

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

MONIQUE VIEIRA MIRANDA

PROCESSOS VERBAIS EM ARTIGOS CIENTÍFICOS:
UMA ANÁLISE COM BASE NA LÍNGUA EM USO

BELO HORIZONTE
2016

MONIQUE VIEIRA MIRANDA

PROCESSOS VERBAIS EM ARTIGOS CIENTÍFICOS:

UMA ANÁLISE COM BASE NA LÍNGUA EM USO

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de mestre 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

FACULDADE DE LETRAS DA UFMG

2016

M672p

Miranda, Monique Vieira.

Processos verbais em artigos científicos [manuscrito] : uma análise com base na língua em uso / Monique Vieira Miranda. – 2016.

105 f., enc. : il., tabs., color., p&b.

Orientadora: Ana Larissa Adorno Marciotto Oliveira.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos da Língua em Uso.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 88-91.

Apêndices: f. 92-105.

1. Linguística aplicada – Teses. 2. Funcionalismo (Linguística) – Teses. 3. Redação acadêmica – Teses. 4. Comunicação escrita – 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
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS



FOLHA DE APROVAÇÃO

Processos verbais em artigos científicos: uma análise com base na língua em uso

MONIQUE VIEIRA MIRANDA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos da Língua em Uso.

Aprovada em 30 de novembro de 2016, pela banca constituída pelos membros:

Prof(a). Ana Larissa A. Marciotto Oliveira - Orientador
UFMG

Prof(a). Cláudio Márcio do Carmo
UFESJ

Prof(a). Bárbara Malveira Orfanó
UFMG

Belo Horizonte, 30 de novembro de 2016.

Para meus pais
Para meu compz
Para Cida e Ana Larissa

RESUMO

Esta pesquisa teve como objetivo elaborar uma lista dos lexemas verbais acadêmicos, de acordo com a divisão do vocabulário da língua proposta por Nation (2001), e analisar qualitativamente os processos verbais acadêmicos mais frequentes, com base na Linguística Funcional Centrada no Uso, principalmente Halliday e Matthiessen (2004). Para tanto, foi compilado um *corpus* composto por artigos científicos de todas as áreas de avaliação da CAPES, a fim de observar a existência de diferenças nas escolhas lexicais de cada Grande Área e de particularidades do gênero que pudessem ser percebidas no *corpus*, em escala quantitativa e qualitativa, assim como diferenças entre os usos dos processos verbais mais frequentes. Como resultado apresentou-se duas listas, uma dos 281 lexemas verbais mais frequentes nos artigos científicos, outra dos 156 verbos acadêmicos mais frequentes, baseada na comparação com *corpora* representativos da linguagem não-acadêmica. A elaboração dessas listas permitiu perceber que algumas palavras altamente frequentes na língua, como “olhar”, “dizer” e “achar”, tiveram pouquíssimas ocorrências no *corpus*, demonstrando ser próprio do gênero evitar itens lexicais que carreguem o lugar-comum da palavra frequente (LINDQVIST, 2014). Como alternativa às poucas ocorrências de “dizer”, um dos lexemas verbais mais utilizados entre as palavras de alta frequência na língua geral, seis processos verbais acadêmicos frequentes foram analisados qualitativamente. Desses, metade foram prototipicamente processos verbais, “destacar”, “sugerir” e “descrever”, enquanto na outra parte, “demonstrar”, “determinar” e “estabelecer”, outros tipos de processos também ocorreram, sendo bastante frequentes. Apesar dessa diferença, todos os processos verbais foram empregados de forma similar ao longo dos trechos analisados, havendo predominância de formas na terceira pessoa do presente do indicativo, que têm a capacidade de fazer com que uma ação no passado seja trazida para o presente do leitor, causando um efeito de atemporalidade (FUZER, 2012), ou omitindo-se o sujeito da oração através da indeterminação, o que distancia o autor de seu objeto de estudo. Esse mesmo propósito foi alcançado com o uso da passiva, a fim de omitir a participação efetiva do autor ou de dar destaque à mensagem expressa pelo processo verbal. Tais mecanismos contribuíram para tornar o discurso científico mais objetivo e impessoal. Mesmo sendo um participante obrigatório dos processos verbais, o Dizente muitas vezes não estava presente, devendo ser inferido pelo contexto, porém, em todos os casos analisados, a mensagem comunicada estava sempre presente, seja como Verbiagem, seja como oração projetada. Em geral, os resultados obtidos permitiram perceber que há diferenças entre as Grandes Áreas, principalmente nas palavras acadêmicas preferidas por cada, mas, apesar de existirem, tais particularidades não interferiram nos dados quantitativamente, sendo apenas observáveis quando comparadas entre si, como as preferências de cada Grande Área por certos processos em detrimento de outros.

