitem trabalhar com os dados de forma científica (*analyse, assess, concept, definition, categories e seek*), por exemplo.

Em língua inglesa, há diferentes listas que retratam tais palavras, sendo uma das mais conhecidas a *Academic Word List* (AWL) (COXHEAD, 2000), que apresenta 570 famílias de palavras de diferentes disciplinas que não estão na GSL. Tal lista, segundo a autora, cobre cerca de 10% dos textos acadêmicos e 1,4% de textos não acadêmicos. Embora tal número não pareça significativo, a cada página (cerca de 400 palavras), 40 são consideradas acadêmicas, o que pode afetar diretamente sua compreensão pelo aprendiz, dado que representam processos fundamentais do trabalho de pesquisa. Na AWL, as palavras são classificadas de acordo com seu alcance ao longo das disciplinas, assim como sua frequência, e organizadas em sub-listas por ordem alfabética, sendo que as primeiras sub-listas contêm as palavras mais frequentes.

Apesar de muito utilizada, essa lista possui algumas importantes limitações tais qual o fato de ter sido baseada na GSL, que, como dito anteriormente, foi criada em 1953 e está desatualizada. Além disso, alguns de seus itens, antes acadêmicos, agora são considerados palavras de alta frequência, de acordo com listas mais recentes (COBB, 2010). Outra limitação é cobertura das palavras ao longo de diferentes discursos acadêmicos que tem seu valor restrito à 4% de discursos acadêmicos falados (DANG e WEBB, 2014). Uma quarta limitação é sua

organização em famílias de palavras, que agrupam palavras como se fossem derivadas em um mesmo grupo, mesmo que uma seja menos frequente que a outra.

Tendo em vista tais limitações, Gardner e Davies (2014) criaram a *Academic Vocabulary List* (AVL), a partir do COCA. São fornecidas aproximadamente 3.000 palavras organizadas por sua posição no discurso acadêmico, incluindo entre elas palavras de alta frequência como *study*, *group* e *level*. Por esse motivo, uma possível limitação da AVL é seu tamanho extenso que pode dificultar levá-la para sala de aula.

É importante observar que essas listas são ainda mais úteis quando levam em consideração possíveis combinação de palavras, afinal, se não seguirem o padrão típico de estrutura utilizado na língua, poderá haver problemas de compreensão para leitores e ouvintes (WEBB e NATION, 2017). Desse modo, conhecer tais itens lexicais em uso contribui para tornar o discurso do aprendiz mais natural, empregando palavras conforme são frequentemente utilizadas na língua. Tais combinações abrangem tanto colocações quanto fórmulas. As **colocações** são itens lexicais que co-ocorrem com frequência em relação sintagmática, sendo linguisticamente previsíveis em maior ou menor grau (CRYSTAL, 2008, p. 86). Um exemplo de listas com colocações é a *Academic Collocations List* (ACL) (ACKERMAN e CHEN, 2013), contendo por volta de 2.500 colocações de adjetivos/nomes, nomes/nomes, verbo/nome, verbo/adjetivo, verbo/advérbio e advérbio/verbo que são frequentes em diversas disciplinas e englobam itens como *further analysis*, *apply (a) method* e *comprehensive approach*.

Por sua vez, as **fórmulas** têm como representante a lista criada por Simpson-Vlach e Ellis (2010), formada a partir de itens com três a cinco palavras frequentes em textos acadêmicos escritos e falados. Exemplos incluem *due to the fact that*, *take into account the* e *as can be seen*.

Destarte, o estudo desse vocabulário justifica-se por ser comum a uma ampla gama de disciplinas, e nem tão comum em textos não-acadêmicos, ocorrendo com frequência relevante em textos do gênero. Sutarsyah *et al.* (1994) demonstram que tais palavras abrangem 8,4% dos *tokens* de textos de divulgação científica e 8,7% de um texto de Economia, tendo como base a *University Word List* (predecessora da AWL). Já Coxhead (2011) estabeleceu que a AWL cobre cerca de 10% dos *tokens* de um *corpus* acadêmico de 3,5 milhões de palavras.

Baker (1988) sugere que maior ênfase deve ser dada ao ensino desse vocabulário, pois tais itens lexicais possuem função retórica e organizacional no gênero. Para a autora, o vocabulário é dividido em léxico geral, léxico especializado (palavras técnicas, no âmbito deste trabalho) e léxico subtécnico ou itens retóricos (vocabulário acadêmico), de modo que as palavras que não se enquadram no léxico geral ou no especializado correspondem aos itens que

desempenham importante papel na construção da argumentação do autor. Por meio da análise de um *corpus* de artigos científicos da área de medicina, ela observou que tais itens acadêmicos/subtécnicos fornecem aos leitores informações sobre de qual modo interpretar as intenções e avaliações do autor do texto a respeito do tópico. Por fim, a autora sugere que tais itens sejam ensinados não apenas por meio de uma lista de exercícios, mas estimulando o aprendiz a estender seu horizonte para além das sentenças, procurando compreender como essas palavras são empregadas na construção do texto como um todo e interligando ideias entre parágrafos e seções retóricas.

Além dessa justificativa para seu ensino e aprendizagem, cabe observar que as palavras acadêmicas não são tão conhecidas quanto o vocabulário técnico, havendo a possibilidade de um mesmo item de uso geral ou acadêmico ser empregado com significado técnico – e nem sempre os aprendizes estão cientes disso (COHEN *et al.*, 1979 apud NATION, 2001, p. 306). Ao mesmo tempo, aprendizes nem sempre reconhecem itens distintos sendo utilizados para se referir a mesma coisa, ou seja, não conseguem assinalar as nuances das instâncias de coesão lexical através de paráfrase (NATION, 2001, p. 307). Concomitantemente, diferente do vocabulário técnico, o acadêmico pode ser ensinado pelo professor de línguas com maior facilidade, pois não requer conhecimento prévio de determinada área do conhecimento e pode atender a alunos de diferentes áreas.

Contudo, cabe reconhecer algumas limitações desse vocabulário, como o fato de que as palavras acadêmicas podem ser mais difíceis de generalizar seus significados ao longo de diferentes áreas. Nesse sentido, Hyland e Tse (2007) discutem que apesar da utilidade de listas de vocabulário acadêmico, como a AWL, analisada pormenorizadamente pelos autores, nem sempre é possível determinar o quanto a cobertura de tais listas é suficiente para a compreensão de textos em uma área. Isto é, os autores argumentam que para sua elaboração não são levadas em consideração as diferenças no alcance das palavras ao longo das disciplinas, no favorecimento de determinadas formas e significados em cada uma delas, assim como as colocações específicas de uma área ou de um pequeno conjunto de disciplinas.

Durrant (2014) também critica essa abordagem ao investigar a diferença no uso de itens lexicais acadêmicos entre várias disciplinas. O autor observou que a variação existente entre disciplinas pode ser diminuída ao se agrupar determinadas áreas, de modo que seus resultados apontaram para uma homogeneidade inerente às disciplinas. Dessa forma, o ensino de gêneros acadêmicos, para o autor, deveria levar em consideração as áreas (com as Grandes Áreas aqui adotadas) em que os alunos estão inseridos, considerando que há menor variação de significado de uma mesma palavra dentro delas.

Resultados similares foram encontrados por Durrant e Lu (2017) ao analisarem palavras acadêmicas em artigos científicos de medicina comparando-as com as listas AWL e AVL. Os pesquisadores constataram que os 100 primeiros itens da AWL cobriram 6,74% do seu *corpus* e os 100 primeiros itens da AVL abrangeram 21,17%. A partir desse resultado, os 100 primeiros itens lexicais restantes, que não pertenciam a nenhuma das listas, cobriram 5,99% – praticamente a mesma quantidade da AWL – e diziam respeito, em grande parte, a palavras relacionadas à área de medicina. Desse modo, os autores advogam em favor de listas baseadas em disciplinas específicas ensinadas concomitantemente com listas acadêmicas e de uso geral, a fim de facilitar o ensino voltado para propósitos específicos.

Por fim, cabe retomar a ideia de que desenvolver o conhecimento lexical é fundamental no desenvolvimento das competências linguísticas. Assim, informações sobre a frequência podem demonstrar o quão relevante o ensino de determinadas palavras pode ser para o aprendiz, tanto de língua materna quanto língua estrangeira. Conforme discutido, conhecer os itens lexicais de alta frequência pode ser mais útil do que os de baixa frequência, pois correspondem a um pequeno número de palavras muito frequentes tanto na escrita quanto na fala. Além disso, se o aprendiz tiver propósitos específicos, analisar a frequência dentro de um tipo de discurso que lhe seja útil pode contribuir para sua aprendizagem. No caso daqueles com fins acadêmicos, dominar o vocabulário acadêmico, assim como técnico de sua área, é fundamental para o sucesso em suas interações nesse contexto. Para auxiliar o ensino, como visto, há diferentes listas que detalham tais palavras, assim como possíveis combinações e fórmulas frequentes nesse ambiente, embora sejam majoritariamente voltadas para língua inglesa.

Destarte, o presente trabalho avança nesse quadro teórico em língua portuguesa, fornecendo uma lista de verbos acadêmicos em artigos científicos, frequentes em diferentes áreas do conhecimento, seguindo critérios inéditos em PB. Além disso, análise a representação de uma amostra do sistema da TRANSITIVIDADE, de seus usos mais frequentes, relacionando-os às especificidades de cada Grande Área. Para tanto, o próximo capítulo apresenta a perspectiva da linguística sistêmico-funcional, com atenção principalmente ao sistema da TRANSITIVIDADE.

3 LINGUÍSTICA SISTÊMICO-FUNCIONAL

Para a análise do sistema da TRANSITIVIDADE instanciado pelos verbos mais relevantes em artigos científicos, primeiramente é necessário revisar os conceitos teóricos da perspectiva de análise da linguística sistêmico-funcional. Nas próximas seções, tais conceitos básicos dessa teoria são revisados, percorrendo o caminho até o estudo do sistema da TRANSITIVIDADE e dos diferentes tipos de processos e circunstâncias que instanciam. Em seguida, a fim de apresentar alguns dos critérios da análise realizada do Capítulo 5, resenha-se introdutoriamente o modelo da ergatividade que, embora não seja a perspectiva principal desta análise, complementou os estudos realizados.

3.1 Conceitos básicos da linguística sistêmico-funcional

A fim de iniciar a revisão teórica sobre os conceitos da LSF necessários para a análise realizada por esta pesquisa, primeiramente é preciso distinguir a relação entre **teoria** e **descrição**. Nessa perspectiva, compreende-se teoria e descrição como recursos, um para construir a linguagem (teoria) e outro, as línguas (descrição). A teoria concebe a linguagem como um tipo de sistema semiótico que abarca simultaneamente os sistemas social, biológico e físico, responsáveis por criar significados, cujo princípio de organização é a estratificação.

A teoria sistêmica diferencia a linguagem de outros tipos de sistemas semióticos, considerando-a um sistema distinto em termos de estratificação e metafunções (conceitos explicados posteriormente nesta seção). Por outro lado, a teoria não diferencia entre variações da linguagem, como Português ou Inglês, pois tal trabalho cabe à descrição sistêmica das línguas. Assim, distingue-se teoria e descrição de modo que as pressuposições da teoria são bastante amplas e gerais enquanto todas as categorias (descritivas) de uma determinada língua pertencem à descrição. A relação entre ambas está no nível de abstração, associada à realização, isto é, a descrição de uma língua em particular é a realização da teoria geral.

Continuando no âmbito da descrição, há a possibilidade de caracterizar temas descritivos e generalizações, entretanto, a última é considerada apenas descritiva e não teórica, pois diz respeito a generalizações sobre categorias descritivas de uma determinada língua. Dito isso, cabe esclarecer que nomeia-se “categorias”, por exemplo, as categorias descritivas de tempo, que abarcam marcas como “passado”, “presente” e “futuro” e são descritivas de uma

língua, não teóricas. Assim, para determinar e justificar as categorias de uma língua é necessário observar os padrões de realização dessa língua. Há dois princípios metodológicos importantes para descrição e, portanto, importantes também para este trabalho. O primeiro princípio é ter como base para interpretação a língua sendo descrita – em vez de tomar emprestadas categorias de outras línguas. O segundo princípio refere-se a proceder a descrição a partir de evidências de instâncias textuais (CAFFAREL; MARTIN; MATTHIESSEN, 2004, p.15).

Retornando à questão da linguagem, a teoria a interpreta como um sistema multidimensional semiótico, cuja organização projeta-se nas descrições das línguas, com potencial para construção de significados. Nessa perspectiva, a abordagem semiótica é holística, e não componencial, procurando centrar-se no contexto e na proeminência da organização sistêmica. Constrói-se a linguagem ao redor de diferentes dimensões semióticas, que, além de a diferenciarem de outros sistemas semióticos, interseccionam entre si mesmas e com a língua em contexto. Tais dimensões, seus princípios e ordens constituidoras estão representados no Quadro 4 a seguir.

Quadro 4 – As dimensões da linguagem, seus princípios e ordens de organização

	Dimensão	Princípio	Ordem
1.	estrutura (ordem sintagmática)	rank	oração ~ grupo/fase ~ palavra ~ morfema [léxico-gramática] ~ grupo tonal ~ sílaba ~ fonema [fonologia]
2.	sistema (ordem paradigmática)	delicadeza	gramática ~ léxico [léxico-gramática]
3.	estratificação	realização	semântica ~ léxico-gramática ~ fonologia ~ fonética
4.	instanciação	instanciação	potencial ~ sub-potencial/ tipo de instância ~ instância
5.	metafunção	metafunção	ideacional [lógica ~ experiencial] ~ interpessoal ~ textual

Fonte: Traduzido de Halliday e Matthiessen (2014, p. 20)

Na dimensão da **estrutura**, distribuem-se recursos sistêmicos de cada estrato (semântico, léxico-gramatical e fonológico), mantendo maleável a complexidade sistêmica e organizando unidades em hierarquias (CAFFAREL; MARTIN; MATTHIESSEN, 2004, p. 23). Ao **rank** compete a relação de ‘ser parte de’ entre as camadas composicionais hierárquicas como, por exemplo, no domínio da escrita, a sentença, a sub-sentença, a palavra escrita e a letra;

no domínio da gramática, a oração, o grupo, a palavra e o morfema (HALLIDAY e MATTHIESSEN, 2014, p. 21).

Tal domínio da estrutura corresponde às relações sintagmáticas na linguagem, a padrões ou regularidades que se combinam. Em oposição, a dimensão do **sistema** organiza as relações paradigmáticas, de padrões que poderiam ter sido utilizados, mas não foram. Desse modo, qualquer conjunto de alternativas configura o sistema, contrastando as escolhas realizadas das demais que poderiam ter sido feitas. Essas escolhas possíveis representam um aspecto característico da linguagem, o **significado potencial**, que é a capacidade de os usuários criarem inúmeros atos significativos (instâncias) a partir do sistema linguístico.

Nessa perspectiva, segundo Caffarel, Martin e Matthiessen (2004, p. 23-24), as análises priorizam o sistema (relações paradigmáticas) sobre a estrutura (relações sintagmáticas) como ponto de partida para a descrição linguística, interessando-se por como o que foi dito (instância) se relaciona, em oposição, ao que poderia ter sido (o potencial). A partir dessa premissa, pode-se chegar à análise da estrutura (sintagmática) de como as partes do dito se associam umas com as outras. Ou seja, compreende-se as relações estruturais (sintagmáticas) como derivadas das sistêmicas (paradigmáticas), observando a estrutura como meio/forma de expressar a escolha linguística, ambas sendo, ao mesmo tempo, mutuamente definidoras.

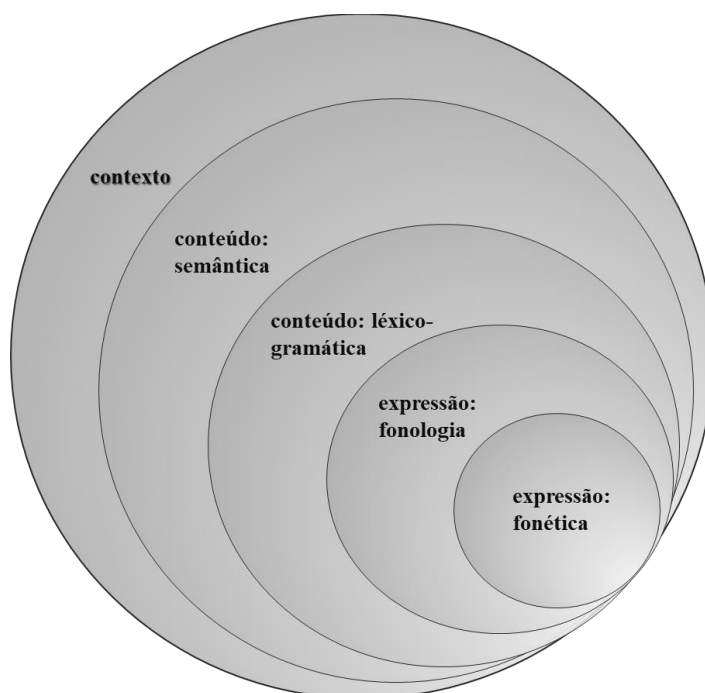
Cabe mencionar que a LSF é assim nomeada devido à sua percepção de que a gramática de uma língua é representada sob a forma de uma rede de sistemas, e não como um inventário de estruturas (HALLIDAY e MATTHIESSEN, 2014, p. 23). Embora a estrutura seja essencial para a descrição, ela é interpretada como a forma tomada pelas escolhas sistêmicas, e não como a característica definidora da linguagem. Voltando à premissa anterior, a linguagem é tida como um recurso de fazer significados, sendo que tais significados residem em padrões sistêmicos de escolha. Assim, a escolha de cada sistema contribui para a formação da estrutura – tais operações estruturais, como inserção e ordenação de elementos, são compreendidas como **realizações** de escolhas sistêmicas. Desse modo, na análise de um texto, como um artigo acadêmico, procura-se demonstrar a organização funcional da estrutura e quais as escolhas significativas realizadas, cada uma em relação ao contexto do que poderia ter sido.

Diferente da visão componencial e modular da linguística formal, nessa abordagem, o sistema molda as relações entre gramática e léxico, em vez da estrutura. A relação entre eles é de **delicadeza**, na qual o léxico (parte mais delicada) herda as características gramaticais mais gerais. Neste ponto vale notar que gramática e vocabulário não pertencem a níveis (estratos) diferentes, sendo considerados dois polos de um único contínuo, denominado **léxico-**

gramática. Da mesma forma que, por exemplo, sintaxe e morfologia são consideradas ambas partes da gramática (HALLIDAY e MATTHIESSEN, 2014, p.24).

Desse modo, ao determinar que a estrutura *realiza* as escolhas sistêmicas, manifesta-se uma outra dimensão da linguagem, a **estratificação**, a qual estabelece que essa realização provém do fato de que a língua é um sistema estratificado organizado em quatro estratos: semântica, léxico-gramática, fonologia e fonética; estando esses agrupados em dois planos: de conteúdo e de expressão. Nessa dimensão, a relação entre os estratos, marcada pela ligação de um nível de organização ao outro, é denominada **realização**. A Figura 3 a seguir representa como a estratificação do sistema linguístico está inserida no contexto.

Figura 3 - Estratificação



Fonte: Traduzido de Halliday e Matthiessen (2014, p. 26)

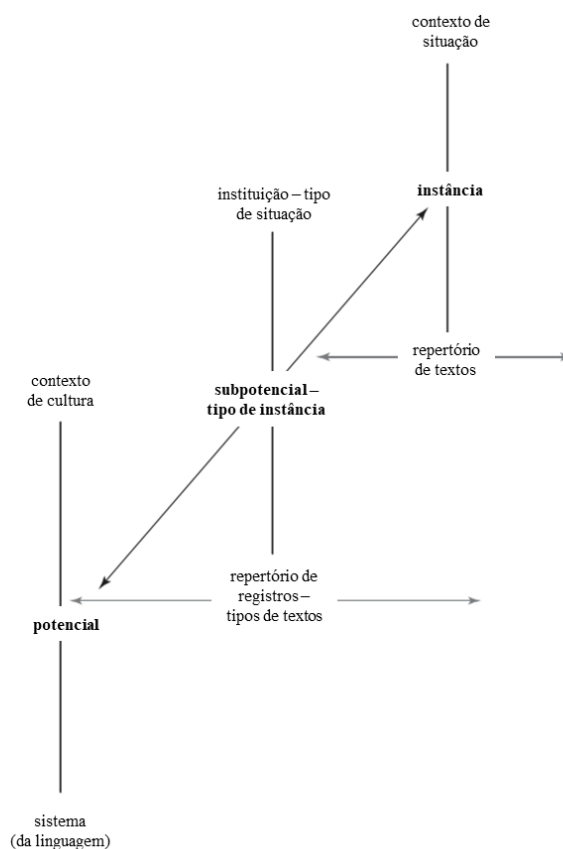
Segundo Halliday e Matthiessen (2014, p.25), o estrato do conteúdo se expande em *léxico-gramática* e *semântica*, permitindo que o significado potencial de uma língua se expanda quase indefinitivamente, já que as funções da linguagem servem aos propósitos comunicativos de seus usuários. Isso implica dizer que a gramática se relaciona com o que ocorre fora da linguagem, isto é, com os acontecimentos e contextos do mundo, assim como com as relações sociais nas quais se envolvem os usuários, além de organizar como a construção da experiência

pode ser transformada em expressão. É desse modo, conforme representado na Figura 3, anterior, que a estratificação do sistema linguístico está inserida no contexto.

A fim de compreender como a linguagem se organiza em função da necessidade do usuário, é preciso ter duas perspectivas em mente: perceber a língua como um sistema e perceber a linguagem como texto (HALLIDAY e MATTHIESSEN, 2014, p. 27). O conceito que relaciona ambas perspectivas é a **instanciação**, no qual o sistema da linguagem é *instanciado* na forma de um texto, que só tem valor existencial dentro do sistema de uma determinada língua – ou seja, não faz sentido para um usuário que não conheça a língua.

Segundo os autores (HALLIDAY e MATTHIESSEN, 2014, p. 27), o sistema é o potencial subjacente de uma língua, um recurso capaz de criar significados. Nessa perspectiva, a linguagem é tida como algo abstrato, não todo o conjunto de textos possíveis de uma língua, mas uma entidade teórica a qual se pode determinar certas propriedades e explicações. Dessa forma, sistema e texto se relacionam através do contínuo de instanciação, ilustrado na Figura 4, no qual ambos pertencem a polos opostos, havendo padrões intermediários entre eles.

Figura 4 - O contínuo de instanciação



Fonte: Adaptado de Halliday e Matthiessen (2014, p. 27)

Como forma de exemplificação, o estudo de um único texto pertence ao polo da instância, considerando-se que o **texto** instancia o sistema linguístico, e este potencializa inúmeros textos. Nesse tipo de análise, os significados de um texto não se restringem às suas particularidades, mas também levam em consideração seu contexto de produção, assim como o potencial do sistema linguístico que o produziu (FIGUEREDO, 2011). Por sua vez, por meio do estudo de um conjunto de textos, pode-se identificar padrões comuns a eles e, então, descrevê-los em termos de **tipos de textos**. Tal é o propósito desta pesquisa, que procura observar padrões comuns a artigos acadêmicos, observando possíveis semelhanças e diferenças de realização entre as diferentes Grandes Áreas do conhecimento. Esse objetivo específico – procurar padrões entre as Grandes Áreas –, provém também do fato de que os textos variam de acordo com a natureza do contexto em que foram usados, havendo inúmeros tipos de textos no cotidiano (como histórias de ficção, reportagens jornalísticas ou horóscopos), assim como dentro da academia, escopo deste trabalho, tais quais monografias, dissertações, resenhas, resumos etc. – que correspondem ao **registro**. O registro é a variação funcional da linguagem, correspondendo aos padrões de instanciação do sistema dentro de determinado contexto, denominado **tipo de situação**.

Segundo Halliday e Matthiessen (2014, p. 29), tais padrões de instanciação apresentam-se quantitativamente em probabilidades sistêmicas da língua, nas quais o registro pode representar uma determinada configuração de probabilidades sistêmicas, isto é, essa “probabilidade nada mais é que a ‘potencialização’ da frequência relativa em um texto, de modo que qualquer nova instanciação de um termo ou outro no desenvolvimento do texto afetará o sistema”¹⁴ (CAFFAREL; MARTIN; MATTHIESSEN, 2004, p. 20). No caso desta pesquisa, os procedimentos metodológicos detalhados no Capítulo 4 buscam procurar tais padrões relacionados às escolhas verbais (denominados *processos*, na perspectiva LSF) dentro de um tipo de texto, por meio da coleta de amostras suficientes para serem representativas de uma parte mais alta do contínuo – de um tipo de texto. Avançando a indagação, procura-se verificar se tais padrões são comuns à maior parte das instâncias textuais, ou se há variação conforme a área do conhecimento a que pertencem.

Retomando a Figura 4, a instância mais ampla de variação tipológica é a do sistema – que corresponde ao significado potencial de uma língua e ao potencial contextual da sua cultura

¹⁴ “probability is nothing but the “potentialization” of relative frequency in text, so any new instantiation of one term or another in the unfolding of text will perturb the system. (CAFFAREL; MARTIN; MATTHIESSEN, 2004, p. 20)

(CAFFAREL; MARTIN; MATTHIESSEN, 2004, p. 37); enquanto a instância mais estreita é a de um texto em um determinado contexto de situação, ponto no qual os estudos tipológicos e de tradução se encontram. Partindo dessa perspectiva, cada instância de texto pode ser compreendida tanto como uma realização da cultura na qual está inserida, quanto como uma instanciação daquela cultura. Este trabalho, por exemplo, se encontra em um ponto intermediário, correspondendo ao estudo de um tipo textual.

De acordo com Caffarel, Martin e Matthiessen (2004, p. 20), o contínuo de instanciação é uma importante dimensão teórica, pois permite localizar um texto como fonte de dados em relação a generalizações sobre tipos de texto e o sistema em geral, além de possibilitar a comparação de registros entre diferentes línguas, sincrônica e diacronicamente, ao mesmo tempo em que assegura que as línguas se apresentem em termos de registros comparáveis.

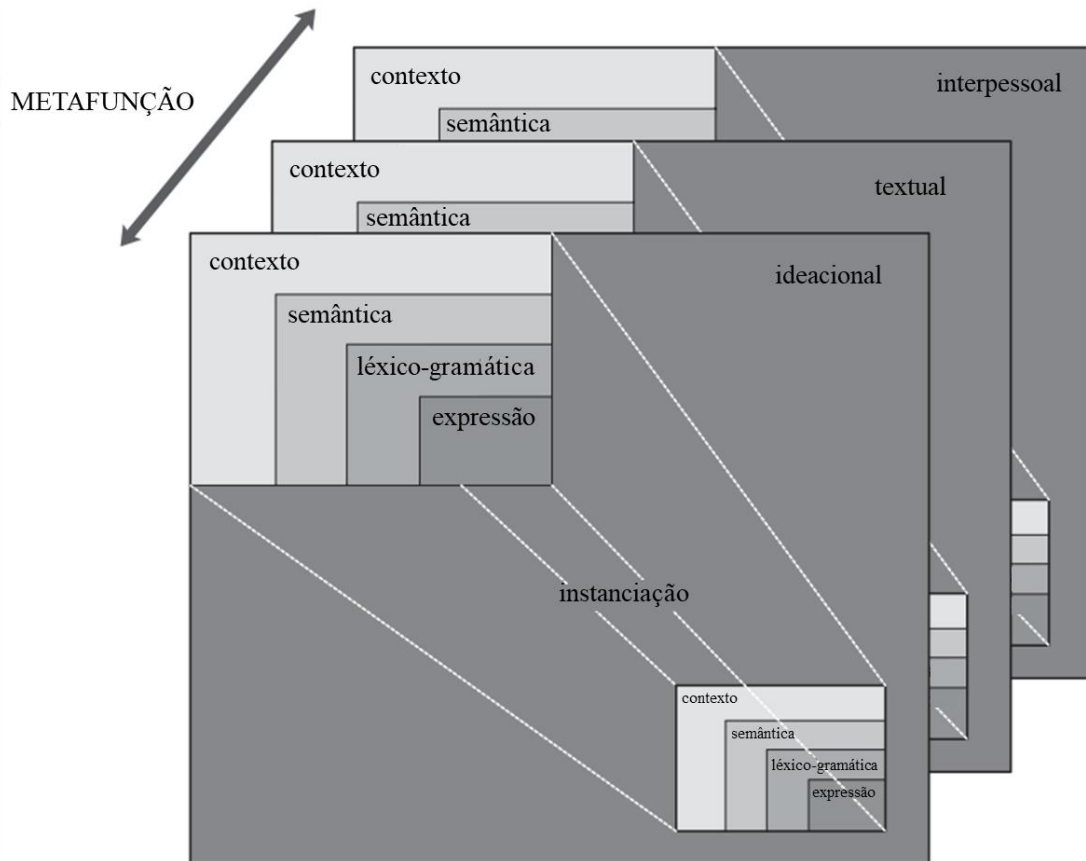
Quanto à instanciação e à estratificação, ambas dimensões são bidirecionais caracterizando, para Caffarel (2006, p. 2), a LSF como inovadora e distinta das demais. Além disso, suas múltiplas perspectivas fornecem teoria para priorizar tanto a análise de recursos específicos de um registro, quanto de recursos mais amplos da língua. Também fornece vários caminhos para explorar como o sentido é, e pode ser, construído por padrões léxico-gramaticais em textos – que é foco da presente pesquisa.

A quinta dimensão diz respeito às **metafunções** (ideacional, interpessoal e textual), que são essenciais para compreender o fenômeno linguístico e representam duas funções fundamentais da linguagem, ligadas ao ambiente físico e social humano: construir a experiência dentro desse ambiente e encenar interações sociais (HALLIDAY e MATTHIESSEN, 2014, p. 30). Em relação à primeira função, a linguagem fornece uma teoria sobre a construção da experiência humana, de modo que existem determinados recursos da léxico-gramática dedicados a essa função: a metafunção ideacional, distinguível em experiencial e lógica (cf. seção 3.2). Relacionada à segunda função, ao mesmo tempo que constrói, a linguagem sempre também encena as interações pessoais e sociais dos seres humanos, correspondendo à metafunção interpessoal. Há ainda um terceiro componente da gramática que permite e facilita a organização do fluxo de informação do texto, a metafunção textual, que ocorre concomitantemente às demais.

Sobre a natureza das metafunções, vistas enquanto sistema, elas são previsões sobre a in(ter)dependência relativa dos sistemas, se relacionando com amontoados de escolhas relativamente independentes; enquanto estrutura, são previsões sobre a natureza de suas realizações estruturais, relacionando-se à realização de diferentes tipos de padrões (particular, prosódico e periódico) (CAFFAREL; MARTIN; MATTHIESSEN, 2004, p. 26). Assim,

segundo Halliday e Matthiessen (2014, p. 31), toda arquitetura da linguagem se organiza funcionalmente, pois tais (meta)funções se desenvolveram junto com a espécie humana. Por causa dessa característica optou-se pelo uso do termo “metafunção” a fim de implicar que este é um componente integral da teoria, conforme pode ser observado na Figura 5.

Figura 5 - Metafunção



Fonte: Traduzido de Halliday e Matthiessen (2014, p. 31)

Determinadas as dimensões da linguagem, cabe ressaltar que ela sempre atua em contexto, ou seja, é analisada, teorizada e descrita dentro do ambiente de significado em que se realiza. De modo similar, o contexto é funcionalmente diversificado, assim como se estende ao longo do contínuo de instanciação (cf. Figura 4) desde instâncias envolvendo interações particulares de pessoas em ocasiões específicas, denominado **contexto de situação**, até o potencial contextual de uma comunidade de fala, denominado **contexto de cultura** (HALLIDAY e MATTHIESEN, 2014, p. 32), por meio de regiões intermediárias de subpotencial/tipos de instância (instituição/tipo de situação) (MATTHIESEN; LAM; TERUYA, 2011, p.77). Apesar de não haver descrições abrangentes e detalhadas de contexto

de cultura, tendo em vista sua natureza, há três categorias nas quais qualquer tipo de situação pode ser caracterizado e que têm impacto direto e significativo no tipo de linguagem que será empregada: campo, relação e modo.

O **campo** diz respeito à atividade que está acontecendo na situação, reconhecida na cultura realizada pelos participantes, e à natureza dessa ação social. A **relação** remete a quem está participando da situação e à natureza dos papéis sociais que desempenham, além do grau de controle de um sobre o outro, da possível hierarquia e do grau de formalidade. Já o **modo** corresponde à função exercida pela linguagem, ao veículo comunicativo utilizado naquela situação ou ao que se espera que a linguagem faça pelos participantes em determinado contexto, referindo-se, portanto, ao papel da linguagem, do compartilhamento entre participantes (monológico ou dialógico) e do meio (oral ou escrito, por exemplo) (EGGINS, 2004).

Definidos tais conceitos, o contexto de situação é o ambiente em que o texto está em funcionamento e explica o motivo de certas coisas terem sido escritas ou ditas de tal forma em determinada ocasião, assim como o que deixou de ser escrito ou dito (HALLIDAY e HASAN, 1989). Por essa razão, usuários conseguem prever ou produzir expectativas sobre o que será dito em um dado texto e o que virá a seguir com base no contexto de situação. Abaixo, os dois trechos ilustram tal característica:

- (2) Amamentar traz muitos benefícios tanto para a mãe quanto para o bebê. É uma incrível maneira de criar intimidade e construir o vínculo afetivo entre vocês. / O leite materno possui todos os nutrientes de que o bebê precisa para se desenvolver bem, tanto que a criança não precisa de mais nada, nem água, nos primeiros seis meses de vida. Mesmo depois da introdução dos alimentos sólidos, o leite da mãe continua a ser cheio de propriedades benéficas para a saúde do bebê. / Estudos mostram que o leite materno possui anticorpos que protegem a criança contra infecções, como gastroenterites (doenças com ocorrência de diarreia), doenças respiratórias, infecções urinárias e otites. [...] As autoridades em saúde como a Organização Mundial da Saúde (OMS) e o Ministério da Saúde, com base em estudos científicos, recomendam o aleitamento materno exclusivo até os 6 meses e a manutenção da amamentação até acima de 2 anos de idade.¹⁵
- (3) A Organização Mundial da Saúde (OMS) recomenda que a criança seja alimentada exclusivamente com o leite materno até os seis meses de idade,⁶ o que tem impacto positivo na sobrevivência e na saúde nessa fase e na vida adulta.^{5,7} Crianças submetidas ao aleitamento materno exclusivo (AME) durante os seis primeiros meses de idade têm menos chances de desenvolver doenças crônicas não transmissíveis na infância, adolescência e vida adulta.⁷ O leite materno contém energia e nutrientes adequados ao grau de maturidade fisiológica do lactente, além de fatores de proteção contra doenças, o que o torna ideal para os primeiros meses de vida.^{5,7} A partir dos seis meses deve-se dar início à alimentação complementar, uma vez que a quantidade e a composição do leite materno já não são suficientes para atender às necessidades nutricionais da criança.⁸ (Sa_RPP_2018_2_AA_006)

¹⁵ Disponível em: <<https://brasil.babycenter.com/x2400130/amamentar-faz-tanta-diferen%C3%A7a-assim>>. Acesso em 15 de jan. de 2019.

Nas amostras linguísticas anteriores, pode-se perceber o diferente lugar social de onde partem os autores do texto, mesmo se a fonte de cada um fosse desconhecida. Ambos discutem sobre a alimentação infantil, de modo que esse é o *campo* do contexto situacional nos dois trechos, assim como terem sido escritos para serem lidos – o *modo*. Porém, quanto ao *teor*, o primeiro é um texto elaborado por jornalistas para o público-alvo do *site*, geralmente pais ou cuidadores, a fim de divulgar informações sobre determinado assunto à população interessada. Já o segundo texto é um artigo científico, retirado do *corpus* desta pesquisa, elaborado por pesquisadores para, em geral, colegas de área.

Assim, apesar de discutirem o mesmo assunto (alimentação infantil com exclusividade de aleitamento materno) e descreverem os mesmos argumentos (leite materno contém todos os nutrientes adequados para a criança, além de prevenir doenças) os dois textos são diferentes entre si por causa das escolhas lexicais de cada um que levam em consideração as variáveis de contexto – *o que, para quem e como o texto será escrito*. Isso demonstra que é possível identificar aspectos do contexto situacional a partir de certos elementos linguísticos empregados em cada um, conforme detalhado no Quadro 5, a seguir.

Quadro 5 – Exemplos da configuração do contexto situacional em textos

	Trecho (2)	Trecho (3)
Campo	Divulgar informações a favor da amamentação.	Expor argumentos a favor da amamentação infantil.
Exemplos	“Amamentar traz muitos benefícios”, “O leite materno possui todos os nutrientes”, “possui anticorpos que protegem a criança”	“a criança seja alimentada exclusivamente com o leite materno”, “menos chances de desenvolver doenças crônicas”, “O leite materno contém energia e nutrientes adequados”.
Teor	Participantes da situação: jornalista responsável e os leitores da página no site online – que não necessariamente se conhecem. Participantes do texto: mães e bebês, o leitor (presume-se uma mãe com seu filho), instituições.	Participantes da situação: os pesquisadores e os leitores do periódico em que foi publicado, geralmente seus pares na área, que podem ou não se conhecerem. Participantes do texto: mães e bebês, instituições.

Exemplos	Mães e bebês: “a mãe e o bebê”. Leitor: “entre vocês”. Instituições: “Organização Mundial da Saúde”, “Ministério da Saúde”.	Mães: omitido pelo uso da passiva, como em “a criança seja alimentada”, ou por referências adjetivais, como em “aleitamento materno”. Bebês: “lactente”. Instituições: “Organização Mundial da Saúde”.
Modo	Verbal escrito, organizado de modo argumentativo.	Verbal escrito, organizado de modo expositivo.
Exemplos	“muitos benefícios”, “incrível maneira”.	“recomenda”, “deve-se dar início”.

Fonte: Elaboração própria a partir dos trechos (2) e (3).

Halliday e Matthiessen (2014) afirmam que essas três variáveis são importantes porque se relacionam com os três tipos de significados estruturados pela língua, exercendo diferentes funções: expressar a experiência, criar relações interpessoais e organizar informações. Cada tipo de significado pode ser associado “para cima” – com o contexto – ou “para baixo” – com a léxico-gramática (EGGINS, 2004, p. 111). Na relação “para cima”, o elo está no tipo de registro associado a um desses significados. Portanto, *campo* pode ser expresso por meio de padrões ideacionais no texto, *modo* pode ser expresso através de significados textuais e a *relação* através de significados interpessoais. Já “para baixo”, a associação está nos tipos de significados sendo realizados por meio da combinação de padrões léxico-gramaticais.

Dessa forma, cada significado se relaciona de forma previsível a uma variável do contexto situacional, não sendo por acaso que cada uma se associa a um significado da linguagem: ideacional, textual e interpessoal. O campo ressoa com significado ideacional, o modo com o significado textual e a relação com o significado interpessoal. Ou seja, a correspondência entre o contexto e a linguagem baseia-se na organização funcional da ordem de ambos significados (HALLIDAY e MATTHIESSEN, 2014, p. 34), na qual tais funções são denominadas metafunções.

Em outras palavras, a linguagem fornece uma teoria sobre a experiência humana, havendo, de tal modo, recursos da léxico-gramática dedicados à cada função. Assim, a metafunção ideacional compreende a expressão de ideias, pensamentos, sentimentos e todo tipo de emoção; a metafunção interpessoal relaciona tais expressões à interação com outros usuários;

e a metafunção textual organiza essas informações em textos. Cada metafunção corresponde a uma variável de contexto, possuindo um sistema próprio do estrato léxico-gramatical – abaixo do estrato semântico (cf. Figura 5) –, conforme representando no Quadro 6 a seguir.

Quadro 6 – Variáveis de registro relacionadas às metafunções e seus sistemas léxico-gramaticais

Descrição	Variáveis	Metafunção	Sistema
A ação social, o assunto sobre o qual se fala, a natureza da ação.	Campo	Ideacional	TRANSITIVIDADE
A estrutura dos papéis dos participantes e suas relações na situação de comunicação.	Relação	Interpessoal	MODO, MODALIDADE e POLARIDADE
A organização simbólica, o canal (escrito ou falado) e o modo retórico da linguagem.	Modo	Textual	TEMA

Fonte: Adaptado de Gouveia (2009, p. 28)

Assim, as metafunções são instanciadas de forma entrelaçada e simultânea na linguagem. Conforme Halliday e Hasan (1989), para compreender de fato um texto, o usuário deve ser capaz de interpretá-lo no âmbito de todas as metafunções.

A **metafunção ideacional**, na qual a oração é vista como representação, é realizada por duas funções distintas: a **experencial**, que compreende os processos que estão sendo instanciados, os seus participantes e as circunstâncias associadas a eles, sendo responsável pela construção de um modelo de representação do mundo ao nível da oração; e a **lógica**, que compreende a relação entre um processo e outro, ou entre participantes, que compartilham a mesma função no texto, ou seja, que é responsável pela combinação de grupos lexicais e oracionais. Ao analisar a oração, o sistema empregado é o da TRANSITIVIDADE por meio do qual representa-se a experiência em termos de configuração de processos, participantes e circunstâncias.

Por sua vez, a **metafunção interpessoal** reconhece a função da fala e os tipos de características retóricas que constituem a linguagem como um ato simbólico (HALLIDAY e

HASAN, 1989) através do sistema de MODO, MODALIDADE e POLARIDADE. Esse sistema é o recurso gramatical empregado para expressar a interação entre participantes do evento comunicativo, considerando-se as funções dos elementos constituintes da oração: Sujeito, Finito, Complemento, Predicador ou Adjunto (FUZER e CABRAL, 2011, p. 33). Nesse âmbito, a oração é entendida como troca de informações, dizendo respeito ao tempo (passado, presente e futuro), à modalidade (probabilidade, usualidade, obrigação e inclinação) e à polaridade (positiva ou negativa).

Por último, a **metafunção textual** abrange a topicalidade da mensagem, o valor de informação nova ou dada, a coerência entre as partes do texto, tendo em vista a oração como mensagem realizada através da estrutura temática – que consiste de um Tema, o ponto de partida da mensagem, acompanhado por um Rema. Relacionada à variável modo do contexto situacional e ao sistema de TEMA, essa metafunção determina as formas de coesão, os padrões de voz e tema, de continuidade e os dêiticos.

Cada sistema descrito anteriormente possui um foco de análise léxico-gramatical distinto, de modo que uma mesma oração pode ser analisada por diferentes perspectivas, embora estejam, ao mesmo tempo, sistematicamente relacionadas. Em vista disso, conforme apresentado no Quadro 7 a seguir, um único item gramatical pode representar distintos componentes dos respectivos sistemas das três metafunções:

Quadro 7 – Exemplificação dos sistemas das metafunções

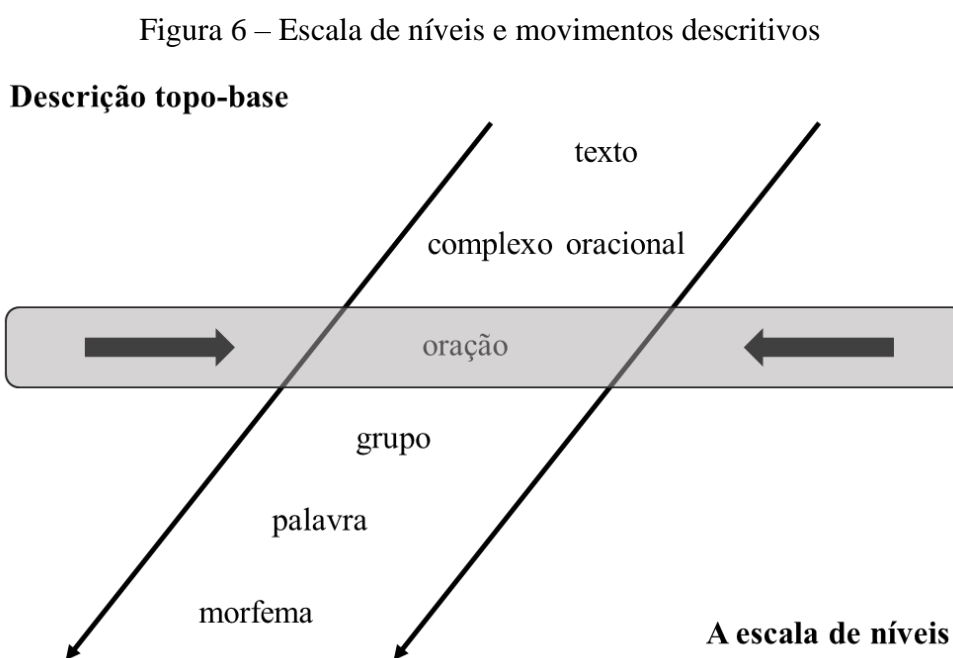
Metafunção (sistema)	83,4% das crianças ingeriram vitamina A acima do valor da RDA. (Sa_JOPE_2014_0_AA_002)			
Experiencial (TRANSITIVIDADE)	Participante	Processo	Participante	Circunstância
Interpessoal (MODO)	Sujeito	Resíduo		
Textual (TEMA)	Tema	Rema		

Fonte: Elaboração própria.

Nessa perspectiva, a oração é a unidade de processamento da léxico-gramática por ser o lugar em que os significados são produzidos em função da estrutura gramatical, sendo simultaneamente uma representação (metafunção ideacional), uma troca (interpessoal) e uma mensagem (textual). Isto é, retomando conceitos expostos anteriormente, enquanto a oração é

a unidade de processamento da gramática, o texto é a unidade de descrição, instanciando o sistema.

Desse modo, tudo se desenvolve ao redor da oração: acima, abaixo e ao redor. De acordo com Gouveia (2009, p. 20), o fato de o texto ser a unidade da descrição que permite o acesso ao sistema, através da oração, faz com que essa seja uma abordagem topo-base: da unidade maior para a menor, em vez de base-topo, como é comum aos estudos formais de descrição da gramática. Conseqüentemente, a léxico-gramática toma por base uma escala de níveis que começa na oração e termina no morfema, passando pelo grupo sintagmático e pela palavra. Essa hierarquia pode ser observada na Figura 6 abaixo, que ilustra a oração como unidade central de análise:



Fonte: Adaptado de Gouveia (2009, p. 21)

Uma parte da escala de níveis da Figura 6 foi representada no Quadro 8, a seguir, que exemplifica um complexo oracional e seus constituintes, os **grupos** (*sintagmas* em outras perspectivas teórica), compostos por uma ou mais palavras, essas se constituindo em morfemas. Como pode ser observado, cada grupo desempenha uma função distinta no estrato da léxico-gramática, outra vez ressaltando-se a co-ocorrência das metafunções:

Quadro 8 – Exemplificação do complexo oracional sob diferentes perspectivas metafuncionais

Oração		<i>Adicionalmente,</i>	<i>alguns autores</i>	<i>utilizaram</i>	<i>genes do rRNA, 16S, 23S e 5S,</i>	<i>para o estudo de Map (A_PVB_2013_0_AA_024)</i>
Grupos		Adverbial	Nominal	Verbal	Nominal	Preposicional
Funções léxico-gramaticais	TRANSITIVIDADE	Circunstância	Participante	Processo	Participante	Circunstância
	MODO	Adjunto	Sujeito	Finito e Predicador	Complemento	Participante
	TEMA	Tema		Rema		

Fonte: Elaboração própria.

Levando em consideração o referido sistema linguístico organizado em estratos, o tipo de análise proposta pela LSF denomina-se **perspectiva trinocular** por compreender a língua em termos “acima” do fenômeno analisado, “abaixo” do estrato em análise e “ao redor”, no mesmo estrato em que o fenômeno em foco ocorre. Pois, para Halliday e Matthiessen (2014, p. 48), não se pode esperar compreender um fenômeno apenas observando-o do seu mesmo nível. Entretanto, também não se pode esperar uma descrição completa de todos os ângulos, visto que sempre haverá um meio termo. Ainda segundo os autores, por ser essa uma abordagem funcionalista da gramática, deve-se priorizar a perspectiva “acima”, observando a gramática enquanto recurso para criar significados, assim como prevalecendo uma análise mais semântica da gramática, mesmo que a última ainda continue a ser foco (idem, 2014, p. 49)

Para os autores, priorizar a perspectiva “acima” implica favorecer o sistema, no qual, conforme mencionado, a língua é entendida como uma rede de escolhas significativas interligadas. Relembrando, as relações aqui são paradigmáticas, em que os componentes da gramática são conjuntos de características mutuamente opostas. Desse modo, a explicação de um fenômeno baseia-se não apenas em demonstrar como se estrutura, mas também como se relaciona com outras possibilidades em termos de relações sistêmicas ou ágatas.

Seguindo a definição de Halliday e Matthiessen (2000, p. 504), sob a perspectiva trinocular, observar um fenômeno “de cima” implica perceber como determinado estrato expressa certo conteúdo; “de baixo” observa-se o fenômeno como conteúdo de certa expressão; e “ao redor” implica observá-lo no contexto de (e em relação a) seu próprio estrato. Transpondo tais definições para esta pesquisa, pretende-se identificar os diferentes tipos de processos em termos de sua transitividade nuclear, levando em consideração o que significam, como foram

expressos e seu qual o potencial sistêmico. Desse modo, analisa-se o sistema da TRANSITIVIDADE sob tal perspectiva: 1) ‘de cima’: como os significados dos processos são construídos ao longo dos *subcorpora*, se é possível perceber generalizações sobre os significados; 2) ‘de baixo’: os grupos verbais que realizam tais processos, se há regularidades na realização das escolhas sistêmicas; 3) ‘ao redor’: os tipos de participantes e circunstâncias que acompanham os processos, assim como as consequências da agnação, delicadeza e pressuposições sobre o léxico empregado.

Iniciada a revisão de literatura sobre alguns dos conceitos teóricos básicos da LSF, a seguir, a metafunção ideacional será revisitada, atentando em maiores detalhes para o construto experiencial, a fim de fornecer a fundamentação teórica necessária para a posterior discussão dos resultados encontrados por esta pesquisa.

3.2 Oração enquanto representação: a metafunção experiencial

Das três funções semânticas apresentadas anteriormente, a presente revisão teórica afunila a discussão para a metafunção ideacional, que oferece recursos semânticos para construção da experiência interna e externa no mundo, a partir de dois construtos: o lógico e o experiencial. Na **metafunção lógica**, a experiência é moldada em uma cadeia de eventos, em sequências de processos realizados pelo sistema de COMPLEXO ORACIONAL. Por sua vez, a **metafunção experiencial** constrói os fenômenos experienciais enquanto configurações de Processos, Participantes e Circunstâncias, por meio do sistema de TRANSITIVIDADE.