Palavras-chave: processos verbais; lexemas verbais; vocabulário acadêmico; artigos científicos.

ABSTRACT

This research aimed to develop a list of academic verbal lexemes, according to the vocabulary division proposed by Nation (2001), and qualitatively analyze the most frequent academic verbal processes, from a Usage-Based Linguistics perspective, especially Halliday and Matthiessen (2004). Thus, a corpus of research articles in all evaluation areas of CAPES was compiled in order to observe the existence of differences in lexical choices of each subcorpora and genre characteristics, as well as differences between uses of most frequent verbal processes. As a result, we presented two lists, one of 281 most frequent verbal lexemes in research articles, and one of 156 most common academic verbs, based on comparison with representative corpora of non-academic language. These lists confirm that some highly frequent words in the Portuguese, such as "olhar", "dizer" and "achar", had very few occurrences in our corpus, showing that academic vocabulary avoids high frequency words (LINDQVIST, 2014). As an alternative to the few tokens of "dizer", we analyzed qualitatively six most frequent verbal processes, of which half were prototypically verbal processes, "destacar", "sugerir" and "descrever" while the other part, "demonstrar", "determinar" and "estabelecer", were frequently associated with other types of processes, such as material processes. Despite this difference, all verbal processes were used in a similar manner over the analyzed passages, with predominance of tokens in simple present tense, which have the ability to bring an action from the past to the reader's present, causing a timelessness effect (FUZER, 2012), or omitting the subject of a sentence, which tends to distance the author from its study. This same purpose has been achieved by using passive voice, in order to omit the author's participation or to highlight the message from verbal processes. Such mechanisms contribute to making the text more objective and its discourse impersonal. Even though Sayer is directly involved participant on verbal processes, it was often not explicit, and needed to be inferred from context. However, in all analyzed cases, there was a message conveyed either as a Verbiage or as projected sentence. Overall, the results allowed us to notice that there were differences between each discipline, especially in preferred academic words. Nevertheless, these characteristics did not interfere with our data quantitatively, being only observed when compared to preferences from each subcorpus ranking.

Keywords: verbal processes; verbal lexemes; academic vocabulary; research articles.

LISTA DE ILUSTRAÇÕES

Quadro 1: Relação entre contexto de situação, metafunção e os principais sistemas de cada metafunção.....	32
Quadro 2: As principais categorias dos processos relacionais exemplificadas.....	42
Quadro 3: Exemplo do sistema de classificação da CAPES.....	48
Figura 1: Tipos de processos	35
Figura 2: Sistema de transitividade	36

LISTA DE TABELAS

Tabela 1: Tamanho e cobertura do vocabulário (CARROLL; DAVIES; RICHMAN, 1971)	19
Tabela 2: Tamanho e cobertura do vocabulário desta pesquisa	20
Tabela 3: Composição do corpus da pesquisa	50
Tabela 4: Média de palavras em artigos científicos por Grande Área.....	53
Tabela 5: Lista dos lexemas verbais acadêmicos mais frequentes	60
Tabela 6: Formas linguísticas dos processos verbais mais frequentes analisados.....	66
Tabela 7: Número total de ocorrências de cada processo por Grande Área.....	71

LISTA DE ABREVIATURAS E SIGLAS

- AVL** *Academic Vocabulary List*
- AWL** *Academic Word List*
- CAPB** Corpus Acadêmico do Português Brasileiro
- GSL** *General Service List*
- LCFU** Linguística Funcional Centrada no Uso
- LE** Língua Estrangeira
- PB** Português Brasileiro
- UWL** *University Word List*