A metafunção experiencial fornece recursos léxico-gramaticais para construir a experiência no mundo externo e interno enquanto significado por meio da linguagem. Sob esse viés, a experiência externa ao indivíduo corresponde a ações ou eventos que acontecem e seus participantes, e a experiência interna constitui-se de processos cognitivos ao nível da consciência (THOMPSON, 2014, p. 91). Em complemento, o indivíduo também é capaz de relacionar fragmentos de sua experiência, identificando-os ou caracterizando-os.

Essa metafunção corresponde à organização funcional da estrutura dentro da oração, de modo que experiências são gramaticalizadas em figuras de acontecimentos. Nessa perspectiva, de acordo com Thompson (2014), a fim de denominar cada uma dessas representações, são utilizadas etiquetas funcionais que indicam o papel interpretado por cada um desses elementos: *processos* que envolvem *participantes* em certas *circunstâncias*, conforme exemplificado no Quadro 9.

Quadro 9 - Exemplos de processos, participantes e circunstâncias

Em um estudo nos Estados Unidos,	36%	apresentaram	supressão viral. (Sa_JOPE_2014_0_AA_007)	
	Tal achado	pode refletir	uma tendência à diminuição do uso da substância	com o avançar da idade (Sa_JBP_2018_2_AA_003)
Inicialmente		foi avaliado	o efeito da concentração de HCl	como eletrólito suporte da amostra (E_QN_2018_1_AA_004)
Grupo adverbial	Grupo Nominal	Grupo Verbal	Grupo Nominal	Grupo adverbial
Circunstância	Participante	Processo	Participante	Circunstância

Fonte: Elaboração própria.

Sob a ótica da metafunção experiencial, **processos** são o núcleo da oração, considerando que seu significado gira ao redor de uma ação, evento ou estado, e são realizados tipicamente pelo grupo verbal. Em alguns casos, como exemplificados na segunda e terceira oração do Quadro 9, o grupo verbal pode constituir outro componente além do verbo nuclear – que é o centro da análise. Ou seja, isso significa que o foco está no verbo principal, não levando em consideração, por exemplo, os elementos interpessoais como o “pode” na segunda oração.

Cada oração costuma possuir ao menos um **participante**, que é normalmente realizado por um grupo nominal, podendo ou não estar diretamente envolvido/impactado pelo processo. Em alguns casos, o participante pode não estar mencionado explicitamente, mas é compreendido como parte do sentido do verbo. Por exemplo, na quarta oração do Quadro 1, o participante que executou a ação não está explícito, mas pode ser retomado pelo contexto como sendo o(s) autor(es) do texto. Além disso, com verbos que se referem a condições climáticas, pode não haver participante algum.

Por sua vez, as **circunstâncias** são geralmente realizadas por grupos adverbiais ou frases preposicionadas. Dentro da oração, costumam ser opcionais, um reflexo de sua função de “pano de fundo” (THOMPSON, 2014, p. 93), embora, com certos processos, possam ser fundamentais para a completude do significado da frase, ampliando a configuração de processo e participantes

envolvidos por meio de relações lógico-semânticas de projeção e expansão. O Quadro 10 a seguir exemplifica tais usos, que estão discutidos com maior detalhe na seção 3.3.7.

Quadro 10 - Exemplo de Participante, Processo e Circunstância, a última mais obrigatória.

o sistema	localiza-se	a cerca de 66° W, (E_RBM_2017_1_AA_008)
A equação de Scherrer (equação 1)	foi empregada	nos cálculos (E_QN_2018_7_AA_003)
Participante	Processo	Circunstância

Fonte: Elaboração própria.

Essa configuração da oração em processo, participantes e circunstâncias denomina-se sistema da TRANSITIVIDADE, que está detalhado na próxima seção.

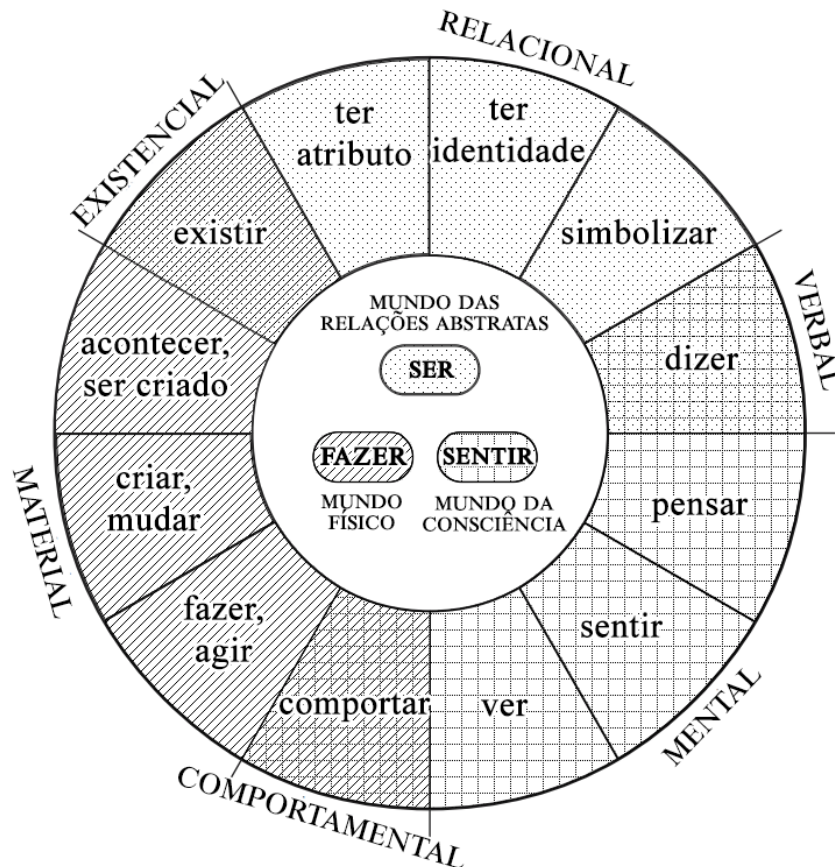
3.3 O sistema da TRANSITIVIDADE: processos em foco

Existem múltiplas atividades e ações que acontecem no mundo ao mesmo tempo em que há múltiplas formas de representá-las linguisticamente em uma cadeia de eventos ou acontecimentos, nessa perspectiva, sob a forma de figura. As **figuras** consistem de um processo se desenvolvendo ao longo do tempo em relação aos participantes diretamente envolvidos, podendo haver ou não a presença de circunstâncias – que, por sua vez, podem ou não estar diretamente envolvidas no processo. De acordo com Halliday e Matthiessen (2014), tais figuras – de fazer, acontecer, sentir, dizer, ser ou ter – organizam-se na gramática da oração, por meio do sistema de TRANSITIVIDADE. Esse sistema se constitui como um mecanismo do estrato da léxico-gramática responsável por representar ações e atividades em configuração de processos, participantes envolvidos e circunstâncias (figuras). Por sua vez, os processos constroem-se em um conjunto de TIPOS DE PROCESSOS, no qual cada tipo configura um modelo distinto para a construção de um determinado domínio da experiência (HALLIDAY e MATTHIESSEN, 2014).

A seguir, a Figura 7 ilustra os tipos de processos que refletem a experiência externa e interna do mundo, sendo três primários (e mais frequentes no sistema da TRANSITIVIDADE da língua inglesa), conforme Halliday e Matthiessen (2014): materiais (figuras de fazer), mentais

(figuras de sentir), relacionais (figuras de ser); e três secundários: verbais, comportamentais e existenciais, que estão na fronteira entre os primeiros e parecem ser mais suscetíveis à variação entre diferentes línguas.

Figura 7 – TIPOS DE PROCESSOS do sistema da TRANSITIVIDADE



Fonte: Miranda (2016), adaptado de Halliday e Matthiessen (2014, p. 216) e Souza (2006, p.54)

A Figura 7 procura demonstrar, conforme Halliday e Matthiessen (2014, p. 216), que não há prioridade entre os tipos de processos, embora sejam ordenados. Concomitantemente, é de suma importância que sua representação visual seja em forma circular, em vez de linear, por exemplo. Os autores ainda acrescentam que, para ser mais precisa, a representação deveria se aproximar de uma esfera, havendo continuidade ao longo do *loop*. Na Figura 7 apresentam-se os seis tipos de processos, relacionando-os a suas respectivas regiões, sendo as áreas centrais aquelas cujos membros são prototípicos dos tipos de processos e as regiões limítrofes, uma mescla, expressando aqueles cuja categorização costuma ser obscura. A fim de exemplificação, seguem os trechos a seguir:

- (4) Neste papel, Alexandra Lucas Coelho **disse** *que o corpo do jornalista é para ser trespassado por uma experiência sensorial, cognitiva, com o cheiro, com o quê vê, com o que ouve.* (So_GSP_2018_1_AA_003)
- (5) O Evangelho eterno não **indica** *uma nova escritura ou um novo testamento para o terceiro estado do mundo.* (H_TFA_2018_1_AA_002)
- (6) Destarte, os resultados **indicam** *que os fluxos de capitais para esse grupo de países possui um caráter mais especulativo e dependente da disposição ao risco desses agentes.* (So_NE_2017_2_AA_003)

Conforme Halliday e Matthiessen (2014, p. 217) justificam, os exemplos (4) e (5) são prototípicos de processos verbais e relacionais, respectivamente. O primeiro possui um participante que realiza o processo verbal, ou seja, que *diz* algo (“Alexandra Lucas Coelho”) e uma oração projetada (em itálico no exemplo) sobre aquilo que foi dito. Já o segundo trecho relaciona uma expressão (“O Evangelho eterno”) e seu conteúdo (“uma nova escritura ou um novo testamento para o terceiro estado do mundo”) através do processo relacional. Por sua vez, o exemplo (6) diz respeito à área intermediária entre os processos (observável na Figura 7, anterior), estando mais para o prototípico relacional, pois não pode, em um teste de reelaboração para auxiliar a caracterizar o tipo de processo, por exemplo, colocar o segundo participante (destacado em itálico) em discurso direto, sem perder parte de seu significado (*Os resultados indicam: “o fluxo de capitais para esse grupo de países possui um caráter mais especulativo e dependente da disposição ao risco desses agentes”*).

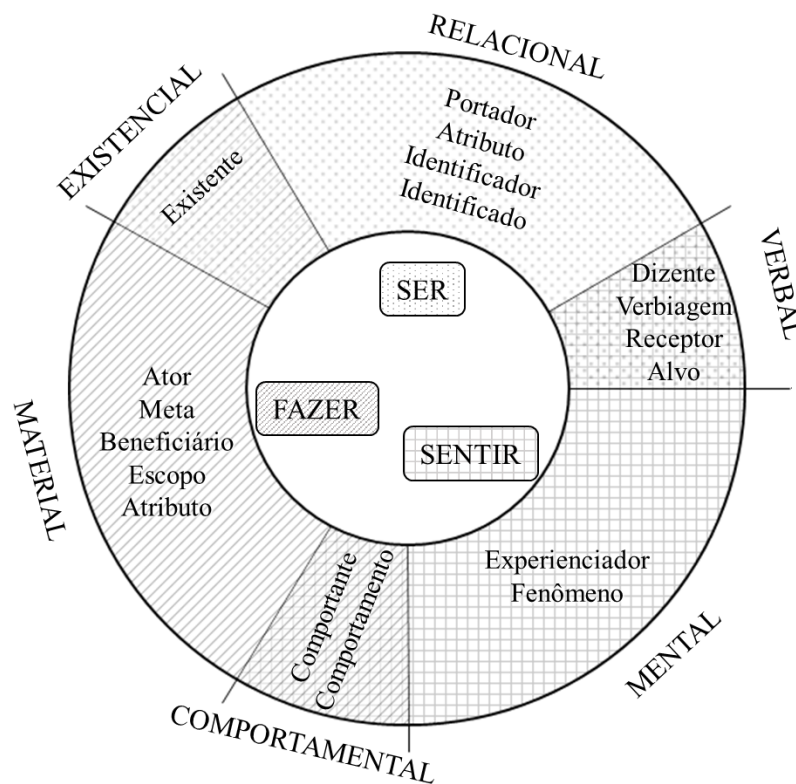
Essa propriedade dos processos, de acordo com os autores, é um princípio fundamental do sistema, denominado **indeterminação sistemática**, na qual o mundo da nossa experiência é altamente indeterminado, refletindo no modo como a gramática constrói seus tipos de processos. Ou seja, para a identificação das categorias, é necessário levar em consideração tanto o contexto quanto as relações semânticas, pois, como visto nos exemplos, um mesmo grupo verbal pode realizar mais de um tipo de processo em função de suas combinações léxico-gramaticais (FUZER e CABRAL, 2011, p. 43).

A partir dessa introdução, e conforme ilustrado na Figura 7, os processos primários podem ser resumidos em: **processos materiais**, que representam a experiência externa – ações e eventos –, realizados por verbos como *fazer, construir e agir*; **processos mentais**, que representam a experiência interna – lembranças, reações e sentimentos –, realizados por verbos como *pensar, imaginar, gostar e querer*; e **processos relacionais**, que representam relações – identificação e caracterização –, realizadas por verbos como *ser, estar, parecer e ter*.

Nas fronteiras entre os processos primários estão os secundários, cujas características podem ser resumidas em: **processos verbais**, que representam dizeres através da atividade linguística dos participantes, localizados entre os processos mentais e relacionais, e podem ser exemplificados por *dizer, afirmar e responder*; **processos comportamentais**, que representam comportamentos, manifestando atividades psicológicas ou fisiológicas do ser humano, situados entre os processos materiais e mentais, e têm como exemplos *dormir, tossir e sonhar*; e **processos existenciais**, situados entre os relacionais e os materiais, que representam a existência de um participante, como *haver e existir*.

Cada processo possui particularidades que implicam a interação de certas entidades, seus participantes, cuja natureza se dá em função do tipo de processo com os quais ocorrem. Assim, cada processo é executado por participantes distintos, de acordo com as escolhas realizadas pelos usuários da língua. A Figura 8 a seguir ilustra os tipos de participantes nas orações, segundo os tipos de processos.

Figura 8 - Tipos de participantes em orações



Fonte: Adaptado de Cabral e Fuzer (2014)

A partir dessa introdução, as próximas seções detalham os seis tipos de processos que constituem o sistema da TRANSITIVIDADE, exemplificando-os com exemplos extraídos do *corpus* da presente pesquisa e fornecendo testes e características que os definem e distinguem, de modo a esclarecer a posterior análise qualitativa realizada dos processos mais frequentes em artigos acadêmicos.

3.3.1 Orações materiais

Orações com processos materiais são representações de “fazer e acontecer”, em que a sentença material desencadeia uma mudança no curso de eventos em acontecimento através do investimento de energia de um participante. Aquele que *faz* a ação é chamado **Ator**, sendo este obrigatório nos processos materiais. Já o participante que possui uma de suas características criada ou alterada pelo processo denomina-se **Meta**.

Segundo Halliday e Matthiessen (2014), em inglês, a forma não-marcada das orações materiais é o presente-no-presente, que distingue algo que está acontecendo no momento (como “Estamos todos comendo agora”), de um hábito (como em “Comemos fora aos Sábados”) ou de uma ocorrência generalizada (denominada “*general truths*”, no original, como em “O progresso aprimora nossas habilidades”), embora a forma não-marcada desses dois últimos sejam o presente simples. Por sua vez, segundo Caffarel (2006), em francês, o pretérito simples ou composto é a forma preferida não-marcada, como em “*Pierre a envoyé une lettre recommandée*”.

Por serem processos que envolvem a feitura de algo, além de observar a escolha temporal do uso do processo, outro critério gramatical que possibilita a distinção entre este e os demais processos primários (mentais e relacionais) é verificar se podem ser indagados com processos materiais prototípicos como *fazer* ou *acontecer* (HALLIDAY e MATTHIESEN, 2014; CAFFAREL, 2006). Nos exemplos abaixo, em (7), pode-se perguntar ‘Quem *fez* o segundo trabalho’ ou, em (8), ‘O que *aconteceu* com “cada amostra sérica”’. Entretanto, tais perguntas não funcionam em (9), pois estamos lidando com um processo mental, no qual a pergunta perde o sentido, “Hume” não agiu sobre sua realidade, ou algo aconteceu a ele – o evento (o ator de *imaginar* algo) ocorreu internamente.

(7) O segundo trabalho **empregou** a determinação da concentração do princípio ativo paracetamol em comprimidos com espectroscopia no infravermelho próximo e regressão por mínimos quadrados

parciais (PLS, do inglês Partial Least Squares) para o ensino de calibração multivariada. (E_QN_2018_3_EC_001.txt)

- (8) Cada amostra sérica, incluindo os controles, **foi submetida** às diluições de 1:25, 1:50 e 1:500 (Andrade *et al.*, 2016), sendo 1:25 o ponto de corte considerado.
- (9) A experiência do ensaio é aquela para a qual Hume **imagina** a sua ciência da natureza humana, a ciência;

As orações materiais podem ser divididas entre aquelas cujas ações envolvem apenas um participante, o Ator, chamadas **intransitivas**, e aquelas que envolvem mais de um participante, as **transitivas**. O Quadro 11 abaixo exemplifica esses dois tipos de orações.

Quadro 11 – Orações materiais transitivas e intransitivas

Intransitiva	Neville d’Almeida		desenhava	com trilhas de cocaína sobre os objetos escolhidos (L_ASP_2017_2_AA_008)
	Ator		Processo	Circunstância
Transitiva	Nos caminhos verticais do olhar pedestre,	Hélio	desenha	o skyline abissal da metrópole (L_ASP_2017_2_AA_008)
	Circunstância	Ator	Processo	Meta

Fonte: Elaboração própria.

Segundo Thompson (2014), processos materiais compõem a mais ampla categoria dentro do sistema da transitividade, havendo variadas sugestões sobre como esse processo pode ser subcategorizado em níveis mais detalhados. Entretanto, um agrupamento importante separa as orações materiais em processos que *fazem surgir* a Meta – processos materiais criativos – e em processos que *afetam* um dos participantes – processos materiais transformativos.

Assim, nas orações materiais **criativas**, o participante passa a existir no mundo interior ou exterior em decorrência do desdobramento do processo, já nas orações materiais **transformativas**, a mudança resulta de algum aspecto do participante presente ser alterado pelo desdobramento do processo (HALLIDAY e MATTHIESSEN, 2014). Tal distinção entre tipos de orações materiais pode envolver tanto a Meta, quanto o Ator, conforme exemplificado no Quadro 12.

Quadro 12 – Orações materiais (in)transitivas criativas e transformativas

Transitiva criativa	No Twitter,	todas as manifestações	foram feitas	por meio das hashtags #GenocídioGuaraniKaiowa, #SouGuaraniKaiowa e #SomosTodosGuaraniKaiowa (L_RBEP_2018_2_AA_002)
	Circunstância	Meta	Processo	Circunstância
	a legenda	forjou	sua interpretação,	como mostra Knightley (2004) (So_GSP_2017_1_RE_001)
Intransitiva criativa	Ator	Processo	Meta	Circunstância
	Conseqüentemente,	surgem	erros de avaliação	em relação aos múltiplos usos dos recursos hídricos nas bacias hidrográficas. (En_ESA_2018_1_AA_011)
	Circunstância	Processo	Ator	Circunstância
Transformativa transitiva	Com essa medida,	90% dos usuários abastecidos pelo Sistema Cantareira	reduziram	o consumo de água em maio de 2014 (En_ESA_2017_5_AA_004)
	Circunstância	Ator	Processo	Meta
Transformativa intransitiva	A cinematografia sobre os processos transnacionais na Espanha	aumenta	com a percepção social do fenômeno (So_GSP_2017_3_AA_003)	
	Ator	Processo	Circunstância	

Fonte: Elaboração própria.

Cabe ainda mencionar que uma lista dos verbos que tipicamente realizam processos materiais foi elaborada por Halliday e Matthiessen (2014, p. 234-236) em língua inglesa. Os verbos foram agrupados conforme sua natureza, sendo classificados em materiais (in)transitivos criativos: gerais e específicos, e materiais (in)transitivos transformativos de: elaboração, extensão e intensificação, com respectivas subdivisões. Conforme Thompson (2014, p. 120), essa é apenas uma das possibilidades de organização do vasto território dos processos materiais que não pretende ser exaustiva. Em língua portuguesa, uma lista similar foi adaptada por Fuzer e Cabral (2014, p. 48-49), mas, conforme as autoras discutem, está inacabada e há diversas

outras possibilidades de combinações entre as escolhas linguísticas e o contexto que tornam outros verbos, típicos de diferentes processos, passíveis de realizarem processos materiais.

Com relação aos participantes dos processos materiais, além daqueles anteriormente mencionados, o Ator e a Meta, os processos materiais também podem ser acompanhados por Escopo, Beneficiário ou Atributo. Retomando, o **Ator** é o participante obrigatório que pratica a ação, presente nas orações transitivas ou intransitivas, e a **Meta** é o participante afetado pelo processo, obrigatório apenas nas orações transitivas.

O **Escopo** é o participante que não é afetado pelo processo, sendo distinguível em dois tipos: quando diz respeito ao domínio em que o processo se desenrola, é denominado **Escopo-entidade**, pois existe independente do processo (conforme o trecho 10); quando não se refere a uma entidade, mas contribui para a construção do processo, denomina-se **Escopo-processo** (conforme o trecho 11). Ambos participantes estão exemplificados em itálico abaixo.

- (10) “**Siga** *a estrada* a partir da igreja mais ao norte e **dirija-se** *ao replantio de pinheiros mais ao sul*” (E_BCG_2016_2_AA_008)
- (11) “Trouxeram o menino e quando chegou na frente da minha casa **deram** *uma rasteira* no menino, **deram** *um tiro* nas costas e dois tiros na cabeça” (H_CP_2017_3_DO_005)

Em relação ao Escopo-processo, outros exemplos são *tomar banho*, *dar uma olhada*, *dar um sorriso* e *fazer um gesto*. Conforme Halliday e Matthiessen (2014, p. 237), o verbo nesses exemplos são lexicalmente amplos, portanto, o processo é expresso apenas pelo nome que funciona como Escopo. Ainda segundo os autores, esses verbos em tais construções são chamados “verbos leves” (op. cit., p. 241). Na terminologia de Moura Neves (2000 *apud* FUZER e CABRAL, 2014, p. 51), denominam-se “verbos suporte”, cujo significado se encontra bastante esvaziado, de modo que, junto ao complemento, possuem significação distinta, próxima a de outro verbo na língua. No trecho acima, por exemplo, o participante “um tiro” é o que determina o significado do verbo leve *dar*, sendo, portanto, o significado da construção “dar um tiro” similar ao verbo *atirar*.

Além dessa característica, o Escopo tipicamente ocorre em orações intransitivas, que possuem apenas um participante direto, seja Ator ou Meta. Assim, segundo Halliday e Matthiessen (2014, p. 241), pode ser difícil diferenciar semanticamente o Escopo da Meta, por isso, é necessário lembrar que o Escopo não está diretamente envolvido no processo, seja beneficiando-se dele ou sendo afetado por ele, como a Meta, em vez disso, constitui-se como um participante gramatical.

Por sua vez, o participante **Beneficiário** é aquele que se beneficia do processo, embora não esteja intrinsecamente relacionado a conotações positivas. Esse participante pode ser **Recebedor** ou **Cliente**. O primeiro é aquele que recebe bens ou tem bens transferidos pelo Ator, o segundo é aquele a quem bens ou serviços são prestados pelo Ator (FUZER e CABRAL, 2014, p. 51). Abaixo, esses dois tipos de funções exercidas pelo Beneficiário estão exemplificados.

- (12) Plasma [Meta] **foi administrado** [Processo material transitivo transformativo] *em pacientes* [Beneficiário] com hemorragia contínua e TP ou TTP anormal [Adjunto]. (Sa_RBAN_2014_0_AA_007)
- (13) Para evitar que ele contestasse a sua escravidão, seu proprietário [Ator] o [Meta] **vendeu** [Processo material transitivo transformativo] *para comerciantes envolvidos no tráfico transatlântico* [Beneficiário]. (H_TOPO_2014_0_AA_005)

Por último, o **Atributo** diz respeito a uma característica conferida a um dos participantes da oração. Apesar de ser um participante típico de processos relacionais (cf. seção 3.3.3), o Atributo pode ocorrer de modo restrito em orações materiais, sendo classificado em: **Atributivo resultativo**, que constrói um estado qualitativo que resulta do Ator ou da Meta após intervenção do processo, considerado participantes marginais; ou **Atributivo descritivo**, que especifica o estado do Ator ou da Meta ao participar do processo. Diferente dos processos relacionais, nos materiais, o Atributo é sempre opcional.

- (14) Só Carioca [Ator] **saiu** [Processo material intransitivo transformativo] carregado [Atributivo resultativo] da 9ª DP [Circunstância]. (So_GSP_2017_3_AA_005)
- (15) Poucos [Ator] **trabalhavam** [Processo material intransitivo transformativo] como pescadores [Atributivo descritivo], apenas três, anotados como do “serviço de lancha”. (H_TOPO_2014_0_AA_001)

Como mencionado na seção 3.3.1, os processos materiais costumam ter apenas um participante obrigatório, o Ator. Entretanto, por vezes, esse participante não ocupa a posição de sujeito da oração ou não está explícito nela, como nas estruturas de orações passivas. Nesse caso, o participante diretamente relacionado com o processo ainda continua sendo classificado de acordo com sua relação semântica, que permanece inalterada (THOMPSON, 2014, p. 97). Ainda segundo o autor, orações passivas são marcadas em relação às ativas, por estas serem a

escolha natural do usuário. Desse modo, quando empregadas, as passivas geralmente possuem uma razão por trás de sua escolha, como pode ser observado no exemplo abaixo.

- (16) Para indução do banco de sementes do solo as sementes **foram colocadas** em copos de polietileno branco (diâmetro inferior de 4,6 cm, diâmetro superior de 7 cm e 8,3 cm de altura) com capacidade de 200 ml. Os copos **foram perfurados** no fundo e laterais com agulha de $\cong 0,25$ mm de espessura e preenchidos até dois terços com solo do local. Logo após, **foram enterrados** até o nível interno do substrato. Os copos **foram distribuídos** aleatoriamente em grupos de 20 amostras em uma área de 100 m², em local próximo do centro do fragmento florestal. Em cada copo **foram colocadas** 25 sementes cobertas com aproximadamente 1 cm de uma mistura de solo e serrapilheira, seguindo metodologia adaptada proposta por Motta *et al.* (2006). (A_FL_2018_1_AA_002)

No trecho (16), as formas passivas foram utilizadas a fim de omitir o Ator do processo, contudo, esse participante pode ser recuperado através do contexto situacional, referindo-se ao(s) pesquisador(es) responsáveis por conduzir os procedimentos metodológicos descritos nas orações, que representam as etapas de realização da pesquisa. Tal estratégia de omissão serve para tirar o foco do Agente do processo, atentando para a atividade realizada através Meta, uma vez que apesar de omitido o Ator, a relação semântica, ou de representação, ainda é a mesma.

3.3.2 Orações mentais

As orações com processos mentais são aquelas que refletem atividades do mundo interior da consciência humana – em oposição aos processos materiais, que expressam atividades do mundo exterior. Desse modo, os processos mentais formam uma categoria semântica que pode indicar maneiras de pensar, imaginar, gostar, desejar etc., conforme o exemplo a seguir.

- (17) Esses líderes **desejam** genuinamente servir os demais com sua liderança.
(En_GP_2017_1_AA_001)

No exemplo (17) anterior, as pessoas que exercem a ação de *desejar* não estão agindo sobre sua realidade, assim como aquilo que elas desejam não está sofrendo nenhuma ação. Portanto, torna-se inapropriado o uso dos mesmos participantes dos processos materiais – Ator e Meta, por exemplo –, fazendo-se necessário novas denominações para as diferentes funções dos participantes dos processos mentais, conforme apresentado a seguir.

Quadro 13 – Exemplos de processos e participantes em orações mentais

os trabalhadores de enfermagem	perceberam	níveis prejudiciais dos fatores psicossociais	no ambiente de trabalho (Sa_RLAE_2018_0_AA_046)
A mulher	aprecia	a paisagem	de dentro do carro, (So_OS_2017_1_AA_001)
Experienciador	Processo mental	Fenômeno	Circunstância
a empresa	acredita	que é possível com a implantação de ferramentas de gestão ambiental, reduzir impactos ambientais com um investimento mínimo (En_GP_2018_1_AA_010)	
Experienciador	Processo mental	Oração projetada	

Fonte: Elaboração própria.

Conforme apresentado no Quadro 13, o **Experienciador**, nos dois primeiros casos, “os trabalhadores de enfermagem” e “a mulher”, ilustra uma característica típica desse participante, na qual ele geralmente é representado por uma entidade humana ou dotada de algum tipo de consciência, com a capacidade de sentir, pensar e desejar. Entretanto, como observável no terceiro exemplo do Quadro 13, essa função também pode ser exercida por entidades não-humanas, como “a empresa”, desprovidas de consciência, desde que criadas pela mente humana, ou sob a forma de um objeto, uma instituição, produtos da consciência humana (filme, lembrança), metonímias humanas, expressões figurativas (FUZER e CABRAL, 2014, p. 54-55) ou personificações (THOMPSON, 2014, p. 98).

O segundo participante dos processos mentais é o **Fenômeno**, que não possui restrição quanto às entidades que podem participar do processo, podendo ser uma coisa ou uma abstração, geralmente realizadas por grupos nominais (como em 18, a seguir), um ato ou um fato, realizados por orações. Similar aos participantes de orações materiais, em que todos são considerados *coisas*, o Fenômeno pode incluir tanto entidades que representem a experiência humana interna quanto externa, além de poder ser construído enquanto uma metáfora

gramatical¹⁶, por meio de nominalizações (como em 19). Quando se realiza por meio de uma oração não-finita denotando um *ato*, o Fenômeno denomina-se **Macrofenômeno** – exemplificado em (20); quando realizado por uma oração (tipicamente) finita denotando fato, denomina-se **Metafenômeno**, exemplo em (21) (HALLIDAY e MATTHIESSEN, 2014, p. 251-252). Além disso, os processos mentais podem **projetar** uma oração, isto é, nesse caso, o participante Fenômeno é realizado por uma outra oração, conforme demonstrado no exemplo (22) e no último trecho do Quadro 13. Desse modo, as orações projetadas não complementam o processo, mas são consideradas orações em si mesmas. Há ainda a possibilidade de não haver Fenômeno, mas uma Circunstância de Assunto (do pensamento), conforme no trecho (22).

- (18) A crítica nova-iorquina [Experienciador] **reconhece** [Processo mental] o pluralismo das abordagens feministas [Fenômeno] – estruturalistas, psicológicas e até formalistas [Circunstância] (H_CP_2018_1_AA_003)
- (19) Neste trabalho **verificou-se** [Processo mental] o aumento na produção de gás [Fenômeno] com a suplementação de MOS. (A_ABMVZ_2013_0_AA_006)
- (20) Feriu o rosto ao subir uma escadaria com o exemplar de *As mil e uma noites* e então **viu** [Processo mental] a febre crescer [Macrofenômeno] (L_LH_2017_4_AA_001)
- (21) **Atentou-se** [Processo mental] para o fato que o esvaziamento do reservatório para a limpeza seria efetuado no período de chuva [Metafenômeno] (En_ESA_2014_0_AA_009)
- (22) Quando Flaubert escreve, ele **pensa em** Madame Bovary [Circunstância de Assunto]; quando lhe atribui certas reações, ele **pensa que são as mesmas que essa mulher teria** [oração projetada] (L_LH_2018_2_AA_007)

Como ilustrado nos exemplos, no Macrofenômeno o ato é visto, ouvido, sentido ou percebido de alguma forma, associando-se a processos mentais do tipo perceptivo (descritos a seguir nesta seção), mas não é tipicamente sentido por meio de emoções ou desejado. Além disso, a oração não-finita costuma ser expressa por meio de gerúndio, infinitivo ou algum tempo no imperfectivo, podendo ser introduzida por “que”. Todas essas configurações são intercambiáveis, pois o tempo em que decorre o Macrofenômeno pode ser o mesmo do Processo Mental (FIGUEREDO, 2011).

¹⁶ A **metáfora gramatical** é um recurso por meio do qual um processo (ou propriedade) é realizado como um nome, transformando-se em uma coisa ou entidade e tendo sua estrutura linguística recodificada de modo mais incongruente e abstrato. Por meio da metáfora gramatical os participantes se tornam mais complexos devido à reconstrução dos processos como substantivos, empacotando suas informações argumentais, reduzindo seu tamanho e condensando-os em nominalizações.

- (23) Assim como os demais partidos socialistas europeus, os socialistas italianos [Experienciador] **sentiam** [Processo Mental] aumentar as pressões nacionalistas [Macrofenômeno] (H_LN_2017_3_AA_008)
- (23') Os socialistas italianos sentiam as pressões nacionalistas aumentando.
- (23'') Os socialistas italianos sentiam que aumentava as pressões nacionalistas.
- (23''') * Os socialistas italianos sentiam (o fato/a ideia/a noção de) aumentar as pressões nacionalistas.
- (24) Nesse período, **vimos** [Processo Mental] surgir novas doenças [Macrofenômeno] (E_QN_2017_6_AA_015)
- (20') Vimos novas doenças surgindo.
- (20'') Vimos que novas doenças surgiram.
- (20''') * Vimos (o fato/a ideia/a noção de) surgir nova doenças.

Na oração mental com Metafenômeno, este é tipicamente realizado por uma oração finita denotando um *fato*, podendo vir enquadrada por essa mesma palavra ou similar, como “noção”, “ideia” ou “possibilidade” (HALLIDAY e MATTHIESSEN, 2014, p. 253), assim como qualquer outra palavra que realize uma abstração denotadora de emoção, pensamento ou que seja núcleo de elaborações encaixadas (FIGUEREDO, 2011), conforme exemplificado em (25), a seguir. Quando não há presença de tais palavras, é possível inseri-las na frase (conforme as reescritas (25') e (26') a fim de verificar se é um Metafenômeno. De acordo com Halliday e Matthiessen (2014), um *fato* não é um fenômeno material e concreto, mas uma proposição existente dentro de seu próprio campo semiótico, independente de alguém trazê-la à tona ou não. Desse modo, o Metafenômeno costuma ambientar-se em orações mentais emotivas (descritas posteriormente nesta seção), nas quais tais fatos vão de encontro à consciência do Experienciador, devido à sua natureza abstrata, como nos ilustrado nos trechos (25) e (26) a seguir.

- (25) Estes não os teriam rechaçado, como não **estranharam** lutar ao lado de angolas e jejes [Metafenômeno]. (H_TOPO_2014_0_AA_001)
- (25') Não estranharam (o fato/a ideia/a noção de) lutar ao lado de angolas e jejes.
- (25'') * Não estranharam lutando ao lado de angolas e jejes.
- (25''') * Não estranharam que lutava ao lado de angolas e jejes.
- (26) Eu **adoro** comprar presentes [Metafenômeno] para a minha filha [Circunstância].
- (26') Eu adoro (o fato/a ideia/a noção de) comprar presentes.

- (26'') * Eu adoro comprando presentes.
 (26''') * Eu adoro que comprava presentes.

Assim como o Macrofenômeno, pode-se realizar o Metafenômeno por meio de orações com infinitivo, gerúndio ou oração iniciada por “que”, entretanto, diferentemente, tais possibilidades não são alternáveis entre si em todas as situações (FIGUEREDO, 2011), como observável nas reescritas de (25) e (26) anteriores, e de (27), a seguir.

- (27) Por outro lado, muita insegurança permeava o mercado, pois muitos marchands [Experienciador] **temiam** que esse crescimento explosivo não se mantivesse [Metafenômeno] em longo prazo [Circunstância]. (L_ASP_2017_3_AA_005)
- (27') Muitos marchands temiam não se manter esse crescimento explosivo.
- (27'') * Muitos marchands temiam não se mantendo esse crescimento explosivo.
- (27''') Muitos marchands temiam (o fato/a ideia/a noção de) que esse crescimento explosivo não se mantivesse.

Por sua vez, os processos mentais (assim como os verbais, c.f. seção 3.3.4) podem projetar uma oração (ou uma combinação de orações) desvinculada do próprio processo. Tais orações projetadas representam as *ideias* (exemplos em 28 e 29), isto é, o conteúdo do pensamento ou da crença do Experienciador, ou seus *desejos* (trecho 30), não sendo consideradas parte da oração mental, mas combinadas com essa. Segundo Thompson (2014, p. 109), a relação entre a projeção e a oração mental é similar a de um quadro e sua moldura, respectivamente, na qual juntos formam uma unidade complexa, embora não sejam partes um do outro. Cabe ainda mencionar que as projeções de desejo não possuem uma localização no tempo, pois se restringem à opinião do Experienciador e, por outro lado, as projeções de ideias são realizadas por orações indicativas, que têm seu próprio tempo, à parte da oração mental que as projeta (vide destaques em itálico no trecho 28, no qual os processos se realizam em tempos verbais distintos e não relacionados).

- (28) a maioria dos entrevistados [Experienciador] **pensa** que isso só é possível se *deixarem* a floresta intocável [Oração projetada] (A_FL_2017_0_AA_042)
- (28') * A maioria dos entrevistados pensa o fato de isso só é possível se deixarem a flores intocável.
- (29) Tem uns médicos que **acham** que **sabem** tudo [Fenômeno] [Oração projetada]. (Sa_RLAE_2017_0_AA_039)

- (29*) * Tens uns médicos que acham o fato de sabem tudo.
- (30) Considerando a linearização do sistema (3.3) e o teorema 4.1, **queremos** que a perturbação λt tenda a zero ao $t \rightarrow \infty$ [Oração projetada]. (E_TSC_2013_0_AA_018)
- (30*) * Queremos o fato de a perturbação λt tenda a zero ao $t \rightarrow \infty$.

Além da distinção entre os tipos de Fenômenos, as orações mentais podem ser divididas em quatro subcategorias de processos mentais: perceptivos, cognitivos, emotivos e desiderativos. Assim como os outros processos do sistema da transitividade, as subcategorias acima constroem a experiência de modo indeterminado, pois os quatro tipos podem mesclar-se entre si. Por exemplo, nos trechos (31) e (32) o mesmo verbo pertence a duas subcategorias diferentes, na primeira com significado cognitivo de ‘compreender’ e na segunda, perceptivo de ‘enxergar visualmente’.

- (31) Ao disporem de mais informações sobre os produtos, os clientes [Experienciador] **viram** [Processo mental cognitivo] seu poder aumentar no relacionamento com as empresas [Macrofenômeno] (En_GP_2018_1_AA_015)
- (32) Vocês [Experienciador] já **viram** [Processo mental cognitivo] o filme O menino do pijama listrado? [Fenômeno] (L_RBEP_2_AA_004)

Quanto aos tipos de orações mentais, os processos mentais **perceptivos** constroem a percepção do mundo através dos cinco sentidos; os processos mentais **cognitivos** evidenciam atos do pensamento do Experienciador; os processos mentais **emotivos** expressam sentimentos ou afeição; e os processos mentais **desiderativos** exprimem desejo ou interesse em algo. Todas essas categorias estão exemplificados a seguir, nos trechos de (33) a (36).

- (33) Alguns deles [Experienciador] **olham** [Processo mental perceptivo] para a câmera [Fenômeno], percebendo a existência do fotógrafo (L_ASP_2017_3_AA_004)
- (34) Alfaya e Schermann¹⁹ [Experienciador] **inferiram** [Processo mental cognitivo] que a sensibilidade materna é mais favorável em mães de lactentes nascidos a termo e saudáveis (grupo com as maiores taxas de aleitamento materno) do que em mães de lactentes nascidos com risco neonatal [Oração projetada]. (Sa_RPP_2017_2_AA_003)
- (35) Morrison semeia o seu chão-território *nesse processo de construção ambivalente de personagens* [Fenômeno] que ela [Experienciador] **deseja** [Processo mental desiderativo] e **odeia** [Processo mental emotivo] ao mesmo tempo. (L_ELBC_2014_0_AA_002)
- (36) Rawls [Experienciador] **almeja** [Processo mental desiderativo] estender a sua concepção de sociedade bem-ordenada [Macrofenômeno] para o quadro internacional. (H_TFA_2017_4_005)

A distinção entre os diferentes tipos de processos mentais não se dá apenas no âmbito semântico, havendo outras características que marcam cada tipo. Segundo Halliday e Matthiessen (2000, p. 143), os processos mentais emotivos se aproximam mais de uma descrição atributiva (c.f. seção 3.3.3 sobre processos relacionais) do que um mental prototípico, como apresentado nos trechos a seguir, nos quais o processo relacional de (37) foi reescrito em uma alternativa similar realizada por uma oração mental em (37'). Além disso, os processos emotivos e os perceptivos podem surgir de uma projeção, como em (30), mas não costumam criá-las (HALLIDAY e MATTHIESSEN, 2000).

- (37) Ora, se [a ideia de ser caçado por capitães do mato] **era** [assustadora] para qualquer cativo, **podemos imaginar** [que o estado de gravidez avançada de Felícia agravou sua condição de vulnerabilidade] (H_CP_2017_2_AA_001)
- (37') A ideia de ser caçado por capitães do mato assustava qualquer cativo.
- (38) “Eu acredito que **odeio** ideologias tanto quanto você” (H_LN_2-17_2_AA_001)

Por sua vez, os processos mentais perceptivos portam-se similarmente aos processos comportamentais, podendo apresentar um único participante (Experienciador) consciente e passivo (FIGUEREDO, 2011), como em (39) a seguir, em que o Experienciador apenas executa a ação de pensar, sem identificar o conteúdo do Processo. Nesse caso, denomina-se Processo não-fenomenalizado (ou Fenômeno não-especificado). Já os processos mentais cognitivos e desiderativos, diferente dos demais, podem ambos projetar orações enquadradas por modalizações, como em (37) anterior e, costumeiramente, são considerados mentais prototípicos.

- (39) Mas eu [Experienciador] não reagia, só **pensava** [Processo mental não-fenomenalizado]. (H_CP_2018_1_AA_006)

Cabe ainda ressaltar que os processos mentais podem ser bidirecionais, isto é, podem ser construídos como um processo consciente que parte do Experienciador e se estende para o Fenômeno (denominado **emanente**), ou como um processo externo que provoca uma reação no Experienciador (**impingente**), ambas direcionalidades estão ilustradas nos trechos (40) e (41), a seguir. Tal bidirecionalidade é própria apenas dos processos mentais, podendo ser considerada

como uma característica desse processo em PB, pois ocorre em grande número de eventos e se estende para mais de um tipo de processo mental, de acordo com Figueredo (2011, p. 266).

- (40) Porém, uma vez que nos [Experienciador] **interessa** [Processo mental impingente] apenas as distâncias calculadas [Fenômeno] (E_BCG_2016_4_AA_011)
- (41) O corte em cubos [Fenômeno] **agradou** [Processo impingente] os provadores [Experienciador] no início do armazenamento [Circunstância] (A_BJFT_2017_0_AA_008)

Quanto ao tempo verbal preferido, em inglês, opta-se pelo presente simples (em vez do presente contínuo geralmente empregado nos processos materiais) (HALLIDAY e MATTHIESSEN, 2014). De acordo com Caffarel (2006), em francês, o pretérito não-marcado é o imperfeito, que se distingue do pretérito simples ou composto comumente utilizado nos processos materiais). Segundo Figueredo (2011), no português brasileiro o tempo verbal típico do “agora” é o presente e o presente-no-presente.

Retomando as propriedades das orações mentais que a distinguem das materiais, conforme Halliday e Matthiessen (2014), essas podem ser resumidas em relação: 1) à presença obrigatória de dois participantes nas orações mentais, diferente das orações materiais em que apenas um é requerido; 2) na oração mental, ao Experienciador é sempre consciente, diferente do Ator na oração material, que pode sê-lo ou não; 3) ao contrário das orações materiais, as mentais podem projetar outras orações; 4) ao tempo verbal não-marcado da oração material é o presente contínuo enquanto o da oração mental é o presente do indicativo, embora cada oração possa ser realizada com os demais tempos verbais; e, por fim, 5) aos processos mentais que não podem ser substituídos ou questionados pelo verbo “fazer”, como podem os processos materiais (como em “O que o [Ator do processo material] fez?”).

3.3.3 Orações relacionais

Os processos relacionais expressam a noção de ser ou estar, servindo para caracterizar e/ou identificar algo, estabelecendo uma relação entre duas entidades diferentes. Em função disso, são utilizadas para representar seres no mundo relacionando-os às suas características e identidades por meio de uma construção cuja mudança/desenvolvimento do processo ocorre em inércia (HALLIDAY e MATTHIESSEN, 2014). Obrigatoriamente haverá dois participantes acompanhando os processos relacionais, cujo comportamento se aproxima do Fenômeno das

orações mentais, por poderem ser não apenas coisas, como também atos e fatos (se distanciando, assim, dos processos materiais), conforme exemplificado nos trechos a seguir.

- (42) Uma alternativa aos ionóforos que surge [Identificado] **é** [processo relacional identificativo] a utilização do conceito de imunização passiva [Identificador] como nova ferramenta para manipular a fermentação ruminal (AGCC) (A_ABMVZ_2013_0_AA_002)
- (43) Nesta turnê, Milton [Portador] **aparecia** [processo relacional atributivo] de cabelo black power, calças boca-de-sino e colares [Atributo], acompanhando o Som Imaginário em improvisações sobre *With a little help from my friends*, dos Beatles. (L_PEMU_2014_0_AA_006)
- (44) O diabetes mellitus [Identificado] **constitui** [processo relacional identificativo] o grupo das doenças crônicas responsável pelas principais causas de morte no mundo [Identificador] e **é** considerado um dos problemas de saúde de maior magnitude¹. (Sa_RLAE_2017_0_AA_010)

As orações relacionais se apresentam de dois modos: atributivo (relações de filiação) ou identificativo (relações de identificação), sendo que cada um pode ser organizado em três categorias distintas: intensivas, possessivas ou circunstanciais.

Quanto ao modo, as **orações relacionais atributivas** podem construir relações abstratas entre membros de uma classe, sendo seus participantes o Portador e o Atributo. Nelas, emprega-se tipicamente o verbo *ser*, mas também podem ser usados outros verbos atributivos como *estar*, *fazer-se*, *manter-se*, *parecer*, *permanecer*, *resultar*, *soar*, *tornar-se* e *virar* (FUZER e CABRAL, 2014, p. 68). Acompanhando o processo, o **Atributo** é a característica de um dos participantes, realizada por um grupo que constrói uma classe de coisas geralmente indefinidas, podendo apresentar um adjetivo ou substantivo comum (mas não um nome próprio ou pronome, pois esses não constroem classes). Já o **Portador** é o participante que recebe essa característica expressa pelo Atributo, relacionando-se à noção de pertencimento à classe típica representada pelo Atributo.

O significado de uma oração atributiva é o de pertencimento à uma classe, ou seja, de filiação entre os participantes. Desse modo, segundo Halliday e Matthiessen (2014), na categoria *intensiva* (detalhada mais adiante nesta seção) tal filiação pode ser classificada em *entidade* ou *qualidade* do Atributo, de acordo com a entidade que constitui a classe (como em 45 e 46), ou o critério nomeado ao referenciar as qualidades que caracterizam a classe (como de 47 a 49).

- (45) Como em 2007 o Parque Beija-Flor [Portador] **era um ambiente novo** [Atributo: Entidade] (B_HOE_2017_3_AA_004)

- (46) O ganho de peso por área [Portador] é um indicativo da eficiência de utilização das pastagens [Atributo: Entidade] (A_ABMVZ_2018_3_AA_020)
- (47) Uma alternativa [Portador] que se tornou viável [Atributo: Qualidade] atualmente é a aquisição de dados por meio de sistemas de VLA. (E_BCG_2017_1_AA_009)
- (48) 87% do território capixaba [Portador] era coberto [Atributo: Qualidade] por florestas (B_R_2017_5_AA_001)
- (49) Essa massa [Portador] era muito menor do que aquela utilizada nos experimentos de troca iônica [Atributo: Qualidade] com a PAni sintetizada quimicamente (E_QN_2014_AA_008)

Como exemplificado nos trechos (45) e (46), anteriores, o Atributo: Entidade é realizado por grupos nominais com uma *coisa* enquanto núcleo, “um ambiente novo”, que é algo concreto, e “um indicativo da eficiência de utilização das pastagens”, que é mais abstrato. Por sua vez, conforme os trechos (47) e (48), o Atributo: Qualidade é realizado por um grupo com um Epíteto como núcleo, podendo ter tanto um adjetivo – como em (47) – quanto um verbo no particípio passado, em (48), como núcleo. O Atributo: Qualidade pode ainda ser modificado por advérbios de grau e comparativo, conforme (49).

Além dessa propriedade do Atributo, de acordo com Halliday e Matthiessen (2014), as orações relacionais atributivas também podem ser identificadas quanto a fase da atribuição, isto é, quanto ao tempo em que ocorre o processo. Na forma não-marcada, a fase é considerada *neutra* e não-especificada (como em grande parte das orações com o verbo *ser*); na marcada, leva-se em consideração a (1) fase temporal (a) incoativa (que designa o começo ou aumento progressivo de ação, e.g. *tornar-se* e *crescer*) ou (b) durativa (e.g. *continuar* e *permanecer*); ou (2) fase de realidade (a) aparente (e.g. *aparecer* e *aparentar*), (b) perceptiva (e.g. *soar* e *parece*) ou (c) realizada (e.g. *provar* e *tornar-se*). O Quadro 14 a seguir exemplifica tais possíveis classificações do processo relacional atributivo.

Quadro 14 - Classificação dos tipos de processos atributivos

Fases		Exemplos
Temporal	Incoativa	Esses se tornam susceptíveis às infecções primárias (parasitárias) e secundárias (bacterianas) (A_ABMVZ_2018_1_AA_001)
	Durativa	Devido ao sistema reprodutivo debilitado, susceptível às infecções, a égua permanece improdutiva por mais anos consecutivos (A_ABMVZ_2014_0_AA_002)

Realidade	Aparente	“Significa ‘corra o mais rápido que puder’”, afirma a menina que aparenta cerca de seis anos de idade. (H_LN_2018_1_AA_05)
	Perceptiva	o texto soava ainda mais desafiador quando se levava em conta que seu autor era um scholar ainda na casa dos vinte anos de idade. (H_LN_2017_3_AA_004)
	Realizada	a impressão de movimento gerada pela sucessão de imagens estáticas logo se provou capaz de animar as mais estranhas criaturas (So_GSP_2018_3_AA_005)

Fonte: Elaboração própria.