SUMÁRIO

1	CONSIDERAÇÕES INICIAIS	13
2	TIPOS DE VOCABULÁRIO	16
2.1	Palavras de alta frequência.....	19
2.2	Palavras de baixa frequência.....	21
2.3	Vocabulário técnico.....	23
2.4	Vocabulário acadêmico	26
3	O SISTEMA DA TRANSITIVIDADE	29
3.1	As metafunções da linguagem.....	31
3.2	Metafunção ideacional: o sistema de transitividade.....	34
3.2.1	Processos materiais	37
3.2.2	Processos mentais	39
3.2.3	Processos relacionais	41
3.2.4	Processos comportamentais	43
3.2.5	Processos verbais	44
3.2.6	Processos existenciais	45
4	METODOLOGIA	47
4.1	Composição do <i>corpus</i>	47
4.2	Procedimentos de coleta.....	52
4.3	Procedimentos de análise	54
5	ANÁLISE QUANTITATIVA: lista dos verbos acadêmicos mais frequentes	57
5.1	Identificando o vocabulário acadêmico.....	57
5.2	Lista dos verbos acadêmicos mais frequentes.....	59
6	ANÁLISE QUALITATIVA: processos verbais em destaque	66
6.1	Panorama geral dos processos verbais mais frequentes	66
6.2	A transitividade em processos verbais acadêmicos.....	70
6.2.1	DEMONSTRAR.....	73

6.2.2	DETERMINAR	75
6.2.3	DESTACAR	77
6.2.4	ESTABELEECER.....	78
6.2.5	SUGERIR	80
6.2.6	DESCREVER	82
7	CONSIDERAÇÕES FINAIS	85
	REFERÊNCIAS	88
	APÊNDICE A: Lista dos lexemas verbais mais frequentes no corpus, contrastados com os lexemas verbais pertencentes ao vocabulário de alta frequência (BERBER SARDINHA, 2004; BIDERMAN, 2000)	92
	APÊNDICE B: Lista dos lexemas verbais mais frequentes na lista de palavras mais frequentes do Banco do Português (BERBER SARDINHA, 2004), contrastado com os mais frequentes em Biderman (2000).....	102

1 CONSIDERAÇÕES INICIAIS

Esta pesquisa teve como objetivo elaborar uma lista dos lexemas verbais acadêmicos, de acordo com a divisão do vocabulário da língua proposta por Nation (2001), e analisar qualitativamente os processos verbais mais frequentes com base no quadro teórico da Linguística Funcional Centrada no Uso (LFCU) e em Halliday e Matthiessen (2004; 2014), Thompson (2014) e Eggins (2004).

Para a análise, foi compilado um *corpus* composto por artigos científicos de todas as áreas de avaliação da CAPES¹, subdividido nas respectivas nove Grandes Áreas de conhecimento: Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências da Saúde, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas, Linguística, Letras e Artes e Multidisciplinar, num total aproximado de 4,8 milhões de palavras. O *corpus* foi assim elaborado a fim de permitir identificar os verbos acadêmicos mais frequentes, sem ser tendencioso para uma determinada área.

Esse mesmo *corpus* serviu como base para o projeto interinstitucional Corpus Acadêmico do Português Brasileiro (CAPB)², entre a Universidade Federal de Minas Gerais e a Universidade Federal de Viçosa, que objetivou coletar textos de diferentes gêneros acadêmicos como trabalhos de conclusão de curso, dissertações e teses disponíveis online, representantes do domínio, com a finalidade de facilitar futuras pesquisas linguísticas na área.

A necessidade de compilar um *corpus* com tamanha extensão e variedade de artigos surgiu durante o projeto de Iniciação Científica e de Trabalho de Conclusão de Curso, ambos sob a orientação da professora Aparecida de Araújo Oliveira, em que foram levantados *corpora* especializados com 300.000 palavras e 900.000 mil palavras, em algumas poucas áreas de conhecimento – duas áreas no primeiro projeto, seis no segundo. Por serem muito específicos dessas áreas, os resultados da análise dos substantivos mais frequentes não foram suficientes para observar um vocabulário acadêmico de qualidade, surgindo então a motivação para ampliar tanto o *corpus*, para ser representativo do gênero, quanto a análise, para focar nos verbos acadêmicos.