Já no segundo tipo, as **orações relacionais identificativas**, caracteriza-se um dos participantes em termos de outro, sendo esses denominados Identificado e Identificador (ou Valor e Característica, respectivamente). Essencialmente, a relação apresentada relaciona uma realização específica em uma categoria mais abrangente e abstrata (THOMPSON, 2014, p.102), em que ‘x é identificado por y’ ou ‘y serve para definir a identidade de x’ (HALLIDAY e MATTHIESSEN, 2014, p. 276). Assim como as atributivas, as orações identificativas normalmente são realizadas pelo verbo *ser*, mas do tipo equativo, como *exemplificar*, *indicar*, *significar* e *representar*. Quanto aos participantes, **Identificado** é a entidade que recebe a identificação enquanto o **Identificador** é a identidade que se atribui ao Identificado. Tipicamente, o grupo nominal que realiza a função de Identificador é um substantivo acompanhado por artigo definido ou outro determinante específico como dêitico, um nome próprio ou pronome, conforme os trechos de (50) a (53) a seguir. Cabe ainda observar nesses exemplos que os processos são de classe equativa, ou seja, expressam uma relação de igualdade entre os termos.

- (50) Os principais recursos utilizados na produção do serviço [Identificado] **são** [processo relacional identificativo] as competências dos peritos criminais [Identificador] (engenheiros, físicos, químicos, farmacêuticos, contadores, biólogos, dentistas etc.), enquanto os artefatos tecnológicos **são** os recursos auxiliares. (En_GP_2017_3_AA_010)
- (51) Os motivos colocados [Identificado] **refletem** [processo relacional identificativo] a ação reflexiva desses atores sociais [Identificador] (So_OS_2018_2_AA_004)
- (52) Por questão de escala a Figura 1b [Identificado] **representa** [processo relacional identificativo] a zona do litoral [Identificador] sem mostrar suas subdivisões. (B_HOE_2014_0_AA_001)
- (53) De acordo com Vidal e Capelo Neto (2014), a eutrofização das águas [Identificado] **significa** [processo relacional identificativo] seu enriquecimento [Identificador] por nutrientes (En_ESA_2018_2_AA_004)

Para Halliday e Matthiessen (2014, 284), a estrutura do processo relacional identificativo é provavelmente a mais difícil de se classificar no sistema da TRANSITIVIDADE, sendo também a mais importante e que tende a predominar em certos gêneros como científico, político e burocrático, pois neles o significado construído é tipicamente simbólico. Isso posto, os autores chamam atenção para alguns subtipos de orações identificativas, que estão resumidas no Quadro 15 a seguir.

Quadro 15 - Subtipos de orações relacionais identificativas

Subtipo de oração relacional identificativa	Exemplo
Equação	Neste artigo utiliza-se n [Identificador] = [Processo] 99 [Identificado]. (E_BCG_2014_0_AA_005)
Equivalência	Cada diagrama corresponde à média dos resultados da amostra considerada. (En_RIEM_2018_3_AA_007)
Papel	Algumas pessoas me perguntam: Mas você já não é mãe? (H_CP_2017_1_AA_001)
Nomeação	Ele então se revela: “Electra, eu sou Orestes.” (H_TFA_2017_4_AA_004)
Definição	O enduro equestre é a competição que visa testar a habilidade do atleta em administrar com segurança a resistência e o condicionamento de seu cavalo contra o trajeto (A_ABMVZ_2018_2_AA_001)
Simbolização	os valores da energia de ativação indicam a sensibilidade da viscosidade aparente à alteração da temperatura (A_BJFT_2014_0_AA_004)
Exemplificação	Alguns critérios nortearam a escolha das palavras de formação do corpus. Um deles é o da familiaridade (L_LH_2017_1_AA_006)
Demonstração	As Figuras 4 e 5 mostram os espectros de excitação para as amostras A4, 5, 6, 7, 8 e 9 (E_QN_2018_8_AA_002)

	Evidências filogenéticas sugerem [[que a transmissão horizontal é particularmente a rota mais importante para a manutenção do vírus em rebanhos ovinos]] (A_AIB_2017_0_AA_00)
--	--

Fonte: Elaboração própria.

É importante frisar que o último exemplo do quadro, com o processo relacional *sugerir*, se aproxima dos processos verbais (cf. seção 3.3.4), pois têm como característica possibilidade de mesclar dois participantes envolvidos em uma relação simbólica (HAO, 2018). Entretanto, no trecho mencionado, o que poderia ser considerado a Locução de um processo verbal, não pode ser citado, apenas reportado (afastando-se, assim, da classificação de processo verbal). Além disso, há ainda a possibilidade de acrescentar-se “o fato” ao Identificador, tornando a oração reportada “rebaixada de nível” (“*down-ranked clause*”), conforme apresentado nos trechos do Quadro 16 a seguir. Tais orações são fundamentais para o discurso acadêmico, pois são bastante frequentes (HALLIDAY e MATHIESSEN, 2014; HAO, 2018) e permitem que o autor do texto exponha um ponto de vista de forma aparentemente objetiva, conforme pode-se observar na análise deste estudo (Capítulo 5).

Quadro 16 - Exemplos de oração relacional do subtipo *demonstrativo*

Evidências filogenéticas	sugerem	(o fato de) [[que a transmissão horizontal é particularmente a rota mais importante para a manutenção do vírus em rebanhos ovinos]] (A_AIB_2017_0_AA_00)
a história	mostra	(o fato de) [[que grandes eventos esportivos têm sido escolhidos de forma frequente como teatro de operações por grupos terroristas]] (En_GP_2017_3_AA_003)
os dados	indicam	(o fato de) [[que esta estratégia está associada a um desempenho cada vez melhor em termos de público e renda de bilheteria]] (So_GSP_2018_1_AA_005)
Identificado	Processo relacional identificativo	Identificador

Fonte: Elaboração própria.

Cabe mencionar ainda que as orações identificativas são reversíveis (HALLIDAY e MATTHIESSEN, 2014; THOMPSON, 2014), isto é, podem ter a ordem de seus elementos invertida sem que a alternativa se torne marcada ou provoque estranhamento no leitor, como demonstram as reescritas dos trechos (50) e (51), a seguir. Por sua vez, inverter os termos das orações atributivas é possível, embora as torne marcadas, conforme as reescritas (48') e (49'), a seguir.

- (50) Os principais recursos utilizados na produção do serviço [Identificado] são [processo relacional identificativo] as competências dos peritos criminais [Identificador] (engenheiros, físicos, químicos, farmacêuticos, contadores, biólogos, dentistas etc.), enquanto os artefatos tecnológicos são os recursos auxiliares. (En_GP_2017_3_AA_010)
- (50') As competências dos peritos criminais são os principais recursos utilizados na produção do serviço [...] os recursos auxiliares são os artefatos tecnológicos.
- (51) Os motivos colocados [Identificado] refletem [processo relacional identificativo] a ação reflexiva desses atores sociais [Identificador] (So_OS_2018_2_AA_004)
- (51') A ação reflexiva desses atores sociais **reflete** os motivos colocados.
- (48) 87% do território capixaba [Portador] **era** coberto [Atributo] por florestas (B_R_2017_5_AA_001)
- (48') *Coberto era 87% do território capixaba por florestas.
- (49) Essa massa [Portador] **era muito menor do que aquela utilizada nos experimentos de troca iônica** [Atributo] com a PAni sintetizada quimicamente (E_QN_2014_AA_008)
- (49') Muito menor do que aquela utilizada nos experimentos de troca iônica era essa massa com a PAni sintetizada quimicamente (E_QN_2014_AA_008)

Além da possibilidade de reversibilidade nas orações relacionais identificativas, há outras propriedades típicas inerentes aos modos atributivos e identificativo que contribuem a distinção de ambos. Em uma combinação entre os quesitos apontados por Thompson (2014, p.104-105) e Halliday e Matthiessen (2014), podemos destacar:

- i) Se o segundo participante da oração for um adjetivo, ou realizado por um grupo nominal indefinido, então será provavelmente um Atributo e o processo relacional, atributivo em vez de identificativo.
- ii) Por sua vez, em oposição, se o segundo participante da oração for realizado por um grupo nominal definido, um pronome, ou um adjetivo superlativo, então o processo provavelmente será identificativo. Por exemplo, essa oração traduzida

de Halliday e Matthiessen (2014, p. 268), “Ele é Charles Darwin” é Identificativa, entretanto, “Ele é *outro* Charles Darwin”, é atributiva, já que o termo *outro* ressignifica “Charles Darwin”, criando uma classe de pessoas que são como Charles Darwin.

- iii) Se um ou ambos os participantes forem uma oração encaixada (“*embedded clause*”), é provável que o processo seja identificativo em vez de atributivo, conforme exemplificado a seguir, em (54).

(54) O objetivo desta seção [Identificado] é [processo relacional identificativo] apresentar as técnicas de processamento usadas para avaliar a câmera MUX do CBERS-4 [Identificador] (E_BCG_2014_0_AA_004)

(54') Apresentar as técnicas de processamento usadas para avaliar a câmera MUX do CBERS-4 é o objetivo desta seção

(55) para o pensamento feminista, a questão não [Identificado] é [processo relacional identificativo] destruir o cânone [Identificador], mas repensá-lo, (H_CP_2018_1_AA_003)

(55') Destruir o cânone não é a questão

- iv) Os verbos lexicais que realizam os processos relacionais atributivos pertencem à classe de “descritivos” enquanto os identificativos são realizados pela classe de “equativos”.

- v) Em orações relacionais identificativas é possível substituir o verbo por “representa/é representado por”, o que não é possível em orações atributivas, de acordo com as rescritas abaixo.

(56) Minas Gerais é o estado brasileiro, atualmente, com maior produção de leite. (So_NE_2018_1_AA_009)

(56') Minas Gerais **representa** o estado brasileiro, atualmente, com maior produção de leite.

(57) Esta nova forma de legitimação é **definida** essencialmente pela qualidade da atuação (So_RDGV_2017_1_AA_006)

(57') Esta nova forma de legitimação é **representada** essencialmente pela qualidade da atuação

- vi) Por último, os dois tipos de processos relacionais podem ser testados por diferentes perguntas: em atributivos, “Como/O que [Portador] é?”, conforme

(48), (49) e suas reescritas; em identificativos, “Qual/Quem [Identificado] é?”, em (54) e (58) e suas reescritas.

- (48) 87% do território capixaba era coberto por florestas (B_R_2017_5_AA_001)
- (48'') Como era 87% do território capixaba? Coberto por florestas.
- (49) Essa massa **era muito menor do que aquela utilizada nos experimentos de troca iônica** [Atributo: qualidade] com a PAni sintetizada quimicamente (E_QN_2014_AA_008)
- (49'') O que era muito menor do que aquela utilizada nos experimentos de troca iônica? Essa massa.
- (54) O objetivo desta seção é apresentar as técnicas de processamento usadas para avaliar a câmera MUX do CBERS-4 (E_BCG_2014_0_AA_004)
- (54'') Qual é o objetivo desta seção? Apresentar as técnicas de processamento usadas.
- (58) No esquema anterior, o cliente **é** o iniciador (INI) do processo tradutório e **é**, em princípio, quem define a encomenda de tradução (L_RD_2018_1_AA_002)
- (58'') Quem é o cliente? O iniciador (INI) do processo tradutório | quem define a encomenda de tradução.

Além da distinção de orações relacionais em atributivas e identificativas, é possível identificar cada uma em três categorias principais de relações: intensivas, possessivas e circunstanciais. No tipo **intensivo**, as orações relacionais servem para caracterizar uma entidade em relações de generalidade (delicadeza), de abstração (realização) ou de uma ocorrência de um tipo (instanciação) (HALLIDAY e MATTHIESSEN, 2000). Além disso, são realizadas tipicamente pelos verbos *ser* e *estar*, assim como *parecer*, *permanecer*, *ficar*, *tornar-se*, *representar* entre outros (FUZER e CABRAL, 2014, p. 65), em função do modo, seja atributivo ou identificativo, em que ocorrem, conforme exemplificam os trechos (59) a (64) a seguir.

- (59) Neste caso, os parâmetros dos modelos B e C **parecem** [processo relacional atributivo intensivo] mais indicados. (En_ESA_2018_3_AA_004)
- (60) as atividades enzimáticas **tornaram-se** [processo relacional atributivo intensivo] independentes da concentração do substrato, seguindo uma cinética do tipo Michaelis-Menten. (E_QN_2017_4_AA_004)
- (61) a carga última obtida no modelo numérico **ficou** [processo relacional atributivo intensivo] um pouco superior à experimental, e para o pórtico A40 o deslocamento último **ficou** [processo relacional atributivo intensivo] um pouco menor em relação ao valor experimental. (En_RIEM_2017_2_AA_006)
- (62) Um caso paradigmático **é** [processo relacional identificativo intensivo] o Programa de Agentes Comunitários de Saúde (PACS) (As_SASO_2014_0_AA_002)

(63) A Tabela 2 **ilustra** [processo relacional identificativo intensivo] os resultados obtidos durante a otimização do pH (B_RAA_3_AA_004)

(64) O nervo isquiático **é** [processo relacional identificativo intensivo] o maior nervo do corpo, continuando até a extremidade distal do membro. (A_ABMVZ_2018_3_AA_007)

Cabe ainda mencionar uma característica própria das orações intensivas, sejam identificativas ou atributivas, que é a possibilidade de *Designação* (*assignment*), em que há um terceiro participante representando a entidade que determina a relação da identificação ou atribuição. Os trechos (65) e (66) a seguir exemplificam tal propriedade. Em (62), a “a rainha da Holanda” é responsável por identificar “ele” (“Bernard L. Feringa”) com seu Identificador “cavaleiro”. Já em (66), o Portador “personagem da novela de Robert Silverbeg”, ao qual o pronome oblíquo “o” refere-se, ficou “exposto” (Atributo) devido a “não foi devidamente fechado”, que é a Designação: Atributor.

(65) Bernard L. Feringa obteve seu PhD na Universidade de Groningen na Holanda, sob a supervisão do professor Hans Wynberg. (...) Em 2008, ele [Identificado] **foi nomeado** cavaleiro [Identificador] pela rainha da Holanda [Designador]. (E_QN_2017_1_EC_002)

(66) A identificação da feição que seria desenhada [Portador] **foi considerada** pela maioria dos usuários do OSM [Atributor] como moderadamente difícil à difícil [Atributo], enquanto o mesmo número de usuários do Wikimapia [Atributor] **considerou** fácil à extremamente fácil [Atributo] encontrar a feição [Portador]. (E_BCG_2016_3_AA_003)

Por sua vez, as orações relacionais **circunstanciais** relacionam elementos em termos de tempo, lugar, modo, causa, acompanhamento, papel, assunto ou ângulo que são os diferentes tipos de Circunstâncias (cf. seção 3.3.7). Em relação aos modos, nas orações atributivas, o elemento circunstancial é um atributo que descreve uma entidade, podendo a Circunstância ser construída na forma de Atributo, como em (67), ou podendo o elemento circunstancial estar inscrito no Processo, como em (68). Diferente das orações relacionais atributivas intensivas, os Atributos das circunstanciais frequentemente são formados por um grupo nominal definido, como nos exemplos a seguir.

(67) O terceiro tópico **é** [processo relacional atributivo circunstancial] sobre a capacidade dos serviços logísticos com enfoque estratégico na RBV. (Em_GP_2017_2_AA_011)

(68) O primeiro caso **concerne** [processo relacional atributivo circunstancial], claro, às lutas travadas por Tereza. O segundo caso *pertence*, por exemplo, ao Comitê de Solidariedade Emília. (H_CP_2017_2_DO_005)

No primeiro caso, (67), a relação se constrói por um verbo menor (de acordo HALLIDAY e MATTHIESSEN, 2014) realizado por uma preposição (*ser + sobre*), sendo apenas o Atributo responsável por definir a oração como circunstancial. No segundo, (68), a relação circunstancial realiza-se pelo próprio verbo (*concerne*). Quando o verbo expressa a relação lexical de circunstância (conforme os trechos a seguir, 69 a 71), a característica de não-reversibilidade se torna marcada, como pode ser observado nas reescritas (69'), (70') e (71').

- (69) Os primeiros trabalhos na área **datam** [processo relacional atributivo circunstancial] de não mais do que vinte anos (Em_RIEM_2017_1_AA_007)
- (69') *Não mais do que vinte anos data os primeiros trabalhos na área.
- (70) A artéria femoral e a veia **correm** [processo relacional atributivo circunstancial] superficialmente entre a borda caudal do sartório e os músculos pectíneos ao lado do nervo safena (Fig. 6). (A_ABMVZ_2018_2_AA_007)
- (70') *A borda caudal do sartório e os músculos pectíneos ao lado do nervo safena correm entre a artéria femoral e a veia.
- (71) Muitas das intervenções previstas na lei **dependiam** [processo relacional atributivo circunstancial] de regulamentação posterior. (So_OS_2017_1_AA_007)
- (71') Regulamentação posterior dependia de muitas intervenções previstas na lei.

Já no modo identificativo, assim como no atributivo, a relação circunstancial (de tempo, lugar, modo, entre outros) organiza-se semanticamente de dois modos (HALLIDAY e MATTHIESSEN, 2014): enquanto característica dos participantes (como em 72) ou característica do processo (como em 73).

- (72) Uma forma de remover ruído instrumental **é** [processo relacional identificativo circunstancial] por alisamento, ou seja, pela aplicação de um filtro que possibilitará a melhora da razão sinal/ruído (E_QN_2018_3_EC_001)
- (73) A inflorescência [identificador] **é** curta e **surge** [processo relacional identificativo circunstancial] da axila das folhas [Identificado] (B_HOE_2017_1_AA_003)

Nas orações relacionais identificativas circunstanciais, o Identificado pode ser realizado por uma ampla variedade de classes gramaticais, como grupos nominais ou adverbiais, frases preposicionadas ou orações iniciadas por “que”, enquanto o Identificador costuma ser um grupo

nominal com o nome de uma classe de circunstâncias como Coisa, conforme exemplificado no Quadro 17, baseado em Halliday e Matthiessen (2014, p. 292).

Quadro 17 - Exemplos de orações circunstanciais identificativas

Identificado	Identificador	Exemplo
Grupo adverbial	Grupo nominal: Coisa: (<i>colega</i>) (<i>eu</i>)	“ <u>hoje</u> é o <i>colega</i> , <u>amanhã</u> sou eu” (H_EDRE_2014_0_AA_0010)
Frase preposicionada	Grupo nominal: Coisa: (<i>forma</i>)	<i>A melhor forma de diferenciá-las</i> é pelo <u>isolamento</u> em meios de cultura, testes de patogenicidade e alguns testes bioquímicos e fisiológicos (A_AIB_2017_0_AA_009)
Oração	Grupo nominal: Coisa: (<i>razão</i>)	<i>A razão</i> é <u>que os relatórios podem reunir todo o material elaborado pelo corpo técnico para a instrução processual</u> (So_RDGV_2017_3_AA_005)

Fonte: Adaptado de Halliday e Matthiessen (2014, p. 292).

Além do tipo exemplificado anteriormente – circunstância como participante –, a relação identificativa circunstancial pode ocorrer por meio do processo, com verbos lexicais que expressam tempo, lugar ou outras categorias circunstanciais (como em 74 e 75). Nesse caso, diferente dos atributivos, o verbo pode aparecer em sua forma passiva (como em 74’ e 75’).

(74) A batata-doce (*Ipomoea batatas*) **ocupa** o sétimo lugar na cultura alimentar mundial (A_ABJFT_2018_0_AA_009)

(74’) O sétimo lugar na cultura alimentar mundial é ocupado pela batata-doce.

(75) Não é este o lugar entificado) para explorar exaustivamente o modo como a questão da técnica (Identificador) **permeia** as reflexões de Benjamin (Identificado) sobre a fotografia (H_TFA_2017_1_AA_004)

(75’) As reflexões de Benjamin são permeadas pela questão da técnica.

Segundo Halliday e Matthiessen (2000, p. 147), os processos relacionais circunstanciais podem ser empregados na construção de arranjos dispostos em um tempo e espaço, assim como na construção do conhecimento de modo metafórico, conforme exemplificado no trecho (76) a seguir.

- (76) O nervo tibial **emerge** das raízes sacrais do tronco lombossacral (Fig. 5). **Seguindo-se** à sua separação do nervo fibular, próximo ao meio da coxa, **emite** ramos musculares proximais importantes para as cabeças pélvicas dos músculos caudais da coxa, **atinge** a perna e **passa** entre as duas cabeças do músculo gastrocnêmio. Esse nervo *é* responsável pela inervação do músculo gastrocnêmio, das três cabeças do músculo flexor digital profundo e superficial e do músculo poplíteo. Ele **desce** ainda entre a porção lateral do músculo gastrocnêmio e o músculo flexor superficial dos dedos e no terço distal da perna, **situa-se** cranialmente ao tendão calcâneo. Na dissecação do nervo femoral, incisões na pele foram feitas para expor os músculos e as estruturas na face medial da pelve. O músculo sartório foi cortado no seu ponto de inserção distal e rebatido. Após a remoção de gordura e do tecido conjuntivo, os nervos e os vasos sanguíneos foram separados, identificados e fotografados (Mahler e Adogwa, 2008). O nervo femoral **passou** pelo músculo iliopsoas até o membro pélvico proximal, onde **se localiza** entre o quadríceps femoral, o vasto medial e o reto femoral, **dividindo-se** em ramos que inervam o quadríceps femoral. O nervo **foi** completamente **coberto** pela extremidade proximal do músculo sartório. A artéria femoral e a veia *tinham* ramos para os dois ventres do músculo sartório que **correm** ao lado e mediocranialmente ao nervo femoral, logo após o músculo iliopsoas. A artéria femoral e a veia **correm** superficialmente entre a borda caudal do sartório e os músculos pectíneos ao lado do nervo safena (Fig. 6). (A_ABMVZ_2018_2_AA_007)

O trecho (76) é a expansão do contexto de (70) que descreve o mapeamento dos nervos de um cadáver canino e de um bezerro Holandês e, para tanto, faz uso extensivo de processos relacionais circunstanciais (destacados em negrito) que permitem identificar a localização das partes envolvidas no processo de mapeamento dos nervos, relacionando-as com as partes adjacentes. No trecho ainda estão destacados em itálico outros processos relacionais (intensivos e possessivos) que foram empregados com maior parcimônia, tendo em vista o próprio objetivo de descrever o mapeamento dos nervos (que em seguida, no texto, foram ilustrados com figuras). Além desse propósito, o emprego de processos materiais (tais como “foram feitos”, “foram cortados” e “foram separados”) serviram para detalhar os procedimentos metodológicos realizados para o mapeamento. Cabe notar ainda a variedade de verbos lexicais identificativos utilizados na descrição, em grande parte, de forma metafórica – tais quais “emergir”, “descer”, “passar” e “correr”, permitindo ao leitor localizar as partes descritas, acompanhando-as mentalmente conforme avança a descrição da imagem e, concomitantemente, amplia o conhecimento do leitor sobre a área reportada.

Por último, no que concerne as orações relacionais **possessivas**, essas expressam a posse de uma das entidades em relação à outra, englobando significados de posse (e.g. 77) e composição (e.g. 78) – ou seja, a posse é interpretada em um sentido amplo, de *extensão*, em que uma entidade é construída como extensão de outra. Há também a distinção sistêmica contrastiva entre ‘possessiva enquanto participante’ e ‘possessiva enquanto processo’, sendo tipicamente realizado por verbos como *ter*, *possuir*, *envolver*, *pertencer* e a expressão *ser de* (FUZER e CABRAL, 2014)

(77) A cobertura vegetal nativa **pertence** ao Domínio da Floresta Atlântica (Costa *et al.*, 2013). (A_FL_2017_0_AA_019)

(78) Mais duas tiras foram obtidas das laterais do disco que **contém** a columela central, totalizando oito tiras com dimensões semelhantes (A_BJFT_2017_0_AA_008)

Assim como as intensivas e as circunstanciais, as orações relacionais possessivas podem ser construídas de modo atributivo ou identificativo. No modo atributivo, a relação possessiva é realizada pelo Atributo ou pelo processo. No primeiro caso, o Atributo realiza-se por um grupo nominal possessivo, em que a coisa possuída é o Portador e o possuidor é o Atributo, como pode ser observado nos trechos (79) e (80), a seguir. No segundo caso, o Processo é o responsável pela relação de posse de modo que tanto o Atributo quanto o Portador podem ocupar o papel de Possuidor/Possuído, conforme apresentado nos trechos (81) e (82).

(79) “50% da culpa do patriarcado [Portador: Possuído] é das mulheres [Atributo: Possuidor]” (L_RBLA_2014_0_AA_002)

(80) A pesquisa que corrobora com o aqui encontrado [Portador: Possuído] é dos autores Moreira *et al.* (2011) [Atributo: Possuidor] (En_GP_2017_1_AA_010)

(81) Ao longo da dinâmica espaço-temporal, as paisagens [Portador: Possuidor] **têm** seus elementos modificados e transformados [Atributo: Possuído]. (A_FL_2017_0_AA_013)

(82) O solo da área experimental [Portador: Possuído] **pertence** à unidade de mapeamento São Pedro [Atributo: Possuidor]. (A_ABMVZ_2018_1_AA_023)

No modo identificativo, a posse toma forma de relação entre as entidades, podendo ocorrer como uma característica dos Participantes ou do Processo. Sendo uma propriedade dos Participantes, um significará o dono/Possuidor e outro, o objeto possuído. Entre as entidades o relacionamento é de identidade, em que *um possui o outro* – aqui cabe mencionar que o *ter* não

é utilizado como um verbo identificativo possessivo. Conforme observável nos exemplos a seguir, os termos das orações identificativas possessivas são reversíveis, quando realizados por verbos lexicais, a forma fica na passiva (como em 83'). Além dessa propriedade, os verbos das orações relacionais possessivas identificativas podem expressar contenção, envolvimento ou benefício (como em 84, em que o Possuidor '*possui* algo e o oferece'). Alguns verbos podem combinar a característica semântica de posse com outros aspectos, por exemplo, *faltar* significa 'não ter', e *requerer* e *merecer*, 'precisa ter', como em (85).

- (83) Os dados [Identificado/ Possuidor] **incluíam** [processo relacional identificativo possessivo] o número de passos por hora e a quantidade de leite diária produzida [Identificador/Possuído]. (A_PBV_2013_0_AA_006)
- (83') O número de passos por hora e a quantidade de leite diária produzida foram incluídas nos dados.
- (84) A avaliação multicritério [Identificado/Possuidor] **fornece** [processo relacional identificativo possessivo] um conjunto de técnicas e procedimentos [Identificador/Possuído] para estruturar os problemas de decisão (E_BCG_2016_3_AA_004)
- (84') Um conjunto de técnicas e procedimentos foi fornecido pela avaliação multicritério.
- (85) Os métodos químicos *são baseados* em reações e utilizam precursores moleculares ou atômicos para a obtenção das nanopartículas desejadas. *São* bem difundidos por se *tratarem* de técnicas mais eficazes e possibilitarem maior controle sobre o processo, tais como deposição de vapor químico (CVD) e crescimento epitaxial. No entanto, essas técnicas **requerem** [processo relacional identificativo possessivo] intensivo condições de crescimento severas, como elevada temperatura e alto vácuo, **possuem** [processo relacional identificativo possessivo] limitação na produção de nanolamelas em quantidade, 1,5, 11, 14 **necessitam** [processo relacional identificativo possessivo] de reagentes químicos e a diferença térmica entre o substrato (superfície aquecida onde o filme é depositado) e o filme favorece a presença de falhas e de impurezas devido ao processamento, 8, 15 aumentando a complexidade e o custo dos processos. (E_QN_2017_9_AA_005)

Vale mencionar ainda que os tipos de orações relacionais intensivo e possessivo são ágnatos, um com o outro, pois oferecem modos alternativos de construir um enunciado com pouca diferença aparente entre eles (HALLIDAY e MATTHIESSEN, 2000), conforme (86) e (87) a seguir e suas reescritas.

- (86) A fase móvel **consistiu** [processo relacional identificativo possessivo] de água + ácido fosfórico 0,1% e metanol. (A_BJFT_2017_0_AA_038)
- (86') A fase móvel **foi** [processo relacional identificativo intensivo] água + ácido fosfórico 0,1% e metanol.
- (87) O material empírico analisado **consistiu** [processo relacional identificativo possessivo] de 28 redações elaboradas por alunos da disciplina de Gestão de Marketing, (So_OS_2017_1_AA_008)

- (87) O material empírico analisado **foi** [processo relacional identificativo intensivo] 28 redações elaboradas por alunos.

3.3.4 Orações verbais

As orações verbais expressam uma ação física que reflete operações ocorridas na mente. Como um tipo de processo secundário, vários aspectos dos processos primários tangentes – mentais e relacionais – são compartilhados com os verbais, conforme descrito mais adiante. Apesar disso, os verbais podem ser distinguidos por sua capacidade de veicular mensagens através do ato da linguagem. Acompanhando o processo, há os participantes Dizente, Verbiagem, Receptor e Alvo. O **Dizente** se responsabiliza pela realização externa do processo verbal; tal realização pode ocorrer por meio da **Verbiagem**, que é a mensagem resumida em um grupo nominal, parte do processo. A Verbiagem pode ser tanto o conteúdo do que foi dito (e.g. “nova opções terapêuticas” em 88), quanto o nome do que foi dito (e.g. “desculpas” em 89). Já o **Receptor** é aquele com quem o Dizente está interagindo e o **Alvo** é o participante sobre quem (ou o que) se fala, não necessariamente humano.

- (88) Stanojevic *et al.* (2011) [Dizente], que **afirmaram** a ocorrência de desnaturação proteica [Verbiagem] durante o armazenamento de grãos. (A_BJFT_2018_0_AA_005)
- (89) Dois hospitais, H2 e H3, [Dizente] **disseram** realizar pré-tratamento obrigatório do subgrupo A1 [Oração projetada: Retato]. (En_ESA_2014_0_AA_004)
- (90) Ao verificar uma passagem anterior em outro instituto daquele complexo hospitalar, ela [Dizente] **pediu** desculpas [Verbiagem] e **perguntou** se ele desejava registrar seu nome social [Oração projetada: Relato]. (H_CP_2017_1_AA_003)

Como exemplificado em (89), diferente do Experienciador das orações mentais, o Dizente não requer um participante consciente necessariamente, podendo ser qualquer entidade que o represente. No trecho, “Dois hospitais” é o Dizente que representa aquilo que foi dito pelos funcionários responsáveis por expressar o posicionamento da diretoria dos hospitais. Desse modo, como dito anteriormente na seção 3.3.3, a natureza desse participante pode ser melhor denominada por “processo simbólico” (HALLIDAY e MATTHIESSEN, 2014).

Semelhante aos processos mentais, os verbais podem estar acompanhados por circunstâncias de Assunto (mas não somente, visto que os variados tipos de Circunstâncias

podem ocorrer com os diferentes tipos de processos), como exemplificado em (91) e (92), que apresentam um resumo da mensagem por meio de uma construção preposicionada (THOMPSON, 2014).

- (91) Destaca-se a atuação do parlamentar Elimar Máximo Damasceno [Dizente], que **proferiu** oito discursos [Verbiagem] com a palavra “aborto” e um com a palavra “nascituro”. Em 2004, houve 35 deputados [Dizente] que **discursaram** sobre o aborto [Circunstância: Assunto]. (H_RBCP_2014_0_AA_001)
- (92) Curiosamente, Klein [Dizente] **falava** com frequência [Circunstância: Modo] do pigmento bruto [Circunstância: Assunto] em termos vitalistas semelhantes ao discurso organicista de Oiticica. (L_ASP_2017_2_AA_013)

Os processos verbais também podem projetar a mensagem em orações à parte da verbal. Nesse caso, conforme ocorre com os processos mentais, a sentença projetada não é analisada como participante do processo e, portanto, distingue-se da Verbiagem. Em alguns casos, como discursos indiretos, a oração projetada pode parecer parte da oração verbal, entretanto, cabe ressaltar, conforme Thompson (2014, p. 204), que a projeção traz uma fala que não a do emissor para a interlocução, ao mesmo tempo em que difere da emissão original, pois foi incorporada a uma nova oração, em um contexto distinto de onde proveio.

Em vista disso, segundo Eggins (2004, p. 236), as orações projetadas pelos processos verbais podem ser de dois tipos: Citação ou Relato¹⁷. A **Citação** introduz uma fala externa como emissora do conteúdo, sem mudanças pelo produtor do texto, havendo a possibilidade de alguma marcação gráfica (como aspas ou travessão) introduzindo-a na oração projetante. O **Relato** também é uma oração na qual atribui-se o conteúdo à fala externa, entretanto, não necessariamente com as mesmas palavras ou estrutura da original, podendo ser uma síntese pelo produtor do texto daquilo que foi dito. Em geral, o Relato é introduzido pelas conjunções *que* ou *se*, ou por uma oração não-finita. Os trechos (93) e (94) a seguir exemplificam os processos verbais acompanhados por seus participantes ou orações projetadas.

- (93) O telefone tocava e uma voz desconhecida [Dizente] **anunciava** “vi sua menina aqui, na comunidade tal” [Projeção: Citação]. Tereza então montava na moto e seguia até o local indicado. “Nada era perto do rapaz” [Projeção: Citação], **disse-me** [Receptor] Mariana [Dizente] (H_CP_2_107_2_DO_005)

¹⁷ Termos traduzidos de acordo com Fuzer e Cabral (2014, p. 74)

- (94) Gabriel e Mariana [Dizente] nos [Receptor] **contaram** que, apesar de serem os responsáveis pela equipe, não convivem diariamente com os operadores – trabalham em salas separadas [Oração projetada: Relato]. (H_CP_2017_1_AA_004)

Segundo Halliday e Matthiessen (2014), os processos verbais podem ser divididos em dois tipos: de **atividade** (subdividido em Alvo e Fala) e de **semiose** (subdividido em Neutro, Indicação e Comando). Fuzer e Cabral (2014, p. 72) apresentam um quadro exemplificando os grupos verbais que realizam esses tipos em PB, a partir de um quadro similar de Halliday e Matthiessen (2004, p. 255) da língua inglesa. Abaixo, o Quadro 16 foi adaptado com exemplos de cada tipo de oração com processos verbais.

Quadro 18 - Exemplos de verbos e orações verbais

Tipo		Exemplos de verbos	Exemplos de orações verbais
Atividade	Alvo	Acusar, caluniar, criticar, culpar, difamar, denunciar, elogiar, injuriar, insultar, lisonjear, repreender, xingar	Segundo relatos e observação direta, alguns clientes [Dizente] já elogiaram a nova postura quanto a prazos [Verbiagem]. (En_GP_2018_2_AA_012)
	Fala	Conversar, falar	A cena era de dois adolescentes parados em uma esquina, uma sugestiva encruzilhada, onde os meninos [Dizente] conversavam sobre o valor pago na moto que aparecia na imagem [Circunstância: Assunto]. (H_RSO_2014_0_AA_004)
Semiose	Neutro	Contar, dizer	Muitos [Dizente] disseram que já haviam participado de outras pesquisas [Oração projetada: Relato], mas que ninguém voltava. (So_OS_2018_2_AA_007)
	Indicação	Anunciar, contar (algo a alguém), convencer (alguém de algo), explicar, informar, provar, relatar, persuadir (alguém de algo), prometer (algo a alguém)	Dahl <i>et al.</i> 20 [Dizente] provaram que o bloqueio nervoso epidural antes da incisão na pele apenas com anestesia local não altera de modo significativo a concentração plasmática de hormônios do estresse

			[Oração projetada: Relato]. (Sa_RBAN_2014_0_AA_001)
		Perguntar (a alguém se), interrogar, indagar(-se)	Ao longo da sessão de pré-teste, o experimentador [Dizente] também perguntava se a criança estava conseguindo compreender os problemas escritos [Oração projetada: Relato]. (H_PTP_2013_0_AA_004)
	Comando	Ameaçar (alguém de algo), convencer (alguém a pensar ou fazer algo), exigir, implorar, mandar, pedir (para alguém fazer algo), ordenar, persuadir (alguém a fazer algo), prometer (algo a alguém), rogar, solicitar, suplicar.	De acordo com Galizia (2005), Wilson [Dizente] pedia <u>para que os atores</u> [Receptor] <u>olhassem para o pensamento com</u> <u>o objetivo de evitar a representação</u> [Oração projetada: Relato] (L_RBEP_2017_1_AA_008)

Fonte: Adaptado de Halliday e Matthiessen (2014, p. 305) e Fuzer e Cabral (2014, p. 72) com exemplos do *corpus* da pesquisa.

Por último, cabe ainda observar que os processos verbais não apenas representam a linguagem realizada, mas também possuem papel relevante na construção de textos acadêmicos. Segundo Halliday e Matthiessen (2014), é por meio das orações verbais que é possível citar e relatar diferentes estudos e pesquisas ao mesmo tempo em que se indica a posição desses em relação ao que foi feito, utilizando-se de verbos como *afirmar*, *discutir* e *apontar*, por exemplo.

3.3.5 Orações comportamentais

Os processos comportamentais compartilham características com os processos centrais, podendo ser compreendidos como subcategoria em vez de um tipo distinto de processo, segundo Thompson (2014, p. 109). De acordo com Halliday e Matthiessen (2014), os

comportamentais são o tipo de processo menos distinto, pois não possuem propriedades definidoras próprias, mas as compartilham com os mentais e materiais. Em geral, se relacionam especificamente à fisiologia (tipicamente) humana por meio de verbos como *tossir*, *sonhar*, *respirar* e *olhar*. Envolvidos na oração, o Participante obrigatório é o **Comportante**, tipicamente consciente, similar ao Experienciador dos processos mentais. Concomitantemente, há o **Comportamento**, que se assemelha à natureza do Escopo-processo das orações materiais, realizando-se através de grupos verbais como “dar uma risada” ou “dar um bocejo” (FUZER e CABRAL, 2014, p. 78). Os trechos a seguir ilustram as orações comportamentais.

- (95) Para Caminha, os portugueses [Comportante] **riram** dos índios; para Laerte, os índios [Comportante] é que **riem** dos portugueses. (L_RBLA_2014_0_AA_003)
- (96) **Circulava** os olhos [Comportamento] sem medo nem ansiedade; não sei até se com uma pontinha de riso nos cantos da boca. (So_RDGV_2017_3_AA_004)
- (97) Mensurada a atenção por movimentos do olho, americanos [Comportante] **olham** fixamente [Comportamento] nos objetos focais mais tempo que os asiáticos do Extremo Oriente, enquanto estes [Comportante] **olham** fixamente [Comportamento] para o pano de fundo mais que os americanos. (So_RDGV_2018_1_AA_002)
- (98) Uma vez, quando a terra estava muito seca, eu [Comportante] **tossi** por semanas (L_RBEP_2018_3_AA_005)

Assim, segundo Thompson (2014), a presente categoria surge da necessidade de distinguir entre os processos mentais e os sinais físicos desses processos, separando, por exemplo *ver* (enquanto processo mental perceptivo) e *olhar* (enquanto processo comportamental, exemplificado em 97). Além disso, costumam ser sempre construções médias (c.f. seção 3.4), consistindo geralmente do Comportante, do processo e de uma Circunstância de Modo ou Assunto, como exemplificado em (99).

- (99) Marx defende que o homem [Comportante] produz a religião⁶, **sonha** com um mundo fantasioso [Circunstância: Assunto] (H_TFA_2017_4_AA_003)

Os processos comportamentais ainda se aproximam do verbais, especialmente na narrativa de ficção, introduzindo Citações ou Relatos e compartilhando seu traço comportamental para a oração verbal, conforme o exemplo (100) a seguir.

- (100) O Morsa [Dizente] **choramingou** que, para variar, não tinha sido avisado sobre campeonato nenhum [Oração projetada: Relato].¹⁸

Por último, cabe ainda observar que os processos comportamentais não são frequentes no *corpus* da presente pesquisa. Resultados similares foram vistos em Mendes (2013), que obteve 0,01% de ocorrências desse processo em um *corpus* de artigos científicos da área de Administração e Engenharia, totalizando cerca de 1,3 milhões de palavras. Zhen, Yang e Ge (2014) analisaram 6.158 orações em um *corpus* de 25 artigos científicos da área de Medicina, a partir da perspectiva da LSF, e encontraram apenas 10 ocorrência de processos comportamentais, equivalendo a 0,16% do total. Um outro estudo, de Vathanalaoha e Tangkiengsirisin (2018), também chegou a resultados semelhantes ao analisar o sistema de transitividade em um *corpus* de 120 resumos de artigos científicos da área de Odontologia em língua inglesa. Na pesquisa, que classificou todas as orações dos textos, não foi encontrado nenhum processo comportamental.

3.3.6 Orações existenciais

Por último, os processos existenciais expressam a existência de uma entidade sem predicar algo além, sendo seu único participante obrigatório o **Existente**. Maiores detalhes sobre esse participante podem ser acrescentados por meio das circunstâncias, conforme demonstram os exemplos (101) e (102), a seguir:

- (101) O maior acúmulo de serapilheira em ambas florestas [Existente] **aconteceu** em locais com maior quantidade de indivíduos do estrato de regeneração natural [Circunstância]. (B_R_2018_2_AA_012)
- (102) Apesar dos cuidados metodológicos para garantir a participação no estudo, **houveram** [Processo existencial] recusas [Existente] nesta investigação [Circunstância], semelhante a outras pesquisas realizadas com adolescentes por meio de questionários. (Sa_BJPT_2014_0_AA_003)

De acordo com o ilustrado na Figura 7 da seção 3.3, entre os processos existenciais e os materiais estão localizados aqueles relacionados ao clima: processos meteorológicos. Nesse âmbito, as estruturas oracionais determinam o tipo de processo realizado pela oração, conforme

¹⁸ GALERA, Daniel. Mãos de cavalo. Companhia das Letras: São Paulo, 2006.

ilustrado nos trechos (103) a (105), nos quais (103) e (105) configuram processos existenciais, sendo ambos tipicamente existencial (*haver*). Nesses trechos, o Existente é respectivamente “uma tempestade...” e “precipitação”, expressando fenômenos meteorológicos tais qual aquele expresso pelo processo *chover* em (103) e por “tornou-se a chover”, em (105). No último caso, o processo aparece conjugado, representando uma ação e, conseqüentemente, caracterizando um processo material meteorológico.

- (103) **Havia** uma tempestade eletricamente ativa se desenvolvendo nas proximidades do trecho Foz do Iguaçu - Ivaiporã da LT 765 kV (E_BCG_2016_4_AA_001)
- (104) [...] em menos de uma hora, **choveu** 35,4 milímetros em Campina Grande (En_ESA_2017_5_AA_008)
- (105) Em Manaus (Fig. 2c) observou-se um período de 11 dias (13-23) em que não **houve** precipitação e a partir do dia 24 (dias com chuva e período atuante da friagem) **tornou-se a chover**. (E_RBM_2018_1_AA_002)

Assim como com qualquer outra escolha léxico-gramatical, a distinção entre as orações existenciais e materiais pode ser relacionada ao efeito discursivo que cada uma ocasiona. Por exemplo, a completa omissão de um Agente/Ator responsável pela ação nas orações existenciais permite que o próprio autor, ao empregá-las, omita sua participação ou presença no texto. Em nível textual, para Thompson (2014), Halliday e Matthiessen (2014) e Franchi, Negrão e Viotti (1998), as orações existenciais se associam à introdução de elementos no discurso, sendo chamadas *construções apresentativas*, considerando que sua escolha tende a introduzir um (novo) assunto no texto, conforme o trecho (106). Já sua reescrita, em uma oração material, descompacta as nominalizações (“ampliação” e “redefinição”) no Existente de (106’), transformando a oração em uma configuração distinta.

- (106) Já em 2013, **há** uma ampliação e redefinição do processo transexualizador por parte do Ministério da Saúde a partir da Portaria GM n. 2.803, de 13 de novembro. (H_CP_2018_1_AA_001)
- (106’) Já em 2013, o Ministério da Saúde ampliou e redefiniu o processo transexualizador a partir da Portaria GM n.2.803, de 13 de novembro.

Como pode ser observado, os processos existenciais são prototipicamente realizados pelos verbos *haver* ou *existir*, que não apresentam sujeito. Apesar disso, outros verbos

tipicamente associados a diferentes processos podem realizar orações existenciais, de acordo com seu emprego.

Recapitulando a discussão deste capítulo até o momento, o Quadro 19 apresenta os diferentes processos realizados pelo sistema da TRANSITIVIDADE, assim como seus participantes, circunscritos pela metafunção experiencial, pertencente à variável campo do contexto de situação.

Quadro 19 – Processos, significados e participantes envolvidos

TIPOS DE PROCESSO	Significado da categoria	Participantes diretamente envolvidos	Participantes indiretamente envolvidos
material: transformativo criativo	‘fazer’ ‘acontecer’	Ator, Meta	Escopo Beneficiário Recebedor Cliente Atributo
comportamental	‘comportar’	Comportante	Comportamento
mental: percepção cognição desiderativo emotivo	‘sentir’ ‘perceber’ ‘pensar’ ‘desejar’ ‘sentir’	Experienciador, Fenômeno	Indutor
verbal	‘dizer’	Dizente, Alvo	Recebedor, Verbiagem
relacional: atributivo identificativo	‘ser’ ‘atribuir’ ‘identificar’	Portador, Atributo Identificado, Identificador	Atributor; Beneficiário Designador
existencial	‘existir’	Existente	

Fonte: Adaptado de Halliday e Matthiessen (2014, p. 311)

3.3.7 Circunstâncias

Como observado nos trechos apresentados até o momento, as circunstâncias podem ocorrer com todos os tipos de processos, sendo realizadas por grupos adverbiais ou grupos

preposicionados. Em geral, as circunstâncias codificam o pano de fundo contra o qual o processo se desenvolve (THOMPSON, 2014, p. 114). Assim, conforme o autor, e ao lado de Halliday e Matthiessen (2014, p. 314), existem algumas categorias em que as circunstâncias podem ser classificadas, como Tempo, Lugar e Modo. Cabe ressaltar, entretanto, que tais categorias não são muito distintas e frequentemente se sobrepõem umas às outras, combinando diferentes tipos de significado. Por exemplo, na oração abaixo (THOMPSON, 2014, p. 215), Tempo e Causa se misturam:

- (107) Se meu parecer particular tivesse prevalecido, eu teria feito o que não desejava e **então** não seria livre. (H_TFA_2018_1_AA_004) → (= ‘depois disso’ + ‘por causa disso’)

Além disso, circunstâncias podem, e são, frequentemente empregadas com sentidos metafóricos, como abaixo (THOMPSON, 2014, p. 215):

- (108) Articulada à imaginação, a memória se transforma em peça fundamental no trabalho de leitura. “Isso é de especial importância nos textos literários”, acrescenta o teórico, “pelo fato de que eles não correspondem a nenhuma realidade objetiva **para fora** deles mesmos” (1990, p.276). (H_AETP_2014_0_AA_003) (Circunstância Localização: Lugar)
- (109) Sem a mediação do sexual, isto é, do contingente (plano no qual o luto se sustenta para reelaborar sua perda), o real transborda **para fora** da compacidade da falta de objeto (H_AETP_2014_0_AA_004) (Circunstância Localização: Lugar)

Segundo Halliday e Matthiessen (2014, p. 311), através da perspectiva trinocular da linguagem, as circunstâncias podem ser compreendidas *de cima* como acompanhamento do processo, referindo-se ao “por que, onde, como, quando” aconteceu algo. Na perspectiva da sentença, *ao redor*, as circunstâncias funcionam como Adjuntos, ou seja, não podem ocupar lugar de Sujeito, tampouco são modais, se analisadas pela perspectiva da metafunção interpessoal. *De baixo*, as circunstâncias são tipicamente expressas como grupos adverbiais ou sentenças preposicionadas.

Ainda segundo os autores (2014, p. 313-314), assim como Thompson (2014), ao escolher um conjunto de categorias para classificar as circunstâncias, é necessário ter em mente que não será capaz de dar conta de todas possibilidades de uma língua – vide os exemplos (107) a (109) anteriores, em que se mesclam categorias e são usados metaforicamente. Assim, um modo de explorar e identificar possíveis classificações é perguntando diferentes questões para que a própria circunstância forneça respostas. Com base nessas perguntas, os autores

apresentam nove classes principais para as circunstâncias, listadas e exemplificadas no Quadro 20 a seguir, cada uma com suas subclassificações e todas organizadas em quatro grandes grupos: três deles derivados da classificação dos processos relacionais de expansão (intensificação, extensão e elaboração) e um proveniente dos processos mentais e verbais (projeção).

Quadro 20 – Tipos e exemplos de circunstâncias

Tipo		Pergunta	Exemplo
Intensificação	1 Extensão	Distância	A que distância? Consequentemente, o Espírito Santo teve o seu desenvolvimento restrito <i>a uma estreita faixa de menos de três léguas (pouco menos de 15 km)</i> da costa (B_R_2017_5_AA_001)
		Duração	Quanto tempo? Os participantes foram orientados a caminhar <i>por dois minutos</i> para se adaptar à esteira (As_RPP_2018_3_AA_008)
		Frequência	Quantas vezes? A cavidade medular é preenchida, <i>muitas vezes</i> por completo, por tecido ósseo (A_ABMVZ_2013_0_AA_003)
	2 Localização	Lugar	Onde? Brasileiros só conduziam o que podiam carregar, sobrevivendo aos desafios de abrir caminhos <i>por entre matas virgens</i> (So_SSSO_2014_0_AA_003)
		Tempo	Quando? Os parâmetros temperatura, pH e condutividade elétrica das amostras foram registrados <i>logo após a coleta</i> (E_QN_2017_5_AA_007)
	3 Modo	Meio	Como? Com o quê? a função social da posse deve ser reconhecida como um caminho <i>por meio do qual seria possível garantir a legitimidade (leia-se: a justiça) do apossamento</i> (So_RDGV_2017_3_AA_002)
		Qualidade	Como? Após o descanso, as amostras foram centrifugadas na rotação de 12.000g por 15 minutos e descartou-se <i>calmamente</i> o sobrenadante. (A_ABMVZ_2018_AA_011)
		Comparação	Como é? Com que parece? <i>Diferentemente</i> de proteínas, lipídios e carboidratos, os genes não nasceram de modo algum como entidades físicas (H_TFA_2018_3_AA_008)
		Grau	Quanto? A região da cidade de Passo Fundo não é <i>intensamente</i> industrializada (En_ESA_2017_3_AA_008)
	4 Causa	Razão	Por quê? 60 meninos foram excluídos <i>por causa da baixa qualidade do sinal</i> (Sa_RPP_2017_3_AA_009)
		Finalidade	Para quê? Essa privação se justifica <i>pela expectativa de preservação dos bens, direitos e valores</i> , de modo que é desarrazoado que tais ativos (So_RDGV_2017_2_AA_003)
		Benefício	Por quem? os evangélicos que têm seguido a trajetória eleitoral não são unívocos nos posicionamentos, mas de forma geral trabalham <i>a favor das pautas moralizadoras</i> . (H_CP_2017_2_DO_002)

	5 Contingência	Condição	Por quê?	<i>Em caso de dúvidas, a sugestão é sempre buscar a resposta na fonte mais atualizada. (E_QN_2017_3_EC_003)</i>
		Falta/Omissão		<i>na falta de ensaios para determinação deste valor, pode ser adotada a seguinte equação para verificação de estado limite de deformação excessiva (En_RIEM_2017_2_AA_007)</i>
		Concessão		<i>É possível afirmar que, a despeito das diferenças existentes em cada alternativa proposta e das possíveis lacunas deixadas por cada uma delas, há um elemento comum no que se relaciona à necessidade de autonomia (En_GP_2017_2_AA_003)</i>
Extensão	6 Acompanhamento	Companhia	Com quem? Com o quê?	<i>A área é caracterizada por grandes zonas de mata nativa <i>junto dos morros</i>, como pode ser observado na vista panorâmica da Figura 7. (So_CAME_2014_AA_001)</i>
		Adição	Quem mais? O que mais?	<i>Além das perdas econômicas, essa espécie bacteriana representa potencial risco à saúde pública (A_AIB_2013_0_AA_002)</i>
Elaboração	7 Papel	Estilo	Ser como o quê?	<i>Quando em 1985 os bispos elegeram <i>como presidente do EPS</i> um prelado de perfil claramente contrário ao oficialismo (H_RCBP_2014_0_AA_005)</i>
		Produto	O quê/em que?	<i>As cenouras foram lavadas, descascadas e cortadas <i>em tiras</i> com tamanho de aproximadamente 1 cm (A_FST_2011_0_AA_007)</i>
Projeção	8 Assunto		Sobre o quê?	<i>Frigaard [8] discorre a respeito dos resultados obtidos após a soldagem de ligas de alumínio (En_SI_2018_2_AA_008)</i>
	9 Ângulo	Fonte		<i>Para as análises, utilizaram-se procedimentos <i>de acordo com American Public Health Association (APHA)</i>, descrita no Compendium of Methods for the Microbiological Examination of Foods (B_RAA_2018_2_AA_002)</i>
		Ponto de vista		<i>os processos de organizar são marcados pela formalidade e parecem ser “lentos” e demorados <i>aos olhos dos moradores</i>. (So_OS_2019_1_AA_004)</i>

Fonte: Adaptado de Halliday e Matthiessen (2014, p. 313) e Fuzer e Cabral (2014, p. 53-54), com exemplos retirados do *corpus* desta pesquisa.