¹ Disponível em:

<http://www.capes.gov.br/images/stories/download/avaliacao/TabelaAreasConhecimento_072012.pdf>. Acesso em: 30 maio 2016.

² Gostaria de agradecer a todos os participantes da equipe, Carlos Guedes, Luisa Andresa Guerra, Lucas Marciano, João Paulo Miranda, Fernando Suarez e Dalila da Silva Carvalho, que se dedicaram ao trabalho árduo de coleta dos textos, assim como às professoras Ana Larissa Adorno Marciotto Oliveira e Aparecida de Araújo Oliveira, que possibilitaram a criação do projeto.

Além do propósito de descrição da língua em uso, esta pesquisa também pretendeu observar padrões específicos do gênero, que são observáveis apenas quando comparados com outros *corpora*, como foi realizado, a fim de distinguir o vocabulário acadêmico.

Halliday e Matthiessen (2014, p. xvii) compreendem a linguagem como um processo semiótico, em que o significado se dá através de um sistema de escolhas baseados no contexto de uso. Essa é a perspectiva deste trabalho, que analisou padrões específicos dos processos verbais acadêmicos mais frequentes. Esse tipo de análise contribui para a compreensão das escolhas feitas pelo usuário da língua nesse domínio social.

De tal forma, as seguintes perguntas de pesquisa foram elaboradas:

1. Quais são os verbos mais frequentes em artigos científicos de diferentes áreas de conhecimento?
2. É possível listar um vocabulário acadêmico, representativo do gênero, a partir desse *corpus*?
3. Há particularidades nessa lista que a distingue de outras, formadas por diferentes gêneros textuais?
4. Quais são os verbos utilizados para reportar discursos (processos verbais) mais frequentes no *corpus*?
5. Existem semelhanças ou diferenças entre os usos dos processos verbais, sob a perspectiva funcional?

Os resultados dessa pesquisa, a lista dos principais verbos acadêmicos e a análise dos processos verbais mais frequentes, forneceram bases teóricas empíricas que poderão ser utilizadas para o ensino de Português Brasileiro (PB) com fins acadêmicos.

Estudos similares sobre artigos científicos foram realizados sob a perspectiva da Linguística Sistêmico-Funcional (PORTELA, 2013; PENHA, 2012; MENDES, 2012; FUZER, 2012; VIVAN, 2010). Tais trabalhos trataram majoritariamente de processos do tipo verbal que ocorreram em artigos científicos de uma ou duas áreas de conhecimento. A presente dissertação se difere dos estudos anteriores por fazer uma análise dos principais processos verbais em um *corpus* representativo do gênero como um todo, e não apenas restringido à determinada área temática.

Com relação ao vocabulário acadêmico, um trabalho similar foi realizado por Santos (2006), cuja dissertação objetivou elaborar uma lista do vocabulário acadêmico a partir de teses e dissertações de diferentes áreas. Este estudo diferencia-se daquele tanto em gênero quanto em metodologia, pois Santos (2006) lematizou apenas substantivos e adjetivos, mantendo verbos em suas formas originais (*type*), assim como advérbios e as palavras

gramaticais. Nesta pesquisa, foram listados os 281 lexemas verbais mais frequentes, que resultaram na lista dos 156 verbos acadêmicos, representativos de artigos científicos.

Esta dissertação foi estruturada da seguinte forma: o capítulo 2 descreveu a abordagem teórico-metodológica da divisão do vocabulário adotada nesta pesquisa. O capítulo 3 apresentou a perspectiva funcionalista em que a análise qualitativa se baseia. O capítulo 4 detalhou o processo de coleta, organização e análise do *corpus*. No capítulo 5 foram apresentados os resultados quantitativos da pesquisa, a lista dos lexemas verbais acadêmicos mais frequentes. No capítulo 6 os principais processos verbais, baseados na lista anterior, foram analisados qualitativamente. E, por fim, nas Considerações Finais retomou-se discussões anteriores, surgidas ao longo da análise.

2 TIPOS DE VOCABULÁRIO

Dominar o gênero acadêmico é relevante tanto para estudantes de língua materna quanto aprendizes de uma língua estrangeira (LE) com fins acadêmicos. Segundo Dell’Isolla (2005), a habilidade de compreender plenamente um desses gêneros depende do tamanho do vocabulário que o estudante domina, sendo necessária atenção redobrada ao estudo desta área se o indivíduo for aprendiz de LE. Além do conhecimento básico da língua, o estudante precisará dominar o vocabulário acadêmico e o vocabulário técnico, próprio de sua área de conhecimento.

Essa afirmação não implica que o estudante deva aprender o maior número possível de palavras, mas sim o maior número de palavras frequentes de uma língua. Estudos sobre a frequência do vocabulário (NATION, 2001; CHUNG; NATION, 2003; BIDERMAN, 2001) demonstram que não há necessidade de tratar com igual importância todas as palavras, pois a proporção de seu uso vai ser diferente, assim como a relevância de seu ensino.

Nation (2001) divide as palavras de uma língua em quatro tipos: **palavras de alta frequência, vocabulário acadêmico, vocabulário técnico e palavras de baixa frequência**. Outros pesquisadores trabalham com propostas similares de divisão do vocabulário de acordo com seu contexto de uso, como Baker (1988), por exemplo, que divide o vocabulário em **léxico geral, léxico especializado e léxico sub-técnico** ou **itens retóricos** (se comparados a classificação de Nation, esses termos correspondem respectivamente a palavras de alta frequência, vocabulário técnico e vocabulário acadêmico), mas não leva em consideração as palavras de baixa frequência.

Nesta pesquisa utilizaremos a classificação de Nation (2001) por ser uma das mais completas, em termos de distinção entre as fronteiras dos tipos de vocabulário, e por ser a que mais se aproxima dos objetivos deste trabalho – a descrição de um léxico acadêmico.

Sob essa perspectiva, as palavras de alta frequência correspondem a cerca de 80% dos *tokens*³ de um texto, sendo compostas por palavras gramaticais e palavras de conteúdo, como nomes, adjetivos e verbos. Nation (2001) utiliza a *General Service List* (GSL) como base para as palavras mais frequentes do inglês. Essa lista, elaborada por West (1953), descreve as

³ Um *token* refere-se a cada ocorrência de um *type* (tipo) (BERBER SARDINHA, 2004, p. 94). Assim, a frase “A menina abraçou a avó” tem 4 *types* e 5 *tokens*, considerando que as duas ocorrências (*tokens*) de “a” correspondem ao mesmo *type*.

2.000 famílias de palavras⁴ mais frequentes e, apesar de ter sido proposta a mais meio século, continua atual e cobre entre 75% e 85% das palavras em um texto, de acordo com o gênero, segundo Chung e Nation (2003).

Em língua portuguesa, uma lista similar das palavras mais frequentes do português foi publicada por Berber Sardinha (2004), com base no *corpus* Banco do Português da PUC-SP que, na época, contava com cerca de 223 milhões de palavras⁵. Essa lista se difere da GSL por não agrupar as palavras em famílias, mantendo as ocorrências organizadas apenas pela frequência de *types*.

Biderman (1998; 2001) realizou um estudo relevante sobre as palavras mais frequentes da língua portuguesa com um *corpus* de quase 5 milhões de palavras, retiradas de amostras da linguagem oral e escrita de diferentes gêneros (romances, dramaturgias, acadêmicos e jornalísticos). Além de compilar um *corpus* variado, a pesquisadora preocupou-se em listar as palavras mais frequentes organizando-as por frequência do lexema.

Por serem dois exemplos de fácil acesso em língua portuguesa, baseados em *corpora* bem-estruturados e formados por amostras exclusivas do PB, utilizaremos essas duas listas como base das palavras de alta frequência em língua portuguesa, assim como Nation (2001) se baseou na GSL para exemplificar tal tipo de vocabulário em língua inglesa.