O grupo de circunstâncias de **intensificação** aumenta a configuração dos processos e seus participantes através da especificação de extensão ou localização em tempo ou espaço no decorrer do processo, da maneira como esse se desenvolve ou de sua contingência. Já o grupo de **extensão** amplia a relação entre processo e participantes através da especificação do elemento que se posiciona em relação extensiva de um dos participantes quanto ao processo. Assim, tal elemento vai desde co-participante a um apêndice dele. Por sua vez, o grupo de **elaboração** expande a configuração do processo e participantes através do papel de um dos participantes. Por fim, o grupo de **projeção** se associa aos processos verbais, relacionado ao participante Verbiagem, e também aos mentais, relacionado àquilo que um participante pensa.

Além da possibilidade das Circunstâncias serem empregadas com sentido metafórico ou abstrato, conforme mencionado anteriormente, Halliday e Matthiessen (2014) elencam outras características que podem dificultar a identificação de elementos circunstanciais, como ter uma frase preposicionada como participante da oração; ter uma preposição ligada ao verbo (que implica que o grupo preposicionado possa ser também um participante), funcionando como parte do processo; ocorrer uma frase preposicionada (como Qualificador) dentro de um grupo nominal; ou uma frase preposicional ocorrer como Modal ou Adjunto Conjuntivo.

Percorrido o caminho teórico entre os diferentes processos, participantes e circunstâncias realizados pelo sistema da TRANSITIVIDADE, a próxima seção introduz um modelo alternativo e complementar a esse, o modelo da ERGATIVIDADE.

3.4 O modelo da ergatividade: breve introdução

Nas seções anteriores foram discutidos o sistema de TIPOS DE PROCESSOS, os participantes e as circunstâncias que podem acompanhá-los, todos pertencentes ao sistema da TRANSITIVIDADE. Segundo Halliday e Matthiessen (2014, p. 333), embora cada um dos processos descritos nas seções anteriores tenha sido distinguido e classificado, eles também se aproximam, pois, comparando-os de um outro ponto de vista, todos são equivalentes – isto é, em um outro nível interpretativo, todos possuem uma mesma estrutura gramatical. A esse outro olhar, denomina-se **modelo ergativo**, que é considerado uma propriedade dentro do SISTEMA DA TRANSITIVIDADE, ao lado do modelo transitivo, de forma que ambos são considerados complementares, duas perspectivas modelando o SISTEMA DE TRANSITIVIDADE, conforme resumido no Quadro 21.

Quadro 21 – Modelos complementares do sistema da transitividade

	Modelo Transitivo	Modelo Ergativo
Generalizado (ao longo dos tipos de processos)		Processo + Meio (\pm Agente) [Médio/Efetivo]
Particularizado (para cada tipo de processo)	<p>Material: Ator + Processo \pm Meta [intransitivo/ transitivo]</p> <p>Mental: Experienciador + Processo + Fenômeno</p> <p>Relacional: Portador + Processo + Atributo; Identificado + Processo + Identificador</p> <p>Verbal: Dizente + Processo (\pm Receptor)</p> <p>Comportamental: Comportante + Processo</p> <p>Existencial: Existente + Processo</p>	

Fonte: Adaptado de Halliday e Matthiessen (2014, p. 334)

O modelo da ergatividade sugere uma generalização dos diferentes tipos de processos do modelo da transitividade ao trabalhar com apenas um processo (sem epíteto) junto com seu participante Meio. Segundo Thompson (2014, p. 139), ambos os modelos, transitivo e ergativo, surgiram ao serem observados verbos que sistematicamente alternavam seus usos entre dois padrões, conforme representado no Quadro 22 a seguir, cujas sentenças foram classificadas de acordo com o modelo da transitividade.

Quadro 22 – Orações classificadas de acordo com o modelo da transitividade

(a) O fornecimento de propilenoglicol (PG)	aumenta	a concentração de propionato no rúmen,	de forma indireta (A_ABMVZ_2018_3_A A_032)
Ator	Processo material	Meta	Circunstância

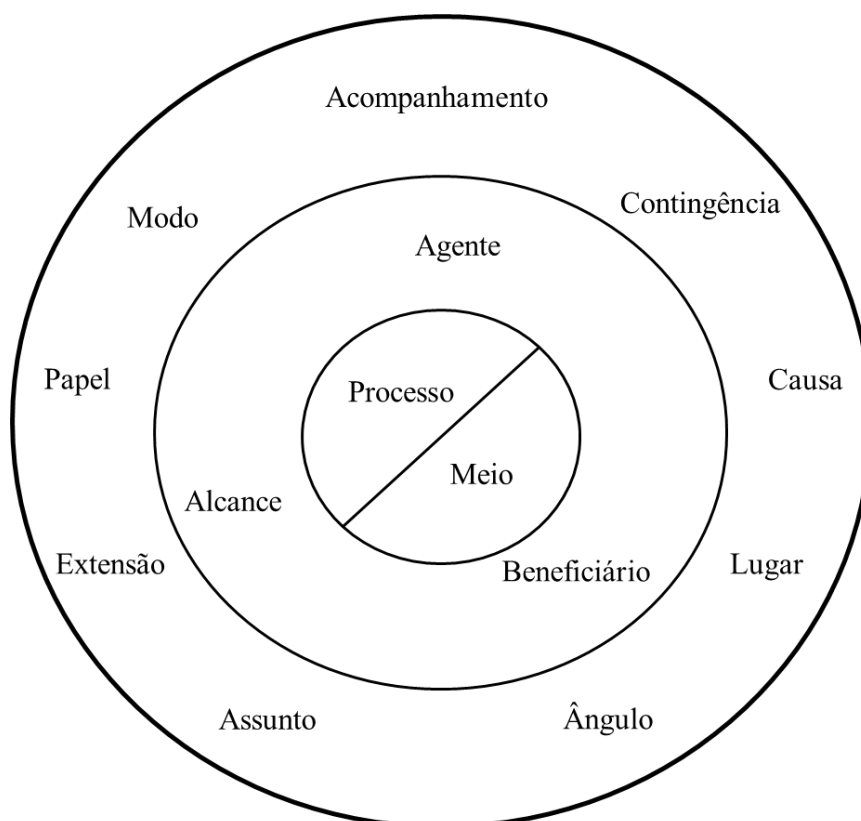
(a') A concentração de propionato no rúmen	aumenta	de forma indireta
Ator	Processo material	Circunstância

Fonte: Elaboração própria.

Em ambos os trechos, o modelo da transitividade obscurece o fato de que o processo afetou o participante “a concentração de propionato no rúmen” nas duas sentenças. Caso a reescrita fosse uma forma passiva (*A concentração de propionato no rúmen foi aumentada*), a Meta mudaria apenas a função sintática para Sujeito, mas continuaria sendo Meta. Entretanto, na segunda oração, (a'), a Meta não apenas mudou de função sintática, mas também sua função dentro do sistema da transitividade para se tornar Ator.

Posto isso, no modelo da ergatividade, Halliday e Matthiessen (2014, p. 336) sugerem uma outra perspectiva, generalizadora, com Processo e Meio centrais à oração e outros participantes mais periféricos, apresentados na Figura 9 a seguir. Nessa perspectiva, o **Meio** é o termo obrigatório pelo qual o processo se concretiza enquanto o **Agente** é a causa externa da ação. Semanticamente, Agente, Beneficiário e Alcance possuem características de participantes e de algumas circunstâncias, isso reflete-se no fato de que são gramaticalmente mistos, podendo interagir na oração diretamente como grupos nominais (similar aos participantes) ou indiretamente como frases preposicionadas (similar às circunstâncias) (op. cit., p. 348). Os demais elementos oracionais (no círculo externo da Figura 9) são representados como as circunstâncias, compostas por grupos adverbiais ou frases preposicionadas.

Figura 9 – Relação entre processo, participantes e circunstâncias no modelo da ergatividade



Fonte: Adaptado de Halliday e Matthiessen (2014, p. 348)

As orações ainda podem ser classificadas quanto à Agentividade: quando a figura construída possui Agente, ela é considerada outra-agentiva (*other-agentive*), e quando construída sem um Agente, ela é auto-agentiva (*self-agentive*) (HALLIDAY e MATTHIESSEN, 2000). Os trechos do Quadro 22, apresentado anteriormente, podem ser classificados da seguinte maneira conforme o modelo da ergatividade, no Quadro 23:

Quadro 23 - Orações classificadas de acordo com o modelo da ergatividade

Oração outra- agentiva	(a) O fornecimento de propilenoglicol (PG)	aumenta	a concentração de propionato no rúmen,	de forma indireta (A_ABMVZ_201 8_3_AA_032)
	Agente	Processo	Meio	Circunstância
	(a'') A concentração de	aumenta	de forma indireta	

Oração auto- agentiva	propionato no rúmen		
	Meio	Processo	Circunstância

Fonte: Elaboração própria.

Semanticamente, o Meio (nos exemplos do Quadro 23, “A concentração do propionato no rúmen”) é o participante pelo qual realiza-se o processo, de modo que a combinação Meio + processo associa um desdobramento temporal (processo) a uma persistência atemporal (o Meio) (HALLIDAY e MATTHIESSEN, 2000). Assim, aos participantes Meio e Agente podem ser atribuídos os diferentes traços dos tipos de processos do modelo particularizado – um ou mais para cada tipo de figura realizada. Por exemplo, em uma figura de dizer (processos verbais), o Meio corresponde ao Dizente enquanto, em uma figura de sentir (processos mentais), corresponde ao Experienciador, conforme exemplificam os trechos (110) e (111) a seguir. Já em (112) e (113), representantes de uma oração material no modelo particularizado, o Ator em (112) corresponde ao Meio, enquanto a Meta em (113) é o Meio, na perspectiva generalizada.

- (110) Joaquim [Meio/Dizente] **contou** uma história bem diferente [Alcance/Verbiagem] (H_CP_2017_2_AA_001)
- (111) ou se é um aluno [Meio/Experienciador] que **gosta** de praticar esportes [Alcance/Macrofenômeno] (H_EDRE_2014_0_AA_006)
- (112) Apagando as unidades habituais do livro e do autor, **foram se formando** “famílias monstruosas” [Meio/Ator] que não pareceram extravagantes a Foucault (So_GSP_2018_1_AA_003)
- (113) A dieta caseira [Meio/Meta] **foi formulada** de acordo com NRC (A_ABMVZ_2018_3_AA_002)

Cabe aqui detalhar um pouco mais sobre como o sistema de voz funciona aplicado a essas orações, de acordo com Halliday e Matthiessen (2014, p. 349). Uma oração sem Agente, não é ativa ou passiva, mas **média**; uma oração com Agente é **efetiva** (quanto à Agentividade). As efetivas podem ser operativas ou receptivas quanto à voz, isto é, nas orações **operativas** o Sujeito é o Agente e o processo se realiza por meio de um grupo verbal ativo; por sua vez, nas **receptivas** o Sujeito é o Meio e o processo é realizado por um grupo verbal passivo. Tal classificação foi representada no Quadro 24.

Quadro 24 – Tipos de construções ergativas

Média	(b) O estudo	se desenvolveu	em três fases (Sa_RLAE_2018_0_AA_051)	
	Meio	Processo	Circunstância	
Efetiva operativa	(b') O pesquisador	desenvolveu	o estudo	em três fases.
	Agente	Processo	Meio	Circunstância
Efetiva receptiva	(c) Os animais	foram mantidos	nas mesmas condições ambientais e de manejo	desde o nascimento (A_ABMVZ_2013_0_AA_009)
	Meio	Processo	Circunstância	Circunstância

Fonte: Elaboração própria.

Em síntese, a oração que representa um processo causado por ele mesmo sem o Agente é média, como em (b); a oração que representa um processo trazido por alguma causa externa, o Agente, é efetiva, conforme a reescrita (b') e o trecho (c). As orações efetivas podem ser operativas, com verbo ativo (como b') ou receptivas, com verbo passivo (como c).

Vale observar que as orações efetivas possuem o traço de Agentividade, mesmo que o Agente esteja implícito (como em c do Quadro 24). É esse traço que diferencia a oração no trecho (b) dos trechos (b') e (c), pois apenas aos últimos pode-se perguntar “Por quem?” (cujas respostas seriam “O pesquisador”/Agente em ambos os casos), enquanto a construção na oração (b) permite apenas um participante. Para Thompson (2014, p. 142), a ergatividade é uma forma gramatical de expressar a diferença entre processos que representam causa ou reflexividade, considerando que a distinção é feita a partir da configuração dos papéis dos participantes em relação ao processo.

De tal modo, segundo Halliday e Matthiessen (2014, p. 340), o sistema de TRANSITIVIDADE, em todas línguas, é uma mistura dos dois modelos semânticos. O transitivo é uma interpretação linear, caracterizada pela ênfase ao distinguir processos e seus participantes. Já o ergativo caracteriza-se por uma interpretação nuclear, de Processo e Meio, podendo haver outros participantes (como nomeados na Figura 9). Do ponto de vista do modelo da transitividade, os participantes do modelo ergativo são circunstanciais, por exemplo, o Agente pode ser um tipo de Modo e o Beneficiário, um tipo de Causa (MORAIS, 2013, p. 32).

Assim, a relação entre os modelos – ergativo e transitivo – ainda é controversa, segundo Thompsom (2014, p. 141). Para Halliday e Matthiessen (2014, p. 347) todos os tipos de processos podem ser analisados sob qualquer uma das duas perspectivas. Já Caffarel (2006, p.

60) analisa a língua francesa a partir dessa perspectiva, compreendendo ambos os modelos como formas complementares de construir a experiência. Entretanto, para Davidse (1992), os dois são mutuamente exclusivos. Por sua vez Thompson (2014, p. 141) defende uma posição menos extrema, na qual a ergatividade se restringe a processos materiais nos quais a mudança, seja causada por si mesma ou externamente, é elemento central, especialmente se o verbo puder ser empregado em ambas estruturas – efetiva ou média.

De acordo com Caffarel (2006, p. 60), a semântica do modelo ergativo é a causação, ou seja, se o processo representa uma mudança causada a si mesma ou por causa externa. Essa primeira forma, conforme sua pesquisa realizada em língua francesa, é marcada no processo pelo uso do clítico *se* que possui natureza produtiva, permitindo que verbos transitivos sejam empregados em orações médias. Esse mesmo resultado Morais (2013) encontrou em sua tese de doutorado ao analisar o clítico *se* em língua portuguesa, na escrita acadêmica.

Cabe também ressaltar que as construções ergativas são mais evidentes em alguns registros ou tipos de textos, conforme apontam Halliday e Matthiessen (2014, p. 337), como em textos científicos ou “*hard news*”. Nesses casos, as construções ergativas são proeminentes, pois, de acordo com Caffarel (2006, p. 61), são tipos de textos cuja interpretação ergativa do processo (enquanto possuidores ou não de uma causa externa/Agente) fornece maior *insight* do que a interpretação transitiva, se estendendo para além do Ator, apenas. Entretanto, ressalta-se que esse tipo de análise realizada pela autora leva em consideração não apenas a oração, mas também o texto em que ocorre. Desse modo, seus resultados giram ao redor do contínuo da gramática até o discurso, não apenas do nível oracional ou léxico-gramatical.

Por fim, este capítulo procurou descrever brevemente a abordagem da Linguística Sistêmico-Funcional sobre a linguagem, desde seus contextos mais amplos até sua realização na sentença através do sistema TRANSITIVIDADE, que é parte da metafunção ideacional – uma das três metafunções da linguagem. Dentro do referido sistema, os modelos transitivo e ergativo foram apresentados, considerando que ambos serviram de bases teóricas para a análise dos processos mais frequentes no *corpus* de artigos acadêmicos desta pesquisa. Delineada essa parte da revisão de literatura, o próximo capítulo discorre sobre a metodologia empregada na construção do *corpus* e sobre os procedimentos para a análise de tais dados quantitativos e qualitativos.

4 METODOLOGIA DE PESQUISA

Até o momento foram apresentadas as discussões sobre as teorias nas quais esta tese se baseia, sendo elas a perspectiva da divisão do vocabulário de Nation (2001) e Nation e Webb (2017), no Capítulo 2, e da Linguística Sistêmico-Funcional de Halliday e Matthiessen (2014) – sendo ambas as perspectivas necessárias para alcançar os objetivos elencados no Capítulo 1. A partir dessas discussões, o presente capítulo discorre sobre os procedimentos metodológicos empregados na análise, desde a construção do *corpus* até os testes realizados para determinar os diferentes tipos de processos, iniciando a seguir por uma breve introdução à Linguística de Corpus.

A Linguística de *Corpus*, a qual tomamos como base para a coleta e a análise quantitativa da presente pesquisa, é um método científico de análise da língua. Tal metodologia requer que o linguista forneça evidências empíricas a partir de amostras reais da língua em uso a fim de fundamentar a análise de hipóteses sobre a língua (BREZINA, 2018). Isso implica que outros pesquisadores devem ser capazes de repetir a metodologia implementada, replicando experimentos e/ou técnicas aplicadas, de preferência naquele mesmo *corpus*. Isso justifica um dos objetivos dessa pesquisa, que é disponibilizar o *corpus* coletado, a fim de facilitar o acesso de outros pesquisadores aos resultados desse trabalho, avançando os estudos nessa área de conhecimento, assim como incentivando diferentes pesquisas sobre a escrita acadêmica, a partir de um conjunto variado de artigos científicos.

Em geral, a linguística de *corpus* trabalha com números que refletem a frequência de palavras e de frases em *corpora*, sendo, portanto, uma metodologia quantitativa cujo uso de recursos estatísticos é fundamental para melhor compreender os dados em análise. Assim sendo, procurou-se pormenorizar os conceitos estatísticos empregados nesta pesquisa, assim como disponibilizar os *scripts* utilizados em anexos (Anexo 1, Anexo 2, Anexo 3, Anexo 4 e Anexo 5), a fim de facilitar a replicabilidade desse estudo, ou um similar.

Nas próximas seções, descreve-se a composição do *corpus*, acompanhada pelos procedimentos de sua coleta e organização, e também os procedimentos de análise a fim de identificar os itens lexicais estudados no Capítulo 5, que corresponde à análise do sistema da TRANSITIVIDADE dos verbos mais relevantes em artigos científicos.

4.1 Composição, procedimentos de coleta e organização do *corpus*

Neste trabalho, procura-se observar os dados em seu contexto, analisando-os minuciosamente enquanto uma pesquisa de descrição da língua em uso. Para tanto, toma-se como base as premissas da LSF, apresentadas no Capítulo 3, na qual a língua é estudada a partir do contexto sociocultural em que ocorre. A fim de analisar a realização do sistema da TRANSITIVIDADE nos verbos relevantes em artigos acadêmicos (vocabulário acadêmico), foi compilado um *corpus* de aproximadamente 12,3 milhões de palavras com textos representantes das diferentes áreas de conhecimento, de modo a evitar que fosse tendencioso a uma única área e, conseqüentemente, tivesse maior proporção de vocabulário técnico.

Adentrando no âmbito de construção deste *corpus*, segundo Brezina (2018, p. 18), não há uma determinação universal sobre o **tamanho** exato que um *corpus* deve ter, pois a resposta depende do fenômeno linguístico que se deseja investigar. O autor cita como exemplo a análise de estruturas passivas, para as quais um *corpus* pequeno, de um milhão de palavras, bastaria, por ser esse um fenômeno relativamente comum em língua inglesa.

Entretanto, vale observar que alguns itens lexicais (e suas combinações) podem ser bastante infrequentes, mesmo em *corpora* grandes. No caso do presente trabalho, em um *corpus* com 12,3 milhões de *tokens* e 174.856 *types*, a palavra *misericórdia* só apareceu 12 vezes – em oposição ao *Corpus Brasileiro*, que apareceu 7.678 vezes, e ao PtTenTen11, 65.025 vezes. Essa variação exemplifica também uma regra geral, de acordo com o autor (BREZINA, 2018, p. 19), na qual a menos que o *corpus* represente toda uma língua (improvável), a falta de evidência não evidencia a ausência. Assim, se uma expressão não ocorrer no *corpus*, não significa que não exista ou que não seja relevante, implica apenas que não há ocorrências naquele *corpus* especificamente.

Vale destacar o quão importante é a construção do *corpus* não apenas em função do tamanho, mas também em termos de representatividade e balanceamento. Nesse quesito, o *corpus* desta pesquisa foi compilado atendendo aos pressupostos da **representatividade** descritos por Brezina (2018, p. 15), em que a escolha completamente aleatória de textos é impraticável na linguística de *corpus*, pois não há como catalogar toda a produção linguística a fim de elencar tais amostras. Assim, para tornar a seleção dos textos mais alcançável, é preciso estabelecer um conjunto de categorias dentro das quais procurar-se-á coletar uma amostra menos enviesada da língua. Tais categorias são denominadas **quadros** e geralmente seguem o padrão de gêneros textuais (conforme o *Corpus Brown* e o *Corpus Brasileiro*). Entretanto, nesta

pesquisa, como apenas um único gênero está sob análise, procurou-se trabalhar com variedade relacionada ao tópico textual, selecionando amostras de diferentes disciplinas. A fim de não ser tendencioso para uma única especialidade, ou um pequeno conjunto, foram coletados artigos científicos de todas as áreas do conhecimento determinadas pelo CNPq.

Em relação às tais áreas, de acordo com o CNPq¹⁹, as várias disciplinas estão organizadas segundo a estrutura da árvore do conhecimento, tendo **Grandes Áreas** como a hierarquia mais abrangente, seguidas pelas **Áreas**, depois pelas **Subáreas** e, por fim, pelas **Especialidades**. Abaixo segue um quadro que ilustra tal hierarquia aplicada à classificação de dois artigos do *corpus* desta pesquisa.

Quadro 25 – Exemplos de classificação de acordo com a hierarquia da divisão de áreas proposta pelo CNPq

Grande Área	Ciências da Saúde	Ciências Agrárias
Área	Medicina	Medicina Veterinária
Subárea	Clínica Médica	Patologia Animal
Especialidade	Pediatria	Anatomia patologia animal
Título do artigo	Consumo de leites em menores de um ano de idade e variáveis associadas ao consumo de leite não materno	Técnicas de citologia aspirativa, biópsia e citobloco de medula óssea para identificação e determinação de intensidade parasitária na leishmaniose visceral canina

Fonte: Elaboração própria.

As Grandes Áreas correspondem ao nível hierárquico superior, que aglomera as diferentes áreas do conhecimento de acordo com seus objetos de estudo, metodologias e recursos instrumentais que refletem contextos sócio-políticos específicos (CORDANI, 2005)²⁰. As Áreas foram definidas como um conjunto de conhecimentos com inter-relação entre si, reunido pelo coletivo em função da natureza do objeto investigado com finalidade de pesquisa, ensino e aplicação à prática (CORDANI, 2005). Já as Subáreas representam a segmentação da

¹⁹ Disponível em: <<http://lattes.cnpq.br/web/dgp/arvore-do-conhecimento>>. Acesso em 15 jul. 2019.

²⁰ Disponível em:

<http://www.sbpcnet.org.br/livro/57ra/programas/CONF_SIMP/textos/umbertocordani.htm>. Acesso em 30 ago. 2019.

área de conhecimento segundo o objeto estudado e os procedimentos metodológicos reconhecidos e utilizados amplamente (CORDANI, 2005). Por fim, a Especialidade corresponde a uma “atividade científica ou tecnológica com um nível maior de especificidade, definida mais pela sua finalidade ou pela sua aplicação” (CORDANI, 2005).

Ainda segundo o autor, Cordani (2005), que foi membro da Comissão Especial de Estudos das Áreas do Conhecimento, apesar de haver necessidade de reelaboração da atual tabela das áreas de conhecimento, é bastante difícil encontrar critérios coerentes e equilibrados para padronizar a classificação das diferentes áreas. Assim, a Comissão optou por manter a estrutura original proposta em 1984 que divide as disciplinas em quatro níveis (listados e exemplificados no Quadro 25, anterior), tendo como unidade a Área, devido à dificuldade de chegar a um consenso entre as diferentes comunidades acadêmicas quanto à criação, extinção ou pertencimento das variadas Especialidades.

Em vista disso, postulou-se que os três primeiros níveis seriam mais estáticos enquanto as Especialidades poderiam sofrer modificações mais radicais, de acordo com a necessidade da disciplina. Atualmente, a lista conta com 8 Grandes Áreas, 76 Áreas, 340 Subáreas e diversas Especialidades. Desse modo, justifica-se a atual organização dos textos analisados por esta pesquisa, que pretendeu compilar um *corpus* que não estivesse sujeito a muitas variações, como de nomenclatura, por exemplo. Assim, subdivide-se em oito *subcorpora* que correspondem às oito Grandes Áreas do conhecimento: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Engenharias, Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes.

Essa organização foi preferida em oposição à classificação das áreas definida pela CAPES (utilizada pela autora desta tese em MIRANDA, 2016) que, de maneira diferente, possui nove Grandes Áreas, embora um número menor de Áreas (48, denominadas Áreas de avaliação). A escolha foi feita não apenas em função do número, mas também por ser a classificação do CNPq a mesma utilizada pelo Portal SciELO – fonte dos textos acadêmicos que foram coletados pela pesquisa.

A opção por artigos acadêmicos disponíveis apenas no portal SciELO se deu por alguns importantes motivos, tais quais a não-restrição de acesso aos artigos, a ampla variedade de periódicos compartilhados pelo site e a licença *Creative Commons*, que abrange todos os textos publicados, permitindo sua distribuição e adaptação, mesmo para fins comerciais, desde que

atribuído o devido crédito ao original²¹. Desse modo, é possível disponibilizar o *corpus* coletado para demais pesquisadores, após a defesa da presente pesquisa, sem sofrer possíveis empecilhos legais – uma das características às quais linguistas devem atentar ao construir um *corpus* (BREZINA, 2018, p. 17).

Além disso, os periódicos disponibilizados pelo portal passam anualmente pelo processo de avaliação da qualidade, no qual são levados em consideração os indicadores de impacto de citações por artigos, de influência na Web e de internacionalização da revista, assim como a pontualidade das publicações, a revisão por pares, o tempo de existência e a representatividade institucional, temática e geográfica²². Todos esses critérios garantem a chancela acadêmica da comunidade científica quanto aos textos publicados e disponíveis no portal SciELO, de modo a justificar seu uso para coleta deste *corpus*.

Apesar de todas características positivas, vale observar que uma limitação da classificação do SciELO é o pertencimento de alguns periódicos a diferentes áreas simultaneamente. Para resolver esse problema, assim como facilitar a uniformização e organização dos textos, foi elaborada uma planilha com todos os periódicos selecionados para a coleta, detalhando as informações pertinentes de cada revista para o presente *corpus* e, por extensão, para o projeto CAPB, ao qual este *corpus* foi incorporado.

A partir desses pressupostos teóricos, a coleta teve como base o *corpus* compilado para a dissertação de mestrado da autora (MIRANDA, 2016), que deu início ao CAPB (MIRANDA, OLIVEIRA; OLIVEIRA, 2018). Entretanto, como um dos propósitos desse trabalho foi disponibilizar o material coletado para demais pesquisadores, parte desse *corpus* foi retirada para que apenas textos sob licença *Creative Commons* fossem incorporados. Além dessa, outras características divergem os *corpora*. O anterior foi subdividido segundo a classificação da CAPES, tendo como unidade as 48 Áreas de Avaliação – o equivalente às Áreas do CNPq – e teve como objetivo coletar 100.000 palavras de cada área, totalizando em um *corpus* de 4,8 milhões de palavras ao longo de aproximadamente 1.200 artigos.

Desse modo, a fim de verificar possíveis diferenças entre disciplinas, o presente *corpus* foi reelaborado, tomando como *subcorpora* as oito Grandes Áreas com a finalidade de se avaliar melhor as diferenças entre as disciplinas. Cabe ressaltar que para esse tipo de análise, uma unidade menor – como Áreas de Avaliação – não seria o ideal, pois, conforme dito por Cordani

²¹ A nota do portal SciELO sobre a licença que abrange seus artigos pode ser encontrada em: <<https://SciELO.org/pt/sobre-o-SciELO/declaracao-de-acesso-aberto>>. Acesso em 1 set. 2019.

²² Os critérios para credenciamento de novos periódicos do portal SciELO pode ser encontrada em: <<http://www.SciELO.br/avaliacao/Criterios%20SciELO%20Brasil.pdf>>. Acesso em 1 set. 2019.

(2005), há muita variação e pouco consenso em sua classificação. A título de exemplo, a CAPES subdivide a área de Saúde em 9 Áreas de Avaliação, a saber, Educação Física, Enfermagem, Farmácia, Medicina I, Medicina II, Medicina III, Nutrição, Odontologia e Saúde Coletiva. As áreas de Medicina I, II e III abrigam várias disciplinas, cerca de 39, como Cardiologia, Oftalmologia, Neurologia etc., sem agrupá-las segundo um padrão exato. Ou seja, não há um bom critério para o agrupamento de disciplinas em determinadas Áreas de Avaliação. O mesmo acontece com a Grande Área de Engenharia, subdivida em apenas quatro Áreas de Avaliação, Engenharias I, II, III e IV.

Por outro lado, o CNPq não agrupa desse modo suas Subáreas e, conseqüentemente, possui um número maior de Áreas (76). Em vez de três categorias para agrupar todas as especialidades de Medicina, há apenas uma categoria que engloba suas Subáreas. Tais observações sobre a organização e divisão das áreas do conhecimento pelas grandes financiadoras de pesquisa em contexto brasileiro são importantes para o trabalho, pois a partir dela procurou-se selecionar os textos para análise. Assim, os periódicos foram escolhidos para que contemplassem, se não todas, a maioria das Subáreas para que os *subcorpora* não fossem tendenciosos para certo tópico. Além disso, diferentemente do trabalho anterior, que se limitou a 100.000 palavras por Área de Avaliação, neste não foi imposto um limite de *tokens*, tendo em vista o número flutuante de Áreas e Subáreas dentro de cada Grande Área.

Obedecendo a esses critérios expostos, um novo *corpus* de 2.898 artigos científicos foi compilado, contendo aproximadamente 12,3 milhões de *tokens*, calculados com auxílio do programa Kitconc© (MOREIRA FILHO, 2008). Os detalhes do *corpus* estão apresentados na Tabela 3 – Composição do *corpus* da pesquisa, a seguir.

Tabela 3 – Composição do *corpus* da pesquisa

Grande Área (Sigla)	Periódicos representantes (Siglas)	Nº de artigos	Nº de tokens
Ciências Agrárias (A)	Arquivo Brasileiro de Medicina Veterinária (ABMVZ), Arquivo Internacional de Biologia (AIB) Brazilian Journal of Food Technology (BJFT), Ciência Rural (CR), Floresta e Ambiente (FL), Food Science and Technology (FST), Pesquisa Veterinária Brasileira	421	1.149.095

Grande Área (Sigla)	Periódicos representantes (Siglas)	Nº de artigos	Nº de tokens
	(PVB), Revista Árvore (RAR), Revista Ceres (RCR).		
Ciências Biológicas (B)	Hoehnea (HOE), Ilheringia, série Zoologia (ISZ), Revista Ambiente e Água (RAA), Revista Ceres (RCR), Rodriguésia (R).	229	674.347
Ciências Exatas e da Terra (E)	Boletim de Ciências Geodésicas (BCG), Química Nova (QN), Revista Brasileira de Meteorologia (RBM), SBA, Controle e Automoção (SBA), Tema, São Carlos (TSC).	370	1.490.360
Engenharias (En)	Ambiente Construído (AMC), Engenharia Sanitária e Ambiental (ESA), Gestão e Produção (GP), Matéria, Rio de Janeiro (MRJ), Polímeros (POL), Revista Ibracon de Estruturas e Materiais (RIEM), Revista Brasileira de Engenharia Biomédica (RBEM), Revista da Escola de Minas (REM), Soldagem e Inspeção (SI).	505	2.121.683
Ciências Humanas (H)	Ágora, Estudos em Teoria Psicanalítica (AETP), Avaliação, Campinas (AC), Caderno CRH (CCRH), Cadernos CEDES (CCED), Cadernos Pagu (CP), Contexto Internacional (COIN), Dados (DADO), Educação e Pesquisa (EDPE), Educação em Revista (EDRE), Kriterion (KRIT), Lua Nova (LN), Psicologia, Ciência e Profissão (PCP), Psicologia, Teoria e Pesquisa (PTP), Religião e Sociedade (RSO), Revista Brasileira de	316	2.126.022

Grande Área (Sigla)	Periódicos representantes (Siglas)	Nº de artigos	Nº de tokens
	Ciência Política (RBCP), Revista Brasileira de Ensino de Física (RBEF), Revista Brasileira de História (RBHI), Topoi (TOPO), TransFormAção (TFA).		
Letras, Linguística e Artes (L)	Alfa, Revista de Linguística (ALFA), ARS, São Paulo (ASP), Bahkitiniana, Revista de Estudos do Discurso (BKTN), Estudos de Literatura Brasileira Comparada (ELBC), Letras de Hoje (LH), Per Musi (PEMU), Revista Brasileira de Estudos da Presença (RBEP), Revista Delta (RD).	244	1.442.676
Ciências da Saúde (Sa)	Acta Paulista de Enfermagem (ACPE), Anais Brasileiros de Dermatologia (ABDE), Arquivos Brasileiros de Oftalmologia (ABOF), Brazilian Journal of Physical Therapy (BJPT), CoDAS (CODA), Coluna, Columna (COLU), Dental Press Journal of Orthodontics (DPJO), International Archives of Otorhinolarhngology (IAOO), International Journal of Cardiovascular Science (IJCS), Jornal Brasileiro de Nefrologia (JBNE), Jornal Brasileiro de Psiquiatria (JBP), Jornal de Pediatria (JOPE), Motriz, Revista de Educação Física (MOTR), Revista Brasileira de Ginecologia Obstetrícia (RBGO), Revista Lationamericana de Enfermagem (RLAE), Revista Brasileira de Enfermagem (RBEN), Revista Brasileira	547	1.604.958

Grande Área (Sigla)	Periódicos representantes (Siglas)	Nº de artigos	Nº de tokens
	de Anestesiologia (RBAN), Revista de Odontologia da UNESP (ROUN), Revista de Saúde Pública (RSAP), Revista Paulista de Pediatria (RPP), Saúde e Sociedade (SASO), Texto e contexto (TECO).		
Ciências Sociais Aplicadas (So)	Ambiente e Sociedade (AMSO), Cadernos Metrópole (CAME), Economia Aplicada (ECAP), Galáxia, São Paulo (GSP), Intercom (INCO), Nova Economia (NE), Organizações e Sociedade (OS), Perspectivas em Ciência da Informação (PPCI), Revista Direito GV (RDGV), Revista Katálysis (RKAT), Sequência, Florianópolis (SEFL), Serviço Social e Sociedade (SSSO), Sociedade e Estado (SOES).	265	1.711.026
	TOTAL	2.898	12.320.167

Fonte: Elaboração própria.

Para a nova coleta, considerou-se quatro novos periódicos de cada Grande Área, de diferentes Áreas, sendo os artigos publicados nos anos de 2017 e 2018. A escolha dos quatro periódicos não foi aleatória, mas baseada no número máximo de periódicos presentes em uma das Grandes Áreas – Ciências Exatas e da Terra. Nessa área, há apenas quatro periódicos atuais cujos artigos publicados estavam em língua portuguesa. Por essa razão, o número de revistas selecionadas para acrescentar cada *subcorpus* foi apenas quatro por área – a fim de diminuir a tendência para uma única Grande Área.

Entretanto, como demonstra a Tabela 4, há uma média maior de artigos publicados por revista em Ciências da Saúde, seguida por Engenharias. A variação entre Grandes Áreas se deve tanto ao número de Áreas dentro de cada Grandes Áreas (listado na Tabela 4 a seguir), assim como a variação própria de cada Grande Área. Por exemplo, na Grande Área de Ciências Exatas

e da Terra, há cerca de 74 artigos por revista, com média de 4.024 palavras por artigo no decorrer das cinco revistas coletadas. Já em Ciências Humanas, a média é de 16 artigos por revista, com 6.707 *tokens* aproximadamente, ao longo de dezenove revistas coletadas. Ou seja, nessa última Grande Área, percebe-se que são publicados menos artigos, porém, mais extensos. Além dessas observações, cabe notar que o número de revistas coletadas acompanha tanto a variedade de Áreas, especificadas na Tabela 4, quanto a disponibilidade dos periódicos no portal SciELO.

Tabela 4 – Médias de artigos por periódicos e de *tokens* por artigos no *corpus*

Grande Área (Sigla)	Nº de Áreas	Média de artigos por periódico	Média de <i>tokens</i> por artigo
Ciências Agrárias (A)	7	46,8	2.729
Ciências Biológicas (B)	13	45,8	2.945
Ciências Exatas e da Terra (E)	8	74,0	4.028
Engenharias (En)	13	56,1	4.201
Ciências Humanas (H)	10	16,7	6.707
Letras, Linguística e Artes (L)	3	27,1	5.913
Ciências da Saúde (Sa)	9	24,9	2.934
Ciências Sociais Aplicadas (So)	13	20,4	6.457

Fonte: Elaboração própria.

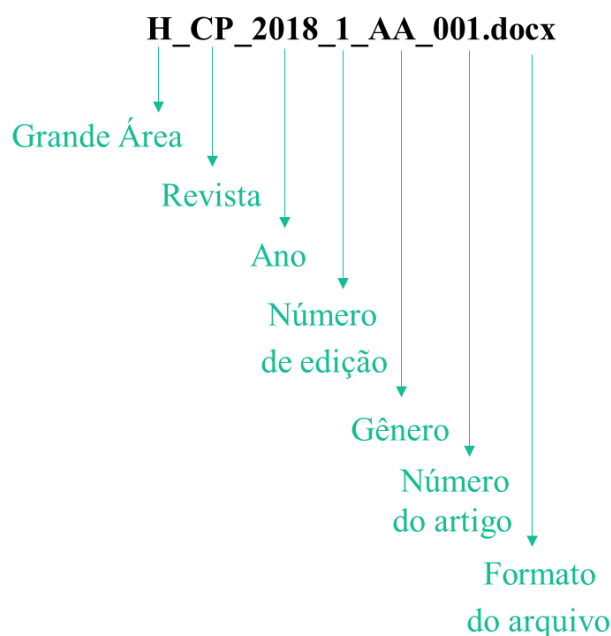
Após a coleta dos textos, de cada arquivo foram retiradas todas as partes menos relevantes para a análise, tais quais título, nome dos autores, resumo, tabelas, quadros, figuras, agradecimentos e referências, e salvos os textos limpos em formato .txt com auxílio do programa gratuito Calibre²³, que converte em massa os textos originais (em .docx) para .txt, formato necessário para utilizar o *corpus* em grande parte dos concordanciadores linguísticos.

Concomitante à limpeza e à conversão dos arquivos, os textos foram nomeados a fim de facilitar a compreensão da origem das amostras linguísticas a serem analisadas. Esse passo de nomenclatura é importante tanto para compreensão posterior dos dados quanto para integração do presente *corpus* ao projeto CAPB, pois unifica os textos coletados àqueles que serão incorporados ao projeto futuramente. Assim, no título de cada arquivo estão resumidas as

²³ Disponível em <<https://calibre-ebook.com/>>. Acesso em 5 set. 2019.

informações estruturais de armazenamento do texto que podem ser relevantes para os usuários, conforme ilustrado na Figura 10 a seguir.

Figura 10 – Nomenclatura dos textos do *corpus*



As abreviações de cada Grande Área, assim como do nome de cada revista, podem ser encontradas na Tabela 3. Em relação ao gênero, na presente pesquisa há apenas um único gênero, artigo acadêmico, sob a sigla AA. Entretanto, essa marca é necessária considerando os demais gêneros acadêmicos que fazem parte do CAPB. Quanto ao número da edição, o “0” denomina edições publicadas em volume único. Por último, o número do artigo foi determinado apenas em função da ordem da coleta dos textos (que reflete a ordem de publicação dos artigos dentro da edição da revista).

Na próxima seção explicam-se os procedimentos de análise dos dados, iniciando pela apresentação das listas de palavras nas quais este trabalho se baseou para criar a lista de verbos acadêmicos em PB, seguida pelo relato das etapas adotadas na pesquisa.

4.2 Procedimentos de análise dos dados

4.2.1 Listas de palavras

Como mencionado anteriormente, na seção 2.2, listas de palavras podem ser ferramentas de grande auxílio para o ensino-aprendizagem de uma língua. Nesta pesquisa, elaborou-se uma lista de verbos acadêmicos em artigos científicos, ou seja, aqueles que aparecem nesse gênero com frequência maior do que na linguagem cotidiana e são comuns a diferentes áreas do conhecimento. A partir dessa lista, relevante para o gênero acadêmico, analisa-se o sistema da TRANSITIVIDADE instanciado pelos verbos mais relevantes em artigos científicos, observando possíveis diferenças e/ou semelhanças em seus usos entre as Grandes Áreas. Desse modo, verificar-se-ia a existência de fato de um vocabulário acadêmico unificado entre as áreas do conhecimento, ou se tais itens lexicais são empregados de maneira distinta entre as diferentes disciplinas.

Para tanto, a presente tese fundamenta-se em métodos anteriores de elaboração de listas acadêmicas, tais como a UWL, AWL, AVL (cf. seção 2.5) e outras de menor porte. Portanto, cabe retomar a diferenciação entre tais listas, iniciada na seção 2.5, descrevendo em maiores detalhes a metodologia de cada uma a fim de estabelecer os procedimentos utilizados nesta pesquisa.

A UWL e a AWL seguem um padrão similar de compilação: a sobreposição de listas. Criada em 1984, por Xue e Nation (1984), a UWL baseou-se na sobreposição de uma lista de vocabulário geral de língua, outra de palavras acadêmicas não pertencentes à anterior e duas listas de palavras cujos aprendizes escreveram traduções em textos acadêmicos. Tal lista foi formulada a fim de não incluir palavras pertencentes a GSL, compondo um total aproximado de 800 famílias de palavras e cobrindo 8,5% de textos acadêmicos, 3,9% de textos jornalísticos e 1,7% de textos ficcionais (HWANG e NATION, 1989). Segundo Nation (2001), a baixa abrangência dessas palavras em gêneros não-acadêmicos demonstra a natureza especializada da lista.

Por sua vez, a AWL (COXHEAD, 2000) substituiu explicitamente a UWL, sendo composta por 570 famílias de palavras compiladas a partir de um *corpus* acadêmico com 3,5 milhões de palavras subdividido nas áreas de Artes, Ciência, Direito e Comércio, englobando 28 disciplinas, como história, psicologia e matemática. Para a elaboração dessa lista, além da sobreposição com a GSL, na qual houve a exclusão das palavras que co-ocorreram, foram considerados tanto a frequência quanto o alcance das palavras ao longo das áreas, de modo que todas as famílias deveriam ter ocorrências nos quatro *subcorpora* e pelo menos cem vezes no *corpus* total. Ainda, as palavras que ocorreram menos de 10 vezes em cada disciplina não foram incluídas, assim como as palavras que não ocorreram em pelo menos 15 disciplinas (das 28).

A fim de verificar se o produto dessa metodologia eram palavras realmente acadêmicas, ou apenas frequentes, mas que não estavam presentes na GSL, comparou-se a lista resultante com um *corpus* de romances de ficção. Obteve-se como resultado uma lista com maior cobertura que a UWL, abrangendo cerca de 10,6% de textos acadêmicos.

Diferente das listas anteriores, que se basearam na exclusão de palavras pertencentes à GSL, a Academic Vocabulary List (GARDNER e DAVIES, 2014) utilizou critérios alternativos para listar as palavras acadêmicas. Além dessa, outra grande diferença está no fato de que as palavras se organizam por lemas – não mais famílias de palavras – e do *corpus* utilizado ser atual – não datado de cinquenta anos atrás, como a GSL. Assim, para a criação da AVL, utilizou-se um *corpus* de 120 milhões de palavras de textos acadêmicos, integrantes do COCA, o qual se subdivide em nove áreas, Educação, Humanidades, História, Ciências sociais, Filosofia, Religião e psicologia, Direito e ciências políticas, Medicina e saúde, Administração e finanças, sendo 70% do *corpus* composto por artigos científicos.

Para a criação dessa lista de palavras acadêmicas, quatro critérios foram adotados: razão, alcance, dispersão e medida entre disciplinas. A *razão* refere-se ao critério empregado para eliminar as palavras de alta frequência da lista através da especificação de que um lema deve ser pelo menos 50% mais frequente no *corpus* acadêmico do que no COCA, isto é, a razão por milhão de palavra é 1.50.

O critério *alcance* estabelece que uma palavra deve ocorrer com pelo menos 20% da frequência esperada em pelo menos sete das nove disciplinas. Isto é, com base no tamanho de uma disciplina, uma dada palavra deve aparecer com ao menos 20% da frequência esperada.

Já o valor para o critério de *dispersão* dos lemas foi de, no mínimo, 0.80 calculado por meio do método Juilland's D (JULLIAND e CHANG-RODRIGUEZ, 1964), que determina o quão igualmente uma palavra ocorre ao longo do *corpus*, podendo variar de 0.01 (baixa dispersão no *corpus*) à 1.0 (distribuição perfeita ao longo das partes do *corpus*). Desse modo, se uma palavra ocorre com frequência 20% acima do esperado em sete das nove disciplinas, mas ocorre com frequência muito maior em duas áreas do que nas demais, não cumpre o critério da dispersão sendo, portanto, eliminada da lista.

Por fim, a *medida de disciplina* foi desenhada para excluir palavras técnicas ou próprias de um conjunto de disciplinas, estabelecendo que um item não pode ocorrer com frequência três vezes maior do que o esperado em nenhuma das nove disciplinas. Cabe observar que todas as medidas utilizadas nos quatro critérios são arbitrárias, definidas após experimentos extensivos com os dados (GARDNER e DAVIES, 2014).

A lista proveniente da aplicação desses critérios possui aproximadamente 3.000 lemas, cobrindo cerca de 14% de textos acadêmicos do COCA e do British National *Corpus* – ou 21.17% de textos acadêmicos da área de medicina, de acordo com Lu e Durrant (2017), provando possuir a melhor cobertura das listas acadêmicas. Por essa razão, a presente pesquisa se baseou na metodologia da AVL, acrescido de um dos critérios da AWL, para identificar os verbos acadêmicos a serem analisados com base no sistema da TRANSITIVIDADE da LSF, detalhado no Capítulo 3. Na seção a seguir, descreve-se os procedimentos utilizados para listar os verbos acadêmicos do *corpus*.

4.2.2 Relato de procedimentos

Para a criação de uma lista lematizada de verbos acadêmicos a partir do *corpus* de artigos acadêmicos da presente pesquisa, adaptou-se os procedimentos metodológicos tanto da AWL quanto AVL, adotando-se cinco critérios distintos: frequência mínima, razão da frequência, razão do alcance, dispersão e medida de disciplinas.

Primeiramente, após os passos de compilação do *corpus* – seleção, coleta e limpeza dos textos –, foi necessário lematizar os *types* do *corpus*. Nesse âmbito, o software #LancsBox foi utilizado, pois apesar de sua interface ser voltada para usuários de língua inglesa, ele fornece suporte a lematização de diferentes línguas ao empregar o etiquetador morfossintático TreeTagger, que reconhece mais de 25 línguas e também é utilizado pelo Sketch Engine²⁴ (*software online* de análises linguísticas). Cabe mencionar que além desse recurso de lematizar o *corpus*, o #LancsBox também possui uma variedade de outras ferramentas relevantes para a análise linguística que o distingue de outros *softwares* citados anteriormente na seção 2.1 (WordSmithTools, AntConc e KitConc), como a variedade de arquivos em que o *corpus* pode ser importado (não apenas .txt, mas também .pdf, .docx, .xls, entre outros), a qualidade no refinamento das pesquisas realizadas e a possibilidade de comparar diferentes *corpora* visualizando-os em uma mesma interface.

Após a lematização dos itens lexicais, que é feita logo ao importar um novo *corpus* no programa, foi utilizada a ferramenta “Words” para criar listas de palavras para cada um dos oito *subcorpora* da pesquisa e, então, organizá-las por lema. Os resultados dessas listas são arquivos .csv, que podem ser abertos em um editor de planilhas (como Microsoft Excel ou Libre

²⁴ Disponível em: < <https://www.sketchengine.co.uk/> > Acesso em 20 de dezembro de 2020.

OfficeCalc) e cujos lemas incluem marcação de etiqueta morfossintática, como pode ser percebido na Figura 11, abaixo.