Em textos da esfera acadêmica, os vocabulários acadêmico e técnico têm frequência em torno de 8,5% e 5% respectivamente (CHUNG; NATION, 2003). Apesar de esses números parecerem pouco significativos, tais itens lexicais são primordiais para o sucesso no entendimento do texto e, de tal forma, requerem maior atenção ao serem ensinados, pois não costumam aparecer em outros gêneros. Grande parte das pesquisas realizadas sobre esses dois tipos de vocabulário tem foco na aprendizagem de língua inglesa.

Assim, além da mencionada GSL, pode-se destacar a *University Word List* (UWL), proposta por Xue e Nation (1984), dos itens lexicais acadêmicos mais frequentes, baseada em famílias de palavras não encontradas na GSL; e a lista proposta por Coxhead (2000), *Academic Word List* (AWL), com 570 famílias de palavras acadêmicas, que igualmente não pertencem à GSL e substituiu explicitamente a UWL (NATION, 2001, p. 12); e a *Academic Vocabulary List* (AVL) (GARDNER; DAVIES, 2014), compilada a partir do *subcorpus* acadêmico do *Corpus of Contemporary American English* (DAVIES, 2012), que se baseia em critérios próprios para determinar o vocabulário acadêmico.

⁴ **Famílias de palavras**, de acordo com Cunha e Cintra (2008, p. 96), consiste na forma básica de uma palavra e todas as suas formas flexionadas, derivadas ou composicionais.

⁵ Banco do Português. Disponível em <<http://www2.lael.pucsp.br/corpora/bp/>>. Acesso em: 10 maio 2016.

Um exemplo de estudo voltado para o ensino de inglês com fins acadêmicos é o de Campoy (2000), sobre verbos e seus colocados em textos da área de Química, que foram classificados de acordo com o tipo de ação expressa. Além desse, Martínez *et al.* (2009) fizeram uma análise quantitativa e qualitativa do vocabulário acadêmico em artigos científicos da área de Agricultura. As autoras observaram que 92 famílias de palavras de seu *corpus* estavam presentes na lista acadêmica de Coxhead (2000), AWL, sendo que algumas destas famílias possuíam significado técnico no *corpus* e outras, de uso geral, que não pertencem à AWL, mas à GSL, possuíam sentido acadêmico. Com base nessas descobertas, as autoras alertaram para a possibilidade de se produzirem listas de vocabulário acadêmico específicas para cada campo de conhecimento.

Com foco no vocabulário técnico, Chung e Nation (2003) propuseram identificá-lo através de uma escala de classificação, utilizando um dicionário especializado ou definições, dicas e glossários que o próprio autor do texto pode fornecer. Essa categoria é marcada por se restringir a determinadas áreas, o que não acontece com o vocabulário acadêmico.

Um importante estudo realizado por Sutarsyah, Nation e Kennedy (1994) comparou dois *corpora*, cada um com 300.000 palavras: um livro de Economia e um conjunto de textos acadêmicos de diversas áreas do conhecimento. Eles observaram que as palavras que compõem o vocabulário técnico, perceptíveis no *corpus* de Economia, constituíram uma pequena lista de palavras muito frequentes enquanto, no segundo *corpus*, devido à sua variedade, o vocabulário acadêmico destacou-se sobre o técnico como mais frequente e mais amplo.

Tais resultados, em conjunto com pesquisas anteriores da autora deste trabalho (MIRANDA; OLIVEIRA, 2016), demonstram que o vocabulário técnico tende a se destacar em listas de frequência de uma mesma área de conhecimento. Consequentemente, o vocabulário acadêmico tende a se destacar em listas com mais de uma área de conhecimento, ao lado das palavras de alta frequência.

Assim, justifica-se e reforça-se a importância da investigação do vocabulário acadêmico, pois seu ensino alcança mais alunos em turmas heterogêneas e, para ele, não há dica do autor ou glossário. Não compreendê-lo pode comprometer o entendimento do texto.