Figura 11 - Lista lematizada dos itens lexicais do *corpus* da pesquisa no programa #LancsBox

Lemma	Frequency: 01 - Freq	Dispersion: 01 - CV
de_other	739666.000000	0.117209
o_other	700040.000000	0.112423
e_con	377554.000000	0.041021
@card@_other	365592.000000	0.400862
que_other	238044.000000	0.312175
de+o_other	207697.000000	0.038046
ser_v	205459.000000	0.185832
de+a_other	205195.000000	0.097179
um_other	176757.000000	0.335434
em_other	170274.000000	0.065046
a_other	136047.000000	0.103098
para_other	125832.000000	0.146251
com_other	121494.000000	0.200888
em+a_other	92811.000000	0.084435
ir_v	92567.000000	0.481365
em+o_other	91796.000000	0.044818
de+os_other	84434.000000	0.120337
por_other	78074.000000	0.067262
como_other	65790.000000	0.365765
de+as_other	64565.000000	0.101082
não_adv	62372.000000	0.300774
seu_other	61315.000000	0.534910
se_other	59044.000000	0.521419
a+o_other	54262.000000	0.158959
a+a_other	50145.000000	0.161345
ou_con	42812.000000	0.245006
entre_other	42423.000000	0.118889
poder_v	42098.000000	0.082788
mais_adv	41077.000000	0.134166
a_n	36732.000000	0.140699
ter_v	36562.000000	0.265138
o_n	34006.000000	0.094256
apresentar_v	32034.000000	0.416202
estudo_n	27830.000000	0.696046

Fonte: *Screenshot* do programa #LancsBox em uso.

A partir dessas listas, a fim de facilitar as próximas etapas, foi preciso manipulá-las, excluindo os itens que não fossem verbos e, depois, removendo a marcação morfossintática de cada lema restante. Essa última etapa é bastante importante, pois permite comparar tais listas com aquela proveniente de um *corpus* representativo da língua geral.

Para remoção de itens lexicais não-verbais e das marcações morfossintáticas dos lemas, utilizou-se um *script* no *software* estatístico R, que pode ser encontrado no Anexo 1 desta tese. Em seguida, cada lista foi editada manualmente a fim de eliminar itens que não fossem verbos e agrupar flexões vde verbos que não foram devidamente lematizadas. Casos assim não foram frequentes, mas podem ocorrer devido ao etiquetamento automático das palavras, cujos avanços na área ainda não permitem cem por cento de acurácia.

Determinados os passos de preparação das listas, as próximas etapas dizem respeito aos critérios (e como foram realizados) de elaboração de uma lista de verbos acadêmicos.

O primeiro critério utilizado proveio da AVL, e estabelece uma **frequência mínima** para as ocorrências dos lemas no *corpus* Geral, a fim de garantir que os verbos resultantes sejam de alta frequência. Coxhead (2000) utilizou 100 *tokens* como ponto de corte em seu *corpus* de 3,5 milhões de palavras. A mesma medida foi utilizada em pesquisas (WANG *et al.* 2008; LEI e LIU, 2016) que exploraram o vocabulário acadêmico (e técnico) da área de medicina e também foi adotada por esta pesquisa. Apesar de parecer um número relativamente significativo, testes foram rodados no *corpus* com e sem esse critério, havendo pouca diferença em relação aos verbos (apenas quatro verbos foram excluídos da lista final, como *despender* e *sedimentar*). Isso ocorre em função do maior rigor metodológico das próximas etapas, que ocorrem simultaneamente. No Anexo 5, o *script* referente à frequência mínima está descrito junto ao critério de medida de Disciplina.

Emprega-se o segundo critério, **razão** da frequência, com o objetivo de eliminar da lista as palavras que são de alta frequência na língua. Para tanto, determina-se que as palavras desta lista devem ocorrer com frequência 50% (razão 1.5) maior neste *corpus* do que no *corpus* de referência, que abrange diferentes amostras da língua. Diferente da AVL, na qual o *corpus* acadêmico (120 milhões de *tokens*) faz parte do *corpus* de referência (COCA, com 425 milhões de *tokens*), este *corpus* não faz parte de um *corpus* maior e representativo da língua. Por isso, como referência adotou-se o PtTenTen11, com 3,7 bilhões de palavras do PB, após testes com esse *corpus* e com o *Corpus* Brasileiro (1 bilhão de palavras). A opção pelo PtTenTen ocorreu pelos seguintes motivos: 1) testes foram feitos com ambos os *corpora*, entretanto, o PtTenTen forneceu o dobro de itens lexicais acadêmicos do que o *Corpus* Brasileiro; 2) tal diferença provavelmente se deve ao número de textos acadêmicos do *Corpus* Brasileiro, cujo total de *tokens* ultrapassa metade do total de aproximadamente um bilhão de palavras.

Cabe mencionar que as listas completas de lemas verbais do PtTenTen e do *Corpus* Brasileiro para comparação foram obtidas por meio do *software* online Sketch Engine, através da compra de acesso a elas, visto que apenas as mil primeiras são fornecidas para *corpora* não-autorais por padrão.

Retomando a medida estabelecida por esse critério, mantivemos a razão 1.5 após experimentos com diferentes valores, como 1.3 e 1.7, que permitiram a entrada de muitas palavras de alta frequência, no primeiro caso, ou a saída de palavras academicamente relevantes. O Anexo 2 relata o *script* utilizado para calcular esse critério.

O terceiro critério, razão do **alcance**, assegura que o lema ocorra em diferentes *subcorpora*. Na composição da AVL, Coxhead (2000) determina que a palavra deveria ter qualquer número de ocorrências em pelo menos metade dos *subcorpora* da pesquisa. Por sua vez, na AVL, Gardner e Davies (2014) estabelecem que cada lema deve ocorrer com pelo menos 20% da frequência esperada em sete dos nove *subcorpora*. Similarmente, neste trabalho optou-se por manter aqueles itens que ocorrem com 20% da frequência esperada em seis ou mais dos oito *subcorpora*. Isto é, comparou-se a frequência da palavra em cada *subcorpus* com a frequência esperada da palavra no *corpus* como um todo. Posteriormente, foram contabilizadas em quantas áreas a palavra alcançava 20% ou mais da frequência esperada (total) e, então, foram consideradas apenas as que ocorreram em, no mínimo, seis *subcorpora*. No Anexo 3 encontra-se o *script* utilizado para medir esse critério.

O quarto critério, a **dispersão**, garante que o lema esteja distribuído igualmente ao longo dos *subcorpora*. O índice Juilland's D foi a medida estatística empregada para calcular a dispersão da palavra que pode variar de 0.01 – itens que ocorrem em pequenas partes do *corpus* – a 1.0 – itens perfeitamente distribuídos ao longo do *corpus*. Na AVL utilizou-se 0.8 como valor arbitrário de dispersão, assim como optou-se por seguir nesta metodologia. O *script* referente a esse critério está no Anexo 4.

Por último, o quinto critério, **medida de disciplina**, foi desenvolvido a fim de eliminar palavras técnicas ou específicas de uma certa área. Assim, determina-se que um lema não pode ocorrer com mais de três vezes a frequência esperada em nenhum dos *subcorpora*. Esse critério foi realizado conforme a AVL, entretanto, não surte efeito em nossos resultados. Acredita-se que isso aconteça em função da natureza da lista – apenas lemas verbais – em vez de variadas categorias gramaticais. Mesmo assim, o *script* referente ao cálculo da medida de disciplina está no Anexo 5 da tese.

Em suma, todos os critérios são medidos internamente ao *corpus*, com exceção do segundo, a razão de frequência, que é calculado com base em um *corpus* de referência, neste caso, o PtTenTen11, a fim de eliminar as palavras de alta frequência da lista. Já os critérios internos foram selecionados de modo a excluir palavra técnicas ou que ocorram majoritariamente em um pequeno conjunto de áreas de conhecimento. Desse modo, a próxima seção discute os resultados da aplicação desses procedimentos metodológicos.

4.3 Lista de verbos acadêmicos do PB

Aplicados os cinco critérios expostos na seção anterior – frequência mínima, razão, alcance, dispersão e medida de disciplina –, obteve-se como resultado uma lista de verbos acadêmicos, relevantes para artigos científicos – gênero em que se baseia o presente *corpus* e tese. A seguir, reproduziu-se essa lista na Tabela 5.

Tabela 5 - Lista de verbos acadêmicos

	Lema	Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
		FR	FA	FR	FA	FR	FA	FR	FA	FR	FA	FR	FA	FR	FA	FR	FA	FR	FA
1	apresentar	379,6	4.520	337,0	2.370	379,5	5.807	358,4	7.783	93,1	1.992	146,8	2.112	304,5	4.976	142,5	2.465	255,4	32.025
2	<i>utilizar</i>	226,5	2.697	155,0	1.090	303,5	4.644	241,4	5.242	49,1	1.051	64,9	934	180,7	2.954	83,5	1.444	159,9	20.056
3	realizar	202,7	2.414	153,7	1.081	166,1	2.542	178,1	3.868	68,6	1.467	85,2	1.226	257,9	4.215	84,1	1.456	145,7	18.269
4	considerar	100,4	1.195	110,5	777	123,9	1.896	147,3	3.199	92,1	1.970	98,4	1.416	150,1	2.454	109,5	1.894	118,0	14.801
5	<i>observar</i>	194,1	2.311	134,1	943	146,8	2.246	107,7	2.339	39,2	839	58,0	835	108,4	1.772	51,3	887	97,1	12.172
6	encontrar	137,6	1.639	154,5	1.087	96,8	1.481	80,0	1.738	60,6	1.297	76,0	1.094	118,3	1.933	62,6	1.084	90,5	11.353
7	ocorrer	112,0	1.334	126,5	890	106,4	1.629	97,3	2.112	49,8	1.065	55,5	799	71,5	1.169	56,6	980	79,6	9.978
8	mostrar	46,7	556	47,5	334	140,0	2.143	101,7	2.209	47,8	1.022	60,9	876	95,7	1.564	49,0	848	76,2	9.552
9	<i>analisar</i>	74,5	887	63,1	444	67,0	1.025	77,5	1.683	42,1	901	48,6	699	78,9	1.290	56,9	985	63,1	7.914
10	<i>relacionar</i>	61,1	727	53,0	373	61,4	939	70,2	1.524	40,7	871	50,3	723	100,8	1.647	60,0	1.039	62,5	7.843
11	<i>determinar</i>	80,2	955	49,3	347	75,5	1.156	69,0	1.499	50,5	1.081	47,3	680	46,3	757	64,5	1.116	60,5	7.591
12	representar	41,2	490	64,0	450	90,4	1.383	74,7	1.622	45,8	979	53,1	764	41,2	673	56,3	975	58,5	7.336
13	permitir	41,6	495	38,1	268	64,0	979	72,6	1.576	56,7	1.213	56,0	806	44,7	731	65,1	1.126	57,4	7.194
14	indicar	71,0	845	69,4	488	81,0	1.240	68,9	1.496	34,1	730	36,4	524	54,6	893	39,9	691	55,1	6.907
15	<i>possuir</i>	48,3	575	92,4	650	73,4	1.124	71,3	1.548	36,7	785	34,3	493	50,7	828	49,6	858	54,7	6.861
16	<i>desenvolver</i>	29,3	349	33,3	234	49,8	762	69,3	1.505	52,1	1.115	41,6	598	53,5	874	54,6	944	50,9	6.381
17	<i>associar</i>	49,6	591	46,2	325	64,4	985	40,3	876	27,9	598	26,2	377	113,7	1.858	33,8	584	49,4	6.194

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
18	identificar	35,4	421	50,0	352	45,1	690	54,4	1.182	40,2	861	33,2	477	75,3	1.230	51,6	892	48,7	6.105
19	propor	22,1	263	24,6	173	68,3	1.045	57,4	1.247	39,3	840	67,5	971	28,0	457	49,5	857	46,7	5.853
20	definir	27,0	322	22,5	158	55,4	848	62,9	1.367	40,3	862	44,7	643	37,3	610	41,3	714	44,1	5.524
21	produzir	41,1	489	22,0	155	41,3	632	41,6	903	52,2	1.116	69,8	1.004	15,4	251	51,4	889	43,4	5.439
22	gerar	27,8	331	25,6	180	72,8	1.114	76,9	1.669	25,0	536	26,8	386	21,3	348	46,3	802	42,8	5.366
23	<i>envolver</i>	20,9	249	14,5	102	39,1	598	47,7	1.035	47,6	1.019	45,4	653	44,4	726	56,2	972	42,7	5.354
24	<i>estabelecer</i>	28,5	339	31,8	224	31,6	484	46,2	1.004	45,9	983	50,4	725	38,0	621	54,0	934	42,4	5.314
25	<i>descrever</i>	77,6	924	51,3	361	57,7	883	39,8	865	26,1	558	40,7	586	40,5	662	27,2	470	42,3	5.309
26	estudar	64,2	765	66,0	464	42,5	650	58,3	1.266	17,7	378	18,0	259	51,0	833	23,5	406	40,0	5.021
27	<i>verificar</i>	69,5	828	51,0	359	51,8	792	53,2	1.155	13,0	279	20,4	294	50,8	830	23,5	406	39,4	4.943
28	<i>demonstrar</i>	55,7	663	28,7	202	28,8	441	31,5	685	29,1	622	24,4	351	73,0	1.193	26,8	463	36,8	4.620
29	aumentar	64,7	770	29,6	208	45,1	690	50,3	1.093	13,8	296	7,6	110	59,2	968	26,9	465	36,7	4.600
30	<i>compor</i>	41,4	493	40,5	285	41,7	638	42,5	924	24,6	527	33,5	482	41,0	670	26,5	458	35,7	4.477
31	<i>resultar</i>	50,1	596	29,3	206	36,4	557	47,4	1.029	24,1	515	23,5	338	39,3	642	30,7	531	35,2	4.414
32	<i>constituir</i>	24,8	295	35,7	251	17,4	266	22,4	487	48,6	1.040	51,3	738	28,8	471	41,3	714	34,0	4.262
33	<i>aplicar</i>	31,3	373	23,6	166	54,0	827	63,9	1.387	14,2	304	12,9	186	32,2	527	25,3	438	33,6	4.208
34	<i>contribuir</i>	33,3	396	32,3	227	30,6	469	35,4	769	28,7	615	21,8	313	43,9	717	38,4	664	33,3	4.170
35	<i>variar</i>	58,6	698	49,5	348	41,7	638	41,7	905	15,2	325	15,2	218	44,3	724	17,2	298	33,1	4.154
36	apontar	9,5	113	19,2	135	13,9	212	27,9	605	41,7	892	37,1	534	43,1	705	45,4	785	31,7	3.981
37	reduzir	45,7	544	32,7	230	35,7	546	45,8	994	16,9	362	11,0	158	37,4	611	29,0	502	31,5	3.947

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
38	<i>sugerir</i>	24,6	293	26,9	189	25,7	393	25,6	556	23,8	509	33,6	483	43,5	711	24,2	419	28,3	3.553
39	<i>adotar</i>	17,6	209	16,5	116	19,8	303	54,3	1.179	20,9	448	<u>15,6</u>	224	25,6	419	37,3	645	28,3	3.543
40	<i>compreender</i>	8,5	101	16,9	119	16,5	252	16,9	368	44,4	951	42,5	612	20,6	337	40,5	700	27,4	3.440
41	<i>adequar</i>	34,2	407	20,6	145	23,7	363	37,9	822	15,4	329	10,8	155	43,6	713	21,8	378	26,4	3.312
42	<i>conter</i>	48,8	581	19,2	135	54,4	832	26,1	567	15,0	321	17,4	251	20,0	327	15,0	260	26,1	3.274
43	<i>basear</i>	13,9	166	25,2	177	38,2	584	33,9	736	16,1	345	21,8	313	24,4	398	30,0	519	25,8	3.238
44	<i>influnciar</i>	42,6	507	29,3	206	26,3	403	31,4	682	12,2	260	13,2	190	31,6	516	23,8	412	25,3	3.176
45	<i>caracterizar</i>	28,4	338	29,6	208	25,1	384	26,1	566	22,9	491	23,8	342	21,2	346	22,9	396	24,5	3.071
46	<i>fornecer</i>	26,7	318	17,1	120	34,3	525	29,4	638	14,0	299	14,9	214	18,8	307	16,8	291	21,6	2.712
47	<i>evidenciar</i>	19,8	236	21,8	153	16,5	253	18,7	406	13,4	286	21,9	315	33,4	546	20,7	358	20,4	2.553
48	<i>possibilitar</i>	15,1	180	16,9	119	22,6	346	24,6	534	21,3	456	18,1	261	18,2	297	19,3	334	20,2	2.527
49	<i>classificar</i>	28,4	338	28,0	197	28,2	431	20,0	435	7,0	149	8,7	125	34,8	569	12,1	210	19,6	2.454
50	<i>depende</i>	16,6	198	12,4	87	22,5	344	24,5	531	18,6	397	15,8	227	12,6	206	23,1	399	19,1	2.389
51	<i>corresponder</i>	18,1	216	19,5	137	32,3	494	21,5	467	14,4	308	18,4	265	18,0	295	11,7	203	19,0	2.385
52	<i>afetar</i>	35,4	422	24,5	172	15,7	241	19,2	416	9,4	202	12,4	179	18,4	300	23,9	414	18,7	2.346
53	<i>selecionar</i>	20,8	248	30,9	217	24,6	376	20,5	446	8,2	176	14,2	205	26,6	434	8,9	154	18,0	2.256
54	<i>atribuir</i>	18,4	219	8,5	60	21,7	332	16,9	367	19,3	413	20,2	291	13,6	222	18,0	312	17,7	2.216
55	<i>consistir</i>	13,5	161	10,0	70	25,8	395	21,4	465	18,9	405	16,1	232	13,1	214	14,6	253	17,5	2.195
56	<i>alcançar</i>	10,7	127	14,1	99	14,7	225	23,0	500	18,9	405	12,6	181	17,2	281	18,2	315	17,0	2.133
57	<i>expor</i>	14,2	169	16,9	119	13,8	211	18,9	411	14,7	314	19,2	276	18,4	300	18,6	321	16,9	2.121

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
58	<i>ressaltar</i>	12,3	147	10,8	76	17,6	269	20,5	446	16,1	344	16,3	234	17,0	278	17,7	307	16,8	2.101
59	<i>refletir</i>	9,2	109	11,4	80	10,9	167	10,1	219	18,6	397	23,8	343	15,7	257	23,2	402	15,7	1.974
60	<i>permanecer</i>	18,5	220	11,7	82	9,8	150	11,8	257	21,2	453	13,3	192	20,8	340	16,1	278	15,7	1.972
61	<i>diminuir</i>	30,6	364	11,1	78	21,6	331	19,3	419	7,2	155	4,0	58	25,0	409	8,8	153	15,7	1.967
62	<i>conduzir</i>	18,5	220	15,2	107	14,9	228	15,8	344	16,9	361	13,4	193	15,8	259	14,4	249	15,6	1.961
63	<i>investigar</i>	5,2	62	4,8	34	15,8	242	15,4	335	13,4	287	19,9	286	26,4	431	16,4	284	15,6	1.961
64	<i>medir</i>	11,3	135	10,8	76	21,0	321	27,7	602	6,7	144	6,3	90	16,7	273	13,5	234	15,0	1.875
65	<i>denominar</i>	8,5	101	8,1	57	21,5	329	16,7	362	13,7	293	15,8	228	10,6	173	17,1	296	14,7	1.839
66	<i>elaborar</i>	14,1	168	11,4	80	8,8	135	12,4	270	19,0	406	17,4	251	13,2	216	16,1	279	14,4	1.805
67	<i>alterar</i>	16,6	198	12,5	88	12,5	192	14,4	312	9,8	210	10,7	154	19,6	321	14,4	249	13,7	1.724
68	<i>inserir</i>	6,7	80	13,4	94	13,6	208	16,3	353	13,3	284	17,2	247	10,0	164	13,8	238	13,3	1.668
69	<i>tender</i>	9,9	118	8,4	59	14,3	219	17,2	373	12,7	272	9,8	141	10,0	164	18,0	312	13,2	1.658
70	<i>mencionar</i>	6,7	80	8,0	56	13,3	204	11,7	253	18,5	395	20,3	292	6,7	109	15,4	266	13,2	1.655
71	<i>limitar</i>	8,1	97	7,5	53	9,7	148	13,9	302	14,9	319	10,8	156	13,6	222	18,3	317	12,9	1.614
72	<i>testar</i>	20,5	244	14,4	101	22,6	346	13,7	298	2,9	61	5,2	75	18,0	295	8,1	141	12,4	1.561
73	<i>justificar</i>	13,2	157	7,0	49	10,8	166	10,7	232	15,0	322	11,3	162	12,4	202	15,4	266	12,4	1.556
74	<i>favorecer</i>	20,1	239	10,7	75	16,6	254	12,1	262	8,8	189	7,4	107	15,9	260	9,1	157	12,3	1.543
75	<i>controlar</i>	11,8	141	9,0	63	13,3	204	15,3	332	8,0	171	6,9	99	17,2	281	12,1	210	12,0	1.501
76	<i>constatar</i>	24,8	295	13,6	96	11,5	176	13,4	290	7,7	164	6,9	99	13,5	220	7,5	129	11,7	1.469
77	<i>requerer</i>	8,1	96	3,8	27	10,5	160	12,8	278	9,8	210	9,7	139	11,0	180	12,9	224	10,5	1.314

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
78	<i>explorar</i>	4,5	54	7,0	49	10,0	153	7,1	154	14,3	305	15,7	226	5,9	96	15,7	271	10,4	1.308
79	<i>reforçar</i>	3,9	47	6,0	42	3,1	47	11,7	255	12,6	269	11,9	171	13,0	212	14,9	258	10,4	1.301
80	<i>diferenciar</i>	10,7	128	13,4	94	6,9	106	8,5	184	11,1	237	12,8	184	7,3	120	13,5	234	10,3	1.287
81	<i>aproximar</i>	7,3	87	5,1	36	14,0	214	11,7	255	11,8	252	13,5	194	4,5	74	9,2	159	10,1	1.271
82	<i>concentrar</i>	18,5	220	11,9	84	10,5	161	10,9	236	8,0	172	6,2	89	3,9	63	10,6	183	9,6	1.208
83	<i>operar</i>	3,1	37	3,3	23	8,5	130	10,6	231	14,8	317	13,8	199	2,9	48	12,7	219	9,6	1.204
84	<i>modificar</i>	12,3	146	8,8	62	14,0	214	8,1	175	7,5	161	11,3	163	8,9	146	7,0	121	9,5	1.188
85	<i>auxiliar</i>	11,9	142	8,8	62	13,4	205	11,8	257	4,6	99	5,3	76	14,1	231	6,5	113	9,5	1.185
86	<i>corroborar</i>	13,7	163	16,1	113	7,1	108	7,9	172	4,5	96	6,3	90	16,9	277	7,3	126	9,1	1.145
87	<i>incorporar</i>	2,3	27	4,8	34	7,8	119	11,7	253	12,7	271	9,9	142	4,3	70	13,1	227	9,1	1.143
88	<i>derivar</i>	11,3	135	4,5	32	17,4	266	5,0	108	8,6	183	10,3	148	3,9	64	7,1	122	8,4	1.058
89	<i>pesquisar</i>	5,4	64	2,3	16	3,9	59	12,5	271	8,0	171	4,9	71	9,9	162	9,1	158	7,8	972
90	<i>apropriar</i>	4,4	52	5,5	39	5,9	91	6,2	134	8,5	182	11,4	164	8,3	135	8,9	154	7,6	951
91	<i>introduzir</i>	4,0	48	7,0	49	8,5	130	5,5	119	8,4	179	11,1	160	4,8	78	6,5	113	7,0	876
92	<i>processar</i>	15,0	179	4,0	28	9,7	149	7,9	171	2,2	48	6,3	91	4,9	80	4,9	85	6,6	831
93	<i>demandar</i>	3,6	43	4,8	34	3,5	54	10,1	220	7,2	155	4,9	70	4,7	77	9,7	168	6,5	821
94	<i>desempenhar</i>	4,6	55	2,6	18	3,4	52	3,8	83	10,7	228	7,4	107	5,9	97	10,4	180	6,5	820
95	<i>abranger</i>	3,6	43	14,9	105	5,4	82	8,2	179	6,3	135	3,6	52	5,8	94	6,8	117	6,4	807
96	<i>dificultar</i>	7,7	92	8,2	58	7,1	109	6,3	136	3,4	72	2,8	40	10,7	175	5,8	101	6,2	783
97	<i>salientar</i>	2,8	33	5,1	36	6,3	97	7,2	157	6,9	148	7,4	107	5,4	89	6,6	115	6,2	782

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
98	<i>induzir</i>	12,6	150	3,7	26	7,5	115	5,7	123	2,8	59	2,6	37	10,4	170	5,6	97	6,2	777
99	detalhar	2,9	34	5,3	37	9,1	140	10,7	232	3,9	84	5,8	84	4,3	71	5,2	90	6,2	772
100	estruturar	2,3	27	1,7	12	4,1	63	8,3	180	5,8	125	5,4	77	7,3	119	9,7	168	6,1	771
101	propiciar	7,9	94	3,7	26	4,6	70	4,9	107	6,1	130	5,2	75	4,7	77	6,5	113	5,5	692
102	acentuar	10,7	128	6,0	42	6,0	92	5,4	117	4,5	97	5,1	73	3,6	59	4,7	82	5,5	690
103	reportar	9,5	113	9,0	63	6,8	104	5,1	110	1,8	39	6,7	97	6,1	100	3,1	53	5,4	679
104	posicionar	3,7	44	1,7	12	6,4	98	8,4	183	4,0	86	4,7	68	4,8	79	4,2	72	5,1	642
105	visualizar	8,9	106	2,3	16	7,0	107	7,8	169	2,6	56	3,3	47	3,3	54	3,5	61	4,9	616
106	correlacionar	6,6	79	7,5	53	5,8	88	5,4	117	1,0	22	1,9	28	9,5	155	3,0	52	4,7	594
107	acarretar	6,6	78	3,8	27	4,2	65	6,5	142	2,3	49	4,0	57	6,9	112	3,0	52	4,6	582
108	advir	2,1	25	2,1	15	3,3	50	3,3	72	6,0	129	6,5	94	3,5	58	7,5	130	4,6	573
109	especificar	3,5	42	3,1	22	5,0	76	8,9	194	3,1	66	3,6	52	2,9	48	3,2	56	4,4	556
110	agrupar	4,5	53	11,1	78	4,9	75	4,3	94	2,5	54	2,2	32	6,4	105	3,5	60	4,4	551
111	inferir	5,2	62	6,5	46	7,8	119	3,5	75	3,2	68	3,8	54	3,5	57	3,7	64	4,3	545
112	absorver	7,2	86	3,6	25	7,4	113	5,9	128	2,8	59	3,7	53	1,1	18	3,5	60	4,3	542
113	conectar	1,8	21	1,7	12	5,4	83	4,4	96	4,5	96	5,9	85	1,6	26	5,6	97	4,1	516
114	<i>descartar</i>	6,3	75	3,7	26	5,2	79	6,9	150	2,3	50	2,6	38	2,6	42	2,6	45	4,0	505
115	originar	6,0	71	3,0	21	4,9	75	4,8	105	4,2	89	2,8	40	2,0	33	3,9	68	4,0	502
116	restringir	2,0	24	2,6	18	2,4	36	3,4	73	5,1	110	4,1	59	3,1	50	7,1	123	3,9	493
117	tanger	1,3	16	1,4	10	1,6	24	3,7	80	4,7	100	4,9	71	4,2	68	6,3	109	3,8	478

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
118	expandir	1,4	17	2,8	20	2,3	35	3,8	83	5,4	116	4,7	68	1,5	24	6,0	103	3,7	466
119	delimitar	2,0	24	4,3	30	2,2	33	2,8	61	4,2	89	7,9	114	1,9	31	4,0	69	3,6	451
120	impactar	1,9	23	4,0	28	3,0	46	6,4	139	2,3	49	1,0	15	3,1	50	5,0	87	3,5	437
121	captar	1,2	14	1,3	9	2,9	44	4,4	96	2,9	62	4,4	63	2,0	33	6,2	108	3,4	429
122	englobar	1,5	18	2,6	18	2,8	43	5,5	120	2,9	63	3,2	46	4,3	70	2,9	51	3,4	429
123	coincidir	3,9	47	2,1	15	4,3	66	2,3	50	3,9	83	4,2	61	2,1	35	2,3	39	3,2	396
124	generalizar	2,3	27	2,6	18	3,2	49	3,2	69	3,5	74	3,2	46	2,8	45	3,8	66	3,1	394
125	desencadear	3,9	47	1,4	10	0,7	11	1,2	25	3,6	76	5,7	82	4,4	72	3,2	56	3,0	379
126	<i>predominar</i>	1,8	21	4,4	31	3,1	48	2,4	52	2,6	56	3,5	51	3,4	56	3,5	60	3,0	375
127	intensificar	1,9	23	1,7	12	5,9	91	2,7	58	2,6	55	3,6	52	1,2	20	3,5	61	3,0	372
128	desconsiderar	1,0	12	3,1	22	2,7	41	2,9	63	3,3	71	1,8	26	1,2	20	4,7	82	2,7	337
129	ponderar	1,9	23	2,3	16	4,2	65	3,0	66	2,1	45	1,4	20	2,8	46	2,8	48	2,6	329
130	reter	4,2	50	5,7	40	3,3	51	3,7	81	2,3	50	1,6	23	0,7	11	1,3	22	2,6	328
131	difundir	2,0	24	0,7	5	2,6	40	2,9	64	3,7	79	2,5	36	1,3	21	3,4	58	2,6	327
132	potencializar	1,8	21	1,1	8	1,0	16	2,4	52	2,8	59	2,8	40	3,1	51	3,3	57	2,4	304
133	atrelar	1,0	12	1,0	7	1,9	29	1,8	40	3,7	80	3,5	50	1,7	28	3,3	57	2,4	303
134	alocar	2,6	31	5,0	35	1,7	26	3,0	65	1,4	29	0,7	10	3,5	57	2,2	38	2,3	291
135	prover	1,0	12	3,1	22	1,5	23	2,9	62	2,9	61	0,9	13	1,7	28	3,4	59	2,2	280
136	sistematizar	1,1	13	1,8	13	0,5	7	1,7	38	2,7	58	2,0	29	2,5	41	3,3	57	2,0	256
137	inverter	2,5	30	1,0	7	1,7	26	1,5	33	3,6	77	2,4	35	1,0	16	1,4	25	2,0	249

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
138	conjugar	1,4	17	1,4	10	4,1	62	1,7	38	2,0	42	2,8	41	0,6	9	1,6	28	2,0	247
139	maximizar	1,3	16	3,7	26	2,4	37	3,3	72	1,2	25	0,6	9	1,0	16	2,4	42	1,9	243
140	assemelhar	1,8	21	2,3	16	1,8	27	1,2	26	1,6	34	4,4	63	1,9	31	1,4	24	1,9	242
141	mapear	0,9	11	2,0	14	3,3	50	2,0	43	1,5	33	1,9	27	1,3	22	2,1	36	1,9	236
142	nortear	0,6	7	0,6	4	1,0	16	1,4	30	1,5	33	2,6	38	3,0	49	2,8	48	1,8	225
143	embasar	1,0	12	1,1	8	0,7	10	1,7	37	1,7	37	2,7	39	2,1	35	2,4	42	1,8	220
144	documentar	1,1	13	2,1	15	1,5	23	0,9	19	1,4	31	2,2	32	3,7	60	1,2	21	1,7	214
145	oscilar	3,1	37	2,6	18	2,2	33	1,5	33	1,4	29	2,0	29	1,0	17	1,0	17	1,7	213
146	estabilizar	1,8	21	1,4	10	3,0	46	2,7	59	1,2	26	1,6	23	0,6	10	0,6	11	1,6	206
147	elucidar	1,4	17	1,7	12	0,7	10	0,9	20	2,0	42	1,3	18	2,6	42	2,4	41	1,6	202
148	supracitar	1,3	15	2,0	14	1,8	28	2,3	50	1,2	26	1,2	17	1,4	23	1,6	27	1,6	200
149	contrastar	1,4	17	1,7	12	0,6	9	0,5	10	2,6	56	3,1	45	1,0	16	1,8	32	1,6	197
150	extrapolar	0,6	7	1,0	7	0,9	14	1,5	32	1,7	36	1,7	25	2,1	34	2,3	39	1,5	194
151	impossibilitar	2,0	24	1,6	11	1,5	23	1,5	33	1,2	26	1,0	15	2,4	40	1,2	20	1,5	192
152	discriminar	0,8	9	2,3	16	2,5	39	0,7	15	1,5	33	0,5	7	2,4	39	1,4	24	1,5	182
153	alternar	1,1	13	2,3	16	2,2	33	1,9	41	0,6	13	2,4	35	0,9	15	0,8	13	1,4	179
154	negligenciar	0,6	7	0,6	4	1,0	16	1,4	30	2,3	50	1,0	14	1,3	21	1,8	31	1,4	173
155	inviabilizar	1,8	21	1,3	9	1,1	17	1,6	35	0,6	12	0,6	8	1,7	28	1,6	28	1,3	158
156	subdividir	2,4	28	2,3	16	1,7	26	1,2	25	0,4	9	1,1	16	1,2	19	0,9	16	1,2	155
157	suprimir	0,6	7	1,4	10	1,2	18	0,7	16	2,4	51	1,4	20	0,5	8	1,2	20	1,2	150

		Agrárias		Biológicas		Exatas		Engenharias		Humanas		Letras		Saúde		Sociais		Geral	
158	averiguar	0,8	9	1,6	11	0,5	8	1,5	33	1,0	21	1,3	19	1,2	20	0,9	16	1,1	137
159	divergir	0,8	9	1,1	8	0,4	6	0,7	16	0,6	13	1,5	21	2,0	33	1,3	23	1,0	129
160	influir	1,1	13	1,0	7	0,3	4	0,9	20	1,3	28	0,3	4	0,7	11	1,2	20	0,9	107

Fonte: Elaboração própria.

Legenda: (FR) Frequência relativa; (FA) Frequência absoluta.

Observa-se na Tabela 5 que, após a aplicação dos procedimentos metodológicos descritos na seção anterior, 160 verbos foram considerados acadêmicos, segundo os critérios estabelecidos. Ao lado de cada lema verbal que passou nos testes apresenta-se também sua frequência relativa e frequência absoluta, a fim de fornecer informações relevantes para a melhor compreensão dos dados. A **frequência absoluta** (FA) refere-se ao valor mais bruto proveniente das estatísticas de um *corpus*: o número total de ocorrências de um *type*, isto é, o número de *tokens*. Por sua vez, a **frequência relativa** (FR), ou *frequência normalizada* (BREZINA, 2018), é utilizada quando se requer a comparação de dois ou mais *corpora*. Para o cálculo da RF é preciso apenas ter o valor da FA e o número total de tokens do *corpus*, conforme apresentado abaixo.

Equação 2 – Fórmula para o cálculo da frequência relativa

$$\text{frequência relativa} = \frac{\text{frequência absoluta}}{\text{número de tokens no corpus}} \times \text{base de normalização}$$

A *base de normalização* é um valor arbitrário e, na elaboração da Tabela 5, utilizamos a base 100.000. Isso significa dizer que o verbo *apresentar* ocorre, em média, 211 vezes a cada cem mil palavras do *corpus*. Desse modo, a RF pode ser considerada a frequência média das ocorrências de uma palavra em função do tamanho daquele (sub)*corpus*. Essa medida é relevante não apenas ao comparar *corpora*, mas também ao facilitar a compreensão dos dados. Por exemplo, em Agrárias, *apresentar* tende a ocorrer 379 vezes a cada cem mil palavras enquanto em Letras ocorre 146 vezes, em média, a cada cem mil palavras. Tais números são mais compreensíveis do que apenas descrever que esse verbo ocorre 4.520 vezes em Agrárias e 2.112 vezes em Letras, sem levar em consideração o tamanho de cada *corpus*.

Além de apresentar quais itens lexicais passaram nos testes, e as respectivas frequências, a Tabela 5 também compara os resultados desta pesquisa com a lista de verbos acadêmicos de Miranda (2016), destacando em itálico os itens comuns a ambas listas.

É pertinente ressaltar que as duas listas foram compiladas por meio de métodos majoritariamente distintos. Embora seus itens lexicais tenham sido contrastados com *corpora* de referência (*Corpus Brasileiro* em MIRANDA, 2016; e PtTenTen11, nesta tese), a metodologia da dissertação da autora (2016) aproxima-se da composição da AWL, pois ambas trabalham com a exclusão de itens que aparecem na lista de referência. Além disso, os lemas do *corpus* de Miranda (2016) foram comparados apenas com as 3.000 primeiras palavras (*types*)

da lista de referência. Por sua vez, este trabalho comparou as listas (deste e do *corpus* de referência) em sua completude.

Diante disso, verificou-se que dos 156 verbos acadêmicos de Miranda (2016), 71 foram bem-sucedidos nos critérios aqui estabelecidos, isto é, apenas 45,5% continuam acadêmicos. Verbos como *perceber, formar, comparar, referir* e *calcular*, que estavam presentes na anterior, não fazem parte da lista atual, tendo sido excluídos por não atingirem os critérios externo e internos estipulados por esta pesquisa.

Em relação à nova lista elaborada, observou-se que todos os itens classificados ocorreram em todos os *subcorpora*, apesar do critério alcance permitir que houvesse dois *subcorpora* sem ocorrências de um determinado item. Acredita-se que isso aconteça em função da rigurosidade do conjunto de critérios, que procuraram garantir que os verbos sejam ao mesmo tempo frequentes e constantemente dispersos ao longo dos *subcorpora*.

Determinada a lista dos verbos acadêmicos, a próxima seção retoma alguns dos fundamentos apresentado no Capítulo 3 ao apresentar os testes gramaticais que norteiam a análise do sistema da transitividade dos verbos mais frequentes em artigos científicos (conforme elencados na Tabela 5).

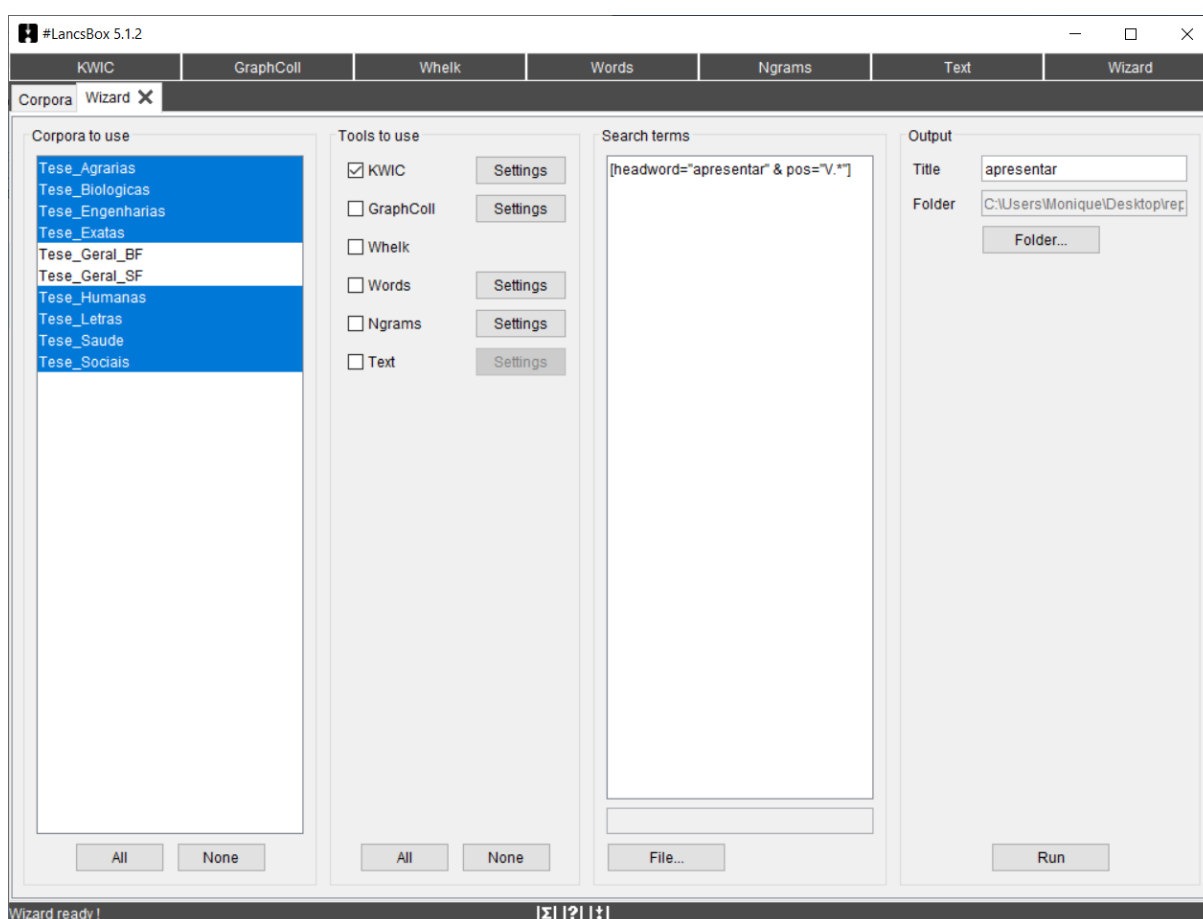
4.4 Parâmetros para distinção dos tipos de processos

Para o estudo qualitativo desta pesquisa, os lemas verbais mais frequentes foram analisados em seu contexto de uso, sendo considerados tanto dentro do próprio texto quanto dentro da Grande Área do conhecimento em que ocorreu. Para tanto, selecionou-se uma amostra aleatória de cinquenta linhas de concordância de cada *subcorpus* da tese, para cada um dos dez verbos acadêmicos mais frequentes, de acordo com os resultados apresentados e listados na Tabela 5 da seção 4.3. Desse modo, analisou-se quatro mil linhas de concordância, cada uma focando na compreensão de como o processo foi empregado no texto.

Em cada oração foram observados uma série de parâmetros correspondentes à sua instanciação, como a realização do grupo verbal quanto à efetividade (sistema de AGÊNCIA), o tempo em que acontece a oração e a presença de modalizador, assim como as características próprias de cada tipo de processo (detalhadas na seção 3.3) e as circunstâncias (cf. seção 3.3.7) acompanhantes.

Para a seleção das linhas de concordância, o programa #LancsBox (BREZINA, 2018) foi empregado, em especial sua ferramenta Wizard, que facilita a extração aleatória de amostras do trecho ao combinar todas as demais ferramentas do programa, vasculhando *corpora* e produzindo relatórios, ao mesmo tempo, resumido e detalhado em diferentes formatos (.doc, .html e .xls). A ferramenta Wizard, ilustrada a seguir na Figura 12, produz tais relatórios de forma automática, bastando selecionar o/s *corpus/corpora* e procedimentos que interessam ao usuário.

Figura 12 - Ferramenta Wizard no software #LancsBox



Fonte: *Screenshot* da ferramenta Wizard do programa #LancsBox em uso

Após a extração das linhas de concordância, os trechos foram analisados segundo os parâmetros mencionados anteriormente em planilhas do Excel, concomitantemente com a ferramenta KWIC (de #LancsBox) a fim de visualizar o contexto expandido de cada oração.

Para a análise das orações na perspectiva do sistema da TRANSITIVIDADE, uma série de critérios distintivos semânticos e gramaticais foram obedecidos (conforme pormenorizado no

Capítulo 3) e resumidos na Tabela 6 abaixo, que segue as propostas de Halliday e Matthiessen (2014) e Caffarel (2006).

Tabela 6 - Critérios para distinguir os diferentes tipos de processos

Testes gramaticais	Tipos de processos						
	material	relacional		mental	verbal	existencial	comportamental
		atributivo	identificativo				
Significado da categoria	fazer (fazer, acontecer, fazer com/a)	ser (atributo)	ser (identidade)	sentir	dizer	ser (existir)	comportar
Número de participantes obrigatórios	1	2	2	1	1	1	1
Natureza do participante obrigatório	coisa	coisa ou fato	coisa ou fato	coisa consciente	fonte simbólica	coisa ou fato	coisa consciente
Natureza do participante secundário	coisa		igual ao obrigatório	coisa ou fato			
Número de participantes passíveis de ocorrer	3	3	3	2	4	1	2
Possibilidade de apresentar Beneficiário/ Receptor	✓ beneficiário				✓ receptor		
Possibilidade de projetar	-	-	-	projeção de ideias	projeção de locuções	-	-
Possibilidade de metafenômeno				✓			
Voz	efetiva ou média	efetiva ou média	efetiva ou média	efetiva ou média	efetiva ou média	média	média

Fonte: Adaptada de Halliday e Matthiessen (2014, p. 354) e Caffarel (2006)

Quanto à distinção entre os tipos de processos, cabe ainda ressaltar que alguns verbos podem ser classificados como um ou mais tipos de processos – conforme demonstrado nas análises a seguir – em função de seu contexto de uso. Halliday e Matthiessen (2014) e Thompson (2013) ilustram essa possibilidade com os verbos *suggest* e *indicate*, definindo-os como processos relacionais quando os elementos nominais na estrutura transitiva da oração são

abstrações ou como processos mentais quando o sentido com que o verbo utilizado é cognitivo – assim como quando o processo *projeta* uma ideia. Exemplos de ocorrências similares em nossos dados foram ilustradas adiante neste trabalho.

Além dessas características, também deve ser observada a importância de analisar as palavras que acompanham os processos como mecanismo para auxiliar sua distinção, em especial, as colocações com clítico *se*. Na literatura, Morais (2013, 2016a, 2016b), Vasconcelos (2013), Ferreira e Vieira (2018) discorrem sobre o tema. Na presente análise, tal estratégia de categorização mostrou-se similarmente frutífera.

Resumidos os critérios que embasam esta análise, o próximo capítulo descreve os resultados encontrados, primeiro de modo abrangente para, em seguida, apresentá-los organizados por tipo de processo, juntos aos seus usos nas diferentes Grandes Áreas de conhecimento.

5 ANÁLISE DOS TIPOS DE PROCESSOS ENTRE OS LEMAS FREQUENTES

O capítulo anterior determina os passos metodológicos para o estabelecimento de uma lista de verbos relevantes e constantes – acadêmicos – em um *corpus* de artigos científicos. Visando analisá-los a partir da perspectiva da transitividade da LSF, as próximas seções fazem uso dos parâmetros apresentados anteriormente, na seção 4.4, para observar como as amostras selecionadas são empregadas nos textos de acordo com os tipos de processos que os referidos verbos instanciam.

5.1 Análise do sistema da transitividade

Conforme introduzido na última seção do capítulo anterior (4.4), aplicados os critérios para seleção dos verbos, assim como os de análise de seus usos por meio de linhas de concordância expandidas, foram estudadas quatro mil amostras aleatórias dos dez verbos considerados acadêmicos mais frequentes: *apresentar, utilizar, realizar, considerar, observar, encontrar, ocorrer, mostrar, analisar e determinar*. Tais verbos realizaram majoritariamente processos materiais e relacionais, seguidos por mentais e existenciais e por algumas poucas ocorrências de verbais. A seguir, a Tabela 7 apresenta a distribuição dos tipos de processos instanciados pelos verbos analisados.

Tabela 7 – Distribuição dos tipos de processos por verbos analisados

	Processos materiais	Processos relacionais	Processos mentais	Processos existenciais	Processos verbais
apresentar	55	344	-	1	-
utilizar	400	-	-	-	-
realizar	389	-	-	11	-
considerar	-	311	89	-	-
observar	-	-	396	-	4
encontrar	215	185	-	-	-
ocorrer	-	-	-	400	-

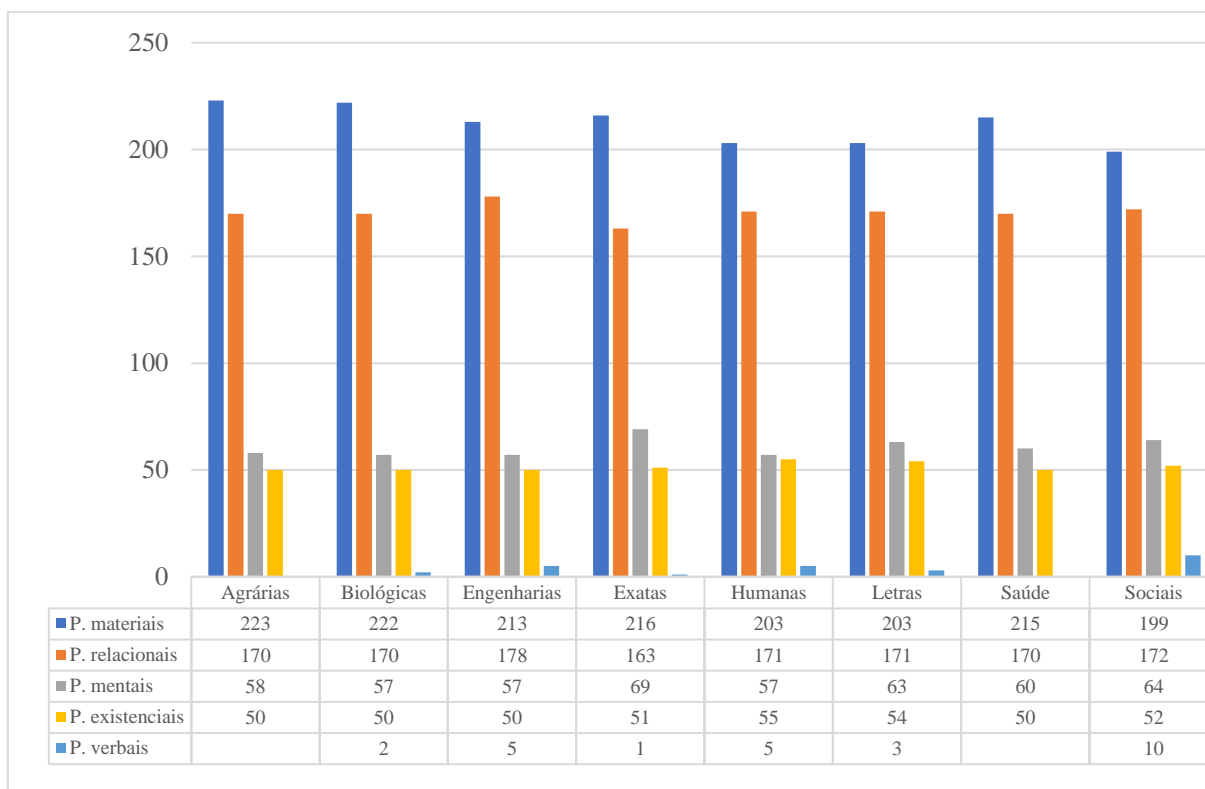
mostrar	-	400	-	-	-
analisar	400	-	-	-	-
determinar	259	123	-	-	18
Total	1718	1363	485	412	22

Fonte: Elaboração própria.

Conforme pode ser observado na Tabela 7, apenas quatro verbos (*analisar, utilizar, mostrar e ocorrer*) instanciaram um único tipo de processo enquanto os demais instanciaram dois ou mais processos distintos.

No computo geral, os resultados encontrados mostram-se em conformidade com os demais estudos voltados à investigação do sistema da TRANSITIVIDADE em diferentes gêneros acadêmicos (MENDES, 2013; ZHEN; YANG; GE, 2014; VATHANALAOHA e TANGKIENGSIRISIN, 2018), nos quais os processos materiais, relacionais e mentais (nessa ordem) são os mais frequentes. A fim de comprovar essa mesma tendência, os dados foram analisados ao longo das diferentes Grandes Áreas, demonstrando que tal padrão também pode ser verificado – nas mesmas proporções – em cada um dos oito *subcorpora* e confirmando, desse modo, que esta é uma característica do gênero. A seguir, o Gráfico 1 ilustra essa tendência ao longo das Grandes Áreas de avaliação.

Gráfico 1 - Contagem dos tipos de processos por Grande Área do conhecimento



Fonte: Elaboração própria.

Concomitantemente aos processos analisados ocorrerem em todos os subcorpora, cabe observar que ocorreram com distribuição similar de frequência nas diferentes Grandes Áreas. Essa característica encontrada nos dados ajuda a corroborar a tese de que tais verbos são relevantes e constantes em todas as subáreas e, à primeira vista, seus usos estão distribuídos similarmente ao longo das disciplinas. Dito isso, nas próximas seções cada tipo de processo é analisado e exemplificado em relação ao contexto em que ocorreram assim como ao *subcorpus*/Grande Área do conhecimento.

5.1.1 Orações materiais

Dentre os tipos que ocorreram nas amostras analisadas, os mais frequentes foram os processos materiais, que expressam a noção de que alguma entidade *fez* algo, incluindo eventos e acontecimentos físicos ou abstratos. As orações materiais correspondem a 43% das orações totais, sendo também as mais frequentes em todos os *subcorpora*. Essa porcentagem elevada

de processos materiais reflete uma propriedade do gênero que é o uso dos processos materiais pelos autores dos textos a fim de descrever e relatar, majoritariamente, procedimentos metodológicos, análises e contribuições suas ou de outros pesquisadores no texto.

Nos dados, os verbos *analisar*, *apresentar*, *determinar*, *encontrar*, *realizar* e *utilizar* instanciaram, ao todo, 1.694 orações materiais – conforme apresentado na Tabela 7 da seção anterior. Todos esses verbos realizaram também ao menos um outro tipo de processo, com exceção de *analisar* e *utilizar*, cujas ocorrências foram todas enquanto processo material.

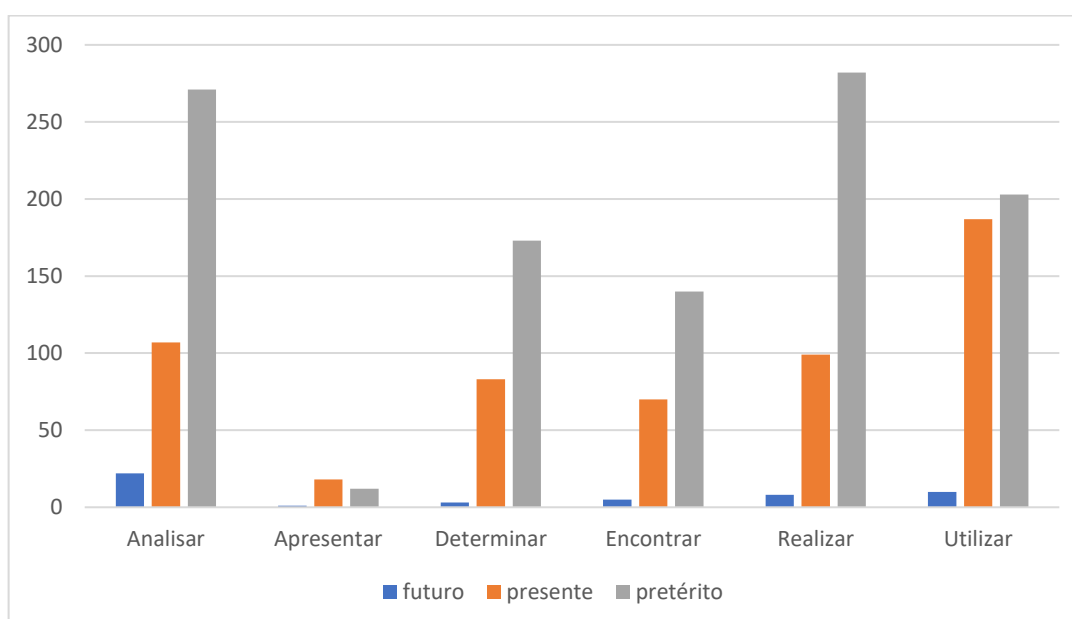
De tais orações, apenas *realizar* instanciou orações materiais criativas enquanto os demais usos corresponderam a orações materiais transformativas. Abaixo, alguns trechos exemplificam os empregos mais comuns dos processos materiais. Em (114), o processo material é empregado com significado de ‘chegar a’, indicando o caminho percorrido pelo Ator (“Esses elétrons”) até alcançar a Meta (“o contra eletrodo”). Já em (115) e (116), os processos materiais realizados pelo verbo *apresentar* significam ‘expor’ algo, a Meta – “uma revisão de literatura” em (115) e “o ‘Plano nacional de estímulo...’” em (116), pelos autores do texto em (115), e por “a comissão” em (116). Em (117), o processo material é empregado a fim de relatar as descobertas de outro estudo, sendo esse (“Panduro (2012)”) o Ator e o que foi descoberto a Meta (“níveis moderados de herdabilidade”). Já em (118), o processo material reporta a ação de um dos participantes envolvidos na pesquisa (“a Lei no 3.459”), significando a imposição da Meta, “a implantação do ensino religioso confessional”. Em (119), a oração material expõe o uso que o Ator (“nosso estudo”) fez da Meta (“as lista (sic) de espécies ameaçadas”), de modo a descrever uma das etapas dos procedimentos realizados pelo/s autor/es do texto. Por sua vez, as orações nos trechos de (120) a (122) omitem o Ator envolvido no processo material (em todos os casos refere-se ao próprio autor do texto), focando naquilo que foi realizado – seja a ação, por meio do processo (*realizar*, *analisar* e *determinar*), ou o que foi feito, por meio da Meta (“o levantamento”, em 120; “a capacidade...”, em 121; e “Os teores totais de ferro e cromo”, em 122).

- (114) Esses elétrons percorrem um circuito externo até **encontrarem** o contra eletrodo (FTO||Pt). (E_QN_2018_5_AA_001)
- (115) Aprofundando essa questão, **apresentamos** a seguir uma revisão da literatura com dois objetivos específicos: (L_ALFA_2014_0_AA_004)
- (116) A comissão **apresentou** o “Plano nacional de estímulo à produção de automóveis e à implantação gradativa da indústria automobilística”, estabelecendo suas diretrizes básicas. (So_OS_2017_2_AA_002)

- (117) Por outro lado, Panduro (2012) **encontrou** níveis moderados de herdabilidade para rendimento de frutos e peso (B_RCR_2014_0_AA_006)
- (118) em setembro de 2000 seria promulgada pelo então governador Antony Garotinho a Lei no 3.459, que **determinou** a implantação do ensino religioso confessional nas escolas públicas do Rio de Janeiro. (H_CP_2017_2_DO_007)
- (119) Além disso, nosso estudo no nível estadual **utilizou** as lista de espécies ameaçadas em nível global e nacional (B_ISZ_2017_sup_AA_051)
- (120) O levantamento **foi realizado** no portal da Câmara de Deputados (H_CP_2017_2_DO_008.txt)
- (121) Nesta seção **é analisada** a capacidade que as funções aproximadoras (Bézier G 1, Phong, Nagata, NLSA e PN) possuem em representar continuidade dos planos tangentes nas interfaces dos elementos triangulares (E_TSC_2018_2_AA_002)
- (122) Os teores totais de ferro e cromo **foram determinados** por extração nítrico-perclórica na fração argila e **analisados** por espectrofotometria (En_ESA_2018_1_AA_004)

Com relação ao tempo, as orações materiais foram classificadas conforme sua realização no pretérito, presente ou futuro – sem adentrar em classificações mais detalhadas – e sua relação com a construção da oração dentro do texto. Como pode ser observado no Gráfico 2 a seguir, quando analisados os resultados do *corpus* como um todo, os processos materiais foram majoritariamente empregados no pretérito, sendo particular e frequentemente associados à descrição das ações dos pesquisadores (o próprio autor ou outros), havendo um número relevante de usos no presente e, com uma frequência menor, no futuro.

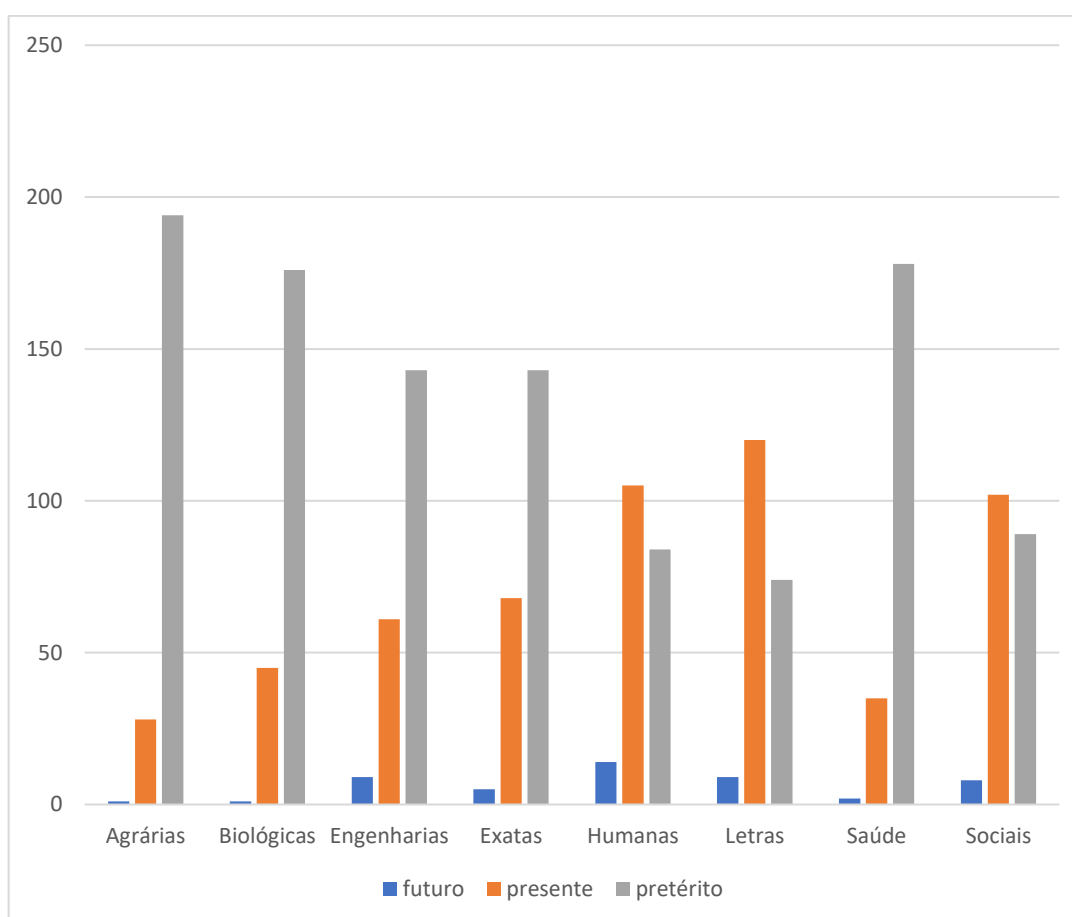
Gráfico 2 - Tempo verbal em orações materiais



Fonte: Elaboração própria.

Quando esses dados são esmiuçados e analisados filtrando-os por *subcorpora*, conforme demonstrado no Gráfico 3 a seguir, é possível perceber que as Grandes Áreas de Ciências Agrárias, Ciências Biológicas, Engenharias, Ciências Exatas e Ciências da Saúde tendem a empregar os processos materiais no pretérito com frequência bastante superior aos usos nas Grandes Áreas de Ciências Humanas, Ciências Sociais e Letras, Linguística e Artes.

Gráfico 3 - Tempo verbal em orações materiais por Grande Área



Fonte: Elaboração própria.

Em geral, ao longo dos *subcorpora*, o uso do processo no tempo presente associou-se a orientações e/ou atitudes tomadas pelo pesquisador ou por terceiros (exemplificado em 123), ao relato de atividades realizadas por terceiros (124), à descrição de resultados esperados em

uma possível replicação da pesquisa (125) ou ao lidar-se com a fundamentação teórica/discussões sobre a teoria empregadas no trabalho (126).

- (123) Mitchel e Collins [7] **determinam** a equação do ângulo de inclinação das bielas, estabelecendo um modelo teórico mais geral, que denominam de teoria do campo de compressão diagonal. (En_RIEM_2017_1_AA_006)
- (124) O decisivo, portanto, consiste nessa “lente dialógica” pela qual Habermas **analisa** a vida política e social. (H_LN_2017_3_AA_006)
- (125) Para redução da sua idade ao primeiro acasalamento, **podem ser utilizadas** pastagens anuais de inverno (A_ABMVZ_2018_3_AA_020)
- (126) Em outras palavras, o capitalismo **utiliza** e se beneficia do confinamento da mulher à esfera reprodutiva, (H_CP_2017_2_DO_010)

Por sua vez, o tempo pretérito (em orações efetivas operativas ou receptivas) associou-se a ações dos pesquisadores (implícitos em efetivas receptivas ou em construções com clítico *se* indeterminador do sujeito), como exemplificado nos trechos (127), (128) e (129); e aos passos metodológicos envolvidos na pesquisa (em 130). Pode-se perceber nos trechos de (123) a (130) que os usos dos processos materiais foram similares, sendo a preferência por um tempo ou outro percebida majoritariamente em função das Grandes Áreas em que ocorreram.

- (127) Os perfis **foram analisados** no aplicativo “GPR Bathymetry”, aplicando ao radargrama a função “set time zero” para a localização e remoção da onda direta (E_BCG_2017_1_AA_002)
- (128) Além disso, não **foram encontrados** estudos desenvolvidos no Brasil. (En_GP_2017_1_AA_011)
- (129) **Realizou-se** um estudo retrospectivo dos casos de necropsia de cães diagnosticados no Setor de Patologia Veterinária da Universidade Federal do Rio Grande do Sul (SPV-UFRGS) no período de 2002-2013. (A_PVB_2014_0_AA_002)
- (130) Para a visualização da reação, **utilizou-se** o cromógeno diaminobenzidina (DAB) (Spring, 3,3'-diaminobenzidina - cod. DAB-060, CA). Para o controle positivo, **foi utilizada** tonsila humana e, para o controle negativo, **foi utilizado** baço de peixe, sem a injeção de HC, a 2,5%. A contracoloração **foi realizada** com hematoxilina de Harris, e a montagem das lâminas com Permout (Fisher Scientific, cod. S70104, CA). A contagem das células imunomarcadas **foi realizada** sob microscopia de luz, com equipamento para fotomicrografia digital (Nikon E200, BR). (A_ABMVZ_2018_1_AA_016)

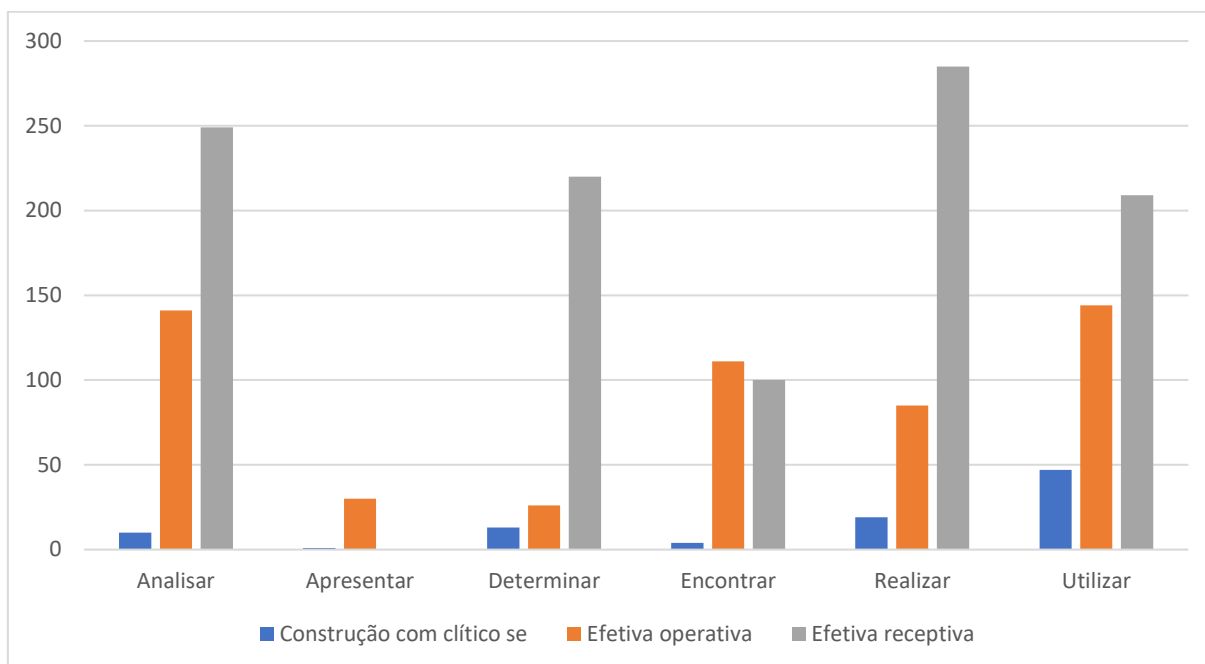
Por último, diferentemente do pretérito e do presente, o tempo futuro foi empregado para apresentar recomendações da pesquisa ou de outras fontes, como em (131), ou para compor a narrativa teórica, como em (132). No trecho (133), os fatos ali expostos descrevem a

metodologia empregada pela pesquisa, de modo que elas já existem/pertencem ao passado. Entretanto, ao colocá-los no futuro, são ressignificados a fim de construir o texto e a metodologia junto com o leitor. Além desses usos, e com frequência consideravelmente maior, o tempo futuro foi empregado para apresentar o trabalho e introduzir as seções do artigo acadêmico, conforme (134) e (135) respectivamente. Por meio do uso do tempo futuro, a estratégia descrita anteriormente é replicada – nesses trechos, que apresentam a proposta e as seções do trabalho como “futuras”, elas apenas são futuras para o leitor do texto, pois pressupõe-se que o autor já as conheça, assim como sua existência.

- (131) A articulação entre o IGC e o CI, no entanto, está explicitada no Parágrafo Único do artigo 33-C da Portaria Normativa nº 40 de 2010: “As Comissões de Avaliação **utilizarão** o CPC e o IGC como referenciais orientadores das avaliações in loco de cursos e instituições, juntamente com os instrumentos referidos no art. 17-J e demais elementos do processo” (BRASIL, 2010). (H_AC_2018_1_AA_009)
- (132) Admite-se que todo o material hidrolisado **encontrará** condições ambientais favoráveis (substrato e microrganismos) para ser completamente eliminado a uma determinada taxa que propuseram no modelo. (En_ESA_2018_2_AA_013)
- (133) A verificação estatística da significância dos tratamentos será feita pela análise de variância (ANOVA) GLM. Para a comparação das médias, **será utilizado** o teste de Tukey, com 5% de probabilidade. Todas as análises **serão realizadas** pelo software estatístico SAS, versão 9.1.2 (SAS, 2009). (A_ABMVZ_2018_4_AA_006)
- (134) INTRODUÇÃO Este trabalho **analisará** a variação regional do português brasileiro (PB) com foco no viés perceptivo dos estudos da linguagem. (L_LH_2017_1_AA_003)
- (135) Nesta seção, **analisaremos** a energia mecânica para obtermos algumas informações prévias do movimento do duplo cone. (H_RBEF_2014_0_AA_008)

Quanto à realização do grupo verbal, analisando em relação a cada processo, a maior parte das orações foram efetivas receptivas, seguidas por efetivas operativas e por orações médias, conforme apresentado a seguir.

Gráfico 4 - Tipos de construções em orações materiais



Fonte: Elaboração própria.

Como pode ser observado no Gráfico 4, à exceção é *encontrar*, a maior parte das orações materiais foram empregadas em construções efetivas receptivas com Ator omitido da oração. Nesses casos, a maior parte das omissões puderam ser facilmente recuperadas seja (majoritariamente) como referentes ao/s próprio/s autor/es do texto ou a outros pesquisadores (expressos por meio de circunstâncias ou em orações adjacentes).

Essa omissão do Ator é uma estratégia que visa desfocar a atenção do autor do texto (quem realizou a ação/Ator) para dar destaque à mensagem/Meta, provocando um efeito de objetividade na construção textual. Assim, por meio da lacuna no preenchimento dos Atores nas orações materiais (que frequentemente fazem referência ao/s próprio/s autor/es), os artigos apresentam ao leitor a informação de modo conciso, direcionando sua atenção à Meta, ou seja, àquilo que foi realizado pelo Ator omitido.

Desse modo, há uma mudança no foco do leitor, retirando-o de quem realiza as ações para a ação realizada, ao mesmo tempo em que essa estratégia distancia o autor de seu texto, atribuindo-lhe a impressão de um texto impessoal e objetivo, mesmo quando relata a descrição dos procedimentos realizados pelos autores para a elaboração da pesquisa. Os trechos (136) a (138) a seguir exemplificam essa estratégia. Neles, o relato da confecção da pesquisa se mostra claro e objetivo, realizado linguisticamente por meio de sentenças efetivas receptivas curtas acompanhadas de uma única circunstância, geralmente de Localização: Tempo, Causa:

Finalidade ou Modo: Meio. Através do uso dessas circunstâncias é possível descrever o que foi feito, qualificando-o ou expandindo-o, mas sem inserir reações pessoais do autor o que foi feito.

- (136) No soro [Circunstância Localização: Lugar], **foram determinados** os valores de proteínas totais, globulinas, albuminas, hemoglobina e ureia (A_ABMVZ_2018_4_AA_016)
- (137) Para a obtenção das medidas [Circunstância de Causa: Finalidade] **foram utilizadas** réguas de escala padrão para as estruturas maiores (B_HOE_2018_2_AA_003)
- (138) A análise de conteúdo **foi realizada** por meio da sua interação com a análise de Keywords [Circunstância de Modo: Meio] (En_GP_2018_2_AA_011)

Ainda dentro do tópico sobre a renúncia da participação no texto pelo autor, observou-se ainda outras estratégias empregadas, como o atenuamento da presença explícita do/s autor/es ao fazer uso de expressões metonímicas como “este estudo” e “a presente pesquisa” (exemplificados nos trechos 139 e 140, a seguir) ou de marcar a autoria (própria ou de outros) por meio de Circunstâncias, como em (141) e (142).

- (139) Portanto, nosso projeto de pesquisa **utilizou** uma combinação de saberes que enfocavam todos os três níveis - mente, corpo e espírito. (L_RBEP_2018_4_AA_003)
- (140) DISCUSSÃO O presente estudo **analisou** a associação entre os valores dos níveis da FeNO e do VEF1 basal e o grau de controle da asma segundo o GINA (Sa_RPP_2018_1_AA_003)
- (141) O tema **foi analisado** em (SILVA, 1998) e (JORION, 1998). (E_SBA_2012_0_AA_001)
- (142) Nesta análise, **foram determinadas** razões de prevalência (Sa_RLAE_2017_0_AA_031)

Além da possibilidade de marcar (ou não) a presença do autor, o Ator também foi comumente empregado para referenciar ações de outros pesquisadores (como em 143), ações de participantes envolvidos na pesquisa (144) ou ações realizadas por outros tipos de entidades – como em (145) e (146).

- (143) O número total de entrevistas **foi determinado** pelo pesquisador (Sa_RLAE_2017_0_AA_013)
- (144) Assim, montamos dois experimentos de percepção que **foram realizados** por nove ouvintes. (L_LH_2017_1_AA_004)
- (145) As imperfeições do mercado fazem com que, ao invés de as empresas **utilizarem** métodos alternativos para expandirem suas atividades além das fronteiras domésticas, como o licenciamento,

a exportação e a venda ou o aluguel de tecnologia, elas preferiram internalizar³ suas operações internacionais. (H_COIN_2014_0_AA_004)

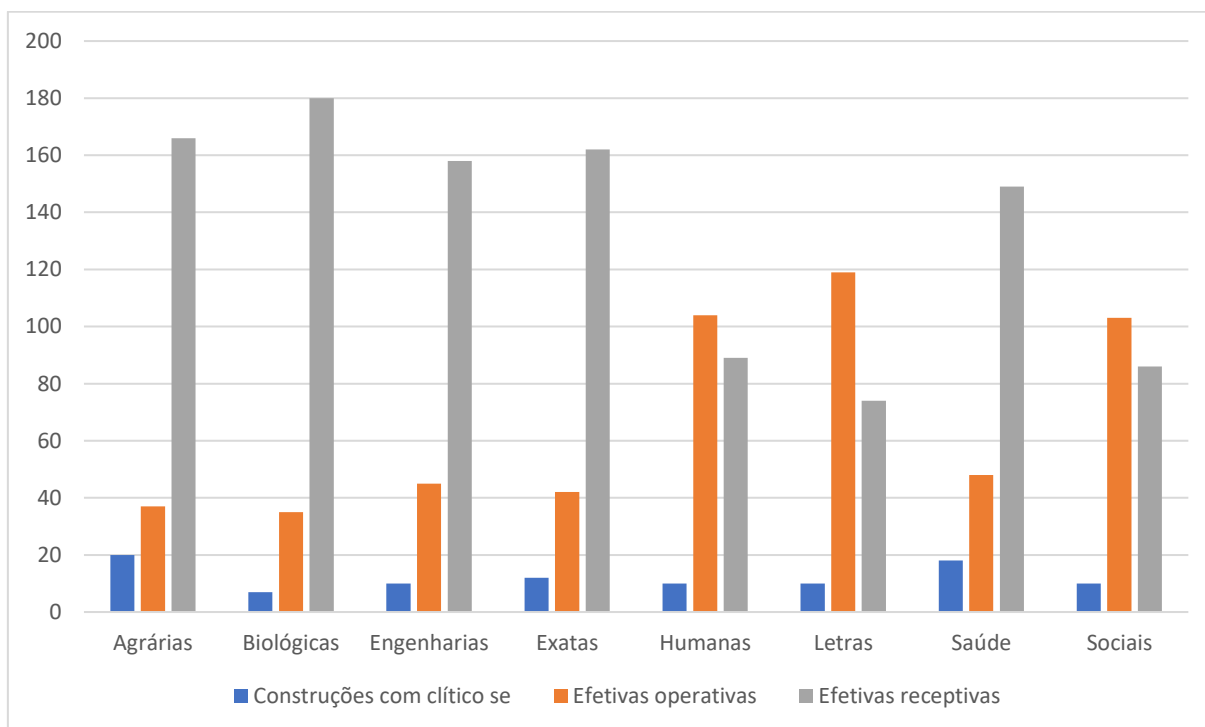
- (146) Mesmo não havendo algum dispositivo legal semelhante, o STJ **poderia encontrar** um artigo de lei “substitutivo” por intermédio de interpretação extensiva (So_RDGV_2014_0_AA_001)

Há também casos em que o Ator foi omitido por meio de uma construção com o clítico *se*. Nesse tipo de construção é possível que haja mudança no significado (e na construção) da oração, de modo que o verbo passa a realizar outro tipo de processo – conforme algumas ocorrências observadas nesta análise. Entretanto, nos casos exemplificados a seguir, a presença do “se” é um índice de indeterminação do Ator.

- (147) Da constelação de cinco satélites RapidEye, **utilizaram-se** as cenas 2226120 e 2226121, datadas de 27 de setembro (E_BCG_2017_3_AA_001)
- (148) Em seguida realizou-se uma avaliação qualitativa sobre a evolução do AR. (H_DADO_2014_0_AA_004)
- (149) Muitos professores deste componente curricular identificaram conteúdos abordados nas questões do ENADE, o que não foi equivalente quando **se analisa** a visão dos alunos. (H_AC_2017_1_AA_006)
- (150) Assim é que **se determina** o nível agregado de investimento. (So_NE_2018_1_AA_005)

Retomando a questão do modo de realização das orações materiais – como efetivas operativas, receptivas ou construções com clítico *se* –, cabe ainda observar que a finalidade de seus usos foi similar ao longo dos diferentes *subcorpora*, embora a frequência dos tipos varie conforme a Grande Área. Como demonstrado no Gráfico 5, seguindo o mesmo padrão observado no Gráfico 3 anterior, Ciências Humanas, Ciências e Letras tendem a empregar os processos materiais em orações efetivas operativas, enquanto as demais Grandes Áreas tendem a optar por orações efetivas receptivas.

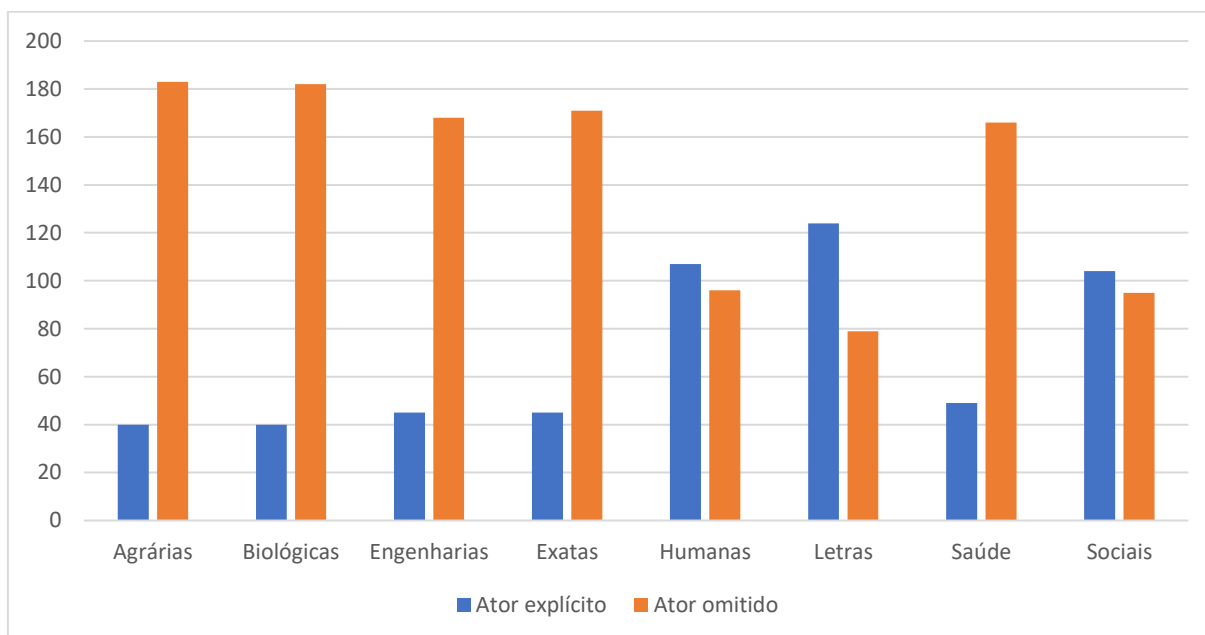
Gráfico 5 - Tipos de construções em orações materiais por Grande Área



Fonte: Elaboração própria.

Por sua vez, a seguir, o Gráfico 6 mostra que os dados ou os objetos de estudo das áreas de Ciências Humanas, Ciências Sociais e Letras, Linguística e Artes tendem a *realizar* mais ações – o que é congruente com o fato de que essas disciplinas são voltadas ao estudo da sociedade como um todo e, portando, lidam com seres humanos, tornando necessário explicitá-los no decorrer das ações realizadas em orações materiais. Ainda cabe observar que o Gráfico 6 segue coerente com o anterior, Gráfico 5, demonstrando que em tais áreas a frequência de orações materiais com o Ator marcado é consideravelmente – o dobro – maior do que nas demais.

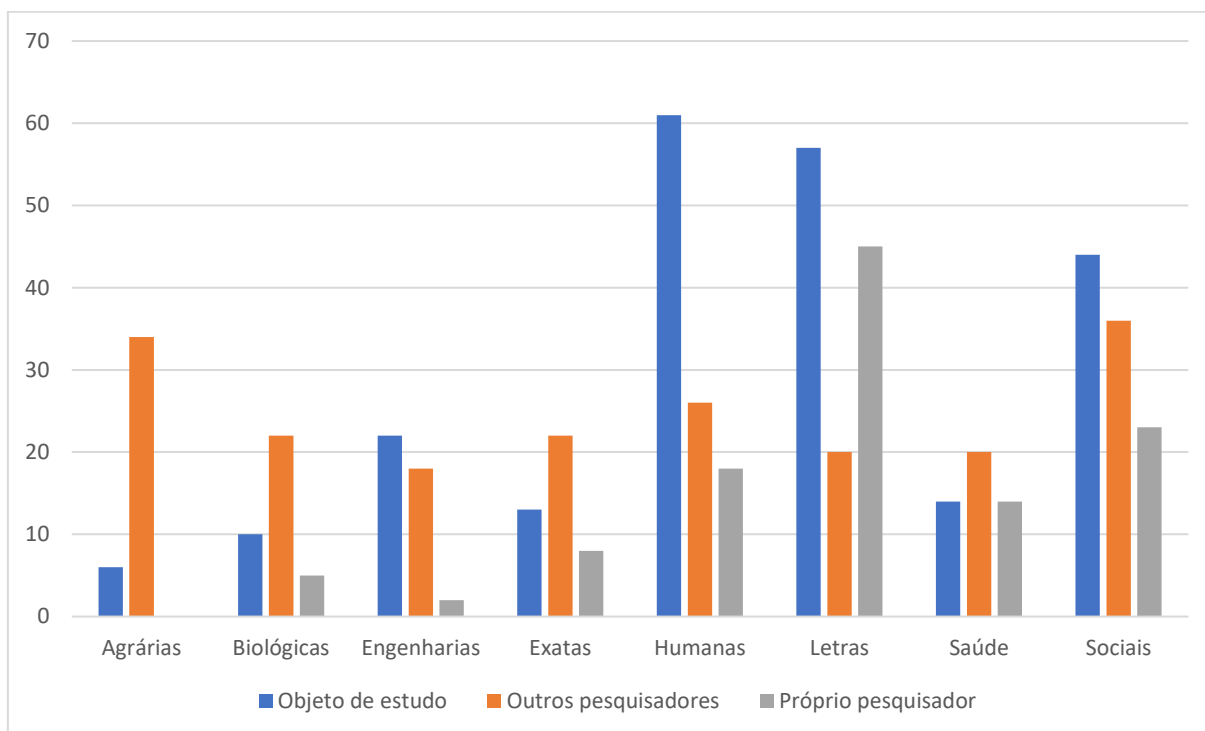
Gráfico 6 – Tipos de Ator em orações materiais por Grande Área



Fonte: Elaboração própria.

Adentrando ainda mais nessa análise, conforme dados apresentados no Gráfico 7 a seguir, pode-se perceber que nas Grandes Áreas mencionadas o Ator representa majoritariamente o objeto em estudo no texto – o que é congruente com os temas abordados por essas áreas. Esse tipo de apresentação dos dados analisados nesta pesquisa também permite verificar que resultado similar – embora em proporção menor – é observado no *subcorpus* de Engenharia. Esse fato pode ser igualmente atribuído à natureza material e operativa do objeto estudado.

Gráfico 7 - Tipos de Ator explícito em orações materiais por Grande Área



Fonte: Elaboração própria.

Ainda é preciso destacar que a classificação “Objeto de estudo” não se limita a entidades humanas ou conscientes (como em 151) responsáveis pela ação realizada na oração, mas também a seres inanimados ou abstratos (como em 152).

- (151) Desses cinco informantes, um **utilizou**, além da pontuação, a pergunta retórica no início do excerto, (L_RD_2017_1_AA_003)
- (152) No caso do escritor Mário de Andrade, por exemplo, a noção de primitivo **realizou** um movimento que perfez o surgimento de uma “atitude estética” até aproximar-se da afinidade entre a beleza e o ideal de coletividade. (L_ASP_2018_1_AA_004)

Em resumo, as orações materiais foram empregadas similarmente ao longo das Grandes Áreas de conhecimento, com a finalidade de descrever e relatar experimentos, análises, descobertas, recomendações e contribuições. Entretanto, ao analisar esses usos levando em consideração as diferentes Grandes Áreas em que ocorrem, é possível observar que há entre as disciplinas preferências particulares no modo em que cada uma empregou essas orações. Tais diferenças, em nossa amostra, parecem ser motivadas pelo próprio objeto de estudo, assim como

por características próprias daquela área de conhecimento. Na próxima seção, o foco da análise está na instanciação dos processos relacionais.

5.1.2 Orações relacionais

Os processos relacionais foram o segundo tipo mais frequentemente empregados nos dados analisados, com 1.365 ocorrências realizadas pelos verbos *apresentar*, *considerar*, *determinar*, *encontrar* e *mostrar* – destes, apenas *mostrar* teve todos os usos analisados enquanto processo relacional. Conforme a revisão teórica apresentada na seção 3.3.3, as orações relacionais realizam uma figura de ‘ser’ que expressa a relação entre duas coisas ou conceitos abstratos, de modo a caracterizá-los (*atributivos*) ou identificá-los (*identificativos*). Relacionados aos textos acadêmicos, que são textos descritivos, argumentativos e expositivos, esses processos tendem a ser altamente informativos (HALLIDAY, 1985) e, portanto, utilizados com frequência na construção dos textos a fim de contribuir com descrições e especificidades do conteúdo.

Um exemplo prototípico dos usos das orações relacionais está no trecho (153), a seguir, que é a reprodução da conclusão de um dos artigos do *corpus*. No trecho, os processos relacionais funcionam de modo a retomar e apresentar os resultados descobertos pela pesquisa, empregando-os como argumento para a justificar a relevância (e contribuições) do artigo para a área de conhecimento.

- (153) 4 CONCLUSÃO Nossos resultados **mostraram** que é factível a utilização do exsudato cárneo para detecção de IgG anti-T. gondii, em cortes de carne bovina, pela técnica de ELISA, utilizando diluição única e padronizada, sendo esta uma ferramenta útil para o controle da qualidade de T. gondii em cortes comerciais de carne, devido a otimização, confiabilidade e reprodutibilidade dos testes (MECCA *et al.*, 2011). Entretanto, para carnes de sol, nossos resultados demonstram que o processamento da carne pela adição de cloreto de sódio interfere na exsudação da carne, limitando a utilização do exsudato cárneo como material para detecção de IgG anti-T. gondii em carnes processadas, pois tal detecção depende da qualidade e quantidade do exsudato cárneo. Vale ressaltar a importância de avaliar outros métodos de obtenção do exsudato cárneo, como, por exemplo, o uso de solução salina e o maceramento do tecido muscular. Neste contexto, a abordagem proposta em nosso trabalho constitui uma ferramenta bastante promissora para a rotina de Laboratórios de Saúde Pública, para monitoramento de carnes destinadas ao consumo humano. Também pode ser utilizada para elucidação de surtos epidêmicos envolvendo a carne bovina, abrindo perspectivas futuras de aplicação desta metodologia não só para a toxoplasmose, mas também para o controle de outras zoonoses transmitidas pela carne e de importância em Saúde Pública. (A_BJFT_2018_0_AA_014)

Quanto aos processos relacionais atributivos, esses foram frequentemente empregados a fim de elaborar a informação descrita pelo autor, procedimentos ou a existência de descobertas (teóricas ou não) relevantes para o texto. No trecho (154), há dois participantes da oração relacional atributiva intensiva com sentido de ‘ser’, um é o Portador (“Esse marcador de fase líquida”), o outro é o Atributo (“seguro e de fácil aplicação”). No trecho (155) o verbo “encontrar”, acrescido do clítico *se*, é empregado com sentido de ‘estar’ (cujo verbo é empregado similarmente mais adiante no trecho) em uma oração relacional atributiva circunstancial, cujos participantes são o Portador, “O anuro *Lepidobatrachus asper*”, e o Atributo, “sob ameaça”. Já em (156), as orações relacionais em destaque são atributivas possessivas, expressando uma relação de posse entre Portador (“a criança que...” e “as crianças pertencentes...”) e Atributo (“melhor regulação do apetite” e “maior risco de estarem no grupo de GPR”).

- (154) Esse marcador de fase líquida **se mostrou** seguro e de fácil aplicação e permitiu assegurar que os opioides em questão utilizados pela via epidural não prejudicam o trânsito gastrointestinal. (A_ABMVZ_2018_5_AA_008)
- (155) O anuro *Lepidobatrachus asper* **encontra-se** sob ameaça no Paraguai (Quase Ameaçado - NT) devido à sua área de distribuição ser muito restrita (aproximadamente 20.000 km²) e suas populações estarem em declínio, provavelmente devido a incêndios e conversão de áreas naturais em pastagem (Aquino *et al.*, 2004). (B_ISZ_2017_sup_AA_048)
- (156) Além disso, a criança que recebe aleitamento materno frequentemente **apresenta** melhor autorregulação do apetite e, portanto, fica menos suscetível ao estabelecimento de horários rígidos e ao consumo de quantidades determinadas de leite. Adicionalmente, as crianças pertencentes a famílias de classes socioeconômicas mais baixas **apresentaram** maior risco de estarem no grupo de GPR. (Sa_RPP_2018_3_AA_006)

Por sua vez, os processos relacionais identificativos foram utilizados quando houve necessidade de esclarecer certos termos, ou seja, na definição de terminologias ou na descrição de informações importantes para a compreensão do texto. Esse tipo de uso é fundamental para que o leitor compreenda porque algo foi escolhido/lido/experimentado em comparação às demais possibilidades. Desse modo, o uso de orações relacionais identificativas associa-se à apresentação de resultados, tópicos ou ideias fundamentais para a discussão, assim como para enriquecimento da argumentação que está sendo construída. Em (157), a oração relacional identificativa intensiva é empregada de acordo com a finalidade do subtipo “demonstrar” (cf. seção 3.3.3), relacionando o Identificado “os resultados” ao Identificador “a possibilidade de se identificarem...”. Na oração relacional do trecho (158), a relação estabelecida entre

Identificado (“a maior diversidade de espécies”) e Identificador (“no Brasil”) é de identificação circunstancial, em que o Identificador expressa a Localização: Lugar do Identificador. Em (159) a oração relacional identificativa é do tipo possessivo, em que o verbo *apresentar* expressa uma relação de contenção entre duas entidades distintas – o Identificado (“As Narrativas estudadas”) que engloba o Identificador (“as etapas de uma narrativa completa...”).

- (157) Neste estudo, os resultados **mostraram** a possibilidade de se identificarem as alterações relacionadas ao horário do dia em que acontecem o pastejo, a ruminação e o ócio, assim como o tempo exato despendido nessas atividades de acordo com o sistema de produção. (A_ABMVZ_2018_3_AA_026)
- (158) Nessa região a maior diversidade de espécies **encontra-se** no Brasil (Moraes 2007), na Floresta Ombrófila Densa Montana, a Mata Atlântica sensu stricto (Oliveira Filho & Fontes 2000), (B_R_2017_1_AA_008)
- (159) As Narrativas estudadas, neste artigo, **apresentam** as etapas de uma narrativa completa: orientação> complicação> resolução> coda. (L_RD_2018_1_AA_007)

Segundo Halliday e Martin (1993), Halliday (2000) e Martinez (2001), os processos relacionais podem renegar à possibilidade de expressar ações por meio do uso de nominalizações (conforme os trechos 160 e 161), de modo a diminuir a presença da agência humana – seja do próprio autor ou de outros – em relação à proposição na oração. Assim, ao empregar tais processos, os autores renunciam expressar sua participação na oração, mudando o foco para a relação entre as duas entidades e apresentando informações como contribuições factuais, por serem semanticamente impessoais e por mostrarem apenas a relação entre dois conceitos que não envolvem a presença de participantes agentivos em sua estrutura transitiva. Essa ausência de Agente torna a informação apresentada mais objetiva, impessoal e factual, contribuindo para tornar os textos mais precisos e as descrições materiais, menos mecânicas.

- (160) A comparação no Ciclo II **mostrou** desempenho inferior do GP II em relação ao GC II em todas as variáveis estudadas. (Sa_CODA_2014_0_AA_004)
- (161) A análise de agrupamento também **mostrou** a dissimilaridade entre as duas florestas, em relação ao estoque de serapilheira e demais variáveis ambientais avaliadas, com uma correlação cofenética de 0,60. (B_R_2018_2_AA_012)

Além dessas características, os processos relacionais também foram frequentemente usados para exemplificar fragmentos de textos analisados, para apresentar e descrever partes do

artigo acadêmico, como seções, tabelas, gráficos, exemplos e funções. Nos trechos de (162) a (165) abaixo pode-se observar que diferentes verbos realizam processos relacionais com função de apresentar tais partes do texto acadêmico.

- (162) A Fig. 1 **apresenta** a comprovação do DNA após a inoculação viral. (A_ABMVZ_2018_4_AA_006)
- (163) A Tabela 7 **apresenta** as confiabilidades externas relativas à observação com maior MDB (E_BCG_2016_3_AA_002)
- (164) As coordenadas colorimétricas das amostras em pó e das respectivas peças esmaltadas **se encontram** na Tabela 3. (En_ESA_2017_6_AA_012)
- (165) A próxima seção **vai mostrar** como a relação entre esses significados pode ser representada. (L_ALFA_2014_0_AA_001)

Determinados tais usos majoritários dos processos relacionais, cabe ainda ressaltar que, apesar de não ter sido frequente igualmente aos demais usos mencionados, há nas orações relacionais a possibilidade de existir a participação de uma figura agentiva que designa a relação entre os dois participantes obrigatórios nesse tipo de processo (cf. seção 3.3.3). Essa figura é expressa por meio do participante Designador que ocorre tanto nas orações relacionais atributivas (em que é denominado Atributor) ou nas identificativas (denominado Designador).

Esse tipo de oração foi particularmente frequente nos usos de *considerar* relacional, em especial quando associados a orações do tipo atributivas, conforme exemplificado nos trechos a seguir.

- (166) Myers et al. (2000) [Atributor] **consideraram** esse bioma [Portador] como um dos grandes hotspots mundiais prioritários [Atributo] para conservação. (B_R_2017_1_AA_004)
- (167) Na fase de pré-teste não foram assinaladas dificuldades de compreensão e de interpretação das questões, portanto os pesquisadores [Atributor] não **consideraram** necessário [Atributo] o re-teste [Portador], passando imediatamente a aplicação da versão de consenso. (Sa_ACPE_2014_0_AA_009)
- (168) A doutrina tradicional [Atributor] não **considera** o que chamo de obiter dicta exprocessuais [Portador] como parte do gênero obiter dictum [Atributo]. (So_RDGV_2018_2_AA_011)

Nos trechos anteriores de (166) a (168), o terceiro participante das orações relacionais intensivas é o Atributor, responsável por estabelecer a relação entre os dois outros Participantes – nesses casos, por criar uma relação de caracterização entre eles. Por sua vez, os trechos a

seguir ilustram as poucas orações em que houve a presença do Designador. Em (169) estabelece-se uma relação de identificação entre os dois termos, Identificado (“a memória”) e Identificador (“o princípio organizativo”). O mesmo ocorre em (170), em que “nós” (expresso por meio de desinência) atribui uma identidade (“= 0,05”) ao Identificado “alfa”.

- (169) O crítico [Designador] **considera** a memória [Identificado] “o princípio organizativo” [Identificador] (L_LH_2017_2_AA_002)
- (170) Todas as análises foram conduzidas no programa estatístico R, com o uso do pacote MASS (Venables & Ripley, 2002) e **consideramos** alfa [Identificado] = 0,05 [Identificador]. (B_ISZ_2017_0_AA_007)

Cabe ainda estabelecer a distinção entre alguns empregos relacionais de *considerar*, dos usos mentais. Nos trechos a seguir, exemplos de orações mentais (cuja análise apresenta-se na seção seguinte), há de um a dois participantes envolvidos no processo, de acordo com a voz da oração, e/ou a presença de uma oração projetada – características de processos mentais. Em (171), o Experienciador (“Freud”) é um ser humano consciente do processo mental, que projeta a oração típico do processo cognitivo realizado. Já em (172), o Experienciador não está expresso, entretanto sua participação é pressuposta por meio do clítico *se* indeterminador do sujeito. Por sua vez, em (173), o processo mental, com significado de ‘ponderar’, expressa o conteúdo da ação cognitiva do Experienciador (“Traugott e Trousdale (2013)”) por meio de um Fenômeno (“a existência de dois tipos básicos...”).

- (171) Sendo assim, Freud **considera** que toda a sinuosa atividade de pensamento que vai da imagem mnêmica até a identidade perceptiva estabelecida pelo mundo exterior é um percurso indireto para a realização de desejo. (H_AETP_2014_0_AA)
- (172) Deve-se **considerar** que “o crescimento dos programas de defesa do meio ambiente foi concomitante à expansão dos mercados globais” (PREVE; CORRÊA, 2007, p. 206). (So_RKAT_2013_0_AA_004)
- (173) Traugott e Trousdale (2013) **consideram** a existência de dois tipos básicos de mudanças relacionadas às construções. (L_RD_2017_1_AA_004)

Conforme exemplificado nos trechos a seguir, os processos relacionais sempre envolvem dois participantes, que podem ser ambos coisas ou fatos, e não projetam orações. Além disso, entre os participantes está expresso algum tipo de relação de caracterização ou identificação, próprio dos processos relacionais. Essa característica difere dos participantes em

uma oração mental, na qual apenas um é obrigatório (o Experienciador), havendo entre eles há uma relação de *pensar/sentir*. Em (174), o Atributor (“as centrais sindicais”) está expresso, relacionando o Portador (“o resultado”) a seu Atributo (“favorável”). Nesse caso, se um dos participantes obrigatórios da oração relacional (Portador e Atributo) não estivesse presente, a sentença então teria seu significado – e estrutura – alterados, podendo então ser classificada como mental. Entretanto, em (175), há a omissão de um dos participantes, o Atributor – que é opcional – por meio do uso da oração em uma construção efetiva receptiva. Nesse caso, o sentido relacional não é perdido, pois ainda está presente a relação entre o Portador (“o esforço amostral”) e o Atributo (“satisfatório”), mesmo que seu Atributor esteja omitido na sentença.

- (174) Apesar dos limites, as centrais sindicais **consideraram** o resultado favorável. (H_LN_2018_2_AA_003)
- (175) O contrário mostrou-se evidente na primeira elevação, onde o esforço amostral **pode ser considerado** satisfatório. (B_R_2017_1_AA_005)

Cabe ainda observar que foi frequente nos dados as ocorrências de processos relacionais realizados por verbos “*de mostrar*”, conforme denominados por Hao (2017). Segundo a autora, os processos realizados por esse tipo de verbos (como *mostrar, indicar, demonstrar, sugerir e ilustrar*) estão localizados na área fronteira entre os processos verbais e os relacionais, na qual estabelecem-se as relações simbólicas. Em função disso, exemplos limítrofes entre processos verbais e relacionais são frequentes nos gêneros acadêmicos – como comprovou-se nesta pesquisa.