Em língua portuguesa, Santos (2006) propôs uma lista das palavras acadêmicas mais frequentes, tendo como base um *corpus* de 7 milhões de palavras composto por teses e dissertações de diferentes áreas do conhecimento, publicadas em três universidades brasileiras. No trabalho, 512 palavras foram classificadas como vocabulário acadêmico, excluindo-se as 3.000 mais frequentes da língua (segundo o *corpus* do Banco do Português, da PUC-SP). Em tal lista, apenas substantivos e adjetivos foram lematizados (considerando-se apenas suas

formas flexionadas), mantendo verbos em suas formas originais (*type*), assim como advérbios e as palavras gramaticais.

Como expresso no capítulo 1, um dos objetivos da presente dissertação foi elaborar uma lista dos lexemas verbais acadêmicos mais frequentes, o que não foi o foco de Santos (2006), a fim de complementar os estudos nesta área em língua portuguesa. Nas próximas seções, cada um dos quatro tipos de vocabulário foi descrito com maiores detalhes, em especial o vocabulário acadêmico.

2.1 Palavras de alta frequência

As palavras de alta frequência constituem um pequeno grupo de palavras extremamente importantes, pois cobrem grande parte das ocorrências de qualquer texto (falado ou escrito), independente do gênero. Como Nation (2001) observa em seu estudo, e também foi observado nesta dissertação (ver seção 6.2), as palavras mais frequentes podem variar de ranking, conforme o *corpus*, mas costumam ser bastante estáveis, mantendo-se dentro da lista das palavras mais frequentes.

Nation (2001, p. 15) apresenta uma tabela (reproduzida na Tabela 1 abaixo) que ilustra a proporção da cobertura do vocabulário de um *corpus* de cinco milhões de palavras (CARROLL; DAVIES; RICHMAN, 1971). Esses dados permitem chegarmos a importantes generalizações sobre as palavras mais frequentes da língua, tais quais: 1) as 2.000 palavras mais frequentes realmente cobrem cerca de 80% do total de palavras; 2) as palavras de menor frequência são mais variadas (possuem maior número de *types*), mas menor proporção de *tokens* por *type*.

Tabela 1: Tamanho e cobertura do vocabulário (CARROLL; DAVIES; RICHMAN, 1971)

Número de palavras	% cobertura do corpus
86.741	100
43.831	99
12.448	95
5.000	89,4
4.000	87,6
3.000	85,2

Número de palavras	% cobertura do corpus
2.000	81,3
1.000	74,1
100	49
10	23,7

Fonte: Adaptado de Nation (2001, p. 15).

A fim de comparação, elaboramos uma lista similar com os dados do *corpus* deste trabalho (Tabela 2). Pode-se perceber que apesar do número total de *tokens* (aproximadamente 5 milhões) ser similar ao de Carroll, Davies e Richman (1971), nosso *corpus* apresenta número maior de *types* (quase vinte mil a mais). Isso ocorre em função da grande quantidade de palavras técnicas de cada uma das 48 áreas de avaliação que compõem este *corpus*.

Tabela 2: Tamanho e cobertura do vocabulário desta pesquisa

Número de palavras	% cobertura do corpus
106.215	100
63.646	99,1
53.000	98,7
12.000	92,1
5.000	84,9
4.000	82,8
3.000	79,8
2.000	75,5
1.000	67,6
100	45,2
10	24,1

As primeiras duas mil palavras mais frequentes deste *corpus*, mesmo não cobrindo a mesma porcentagem do *corpus* de Carroll, Davies e Richman (1971), se mantêm dentro da média estabelecida por Hwang e Nation (1995). Porém, deve-se ressaltar que tal comparação não deve exceder os limites da mera apresentação dos dados, pois nosso *corpus* não é representativo de diferentes gêneros, apenas do gênero artigo científico, e dada a grande

variedade temática deste *corpus*, é natural essa diferença na proporção do número de *types* por porcentagem da ocorrência.

Além disso, por nosso *corpus* ser representativo de um único gênero, nossa lista de palavras não irá apenas fornecer palavras de alta frequência, mas também palavras pertencentes ao vocabulário acadêmico, como dito na seção anterior (e demonstrado por SUTARSYAH *et al.*, 1994; MIRANDA; OLIVEIRA, 2016). Essa divergência é uma característica do nosso *corpus*, que foi compilado para cumprir o propósito de destacar o vocabulário acadêmico.