Retomando o que foi apresentado na seção 3.3 sobre as áreas limítrofes entre os tipos de processos, segundo Halliday e Matthiessen (2014), os processos que ocorrem em tais áreas pode ser opacos e difíceis de classificar devido ao princípio denominado *indeterminação sistêmica*, que compreende a nossa experiência no mundo como altamente indeterminada – e essa característica se reflete em como a gramática é construída por meio do sistema de transitividade.

Tendo isso em vista, os trechos (176), (177) e (178) exemplificam esse princípio relacionado à área limítrofe entre os processos verbais e os processos relacionais. Em (176), o processo verbal envolve participantes típicos da oração verbal como o Dizente, “Os autores”, e uma Citação projetada (“que esta equação é válida...”). Já em (177), que instancia uma oração

relacional de tipo intensivo, estão envolvidos os dois Participantes obrigatórios, o Identificador e o Identificado, em uma relação do subtipo semiótico de *demonstração*.

- (176) Os autores **determinam** que esta equação é válida independentemente do alinhamento com o referencial de navegação. (E_BCG_2016_4_AA_011)
- (177) O mapa de cluster (Figura 3) **mostra** os focos das infrações ambientais constatadas no período estudado. (A_FL_2017_0_AA_042)

Por sua vez, (178) exemplifica uma oração localizada entre esses dois tipos de processos. De acordo com Hao (2017), isso ocorre devido as características de ambos processos poderem ser mescladas a fim de estabelecer uma relação simbólica entre dois participantes. Assim, essas relações abstratas e simbólicas são tipicamente realizadas por verbos “de mostrar”. O exemplo a seguir, (178), tem características de um processo verbal – como a possibilidade de acrescentar um participante Alvo à oração, conforme o teste de reescrita em (178’). Entretanto, conforme a reescrita em (178’’) a Locução/Identificador pode ser apenas reportado, mas não citado – afastando-o dos processos verbais.

- (178) Valores de PCE **mostram** que os tipos de estrutura esperada para os agregados formados pelos surfactantes são de micelas cilíndricas relativamente grandes ou em formato de bastão e o tipo de emulsão esperada para os mesmos é do tipo óleo em água (O/A). (E_QN_2017_3_AA_003)
- (178’) Valores de PCE nos mostram que os tipos de estrutura esperada...
- (178’’) * Valores de PCE mostram: “os tipos de estrutura esperada para os agregados formados....”

Além dessas propriedades, esse tipo de processo relacional identificativo pode ter seus participantes classificados como em (178’’’), em que uma parte do Identificador é omitida, “o fato”, tornando “que os tipos de estrutura esperada...” uma oração encaixada, rebaixada de nível (“*down-ranked clause*”).

- (178’’’) Valores de PCE **mostram** (o fato de) [[que os tipos de estrutura esperada para os agregados formados pelos surfactantes são de micelas cilíndricas relativamente grandes ou em formato de bastão e o tipo de emulsão esperada para os mesmos é do tipo óleo em água (O/A)].

Com relação ao verbo mais frequente no *corpus* analisado, *apresentar* teve ocorrências em orações materiais (conforme exemplificado na seção anterior), entretanto, a maior parte de seus usos foram em orações relacionais atributivas (88% das ocorrências relacionais) e possessivas (82% das ocorrências relacionais atributivas), conforme exemplificado a seguir em (179) e (180). Em todos os trechos, empregados com sentido de possessivo de ‘ter’, há a elaboração da descrição do tópico em discussão, de modo a caracterizá-lo em termos de, em (179), descrição para o leitor e de, em (180), especificidade das informações apresentadas.

(179) Na margem esquerda (paranaense) o rio **apresenta** relevo ondulado com paredões de Arenito da Formação Caiuá, (B_R_2018_2_AA_009)

(180) Um ponto importante de discussão é sobre os pontos específicos do problema que **apresentam** infactibilidade. (En_GP_2017_4_AA_010)

Como mencionado, os usos anteriores correspondem a maior parte das orações relacionais instanciadas pelo verbo *apresentar*. Entretanto, cabe ainda ilustrar alguns dos usos menos frequentes enquanto processos relacionais identificativos. Em (181), *apresentar* é empregado em uma oração relacional identificativa circunstancial (do tipo Localização: Lugar). Já em (182), é utilizado em uma oração relacional identificativa intensiva. Pode-se perceber que ambos trechos (representativos da maior parte das ocorrências relacionais identificativas) foram empregados a fim de apresentar e descrever partes do trabalho científico, tais quais seções, tabelas e figuras.

(181) Uma breve descrição destes dois métodos é **apresentada** a seguir. (E_TSC_2013_0_AA_0014)

(182) O Gráfico 9 **apresenta** os modelos de contratação, dentre os entrevistados inseridos no mercado (H_AC_2018_1_AA_006)

No que concerne os verbos *encontrar* e *determinar*, além de instanciarem orações materiais (cf. seção 5.1.1), esses também instanciaram orações relacionais. Em (183), há uma oração relacional atributiva intensiva relativamente similar a (184) quanto à finalidade de apresentar algo em uma tabela, mas que diferem em seus empregos. Em (184), o uso ocorre em uma oração relacional identificativa circunstancial, sendo a referência à tabela expressa por meio do participante obrigatório do processo relacional. Já em (183), tal referência é expressa por meio de uma Circunstância.

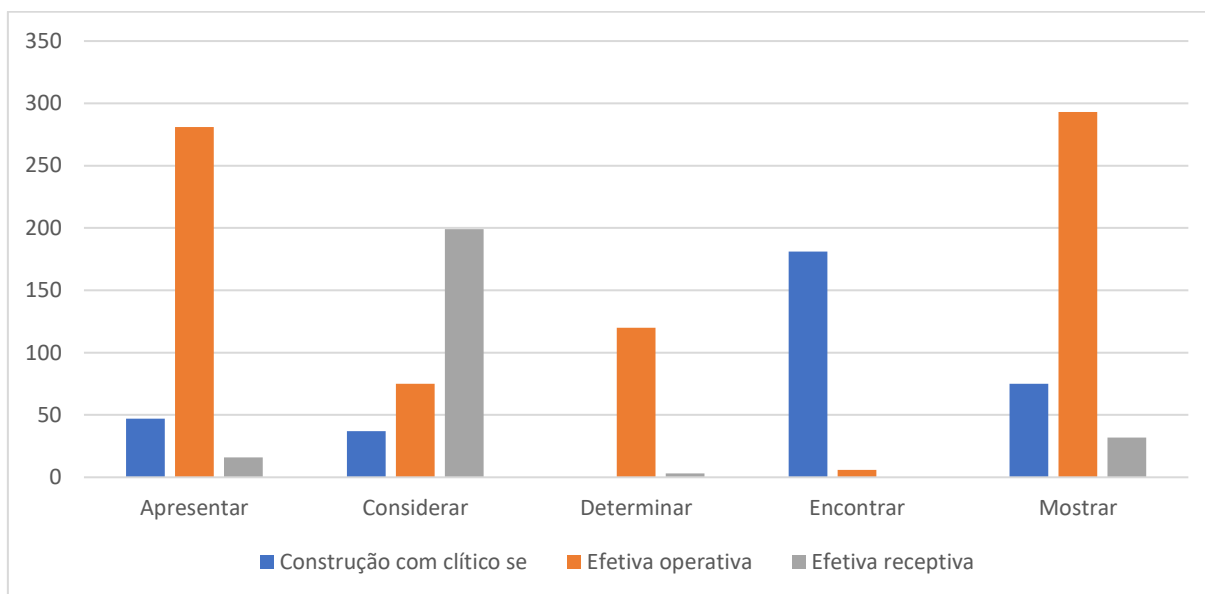
- (183) Os pictogramas do GHS associados às suas respectivas classes de perigo, bem como os códigos utilizados para fins de preenchimento do questionário [Portador], **encontram-se** descritos [Atributo] na Tabela 2 [Circunstância]. (E_QN_2017_3_EC_002)
- (184) As coordenadas colorimétricas das amostras em pó e das respectivas peças esmaltadas **se encontram** na Tabela 3. (En_ESA_2017_6_AA_012)

Por sua vez, (185) e (186) apresentam os usos relacionais de *determinar*, que foi empregado com significado circunstancial, realizado pelo próprio verbo (em vez de pelos participantes, como em 184), de ‘ocasionar, causar’. Conforme Halliday e Matthiessen (2014), e observado nestes dados, tais usos correspondem a orações relacionais identificativas circunstanciais de Causa: Condição.

- (185) Portanto, apesar de a hipoglicemia sintomática ser relatada com mais frequência, sua ausência não **determina** diagnóstico de GSDI. (Sa_JOPE_2014_0_AA_004)
- (186) O mesmo crescimento vertiginoso da oferta e procura de espaços **determina** a elevação das estruturas metálicas das arquibancadas (H_CCRH_2012_0_AA_001)

Passando a uma perspectiva de análise mais ampla das ocorrências dos processos relacionais, conforme pode ser observado no Gráfico 8 a seguir, os verbos *apresentar*, *determinar* e *mostrar* foram empregados majoritariamente em orações efetivas operativas, enquanto *encontrar* ocorreu quase totalmente em construções com o clítico *se* (as demais ocorrências desse verbo instanciaram orações materiais). Por outro lado, em *considerar* o modo preferido foram orações efetivas receptivas, traço que caracteriza a omissão do terceiro possível terceiro participante das orações relacionais, o Atributor/Designador.

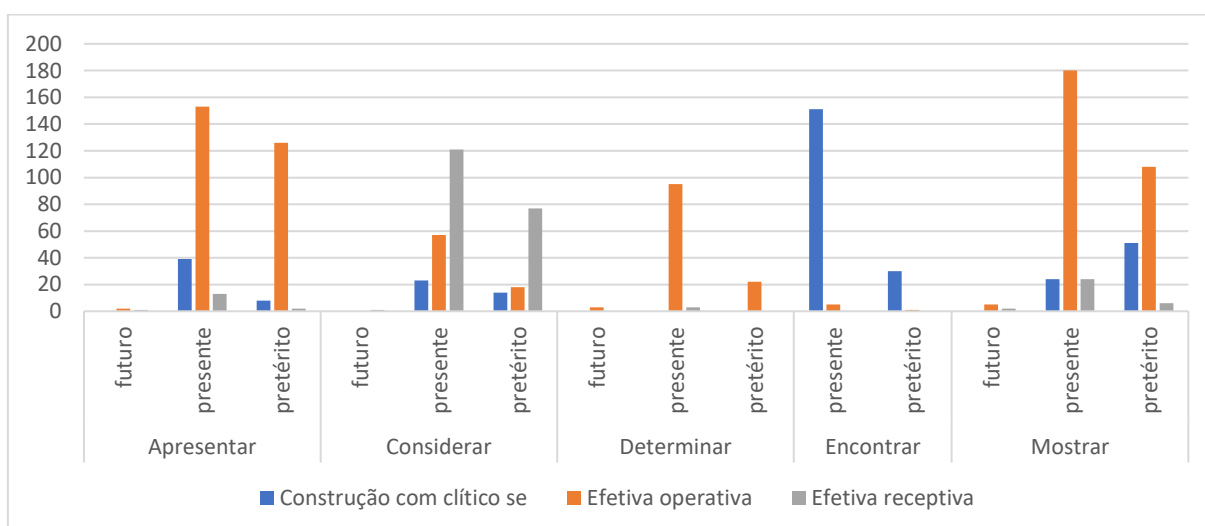
Gráfico 8 – Tipos construções em orações relacionais



Fonte: Elaboração própria.

Quanto ao tempo verbal em que ocorreram as orações relacionais, diferentemente dos processos materiais, as orações relacionais tenderam a serem empregadas no presente, majoritariamente, conforme pode ser observado no Gráfico 9 a seguir.

Gráfico 9 – Tempo verbal em orações relacionais



Fonte: Elaboração própria.

O uso do presente nesses casos pode ser associado ao caráter genérico atribuído à oração relacional ao apresentar conceitos teóricos da pesquisa (exemplificado em 187), assim como ao introduzir resultados (188) e demais informações sob a forma de tabelas ou seções do trabalho (189) ou caracterizar/identificar os objetos do estudo em questão (como em 190, cujos processos relacionais são empregados na identificação das espécies estudadas).

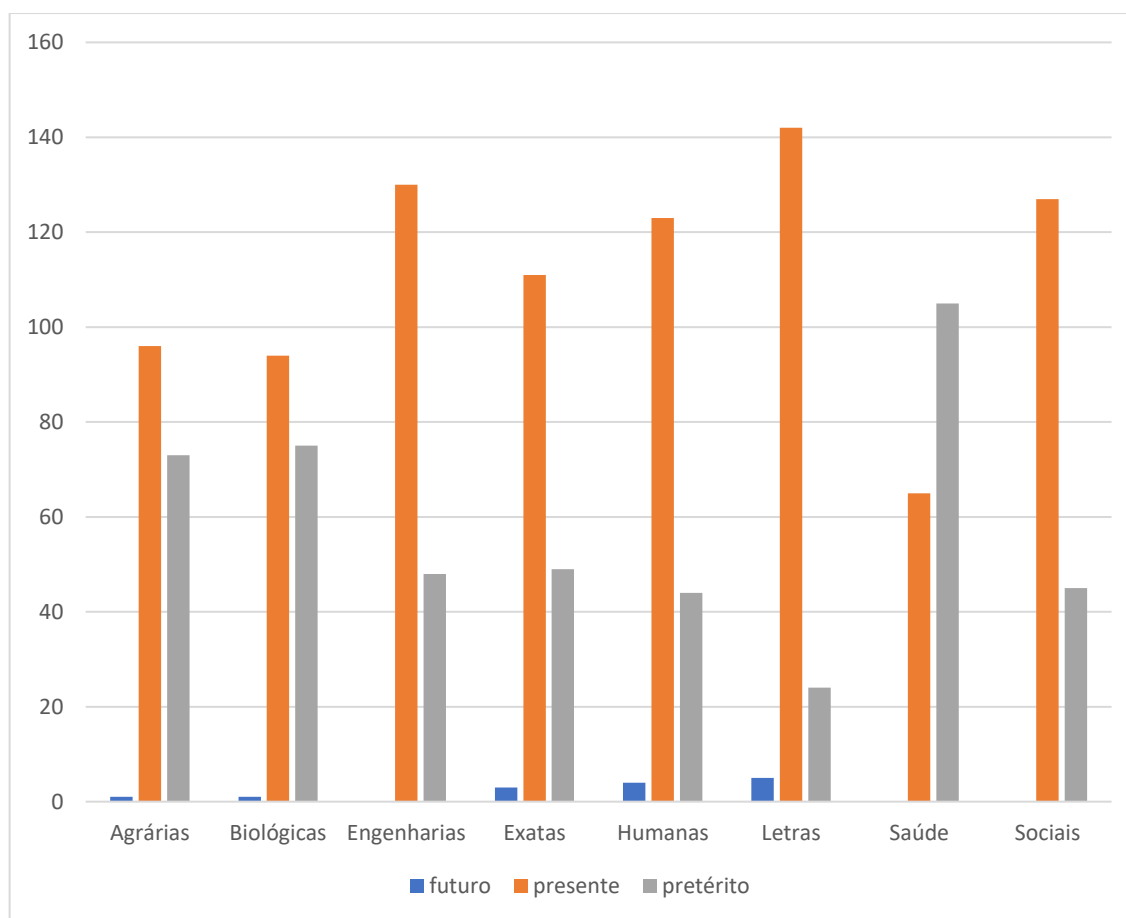
- (187) Esse princípio **determina** o resultado final da montagem, mas também o conteúdo dos planos justapostos. (L_ASP_2018_2_AA_007)
- (188) Os resultados obtidos **mostram** que a água tratada apresenta valores abaixo destes limites, evidenciando (B_RAA_2014_0_AA_003)
- (189) Na Tabela 4, **encontra-se** o índice de aceitabilidade das barras de cereais controle, probiótica e comercial. (A_BJFT_2018_0_AA_021)
- (190) *Pavonia sagittata* é uma espécie próxima de *P. hastata* Cav., diferenciando-se pela forma das bractéolas do epicálice e indumento dos ramos. *Pavonia sagittata* caracteriza-se por apresentar indumento amarelo forte, pelas bractéolas do epicálice muito largas, variando de obovada-espatulada até quase orbiculares e pela nervação muito leve dos mericarpos, enquanto *P. hastata* **apresenta** indumento glauco, bractéolas do epicálice elípticas a estreitamente obovadas e mericarpos com nervação leve a proeminente, sem ou com 2-4 tubérculos em cada lado da nervura média (Fryxell 1999). (B_HOE_2017_4_AA_002)

Por seu turno, o uso do pretérito associou-se majoritariamente a apresentação dos dados da própria pesquisa (como em 191 e 192). Nos exemplos, e nos demais dados similares analisados, é possível perceber como o uso dos processos relacionais permite omitir a presença/agência do autor, focando nos dados analisados e resultados apresentados. O trecho (192) exemplifica tal uso de processos relacionais no pretérito a fim de reportar resultados encontrados – sem haver a inserção do pesquisador/autor em seu texto.

- (191) Estes resultados **mostraram** que os alergênicos tiveram alterações estruturais pelo tratamento de radiação e que a tecnologia de irradiação pode ser aplicada para reduzir a alergenicidade dos alimentos. (A_FST_2011_0_AA_006)
- (192) RESULTADOS Do total de 90 análises microbiológicas, as amostras do lúmen das T3Vs tiveram níveis positivos de crescimento de 40% no meio de cultura TSB. As amostras do curativo **apresentaram** 86,7% de contaminação e a superfície externa das T3Vs **apresentou** 100% (Tabela 1). O tempo de permanência do CVP variou de 2 a 8 dias: 36,7% das amostras indicaram que o cateterismo venoso havia sido realizado há três dias (72 horas) e 26,7% há quatro dias (96 horas). A média e a mediana do tempo de permanência do CVP foram 3,75 dias e três dias, respectivamente, com desvio padrão de 1,48. (Sa_RLAE_2017_0_AA_033)

Por meio da análise do tempo verbal em que ocorreram os processos relacionais ao longo dos *subcorpora*, percebeu-se outra vez que as orações relacionais se comportam de maneira distinta das materiais. Como apresentado Gráfico 10, a seguir, em todas as Grandes Áreas, com exceção de Ciências da Saúde, os processos relacionais foram empregados com mais frequência no presente.

Gráfico 10 - Tempo verbal nas orações relacionais por Grande Área



Fonte: Elaboração própria.

O uso do pretérito nesse *subcorpus* associou-se em grande parte a finalidade com que foram empregados os processos relacionais na Grande Área de Ciência da Saúde: apresentar os resultados da pesquisa de modo impessoal e objetivo, conforme exemplificado em (192), anterior, e nos trechos a seguir (de 193 a 195).

(193) Quando comparados os desempenhos nas tarefas que avaliaram habilidades subjacentes, os escolares do GC II **mostraram** melhor desempenho que o GC I, com exceção das tarefas de

decodificação de texto e compreensão oral, pontuação ponderada de spam de dígitos. A comparação entre os Grupos Pesquisa **mostrou** semelhança de desempenho nas tarefas de decodificação de texto, compreensão oral, memória fonológica de curto prazo (dígitos e pseudopalavras) e compreensão leitora. Nas demais tarefas (decodificação de palavras e pseudopalavras isoladas, memória fonológica operacional e clausura gramatical), o GP II **mostrou** melhor desempenho que o GP I. Os resultados da correlação entre as variáveis de compreensão leitora e as demais variáveis do estudo, no GC I, **mostraram** correlação positiva: entre pontuação total de compreensão leitora e variáveis de fluência leitora de palavras (Sa_CODA_2014_0_AA_004)

- (194) Uma pontuação a partir de 8 na EDG-15 **foi considerada** alterada com a finalidade de priorizar sua especificidade para o rastreo de TDM (Sa_JBP_2017_1_AA_002)
- (195) A única exceção foi um paciente que **apresentou** edema macular cistoide, um mês após a cirurgia, com piora da acuidade visual. (Sa_ABOF_2013_0_AA_002)

Com relação ao emprego no tempo futuro, embora pouco frequentes, as orações relacionais foram assim utilizadas a fim de apresentar informações explicadas em um momento posterior no texto, conforme exemplificado nos trechos (196) e (197).

- (196) A seguir **será apresentado** e discutido o espaço da Química em museus e centros de ciências brasileiros. (E_QN_2018_9_EC_001)
- (197) A próxima seção **vai mostrar** como a relação entre esses significados pode ser representada. (L_ALFA_2014_0_AA_001)

Recapitulando esta seção, os usos das orações relacionais associaram-se majoritariamente a construções efetivas operativas no presente, com intuito de introduzir conceitos ou resultados da pesquisa, caracterizando-os ou identificando-os, de modo descritivo e preciso, e contribuindo para a exposição de informações, teorias e descobertas (próprias ou de outros) de modo impessoal. Além disso, também foram empregadas tais orações a fim de apresentar partes do trabalho, como exemplos, tabelas e gráficos.

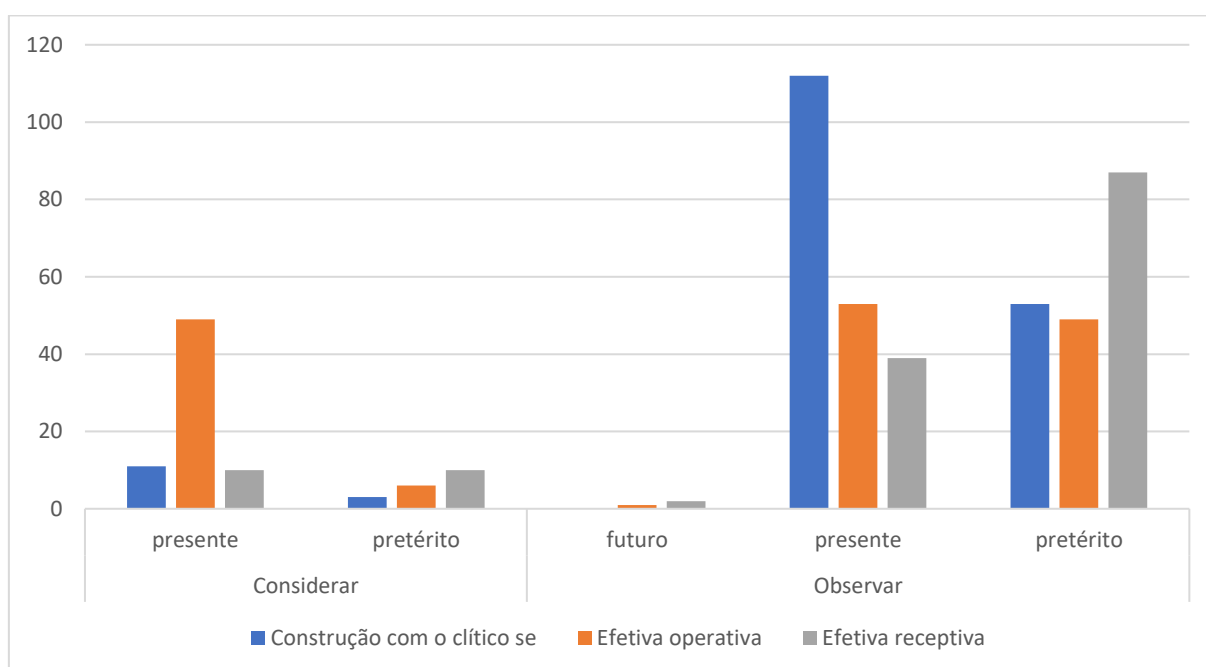
Na sequência, a próxima seção apresenta e discute as ocorrências dos processos mentais analisados nos dados desta pesquisa.

5.1.3 Orações mentais

Os processos mentais foram os terceiros mais frequentes, com 485 ocorrências, instanciados pelos verbos *considerar* e *observar* – ambos também realizaram processos relacionais e verbais, respectivamente. Dentre os dados analisados, as orações mentais

corresponderam a quase três vezes menos ocorrências do que o segundo colocado, e quatro vezes a menos que o primeiro. Como discutido na seção 3.3.2, as orações mentais consistem de um Experienciador consciente (ou ao qual é atribuído consciência) e um Fenômeno, com a possibilidade de projetar de orações. Desse modo, conforme observado nas análises das seções anteriores, os autores tendem a reduzir sua participação no texto e, no caso, como Experienciadores dos processos mentais, omitindo-se por meio de construções efetivas receptivas ou construções médias. Essa característica está demonstrada no Gráfico 11 abaixo, que apresenta os usos do tempo verbal e do sistema de voz em orações mentais.

Gráfico 11 - Tempo verbal e tipos de construções em orações mentais



Fonte: Elaboração própria.

De acordo com os dados analisados e condensados no Gráfico 11, os dois verbos que instanciam orações mentais se comportaram de maneira distinta. Faz-se necessário, então, uma análise individual de cada um.

Quanto a *considerar*, nos trechos a seguir (prototípicos das demais ocorrências encontradas) seus usos se aproximam do significado de ‘compreender’, ‘refletir’ e/ou ‘achar’. Em (198), (199) e (200), congruentemente com os modos de ocorrência mais frequentes (orações efetivas operativas no presente), os processos mentais cognitivo foram majoritariamente empregados com o objetivo de trazer as ideias de outros pesquisadores para o texto, contribuindo para a construção da resenha de literatura.

- (198) Leite *et al.* (2001) **consideram** que a produção média pode variar de 0,3 a 3 L/árvore, podendo-se esperar, ocasionalmente, indivíduos produzindo cerca de 30 litros, em uma coleta. (A_RCR_2014_0_AA_006)
- (199) Nitschke-Pagel e Wohlfahrt (2002), por outro lado, **consideram** que o grau de restrição não necessariamente é equivalente ao tamanho da peça (maior peça tem maior grau de restrição). (En_SI_2018_1_AA_005)
- (200) Harris e Li (2009) **consideram** que P&D é importante para a firma passar a ser exportadora, mas investimentos adicionais apresentam pouco impacto sobre o esforço exportador. (So_NE_2017_3_AA_006)

Já em (201) e (202), o processo mental é empregado de modo a reportar os achados do objeto em estudo (como “o parecer técnico” e o participante da pesquisa “um dos professores...”), por meio de uma proposição projetada. Note-se que, nesses casos, como o texto não lida com outras pesquisas e/ou teorias, os processos tenderam a ocorrer no pretérito.

- (201) Em relação à contaminação das águas subterrâneas e superficiais o parecer técnico **considerou** que as informações contidas no EIA foram suficientes, desde que as propostas de proteção fossem adotadas. (B_RAA_2017_6_AA_006)
- (202) Em um raciocínio similar, um dos professores sem ligação direta à temática em análise **considerou** que uma universidade qualquer que tenha cuidados com as áreas verdes e seja referência na gestão energética e de água, (H_AC_2018_2_AA_007)

Os trechos (203) e (204) a seguir ilustram mais um uso recorrente, o apagamento do Experienciador – e autor – por meio do uso do clítico *se* indeterminador. Nesses casos, o foco da oração recai sobre a ideia projetada pelo processo mental, ou seja, sobre a mensagem que o autor do texto deseja expor, tirando a atenção daquele que realiza a ação cognitiva expressa pelo processo mental.

- (203) Para que não haja influência das decisões de longo prazo nas análises, **considera-se** que a energia proveniente de leilões de longo prazo (A-5 e A-3) será recontratada, caso esses contratos terminem dentro do período de análise considerado. (E_SBA_2012_0_AA_001)
- (204) Este estudante foi selecionado utilizando como critério a compatibilidade de horários das aulas que foram ministradas nas escolas com a disponibilidade dos pesquisadores para acompanhá-lo. Além disso, **considerou-se** a concordância do mesmo, a autorização da escola, dos alunos e seus responsáveis, pois as aulas ministradas por ele, em umas das turmas em que estagiou, foram filmadas por um dos pesquisadores. (E_QN_2017_2_EC_001)

Atrelado a esse uso, nos trechos a seguir exemplificam-se algumas das ocorrências encontradas nas quais os processos mentais são empregados acompanhados dos modalizadores *dever* e *precisar*, atribuindo à oração um caráter de forte sugestão sobre a mensagem exposta.

- (205) Diante desta inferência, **deve-se considerar** que, i) assim como as demais espécies da comunidade residente, *L. lucidum* está colonizando áreas que podem ser ambientalmente favoráveis ao seu recrutamento e ii) tanto *L. lucidum*, quanto as espécies nativas, podem estar atuando como facilitadores no estabelecimento uma das outras. Além disso, **deve ser considerado** o nível de inclusão amostral utilizado (DAP maior ou igual a 5 cm), de forma que, se outras fases ontogenéticas fossem avaliadas, resultados distintos poderiam ser encontrados. (B_R_2018_2_AA_003)
- (206) **Deve-se considerar** que “o crescimento dos programas de defesa do meio ambiente foi concomitante à expansão dos mercados globais” (PREVE; CORRÊA, 2007, p. 206). (So_RKAT_2013_0_AA_004)
- (207) Um tema em destaque e que representa a preocupação de vários pesquisadores é a democratização da Educação Superior, para além da ampliação de vagas, indicando que outros elementos **precisam ser considerados** para que tal ampliação seja identificada, como a democratização desse nível de educação. (H_AC_2018_2_AA_002)

Uma última observação sobre *considerar* em orações mentais é que, embora esparsos, tais usos também fizeram referência marcada linguisticamente ao/s autor/es do texto, conforme exemplificado a seguir. Ocorrências como essas foram raras, as quais o autor está diretamente expresso no participante Experienciador, como em (208), ou quando está inserido indiretamente na oração por meio de metonímia (como “este estudo”, em 209).

- (208) Ele alerta para o fato de que as noções primitivas da alma, do corpo e da união devem ser estudadas separadamente: **Considero** também que toda a ciência dos homens consiste tão somente em distinguir essas noções, e não atribuir cada qual senão às coisas a que pertencem. (H_TFA_2018_3_AA_003)
- (209) Assim sendo, este estudo **considera** que seria recomendável, para uso de forma geral nos pavimentos de garagem, o valor de carregamento acidental distribuído corrigido, associado aos veículos de carga máxima de 6,5 kN por roda, ou seja, q-6,5. (En_RIEM_2014_0_AA_004)

Com relação às ocorrências de *observar* em orações mentais, a maior parte dos usos foram em construções com o clítico *se* indeterminador do sujeito. Nesses casos, o Experienciador pode ser facilmente recuperado como sendo o próprio autor, seja em função do contexto da oração ou porque estava expresso por meio de Circunstâncias. Nos trechos a seguir, (210) e (211) exemplificam ocorrências em que o Experienciador está omitido devido ao emprego do clítico *se*, entretanto, pode ser recuperado pelo contexto textual como sendo o/s

próprio/s autor/es do texto. Já em (212) e (213), a presença do autor pode ser recuperada pelas circunstâncias que acompanham o processo mental, “No presente estudo”, em (212) e “aqui”, (213).

- (210) De acordo com a tabela 2, **observa-se** um crescimento das bolsas Pibic de 2004 a 2016. (H_AC_2017_3_AA_004)
- (211) Em relação ao pico de atividade de vocalização dessa espécie, **observou-se** que os turnos do ocaso e as primeiras horas da noite constituíram os períodos de maior atividade, embora tenha sido registrada em todos os períodos do dia (manhã, tarde, noite). (B_ISZ_2017_0_AA_004)
- (212) No presente estudo, **observou-se** que nenhum grau de hemólise superou o ETA aceito pela ASVCP (Harr, 2013) para albumina, porém o erro relativo médio da albumina plasmática em todos os diferentes graus de hemólise ultrapassou o LEPC calculado com os valores de referência do próprio laboratório, e o mesmo fato ocorreu em relação ao LEPL estimado da literatura quando as concentrações foram maiores ou iguais a 0,36 g/L de hemoglobina. (A_ABMVZ_2018_3_AA_005)
- (213) Como **se pode observar**, aqui a lógica é a mesma que fundamenta a constituição das escalas envolvidas em implicaturas: o elemento à esquerda acarreta o elemento à direita, sendo, por isso, mais informativo. (L_LH_2017_3_AA_005)

Além dos usos mencionados, os trechos acima também exemplificam uma das finalidades mais comuns dos processos analisados: demonstrar o trabalho cognitivo realizado com/sobre os dados da pesquisa em questão. Abaixo seguem mais trechos ilustrativos. Em (214) e (215), os Fenômenos simples (“fato” e “a representação do sistema”) são enfocados pelo processo mental, cujo uso omite os responsáveis pela ação cognitiva (mas que podem ser recuperados na própria oração por meio das circunstâncias sublinhadas) que são os próprios autores. Por sua vez, e mais frequentemente empregados nos dados, os trechos de (216) a (218) trazem para o texto o trabalho (também cognitivo) de outros pesquisadores, sempre marcados linguisticamente, a fim de legitimar o próprio trabalho e/ou resultados.

- (214) Segundo Cecílio *et al.* (2009), a aplicação da Krigagem necessita de uma distribuição mais ou menos uniforme das estações pluviométricas (pontos amostrais), fato que não **foi observado** neste trabalho, o que pode indicar os altos valores de IDE e a baixa dependência espacial entre os coeficientes interpolados. (B_RAA_2018_1_AA_002)
- (215) A representação do sistema **pode ser observada** na Figura 1, na qual foram destacadas as dimensões e as subdimensões. (H_AC_2018_2_AA_004)
- (216) Em estudo de Peres *et al.*, **foi observada** hiperoxalúria em 8,0% dos pacientes com nefrolitíase, sendo o resultado (Sa_JBNE_2014_0_AA_007)
- (217) Similaridade foi percebida por ALI *et al.* (2014), que **observaram** redução no número de linfócitos do tecido sanguíneo de *Fundulus grandis* expostos ao petróleo. (A_AIB_2017_0_AA_014)

- (218) O mesmo **foi observado** por Zorzanello e Weschenfelder (2017) em estudo publicado com rotulagem de pães integrais, em que 16,7% dos rótulos avaliados não apresentaram a declaração para alergênicos. (A_BJFT_2018_0_AA_027)

Em resumo, as orações mentais fornecem um espaço para que sejam inseridas análises e observações dos autores (do texto ou outros pesquisadores), sendo seu modo de realização uma forma de priorizar a mensagem exposta, por meio do Fenômeno/oração projetada, em vez de o Experienciador. Particularmente, quando relacionado ao próprio autor, opta-se por construções na voz efetiva receptiva ou operativas com clítico *se*, omitindo o Experienciador do processo mental. Tais estratégias afastam o autor de suas observações, em uma tentativa de tornar o texto objetivo e impessoal – mesmo ao lidar com questões subjetivas.

A seguir, a próxima seção apresenta as orações existenciais que ocorreram nos dados analisados.

5.1.4 Orações existenciais

Nos dados foram encontrados 412 usos de orações existenciais instanciadas pelos verbos *ocorrer, apresentar e realizar*. Desses, apenas o primeiro foi empregado exclusivamente em orações existenciais. De acordo com a revisão teórica apresentada na seção 3.3.6, os processos existenciais se caracterizam por expressarem a existência de um único participante – o Existente. Enquanto os demais preconizam a relação entre dois ou mais participantes e/ou circunstâncias, os processos existenciais reduzem a agência humana ao evitar a relação entre entidades. Por meio de seu uso, o Existente pode ser apresentado como participante único, que ocorre independente da agência humana – ou do autor do texto, nos casos analisados. Essa propriedade acentua o caráter de objetividade atribuído ao texto por seu emprego, em especial quando associado aos resultados da pesquisa, conforme exemplificado em (219) a seguir. No trecho, a participação do autor/pesquisador está escondida, a fim de manter a objetividade do discurso, mesmo ao lidar com descobertas/observações que sabe-se terem sido feitas pelo próprio pesquisador.

- (219) A espécie **ocorreu** em 3,5% das amostras estudadas em condição mesotrófica. (B_HOE_2017_1_AA_001)

Concomitantemente a essa função, o uso dos processos existenciais contribui para que o foco do leitor permaneça na informação expressa pelo Existente. Nos trechos abaixo, apresenta-se alguns dos outros usos mais frequentes analisados, como com intuito de introduzir informação nova no discurso (220), no caso, o Existente introduz o tema a ser tratado no artigo; ou estabelecer um novo tópico para discussão, (221), mudando a atenção do leitor para outro tema.

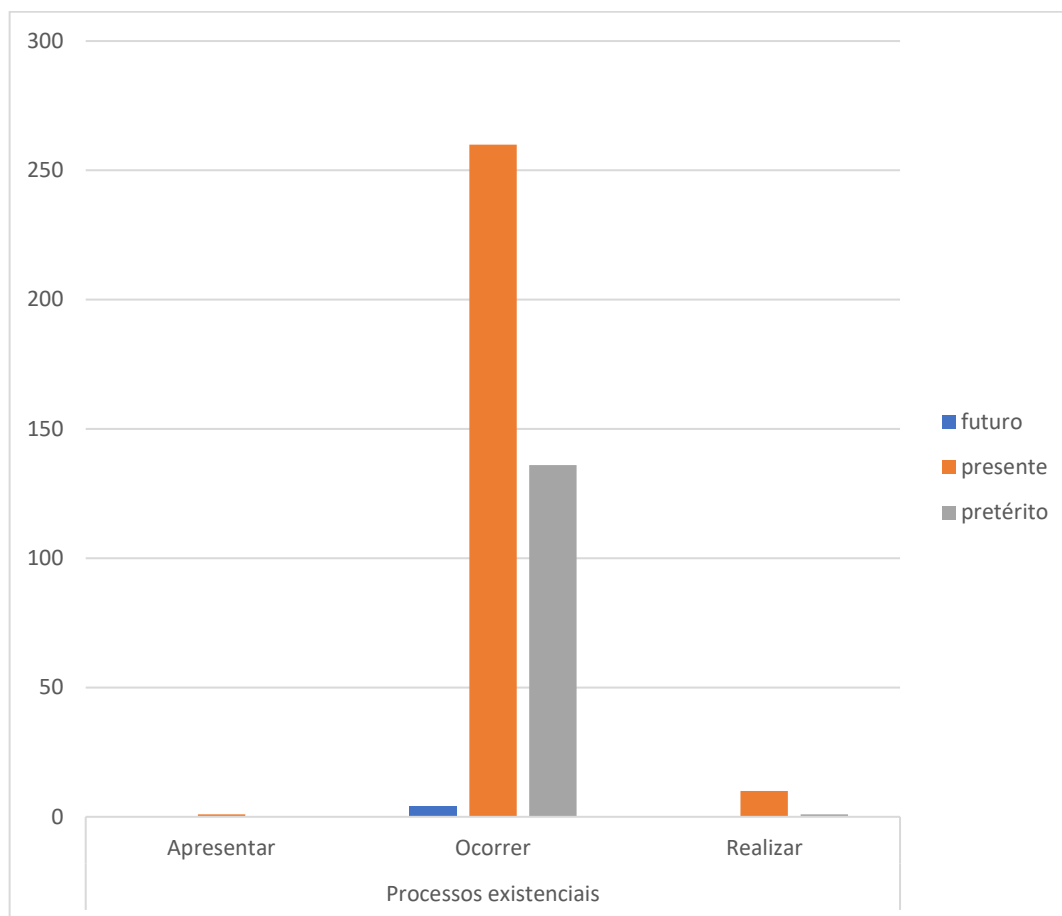
- (220) INTRODUÇÃO A luxação de patela **ocorre** frequentemente em cães e ocasionalmente em gatos, e é um evento comum na prática diária da traumatologia e ortopedia veterinária (Roush, 1993). (A_ABMVZ_2013_0_AA_005)
- (221) Warnekulasuriya *et al.* (1998) já haviam demonstrado a interferência do sal na exsudação de carnes curadas e na detecção de micro-organismos pelo ensaio de PCR, em decorrência da alteração da pressão osmótica das células. Outro efeito que **ocorre** na carne de sol é a mudança de cor, já que, durante o processamento, a utilização de sais transforma a mioglobina (pigmento vermelho) em metamioglobina e ferro hemocromo, ambos de cor castanha (CHASCO *et al.*, 1996). (A_BJFT_2018_0_AA_014)

No exemplo abaixo, os processos existenciais (incluindo aquele sublinhado que não foi alvo desta análise) foram empregados a fim expor os resultados encontrados pelos autores ao mesmo tempo que omite sua participação nesse processo de descoberta científica. No trecho (222), os processos existenciais são usados de modo a enumerar tais descobertas.

- (222) Este resultado indica que existe uma maior absorção de calor na área de pastagem, resultando numa maior taxa de aquecimento do ar sobre a região. Além disso, indicam que **ocorre** uma alteração no microclima local quando a floresta é substituída por pastagem, levando a um maior aquecimento da atmosfera, devido à maior disponibilidade de energia na forma de calor sensível em comparação com a floresta.

Quanto à realização do grupo verbal, como apresentado no Gráfico 12 a seguir, a preferência foi pelo uso do tempo presente. Em relação às ocorrências existenciais de *apresentar* e *realizar*, essas associaram-se ao emprego da partícula *se*, conforme exemplificado nos exemplos (223) e (224), na sequência. Por sua vez, diferentemente, as orações existenciais instanciadas pelo verbo *ocorrer* foram majoritariamente (98%) efetivas operativas.

Gráfico 12 - Tempo verbal em orações existenciais



Fonte: Elaboração própria.

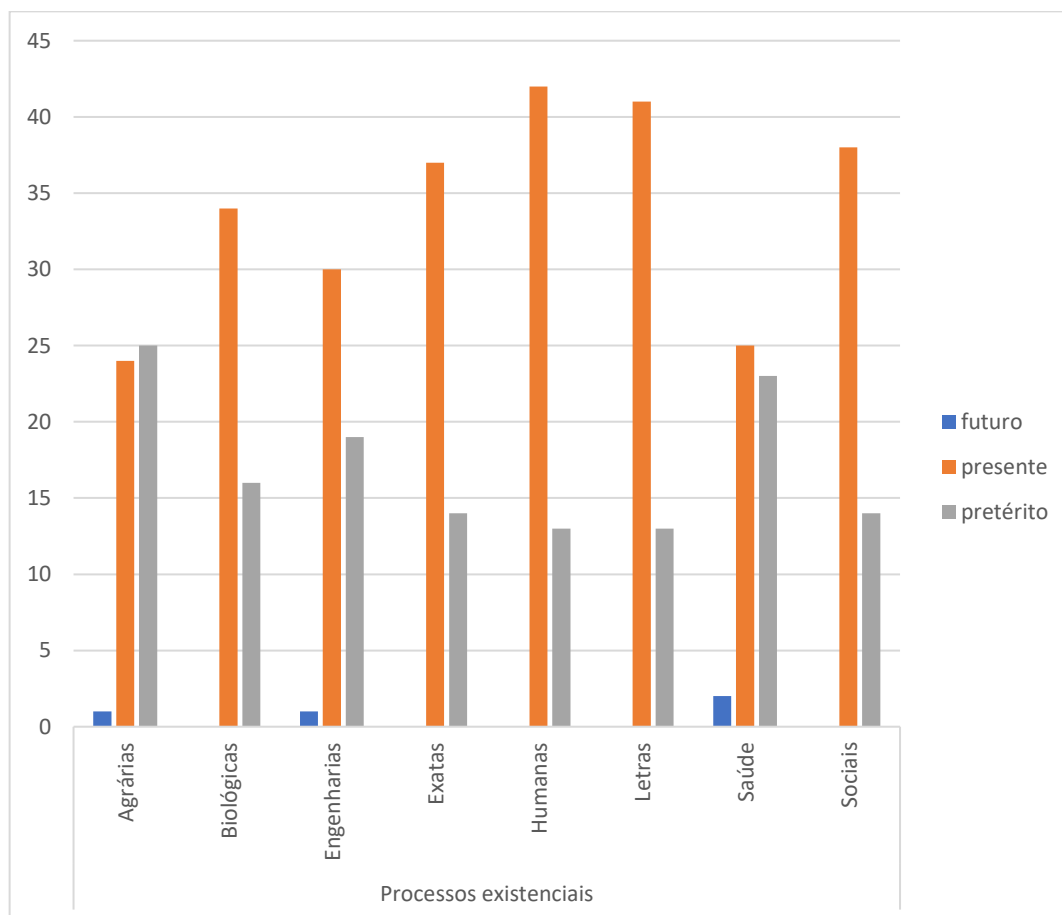
Em (223) e (224), abaixo, estão exemplificados os dois usos em construções com clítico “se”. Em ambos casos a presença da partícula caracteriza o significado da oração como ‘acontecer’.

- (223) É precisamente graças à camuflagem que ela consegue fazer frente às contrariedades que **se apresentam** e permanecer intacta até o final do périplo, credenciando-se a ao casamento com o príncipe Arnesto. H_CP_2017_1_AA_004.txt
- (224) Quando há apenas um cargo a ser preenchido, o procedimento de escolha é um pouco distinto daquele acima descrito. Embora também **se realize** por meio de duas votações, na votação de apresentação, cada eleitor escreverá apenas um nome em seu boletim de votação, H_TFA_2018_1_AA_004.txt

Cabe ainda observar que os processos existenciais com verbo *ocorrer* foram majoritariamente empregados em orações existenciais operativas no presente – conforme ilustrado nos exemplos dessa seção até momento. Entretanto, ao analisar esses dados por

subcorpora, pode-se perceber a preferência de Ciências Agrárias por empregar tais processos no tempo pretérito.

Gráfico 13 - Tempo verbal em orações relacionais por Grande Área



Fonte: Elaboração própria.

Por meio da análise minuciosa das ocorrências, foi possível notar que isso se deu devido à própria função do processo dessas orações. Nessa Grande Área *ocorrer* foi amplamente utilizado para apresentar resultados/dados encontrados, conforme ilustrado nos trechos (225) a (227) a seguir, em que o processo existencial serviu tanto para apresentar as descobertas das pesquisas (em 225 e 226), quanto para reportar uma justificativa para os eventos que se desenrolaram durante o estudo.

- (225) Não obstante o efeito antioxidante do ácido úrico tenha diminuído em ratos tratados com RV e submetidos ao esforço físico (N+R+), nesse grupo **ocorreu** um aumento significativo da TAC do plasma que pode ter contribuído para minimizar o estresse oxidativo. (A_ABMVZ_2018_3_AA_018)

- (226) Entretanto, caso o mencionado **tenha ocorrido**, a interferência **ocorreu** em todas as amostras analisadas, não prejudicando o comparativo entre as mesmas. (A_BJFT_2018_0_AA_026)
- (227) Essa diferença **pode ter ocorrido** devido à adição da farinha mista que, provavelmente, tem maior capacidade de absorção de água e, portanto, menor quantidade de água livre disponível na massa. (A_BJFT_2018_0_AA_017)

O último trecho anterior demonstra a presença de modalizador acompanhando o processo existencial. Esses casos ocorreram ao longo dos *subcorpora* sempre em situações semelhantes em que a modalização servia para atenuar o grau de certeza sobre as afirmações feitas a respeito dos resultados encontrados. Mais exemplos desses usos seguem em (228) e (229):

- (228) O que **pode ter ocorrido** por uma interação negativa com Mg, já que são elementos que competem pelo mesmo sítio de troca nas plantas. (A_FL_2017_0_AA_033)
- (229) As altas remoções de DQO observadas, se comparadas a outros estudos, **podem ter ocorrido** pelo pH básico (entre 8,0 e 9,5) dos ensaios. (En_ESA_2018_1_AA_009)

Já quando associados a orações em tempo presente, a modalização serviu para expressar a possibilidade ou não do Existente acontecer, tal qual exemplificado a seguir.

- (230) Com base no parâmetro da precedência, verifica-se a variação na colocação pronominal, em que o clítico **pode ocorrer** antes ou depois da forma verbal. (L_LH_2017_1_AA_008)
- (231) As complicações vão de brandas a até mesmo fatais. **Podem ocorrer** vômito, taquicardia, palpitações, desmaio e consequente parada cardíaca. (Sa_ROUN_2014_0_AA_0010)

Quanto ao Existente, é preciso mencionar a frequente associação entre nominalizações e processos existenciais, conforme observado nos trabalhos de Morais (2013) e de Miranda e Oliveira (2020). Nos trechos (232) a (235) a seguir, o efeito de impessoalidade atribuído à oração existencial se relaciona com a omissão do autor do texto por meio da não realização do Agente. Nos casos a seguir, tal efeito é alcançado ao utilizar-se uma nominalização encabeçando o participante Existente.

- (232) A associação entre as variáveis Saneamento e Gestão Ambiental e a variável população é de grande importância, pois significa que, de forma geral, o aumento da população é acompanhado de maiores

investimentos na manutenção do equilíbrio ecológico. Nesse caso, essa associação não **ocorreu**. (A_FL_2017_0_AA_080)

- (233) As amostras coletadas nos meses de julho, agosto e setembro de 2014 **ocorreram** em dias de sol entre nuvens, sendo observada a ocorrência de chuva nos dias anteriores. A coleta realizada em outubro de 2014 **ocorreu** em um dia com grande intensidade de vento e umidade relativa do ar mais baixa; a temperatura foi se elevando ao longo do dia, chegando a 40,8°C. (En_ESA_2018_1_AA_008)
- (234) A seleção dessa amostra **ocorreu** sequencialmente nos dois serviços, nos dias do atendimento, desde que os participantes estivessem acompanhados pelos pais ou responsáveis. (Sa_RPP_2017_4_AA_003)

A escolha lexical do uso de nominalizações é comum no gênero acadêmico, conforme estudos em língua inglesa (HAO, 2018; HYLAND, 2011; THOMPSON, 2014) e do português brasileiro (OLIVEIRA; CUNHA; MIRANDA, 2017; OLIVEIRA; LOUREIRO, 2017; MORAIS; BARBARA, 2018;), devido ao efeito de objetividade causado, possibilitando a supressão de um ou mais participantes do processo, mesmo que possam ser recuperados pelo contexto discursivo. Nas reescritas a seguir, a informação condensada por meio da nominalização do Existente em (234) é reelaborada de modo mais congruente.

- (234) A seleção dessa amostra ocorreu sequencialmente nos dois serviços
- (234') As amostras foram selecionadas sequencialmente nos dois serviços
- (234'') Os pesquisadores selecionaram as amostras sequencialmente nos dois serviços.

Em (234), o processo existencial descreve um acontecimento (“A seleção dessa amostra”) de um determinado modo (“sequencialmente”) e em um certo lugar (“nos dois serviços”). Porém, nas sentenças reelaboradas a fim de ‘desempacotar’ a informação condensada pela nominalização, um novo participante surge – o Agente/autor do texto – que em (234') está omitido, mas é de possível recuperação pelo contexto – e que em (234'') está expresso. Nesses dois casos perdeu-se o processo existencial, tornando-se orações materiais em que a presença de um Ator e/ou Agente é pressuposta, implícita (234') ou explicitamente (234'').

Dessa forma, nota-se que o uso da nominalizações em orações existenciais contribuem para tornar o texto mais impessoal e objetivo, redirecionando a atenção do leitor à informação expressa pelo Existente e não a quem os realiza. Daí sua importância e uso frequentes em

determinadas partes do processo científico, como na apresentação dos resultados encontrados pela pesquisa.

Além de encabeçados por nominalizações, os Existentes frequentemente foram expressos por pronomes demonstrativos, conforme exemplificado nos itens (235) e (236). Nesses casos, os pronomes retomam a informação apresentada anteriormente, expandindo-a, como em (235) e (236).

- (235) Diante dessa falta de condições, os passageiros têm que embarcar/desembarcar carregados. Tal procedimento também **ocorre** para realização da transferência da cadeira de rodas para o assento na cabine. (En_GP_2017_1_AA_011.txt)
- (236) As imagens lacônicas e fragmentárias, próprias dos enredos oníricos, são, na verdade, elementos aglomerados, não aleatoriamente, mas no interior desta lógica inconsciente, que a psicanálise visa reconstituir. Esta tarefa **ocorre** por meio de associações que vão mostrando pouco a pouco como elementos ou afetos infantis, restos diurnos, diferentes figuras e palavras dispersas foram congregados ou invertidos nas cenas de um sonho, aparentemente sem sentido. (H_AETP_2014_0_AA_001.txt)

Diante desta análise, em resumo, pode-se perceber que os processos existenciais, por meio da ausência de Agente associado à sua construção, contribuem para tornar o texto mais objetivo e impessoal, focando na mensagem/fatos e não em quem os veicula. Conforme demonstraram os exemplos apresentados, essa propriedade é particularmente útil – e se faz presente – especialmente em certas partes do texto acadêmico.

A seguir, a próxima seção apresenta as ocorrências encontradas de processos verbais nas amostradas analisadas dos *subcorpora*.

5.1.5 Orações verbais

Os processos verbais foram os menos frequentes nos dados analisados, com poucas ocorrências sendo instanciadas pelos verbos *determinar* e *observar*, totalizando 22 orações. Em geral, as orações verbais podem envolver os participantes Dizente, Receptor, Alvo e Verbiagem, além da possibilidade de projetarem orações. Quanto aos seus participantes, todas as ocorrências de orações verbais encontradas tinham o Dizente expresso, referindo-se majoritariamente a outros estudos, conforme exemplificado nos trechos de (237) e (238) a seguir. Nesses casos, trazer a voz de outros pesquisadores para o texto contribui para aumentar

a legitimidade das teorias selecionadas para a pesquisa, como em (237), assim como a validar as observações e posicionamento expressos pelo autor do texto, em (238).

- (237) Um entendimento adequado do Estado como formação histórica não poderia, contudo, se resumir a uma caracterização do lugar que as relações de força e de poder ocupam em sua constituição interna. Isso porque, **observa** Meinecke, “a essência de uma formação vital humana não se compõe apenas da substância que está na sua base, mas também do fim ao qual esta serve” (p. 469). Por isso, seria necessário ter em vista como, na vida política, a “necessidade estatal” se conjuga com a liberdade exigida por toda autêntica escolha moral. (H_LN_2017_1_AA_006.txt)
- (238) Não é de se admirar que as diversas ondas de chineses, japoneses, coreanos e imigrantes do sudeste asiático que vieram para os Estados Unidos durante os séculos XIX e XX também se encontravam no outro lado da linha de cor – separados em termos de não serem considerados branco, como uma outra raça de cor que ameaçava a construção dominante da cor branca e dando origem ao “perigo amarelo” do século XX. De fato, como Chametzky **observa**: “o confinamento e opressão de asiáticos são marcas decisivas de nossa história” (1989, p. 12, grifo do original). (L_ELBC_2014_0_AA_003)

Além de trazer para o texto a voz de outras pesquisas, os processos verbais também foram empregados a fim de relatar regulamentações de procedimentos, como normas governamentais (em 239) e decretos (em 240 e 241).

- (239) A NBR 10151 (ABNT, 2000) **determina** que o nível de ruído ambiente não deve ultrapassar 50 dB(A) em áreas residenciais, escolas e hospitais e 65 dB(A) em áreas voltadas à recreação, no período diurno. (En_ESA_2017_6_AA_006)
- (240) No caso brasileiro – anteriormente à Emenda Constitucional n. 45, de 2004 –, a Constituição não **determinava**, de maneira expressa, a posição hierárquica das normas de direito internacional, exceto dos tratados internacionais sobre direitos humanos. (So_RDGV_2017_2_AA_008)
- (241) Se a Constituição estabelece um sistema de complementaridade entre comunicação privada, estatal e pública, o projeto de lei **determina** que ao menos 33% dos canais deverão integrar o sistema público, tanto no rádio quanto na TV. (So_GSP_2018_2_AA_011)

Cabe ainda mencionar que a maior parte (19 ocorrências) dos processos verbais tiveram a mensagem exposta na forma na forma orações projetadas: 15 enquanto Projeção: Relato (exemplificadas nos trechos 239 a 241) e 4 enquanto Projeção: Citação (exemplificadas nos trechos 237 e 238). As demais ocorrências (3) expressaram o conteúdo da ação de dizer por meio de Verbiagem, conforme apresentado no trecho (242) a seguir.

- (242) Os passos para o gerenciamento ambiental de áreas contaminadas são descritos na Resolução CONAMA n° 420/2009 (Brasil, 2009) que **determina** critérios e valores orientadores de referência

de qualidade, de prevenção e de investigação do solo e valores de investigação para águas subterrâneas. (B_RAA_2017_2_AA_008)

A partir do número limitado de ocorrências, pode-se observar que os processos verbais serviram para legitimar a argumentação e/ou a exposição dos dados encontrados ao trazerem para o texto a fala de outros pesquisadores ou instituições de autoridade sobre o assunto. De tal modo, atribui-se um traço confiável e persuasivo ao texto. Por último, devido ao número restrito de usos, não foi possível fazer considerações sobre o uso de tipo de processo ao longo dos *subcorpora*.

Retomando os principais tópicos apresentados e discutidos nas últimas seções, o presente capítulo analisou os diferentes tipos de processos que os verbos mais relevantes para o gênero artigo científico instanciaram, levando em consideração o contexto textual de sua ocorrência, assim como a Grande Área de conhecimento em que ocorreram. Na análise, não apenas o tipo e uso do processo foi analisado, mas também seus participantes e, quando relevantes, as circunstâncias que o acompanharam.

Foi possível observar que as ocorrências de um mesmo tipo de processo tenderam a ser empregadas de maneira similar – tanto entre si quanto ao longo dos diferentes *subcorpora*. Também observou-se que, apesar da função dentro do texto acadêmico ser parecida, o modo de realização das orações foi diferente, de acordo com a Grande Área em que ocorreu. Por exemplo, os processos materiais foram mais frequentemente empregados em orações efetivas receptivas no pretérito nos *subcorpora* de Ciências Agrárias, Ciências Biológicas, Ciências Exatas, Engenharias e Ciências da Saúde, enquanto nos demais – Ciências Humanas, Ciências Sociais e Letras, Linguística e Artes – a escolha preferida foi orações efetivas operativas no tempo presente.

Por meio da análise detalhada de cada ocorrência, pode-se notar que tais diferenças são motivadas, em parte, devido à natureza dos estudos de cada área, considerando que o segundo grupo lida com objeto social de estudo. Apesar desse elemento distintivo, os processos foram majoritariamente empregados com a mesma função ao longo dos textos. Tais resultados justificam e verificam a validade da lista de verbos acadêmicos apresentados no Capítulo 4: apesar do modo de realização não ser o mesmo, a finalidade com que foram empregados os processos foi constante ao longo dos *subcorpora* (guardadas as devidas limitações de tipo). Pode-se, enfim, comprovar que a lista é de fato relevante para os estudos do gênero acadêmico artigo científico, assim como para a construção das etapas e procedimentos envolvidos nesse tipo de texto.

A próxima seção apresenta as considerações finais do trabalho, retomando o percurso percorrido a fim de responder as perguntas de pesquisa e objetivos propostos no início deste estudo.

6 CONSIDERAÇÕES FINAIS

Conforme exposto no capítulo inicial desta pesquisa, há duas vertentes teóricas distintas na literatura sobre o vocabulário acadêmico: uma afirma que existe de fato um conjunto de palavras centrais e comuns às diferentes áreas do conhecimento de uma língua, enquanto a outra questiona sua existência, alegando que há variações fundamentais no uso de tais palavras de acordo com a disciplina em que ocorrem. Nesse contexto, a presente tese contribui para o aprofundamento de tais questionamentos – se de fato existe um vocabulário central aos gêneros acadêmicos, ou não, – ao mesmo tempo em que preenche uma lacuna de estudos similares sobre vocabulário acadêmico em Português Brasileiro.

A partir dessa premissa, este trabalho teve dois objetivos. O primeiro foi criar uma lista de verbos acadêmicos com base em um *corpus* de artigos científicos abrangente de todas as Grandes Áreas do conhecimento, reconhecidas pelo CNPq. O segundo foi analisar os verbos mais frequentes dessa lista, segundo os pressupostos da LSF, a fim de observar se os usos ao longo das Grandes Áreas foram similares (de modo a corroborar a existência de um vocabulário acadêmico central) ou se foram particularizados em cada área (de modo a corroborar e justificar a segunda vertente mencionada, que preconiza a existência e o estudo de vocabulários acadêmicos específicos de cada disciplina).

Para alcançar tais objetivos, compilou-se um *corpus* de artigos científicos com aproximadamente 12,3 milhões de palavras, baseado no *corpus* de pesquisa anterior da autora (MIRANDA, 2016), com textos representantes de todas as Grandes Áreas e Subáreas de conhecimento, a fim de evitar que o conjunto textual fosse tendencioso para uma determinada disciplina. Com base nesse *corpus*, seguindo uma série de critérios metodológicos rigorosos expostos no Capítulo 4 que levam em consideração a frequência mínima, a razão, o alcance, a dispersão e a medida de disciplina em cada item lexical analisado, foi possível chegar a uma lista de verbos relevantes em artigos científicos de todas as áreas do conhecimento.

Essa lista, apresentada na Tabela 5, não apenas enumera os verbos considerados acadêmicos, mas também relata suas ocorrências em termos de frequência absoluta e de frequência relativa, a fim de facilitar a comparação dos valores entre *subcorpora*. Além disso, a lista aqui elaborada teve seus resultados comparados com o trabalho anterior da autora (MIRANDA, 2016), cujos procedimentos metodológicos (de exclusão de itens lexicais, conforme descrito no Capítulo 4) se aproximam daqueles utilizados em estudos mais antigos

sobre vocabulário acadêmicos (como em XUE e NATION, 1984 e COXHEAD, 1998). A partir da comparação, foi possível perceber que um pouco menos da metade dos resultados anteriores se mantiveram acadêmicos nesta nova lista.

Apesar de todos os critérios estabelecidos para a confecção da lista de verbos acadêmicos, somente sua elaboração não responderia às questões teóricas inicialmente apresentadas por este trabalho. Para determinar se os verbos ali presentes – relevantes em todas as disciplinas – eram empregados de modo similar ou se seus usos eram particulares de acordo com a área de conhecimento em que ocorreram, foram analisadas amostras de quatrocentas orações com contexto expandido para cada um dos dez verbos mais frequentes da lista. Essa etapa resultou na análise total de quatro mil ocorrências desses lemas, com base nos pressupostos teóricos da LSF, especificamente, o sistema de TRANSITIVIDADE.

A partir dessa perspectiva, analisou-se as orações levando em conta como os significados dos processos foram construídos ao longo dos *subcorpora*, se foi possível perceber generalizações sobre os significados, como os grupos verbais realizaram tais processos, se houve regularidade na realização das escolhas sistêmicas, como/quais os participantes (e ocasionalmente circunstâncias) acompanharam os processos, assim como as consequências de agnação e pressuposição a respeito do léxico empregado.

A análise dos verbos mais frequentes *apresentar, utilizar, realizar, considerar, observar, encontrar, ocorrer, mostrar, analisar e determinar* resultou em cinco tipos de processos instanciados nas amostras: processos materiais, relacionais, mentais, existenciais e verbais. Desses, os mais frequentes – e com número mais expressivo de ocorrências – foram os materiais e relacionais, seguidos pelos mentais e existenciais e por esparsas ocorrências de verbais. À parte de *analisar, utilizar, mostrar e ocorrer*, os demais verbos instanciam dois mais tipos de processos.

Apesar do número limitado de orações analisadas, no âmbito geral, os resultados encontrados se mostraram em conformidade com pesquisas anteriores que estudaram a transitividade em textos acadêmicos (MENDES, 2013; ZHEN; YANG; GE, 2014; VATHANALAOHA e TANGKIENG SIRISIN, 2018). Nesses trabalhos, os processos materiais, relacionais e mentais foram os mais frequentes, nessa ordem de ocorrência. Estreitando essa análise para cada Grande Área, percebeu-se que o mesmo padrão de ocorrência de cada tipo de processo repetiu-se em todos os *subcorpora* analisados.

Em sequência, por meio da análise do contexto expandido de cada oração, algumas pressuposições puderam ser feitas a respeito dos usos de cada tipo de processo. Com relação

aos processos materiais, esses foram empregados similarmente ao longo das Grandes Áreas de conhecimento, com intuito de descrever e relatar experimentos, análises, recomendações e contribuições. Entretanto, na análise de cada *subcorpus*, percebeu-se que a diferença entre as Grandes Áreas estava na forma como as disciplinas preferiram empregar tais orações – e nas implicações textuais assim ocasionadas.

Nessa parte da análise, observou-se que as Grandes Áreas de Ciências Agrárias, Ciências Biológicas, Engenharias, Ciências Exatas e Ciências da Saúde tendem a empregar as orações materiais em construções efetivas receptivas no pretérito, enquanto as Grandes Áreas de Ciências Humanas, Ciências Sociais e Letras, Linguística e Artes optaram preferencialmente por construções efetivas operativas no presente. Tal padrão, em termos mais locais de análise, implica que o primeiro grupo optou por omitir o Ator/Agente responsável pela ação com maior frequência do que o segundo grupo.

Por sua vez, a partir das amostras analisadas, a opção preferida do segundo grupo deveu-se ao fato de que essas disciplinas são voltadas para o estudo da sociedade, tornando necessário explicitar os seres e entidades (e suas ações) envolvidos no decorrer no processo de pesquisa realizado.

Com relação às orações relacionais, essas foram amplamente empregadas em construções efetivas operativas no presente, em todas as Grandes Áreas, com intuito de introduzir conceitos ou resultados da pesquisa, caracterizando-os ou identificando-os, de modo descritivo e impessoal. Seu uso contribuiu para a exposição de informações, teorias e descobertas, tanto de outros autores quanto da própria pesquisa, de maneira impessoal. Além disso, seus usos associaram-se frequentemente à apresentação de partes do trabalho acadêmico, como seções, exemplos, tabelas e gráficos.

Quanto às orações mentais, seus empregos forneceram um espaço para a inserção de análises e observações dos autores (do próprio texto ou de outros trabalhos). Sua realização tendeu, em todas as Grandes Áreas, a priorizar a mensagem exposta, por meio de Fenômeno ou de oração projetada, no lugar do Experienciador – responsável por realizar a ação cognitiva expressa pelo processo mental. Quando relacionado ao próprio autor, a preferência ao longo dos *subcorpora* foi pelo emprego de orações efetivas receptivas ou operativas com clítico *se*, omitindo das sentenças o Experienciador do processo mental. Essas estratégias serviram para afastar o autor de suas observações, tornando o texto mais objetivo e impessoal, mesmo ao lidar com um ponto de vista subjetivo.

Por sua vez, quase tão frequente quanto as mentais analisadas, as orações existenciais contribuíram para tornar o texto mais objetivo e impessoal, focando na mensagem ou fatos expostos, e não em quem os veicula. Isso se deve à ausência de um Agente associado à sua construção, tornando seu uso particularmente frutífero em etapas do texto acadêmico, como na introdução de (novos) temas, no estabelecimento de um novo foco ao tópico discutido e no relato dos resultados encontrados pela pesquisa. Para atingir tais fins, os processos existenciais foram majoritariamente empregados em orações efetivas operativas no presente, de modo uniforme nas diferentes Grandes Áreas do conhecimento.

Por último, os processos verbais foram os menos frequentes. Por meio da análise de suas poucas ocorrências, percebeu-se que foram todos empregados a fim de abrir um espaço para trazer outras pesquisas ou reportar regulamentações de normas ou do governo. Tais usos associaram-se a legitimação dos argumentos e/ou dados da própria pesquisa, atribuindo ao texto um caráter mais confiável e persuasivo.

Por meio dessa análise dos empregos dos verbos mais relevantes em artigos científicos, na perspectiva do sistema de TRANSITIVIDADE, foi possível observar que as ocorrências de um mesmo tipo de processo tenderam a ser similares entre si ao longo das Grandes Áreas. Entretanto, apesar da função dentro do texto ser parecida, o modo de realização da oração foi diferente. Dois padrões foram mais claros, como a preferência das Grandes Áreas de Ciências Agrárias, Ciências Biológicas, Ciências Exatas, Engenharias e Ciências Humanas por empregar orações materiais em construções efetivas receptivas no pretérito, enquanto nas demais Grandes Áreas prevaleceu o uso de construções efetivas operativas no presente. Conforme observado anteriormente neste capítulo, as motivações para essa diferença aparentaram estar na própria natureza das disciplinas.

Cabe ainda ressaltar que diferenças desse tipo estiveram majoritariamente associadas às orações materiais. Em menor grau, diferenças ocorreram com relação a áreas específicas, como a preferência do uso do pretérito em processos mentais na Grande Área de Ciências da Saúde; e em processos existenciais em Ciências Agrárias.

A partir dos resultados encontrados por tais análises, é importante voltar aos objetivos iniciais desta pesquisa. Assim como proposto, foi possível elaborar uma lista de verbos pertencentes ao vocabulário acadêmico – cujos usos são relevantes em artigos científicos –, assim como analisar uma amostra das ocorrências mais frequentes com base na perspectiva da perspectiva da LSF, em particular, o sistema de TRANSITIVIDADE.

É possível, desse modo, retomar os objetivos específicos explicitados nas Considerações Iniciais, elaborando-os em resposta aos resultados encontrados:

- a) A identificação dos verbos mais frequentes em um *corpus* de artigos acadêmicos foi realizada, seguindo uma série de critérios internos e externos ao próprio *corpus*, a fim de aprimorar a qualidade dos dados. Como resultado apresentou-se a lista dos verbos acadêmicos em artigos científicos na Tabela 5, da seção 4.3.
- b) Baseado nessa lista, afinou-se a pesquisa para o estudo do sistema da TRANSITIVIDADE instanciado por meio dos verbos acadêmicos mais frequentes. Para tanto, analisou-se quatro mil orações com contexto expandido, correspondendo a quatrocentas amostras aleatórias dos dez primeiros colocados da lista anterior.
- c) A partir dessa análise, pode-se estudar padrões no uso dos diferentes tipos de processos encontrados, em especial quando relacionados às suas funções dentro dos textos. Ao mesmo tempo, quando analisadas ao longo das diferentes Grandes Áreas do conhecimento, percebeu-se preferência por certos usos específicos em alguns casos (por exemplo, o emprego do pretérito nas orações existenciais de Ciências Agrárias).
- d) O último objetivo proposto foi integrar o *corpus* compilado para esta pesquisa ao CAPB. O *corpus* já está pronto para ser integrado ao projeto após a defesa desta tese e estará disponível online²⁵.

Tais objetivos foram norteados por duas perguntas de pesquisa. A primeira diz respeito à possibilidade de elaboração de uma lista de verbos acadêmicos do PB, representativa do gênero artigo científico e das diferentes Grandes Áreas do conhecimento. A resposta parcial dessa questão está associada ao objetivo (a), detalhado anteriormente, e pode ser encontrada Tabela 5, da seção 4.3.

A outra parte da resposta está atrelada à segunda questão, que corresponde à análise do sistema da TRANSITIVIDADE dos verbos mais frequentes, a fim de identificar padrões comuns às Grandes Áreas (e, assim, ao gênero acadêmico) ou se são particulares de cada Grandes Áreas (e, assim, justificando a criação de listas disciplinares).

Conforme observado nas análises do Capítulo 5, embora tenha havido diferenças entre determinados usos ao longo dos *subcorpora*, especialmente relacionadas ao modo como a oração foi construída linguisticamente. Tais diferenças podem ser justificadas devido à natureza

²⁵ Disponível em: <<https://sites.google.com/view/corpusacademico/p%C3%A1gina-inicial>>. Acesso em 20 de dezembro de 2020.

distinta de cada disciplina. Dito isso, e respondendo à questão anterior, em função dos critérios estabelecidos e das análises realizadas, a lista de verbos pode ser considerada acadêmica, mesmo com tais particularidades.

Percebeu-se também que é necessário um estudo mais aprofundado e com fins específicos de cada disciplina, conforme sugerem os trabalhos Martinez (2001), Hyland e Tse (2007) e Durrant (2016), a fim de verificar outras possíveis preferências de cada área. Porém, para fins de ensino desse gênero acadêmico, a presente lista se torna útil e válida, especialmente se aliada à perspectiva da LSF, que pode fornecer bases teóricas para facilitar a descrição de artigos científicos que vão além daquela realizada no Capítulo 5 desta pesquisa, e às ferramentas da Linguística de *Corpus*, como o site Sketch Engine²⁶, que hospedará o *corpus* da presente pesquisa.

Desse modo, com base nos resultados encontrados, o presente trabalho posiciona-se junto a Hyland (2016) e de Dang, Coxhead e Webb (2017), a favor de uma convergência entre as vertentes inicialmente expostas que compreendem esse tipo de vocabulário como um contínuo, desde as palavras e análises tais quais as apresentadas por esse trabalho, até o seu aprofundamento em cada Grande Área e/ou Subáreas.

Determinadas as contribuições da pesquisa, cabe ainda observar três implicações futuras. Em primeiro lugar, os textos analisados não são representantes de todos os gêneros acadêmicos, mas apenas de artigos científicos. Embora esses textos sejam considerados importantes dentro do meio acadêmico, para uma análise do vocabulário acadêmico como um todo, é necessária ainda a inclusão de diferentes gêneros. Em segundo lugar, a análise desta pesquisa limitou-se ao estudo dos itens lexicais verbais, sem adentrar nos substantivos, adjetivos ou advérbios que comporiam o vocabulário acadêmico. Assim, tais áreas podem ser exploradas futuramente, inclusive com auxílio dos procedimentos metodológicos apresentados pela pesquisa. Em terceiro e último lugar, a análise realizada do sistema da TRANSITIVIDADE foi insuficiente para mostrar o funcionamento de todos os tipos de processos diante do recorte proposto, sendo, portanto, interessante ampliar o recorte em próximos estudos.

À conclusão, com a presente pesquisa avançou-se nos estudos teóricos sobre os tipos de vocabulário em PB, assim como nas pesquisas sobre o sistema de TRANSITIVIDADE em artigos acadêmicos de variadas áreas do conhecimento, preenchendo lacunas anteriormente existentes em tais áreas. E, por fim, cabe mencionar o próprio *corpus*, disponibilizado ao público após a conclusão do trabalho, que poderá contribuir para outras pesquisas sobre artigos científicos.

²⁶ Disponível em: < <https://www.sketchengine.co.uk/> >. Acesso em 20 de dezembro de 2020.

REFERÊNCIAS

- BAKER, Mona. Sub-Technical Vocabulary and the ESP Teacher: An Analysis of Some Rhetorical Items in Medical Journal Articles. **Reading in a Foreign Language** v. 4, n. 2. 1988.
- BARBARA, Leila; MACEDO, Célia Maria. Linguística Sistêmico-Funcional para a análise de discurso: um panorama introdutório. **Cadernos de linguagem e sociedade**, v. 10, n. 1, 2009.
- BAUER, L., NATION, P. Word families. **International Journal of Lexicography**, v. 6, p. 253–279, 1993.
- BASTURKMEN, H. **Ideas and options in English for Specific Purposes**. NJ: Lawrence Erlbaum, 2006.
- BERBER SARDINHA, T. Looking at Collocations in Brazilian Portuguese Through the Brazilian Corpus. In: BERBER SARDINHA, T.; FERREIRA, T. L. S. B. **Working with Portuguese corpora**. Londres: Bloomsbury, 2014.
- BREZINA, V. **Statistics in Corpus Linguistics**. Cambridge: Cambridge University Press, 2018.
- BREZINA, V.; WEILL-TESSIER, P.; MCENERY, A. **#LancsBox v. 5.x**. (software). 2020. Disponível em: <<http://corpora.lancs.ac.uk/lancsbox>> . Acesso de 25 dez. 2020.
- BREZINA, V., TIMPERLEY, M., MCENERY, T. **#LancsBox v. 4.x**. (software). 2008. Disponível em: <<http://corpora.lancs.ac.uk/lancsbox>> . Acesso de 25 dez. 2020.
- BREZINA, V.; GABLASOVA, D. Is There a Core General Vocabulary ? Introducing the New General Service List. **Applied Linguistics**, v. 36, n. 1, p. 1–22, 2015.
- CAFFAREL, A.; MARTIN, J. R.; MATTHIESSEN, C. (org.). **Language Typology: A functional perspective**. Amsterdam: John Benjamin Publishing, 2004.
- CAFFAREL, A.; **A systemic functional grammar of French: from grammar to discourse**. London: Continuum, 2006.
- CHUNG, T. M.; NATION, I. S. P. Technical vocabulary in specialised texts. **Reading in a Foreign Language**, v.15, n. 2, p. 103–116, 2003.

COBB, T. Learning about language and learners from computer programs. **Reading in a Foreign Language**, v. 22, n. 1, p. 181–200, 2010.

COHEN, A. D., GLASMAN, H., ROSENBAUM-COHEN, P. R., FERRARA, J. FINE, J. Reading English for specialised purposes: discourse analysis and the use of student informants. In: CARRELL, P.; DEVINE, J.; ESKEY, D. E. (org.). **Interactive Approaches to Second Language Reading**. Cambridge: Cambridge University Press, 1988.

COXHEAD, A. A new academic word list. **TESOL Quarterly**, 34 (2), 213-238, 2000.

COXHEAD, A. The Academic Word List ten years on: Research and teaching implications. **TESOL Quarterly**, v. 45, n. 2, p. 355-362, 2011.

COXHEAD, A. Vocabulary and ESP. In: PALTRIDGE, B.; STARFIELD, S. (org.). **The Handbook of English for Specific Purposes**. Boston: Wiley-Blackwell, 2013.

COXHEAD, A., HIRSH, D. A pilot science word list for EAP. **Revue Française de Linguistique Appliquée**, v. XII, n. 2, p. 65–78, 2007.

DANG T.; COXHEAD, A.; WEBB, S. The academic spoken word list. **Language Learning**, v. 67, n.4, p. 959-997, 2017.

DANG, T; WEBB, S. Evaluating lists of high-frequency words. **International Journal of Applied Linguistics**, v. 167, n. 2, p. 132–158, 2016a.

DANG, T; WEBB, S. (2016b). Making an essential word list for beginners. In: NATION, I. S. P. **Making and using word lists for language learning and testing**. Amsterdam: John Benjamins, 2016b.

DAVIDSE, K. Existential constructions: A systemic perspective. **Leuvense Bijdragen**, n. 81, p. 71-99, 1992

DELLAI, N. L. V. **Ensino de vocabulário acadêmico no processo de compreensão em leitura: uma revisão bibliográfica**. 36 f. Graduação. (Trabalho de conclusão de curso em Letras). 2016. Universidade Federal da Fronteira Sul: Chapecó, 2016.

DURRANT, P. To what extent is the Academic Vocabulary List relevant to university student writing? **English for Specific Purposes**, v. 43, p. 49-61, 2016.

LU, C., DURRANT, P. A corpus-based lexical analysis of Chinese medicine research articles. **The Asian Journal of Applied Linguistics**, v. 4, n. 1, p. 3-15, 2017.

EGGINS, Suzanne. **An Introduction to Systemic Functional Linguistics**. Nova York: Continuum publishing, 2004.

FERREIRA, B. G. P. Ficar, tornar-see virarem construções relacionais: variação e/ou mudança construcional? **Revista da Associação Portuguesa de Linguística**, n. 4, 2018.

FIGUEREDO, G. Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues. 2011. 385 p. (Tese de Doutorado). Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

FLOWERDEW, J. Definitions in science lectures. **Applied Linguistics**, v. 13, e. 2, p. 202–221, 1992.

FRANCHI, C.; NEGRÃO, E.; VIOTTI, E. Sobre a Gramática das Orações Impessoais com Ter/Haver. **DELTA**, v. 14, n. esp.m 1998.

FRUTOSO, A. Os processos relacionais e a construção de sentido no gênero dissertação. Dissertação (Mestrado em Letras). Universidade Do Estado Do Rio Grande Do Norte, 2016.

FUZER, C.; CABRAL, S. R. S. **Introdução à gramática sistêmico-funcional em língua portuguesa**. Campinas: Mercado de Letras, 2014.

GARDNER, D.; DAVIES, M. A new academic vocabulary list. **Applied Linguistics**, v. 35, n. 3, p. 305–327, 2014.

GOULDEN, R.; NATION, I. S. P.; READ, J. How large can a receptive vocabulary be? **Applied Linguistics**, v. 11, n. 4, p. 341–363, 1990.

GOUVEIA, Carlos A. M. Texto e Gramática: uma introdução à Linguística Sistêmico-Funcional. **Matraga**, v. 16, n. 24, p. 13–47, 2009.

HALLIDAY, M. A. K.; HASAN, R. **Language, Context and Text**. Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K.; MATTHIESSEN, C. **Construing Experience Through Meaning**. London: Continuum, 2000.

HALLIDAY, M.; MATTHIESSEN, C. **An Introduction to Functional Grammar**. 4. ed. London: Edward Arnold, 2014.

HAO, J. The discourse semantic meanings of ‘showing’ in scientific texts. **Transformin Contexts**, v. 102, 2017.

- HAO, J. Reconsidering cause inside the clause in scientific discourse - from a discourse semantic perspective in systemic functional linguistics. **Text and Talk**, v. 38, n. 5, p. 525–550, 2018.
- HEBERLE, V. M. Apontamentos sobre linguística sistêmico-funcional, contexto de situação e transitividade com exemplos de livros de literatura infantil. **DELTA**, São Paulo, v. 34, n. 1, p. 81-112, jan./mar. 2018.
- HSU, W. Bridging the vocabulary gap for EFL medical undergraduates: the establishment of a medical word list. **Language Teaching Research**, v. 17, e. 4, p. 454–484, 2013.
- HWANG, K.; NATION, I. S. P. Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. **Reading in a Foreign Language**, v. 6, n. 1, p. 323–335, 1989.
- HYLAND, K. General and specific EAP. In K. Hyland & P. Shaw (org.). **The Routledge handbook of English for academic purposes**. London: Routledge, 2016.
- HYLAND, K. Academic discourse. In: HYLAND, K.; PALTRIDGE, B. (org.) **Continuum Companion to Discourse Analysis**. London: Continuum. pp 171-184, 2011.
- HYLAND, K., TSE, P. Is there an ‘academic vocabulary’? **TESOL Quarterly**, v. 41, n. 2, p. 235–253, 2007.
- JUILLAND, A. G.; CHANG-RODRÍGUEZ, E. **Frequency dictionary of Spanish words**. The Hague: Mouton, 1964.
- KILGARRIFF, A.; BAISA, V.; BUŠTA, J. *et al.* The Sketch Engine: ten years on. Lexicography. **ASIALEX** v. 1, p. 7–36, 2014.
- KUČERA, H.; FRANCIS, W. N. **Computational analysis of Present-Day American English**. Providence, RI: Brown University Press, 1967.
- LAUFER, B.; SIM, D. D. Measuring and explaining the reading threshold needed for English for academic purposes texts. **Foreign Language Annals**, v. 18, n. 5, p. 405–411, 1985.
- LIMA, L. R. Os processos existenciais em reportagens de capa da revista Superinteressante. **Domínios de Lingu@agem**, v. 6, n. 1, p. 261–284, 2012.
- LIMA, L. R. “Aparecer” e “surgir” – processos materiais ou existenciais? **E-escrita**, v. 5, n. 1, p. 316–331, 2014.

MARTIN, A. V. Teaching Academic Vocabulary to Foreign Graduate Students. **TESOL Quarterly**, vol. 10, n. 1. p. 91-98, 1976.

MARTÍNEZ, I. A. Impersonality in the research article as revealed by analysis of the transitivity structure. **English for Specific Purposes**, v. 20, p. 227–247, 2001.

MARTÍNEZ, I. A., BECK, S. C.; PANZA, C. B. Academic vocabulary in agriculture research articles: A corpus-based study. **English for specific purposes**, v. 28, n. 3, p. 183-198, 2009.

MATTHIESSEN, C.; TERUYA, K.; LAM, M. **Key Terms in Systemic Functional Linguistics**. London: Continuum, 2010.

MENDES, J. M. de O. **Processos verbais em artigos científicos de Administração de Empresas e Engenharia**: perspectivas sistêmico-funcionais. 2013. 112 f. Dissertação (Mestrado em Linguística Aplicada a Estudos da Linguagem), PUC-SP, São Paulo, 2013.

MIRANDA, M. V. **Processos verbais em artigos científicos**: uma análise com base na língua em uso. 2016. Dissertação (Mestrado em Estudo Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

MIRANDA, M. V.; OLIVEIRA, A. L. A. M. O uso de processos existenciais no domínio acadêmico: uma análise com base em corpus de artigos científicos. **Revista do GEL**, v. 17, n. 1, p. 160-188, 2020.

MIRANDA; M. V.; OLIVEIRA, A. L. A. M.; OLIVEIRA, A. A. **Corpus de Artigos Acadêmicos do Português Brasileiro** – CAPB. 2018. Disponível em: sites.google.com/view/corpusacademico. Acesso em: 19 jun. 2020

MORAIS, F. B. C. **Entre alhos e bugalhos** - os usos do clítico se na escrita acadêmica. 2013. 183 f. Tese (Doutoramento em Linguística). São Paulo: PUC-SP, 2013.

MORAIS, F. B. C. Os dizentes nos artigos científicos de Linguística – um estudo baseado na Linguística Sistêmico-Funcional e com o auxílio da Linguística de Corpus. **Letras & Letras**, v. 30, n. 2, p. 46-63, 2014.

MORAIS, F. B. C. O uso do processo existencial haver na escrita acadêmica - um estudo com base em um corpus de artigos científicos de diversas áreas do conhecimento. **Revista (Con) Textos Linguísticos**, v. 9, n. 14, p. 142–160, 2015.

MORAIS, F. B. C. Construções impessoais na escrita acadêmica: um estudo das construções com “se” em um corpus formado por artigos científicos. **Blucher Social Sciences Proceedings**, v. 2, n. 3, 2016a.

MORAIS, F. B. C. Variação de usos do clítico “se” na comunidade acadêmica: um estudo descritivo com base na linguística sistêmico-funcional. **Cadernos de Linguagem e Sociedade**, v. 17, n. 1, p. 70–100, 2016b.

MORAIS, F. B. C.; BARBARA, L. O uso de nominalização como recurso de impessoalização em artigos científicos escritos em língua portuguesa: um estudo com base na linguística sistêmico-funcional. **Cadernos de Linguagem e Sociedade**, v. 19, n. 1, 2018.

MOREIRA FILHO, J. L. **Kitconc 4.0**. (software). 2008.

MOURA NEVES, M. H. **Gramática de usos do português**. São Paulo: UNESP, 2000.

NATION, I. S. P. **Learning vocabulary in another language**. Cambridge: Cambridge University Press, 2001.

NATION, I. S. P. (2006). How large a vocabulary is needed for reading and listening? **Canadian Modern Language Review**, v. 63 n. 1, p. 59–82, 2006.

NATION, P. **Teaching Vocabulary: Strategies and Techniques**. Boston: Heinle Cengage, 2008.

NATION, I. S. P. **The BNC/COCA word family lists**. Disponível em: <<http://www.victoria.ac.nz/lals/about/staff/paul-nation>>. Acesso em 17 nov 2015.

OLIVEIRA, A. L. A. M; CUNHA, G. X.; MIRANDA, M. V. Nominalizations as complex strategies of politeness and face-work in scientific papers written in Brazilian Portuguese. **Cadernos de Estudos Lingüísticos**, v. 59, p. 361, 2017.

OLIVEIRA, A. L. A. M; LOUREIRO, G. P. Instâncias de nominalizações e a densidade lexical do texto acadêmico: algumas implicações para o ensino. **Revista do GEL**, v. 14, n. 3, p. 119-134, 2017.

OLIVEIRA, M. M. C. **A transitividade em evidência em notícias jornalísticas: um olhar funcionalista**. 2017. (Dissertação de Mestrado). Universidade Federal Do Espírito Santo, Vitória. 2017.

PENHA, R. F. A transitividade dos Processos Verbais Dizer e Afiramar em artigos científicos de graduandos em Letras. **Revista Ao Pé da Letra**. v. 14, n. 1, 2012.

PORTELA, K. C. A. Um estudo dos processos verbais no gênero artigo científico em revistas do secretariado executivo: uma perspectiva da linguística sistêmico-funcional. **Holos**, ano 29, v. 4, 2013.

PORTELA, K. C. A.; BRAUER, K. C. N.; SCHUMACHER, A. J. Os processos verbais presentes nos artigos científicos da Revista Gestão e Secretariado: uma análise com base na Linguística Sistêmico-Funcional. **Revista Gestão e Secretariado**, v. 5, n. 2, p. 97-116, 2014.

PORTELA, K. C. A.; BRAUER, K. C. N.; Análise de processos verbais em introduções de artigos científicos de três áreas: Economia, Engenharia Sanitária Ambiental E Linguística sob o viés da linguística sistêmico-funcional. **Holos**, v. 104, n.1, 2017.

SANTOS, V. S. **Vocabulário e leitura**: a elaboração de uma lista de palavras de uso acadêmico em Português do Brasil. 2006. 129 f. Dissertação (Mestrado em Linguística Aplicada na área de Língua Estrangeira). Universidade Estadual de Campinas, Campinas, 2006.

SAUTCHUCK, Inez. **Prática de morfossintaxe**: como e por que aprender análise (morfo)sintática. 2. ed. Barueri: Manole, 2010

SIMPSON-VLACH, R., ELLIS, N. C. An academic formulas list: new methods in phraseology research. **Applied Linguistics**, v. 31, n. 4, p. 487–512, 2010.

SOUZA, M. M. **Transitividade e construção de sentido no gênero editorial**. 2006. Tese (Doutorado em Letras) – Centro de Artes e Comunicação, Universidade Federal de Pernambuco, Recife, 2006.

SCHMITT, N.; SCHMITT, D. A reassessment of frequency and vocabulary size in L2 vocabulary teaching. **Language Teaching**, v. 47, n. 4, p. 484–503, 2014.

SCHMITT, N., SCHMITT, D., CLAPHAM, C. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. **Language Testing**, v. 18, n. 1, p. 55–88, 2001.

SUTARSYAH, Cucu; NATION, Paul; KENNEDY, Graeme. How useful is EAP vocabular for ESP? A corpus based study. **RELC Journal**, v. 25, n. 2, p. 34-50. 1994.

THOMPSON, Geoff. **Introducing Functional Grammar**. Abingdon: Routledge, 2014

TOWNSEND, D.; FILIPPINI, A.; COLLINS, P.; BIANCAROSA, G. Evidence for the Importance of Academic Word Knowledge for the Academic Achievement of Diverse Middle School Students. **The Elementary School Journal**, v. 112, n. 3, mar. 2012.

TRASK, R. L. **A dictionary of phonetics and phonology**. London: Routledge, 2004.

- VASCONECLOS, S. I. V. **O clítico SE: entre a norma e a variação**. 2013. 134 f. Dissertação. (Mestrado em Consultoria e Revisão Linguística da Faculdade de Ciências Sociais e Humanas). Universidade nova de Lisboa, 2013.
- VATHANALAOHA, K.; TANGKIENGSIRISIN, S. Genre Analysis of Experiment-based Dental Research Article Abstracts: Thai and International Journals. **3L The Southeast Asian Journal of English Language Studies**, v. 24, p. 1-14, 2018.
- VIVAN, E. G. S. **Principais usos de processos verbais e metáforas interpessoais em artigos de Linguística Aplicada**. Doutorado. (Tese de doutoramento). São Paulo: PUC-SP, 2010.
- WEBB, S. A., CHANG, A. C.-S. Second language vocabulary growth. **RELC Journal**, v. 43, n. 1, p. 113–126, 2012.
- WEBB, S.; NATION, Paul. **How Vocabulary Is Learned**. Oxford: Oxford University Press, 2017.
- WEST, M. **A general service list of English words**. London: Longman, Green & Co, 1953.
- WOODWARD-KRON, R. More than just jargon – the nature and roles of specialist knowledge in learning disciplinary knowledge. **Journal of English for Academic Purposes**, v.7, p. 234 – 249, 2008.
- XUE, G., NATION, I. S. P. A university word list. **Language Learning and Communication**, v. 3, n. 2, p. 215–229, 1984.
- YANG, M. N. A nursing academic word list. **English for Specific Purposes**, v. 37, n. 1, p. 27–38, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.esp.2014.05.003>>.
- ZHENG, S.; YANG, A.; GE, G. Functional Stylistic Analysis: Transitivity in English-Medium Medical Research Articles. **International Journal of English Linguistics**, v. 4, n. 2, p. 12–25, 2014.
- ZIPF, G. **Human behavior and the principle of least effort: an introduction to human ecology**. New York: Hafner, 1949.

ANEXO 1 – Script para remoção de itens não-verbais e marcações morfossintáticas

```

library(tidyverse)

tabela_brutas <- list(
  agrarias = read.csv("../LISTAS LancsBox/agrarias
lemmalist/TESE_Agrarias.csv", comment.char = "#"),
  biologicas = read.csv("../LISTAS LancsBox/biologicas
lemmalist/TESE_Biologicas.csv", comment.char = "#"),
  engenharia = read.csv("../LISTAS LancsBox/engenharias
lemmalist/TESE_Engenharias.csv", comment.char = "#"),
  humanas = read.csv("../LISTAS LancsBox/humanas
lemmalist/TESE_Humanas.csv", comment.char = "#"),
  letras = read.csv("../LISTAS LancsBox/letras
lemmalist/TESE_Letras.csv", comment.char = "#"),
  saude = read.csv("../LISTAS LancsBox/saude lemmalist/TESE_Saude.csv",
comment.char = "#"),
  sociais = read.csv("../LISTAS LancsBox/sociais
lemmalist/TESE_Sociais.csv", comment.char = "#"),
  exatas = read.csv("../LISTAS LancsBox/exatas
lemmalist/TESE_Exatas.csv", comment.char = "#"),
  geral = read.csv("../LISTAS LancsBox/geral lemmalist/TESE_Geral.csv",
comment.char = "#")
)

lapply(tabela_brutas, function(df) sum(df$Frequency..01...Freq) ) ##
confere com o numero de Tokens

head(tabela_brutas$geral, 4)

verbos_prelim <- lapply(tabela_brutas, function(df) dplyr::filter(df,
grepl("\\_v", Lemma)) )

verbos_sorted <- lapply(verbos_prelim, function(df)
df[order(df$Frequency..01...Freq, decreasing = TRUE), ] )

K <- length(verbos_sorted)
for(i in 1:K){
  write.table(verbos_sorted[[i]], row.names = FALSE, sep = "\t",
file = paste("../Outputs_processados/verbos_",
names(verbos_prelim)[i], ".txt", sep = ""))
  write.table(verbos_sorted[[i]], row.names = FALSE, sep = ",",
file = paste("../Outputs_processados/verbos_",
names(verbos_prelim)[i], ".csv", sep = ""))
}

lapply(verbos_sorted, nrow) ## quantos verbos em cada area

lapply(verbos_sorted, tail, 3)

```

ANEXO 2 – Script para calcular Razão

```

library(tidyverse)

area_names <- c("agrarias", "biologicas", "engenharia", "humanas",
"letras", "saude", "sociais", "exatas")

load("../Outputs_processados/verbos_processados.RData")

lapply(verbos, head, 5)

verbos.dt <- comparado <- verbos %>%
  reduce(full_join, by = "Lemma")

head(verbos.dt)

frequencies <- verbos.dt[, c(1, grep("freq.", names(verbos.dt)))]
frequencies$Lemma <- gsub("_v", "", frequencies$Lemma)

head(frequencies)

write.csv(frequencies,
  file = "../Outputs_processados/frequencias.csv",
  row.names = FALSE)

counts.stuff <- verbos.dt[, c(1, grep("count.", names(verbos.dt)))]

counts.stuff$count.total <- rowSums(counts.stuff[, -1], na.rm = TRUE)

sum(counts.stuff[, 10])

N_corpus <- 12539308

counts.stuff$freq.total <- counts.stuff$count.total/N_corpus

head(counts.stuff)

verbos_sem_v <- counts.stuff
verbos_sem_v$Lemma <- gsub("_v", "", verbos_sem_v$Lemma)

head(verbos_sem_v)

ref_corpus <- read.csv("../Outputs_processados/verbos_PT11.csv")

N_ref_corpus <- 4622750491

ref_corpus$ref.freq <- ref_corpus$Freq/N_ref_corpus

com_ref <- full_join(verbos_sem_v, ref_corpus, by = "Lemma")

head(com_ref)

com_ref$ratio <- com_ref$freq.total/com_ref$ref.freq

subset(com_ref, Lemma == "utilizar")

limiar_ratio <- 1.5

```

```
ratio_maior_limiar <- subset(com_ref, ratio >= limiar_ratio)

dim(ratio_maior_limiar)

write.csv(ratio_maior_limiar,
          file = paste("../Outputs_processados/filtrado_por_ratio_",
                        limiar_ratio, ".csv", sep = ""),
          row.names = FALSE)
```


ANEXO 3 – Script para calcular Alcance

```

library(tidyverse)

counts <- read.csv("../Outputs_processados/filtrado_por_ratio_1.5.csv")

head(counts)
dim(counts)

apenas_frequencia <- read.csv("../Outputs_processados/frequencias.csv")

dim(apenas_frequencia)

apenas_frequencia <- left_join(counts[, c("Lemma", "freq.total")],
apenas_frequencia, by = "Lemma")

head(apenas_frequencia)
dim(apenas_frequencia)

ratios_internos <- data.frame(
  Lemma = apenas_frequencia$Lemma,
  t( apply(apenas_frequencia[, -1] , 1, function(x) x/x[1] ) )
)

freq.limiar <- .2

names(ratios_internos) <- gsub("freq.", "internal_ratio.",
names(ratios_internos))

ratios_internos$range <- apply(ratios_internos[, -c(1,
ncol(ratios_internos))], 1,
function(x) sum(na.omit(x) > freq.limiar) )

ratios_internos <- data.frame(ratios_internos, apenas_frequencia[, -1])

head(ratios_internos)

tail(ratios_internos)

names(ratios_internos)

hist(ratios_internos$range, xlab = "Valor do range", ylab = "Frequência",
main = "")

range_6 <- subset(ratios_internos, range > 6)

range_6

dim(range_6)

names(range_6)

range_7 <- subset(ratios_internos, range > 7)

range_6[!(range_6$Lemma %in% range_7$Lemma), ] ## que verbos perdemos se
fizemos range = 7 como limiar

```

```
range_6$perde_range_7 <- ifelse(range_6$Lemma %in% range_7$Lemma, "Não",  
"Sim")  
  
write.csv(range_6, "../Outputs_processados/filtrado_por_range_6.csv",  
row.names = FALSE)
```

ANEXO 4 – Script para calcular Dispersão

```

filtrado_ratio_range <-
read.csv("../Outputs_processados/filtrado_por_range_6.csv")

filtrado_ratio_range

apenas_frequencia <- filtrado_ratio_range[, c(1, grep("freq.",
names(filtrado_ratio_range)))]

apenas_frequencia

apenas_frequencia$media_freq <- rowMeans(apenas_frequencia[, -c(1)], na.rm
= TRUE)

apenas_frequencia$sd_freq <- apply(apenas_frequencia[, -c(1)], 1, sd, na.rm
= TRUE)

apenas_frequencia$cv <-
apenas_frequencia$sd_freq/apenas_frequencia$media_freq

apenas_frequencia$Lemma <- gsub("_v", "", apenas_frequencia$Lemma)

apenas_frequencia

Juilliand_d <- function(cv, K = 8){
  1 - cv/sqrt(K - 1)
}

apenas_frequencia$dispersion <- Juilliand_d(apenas_frequencia$cv)

apenas_frequencia

hist(apenas_frequencia$dispersion, xlab = "Juilliand d", ylab = "", main =
"")

limiar_dispersion <- .80

dispersion_higher <- subset(apenas_frequencia, dispersion >=
limiar_dispersion)

dispersion_higher

write.csv(dispersion_higher,
file = paste("../Outputs_processados/filtrado_por_dispersion_",
limiar_dispersion, ".csv", sep = ""),
row.names = FALSE)

```

ANEXO 5 – Script para calcular Medida de disciplina e Frequência mínima

```

library(tidyverse)

correct_dispersion <- read.csv(file =
"../Outputs_processados/filtrado_por_dispersion_0.8.csv")

head(correct_dispersion)

tail(correct_dispersion, 10)

apenas_frequencia <- correct_dispersion[, c(1, grep("freq.",
names(correct_dispersion))) ]

head(apenas_frequencia )

ratios_internos <- data.frame(
  Lemma = apenas_frequencia$Lemma,
  t( apply(apenas_frequencia[, -1] , 1, function(x) x/x[1] ) )
)

names(ratios_internos) <- gsub("freq.", "internal_ratio.",
names(ratios_internos))

ratios_internos

apply(ratios_internos[, -1], 2, max, na.rm = TRUE)

subset(apenas_frequencia, Lemma == "processar")

apply(ratios_internos[, -1], 2, function(x)
ratios_internos$Lemma[which.max(x)])

discipline_measure_limiar <- 3

## calculando se o ratio excede o limiar em qualquer das áreas
ratios_internos$exceeds_limiar <- apply(ratios_internos[, -1], 1,
function(x) any( na.omit(x) > discipline_measure_limiar))

ratios_internos

good_measure <- subset(ratios_internos, exceeds_limiar == FALSE)

good_measure

good_dispersion_and_measure <- correct_dispersion[ correct_dispersion$Lemma
%in% good_measure$Lemma, ]

good_dispersion_and_measure

write.csv(good_dispersion_and_measure,
file =
"../Outputs_processados/filtrado_por_disclipline_measure.csv",
row.names = FALSE
)

```