Ainda sobre as palavras de alta frequência, mas dessa vez em língua portuguesa, Biderman (1998), em seu *corpus* (introduzido na seção anterior) composto por cinco milhões de palavras retiradas de diferentes gêneros, classifica as palavras com até 500 ocorrências como as de “altíssima frequência”, que correspondem às primeiras 1.078 de seu *corpus* e representam 42% das ocorrências totais. Se aplicarmos seu ponto de corte de ocorrências aos nossos dados, teremos 1.105 *types* representando 68,8% das ocorrências totais do *corpus*.

Apesar de ambos *corpora* comparados (BIDERMAN, 1998; CARROL; DAVIES; RICHMAN, 1971) possuírem a mesma quantidade de palavras e listarem as palavras mais frequentes agrupadas por *types* (em vez de famílias de palavras como a GSL), nossas porcentagens se aproximam das de Biderman (1998). Isso ocorre provavelmente pela composição do *corpus*, por ser o de Biderman (1998) mais variado quanto aos gêneros textuais (e o nosso variado quanto à temática, aumentando assim o vocabulário técnico), assim como a diferença na estrutura de cada uma das línguas, portuguesa e inglesa.

É interessante observar, nas Tabelas 1 e 2, que as primeiras 1.000 palavras são bastante frequentes, cobrindo 74,1% e 67,6% respectivamente, porém, não há diferença drástica entre as primeiras 1.000 palavras e as 2.000 palavras, havendo aumento apenas de 7,2% e 7,9% em cada um dos casos. Esse número vai diminuindo exponencialmente conforme a frequência de *tokens* diminui. A existência desse pequeno número de palavras altamente frequente (independente do gênero) justifica a importância que se deve dar ao seu ensino/aprendizagem.

Na próxima seção, o tipo de vocabulário na extremidade oposta das palavras de alta frequência foi discutido: as palavras de baixa frequência.

2.2 Palavras de baixa frequência

Esse grupo constitui uma enorme gama de palavras infrequentes, que cobrem uma pequena parcela dos textos, em geral, menos de 5% das palavras de um texto acadêmico,

segundo Nation (2001, p. 12). Essa é uma porcentagem fácil de se observar em nosso *corpus*, tomando a Tabela 2 como exemplo, em que metade das ocorrências registradas (53.000 *types*) possui menos de duas ocorrências totais. Na verdade, a porcentagem mencionada por Nation (2001), 5%, corresponde a cerca de 86.000 *types* em nosso *corpus*, ou seja, as primeiras 20.000 palavras cobrem 95% das palavras do *corpus*.

As palavras de baixa frequência são aquelas que não pertencem ao grupo de alta frequência, abrangendo tanto as de frequência média quanto baixa. Nesse ponto cabe dizer que os limites entre alta e baixa frequência é arbitrário, muitas vezes cabendo ao pesquisador definir um número, de acordo com o **alcance** (*range*) da palavra entre os diferentes *subcorpora* utilizados.

Assim, é possível certas palavras serem consideradas como baixa frequência em um *corpus*, e alta frequência em outro. É nessa possibilidade que nosso trabalho se baseia: classificamos como palavras acadêmicas aquelas que 1) seriam consideradas de baixa frequência em *corpora* representativos de outros gêneros; 2) não pertenceriam a uma área específica de conhecimento, pois o alcance da palavra abrange diversas áreas.

A segunda alínea anterior exemplifica outro limite para a definição de palavras de baixa frequência: não devem pertencer exclusivamente a uma determinada área de conhecimento, pois, nesse caso, seriam consideradas vocabulário técnico. Na seção 2.3, foi descrita a metodologia bastante útil de Chung e Nation (2003) para caracterizar o vocabulário técnico, distinguindo-o dos demais.

Tendo tais observações em mente (palavras de baixa frequência são um grupo muito amplo, mas infrequente, podendo variar de acordo com o *corpus* utilizado, mas sem abranger o vocabulário técnico), as palavras de baixa frequência costumam ser formadas por nomes próprios, palavras muito formais, arcaicas, vulgares, estrangeiras, neologismos e gírias.

De tal forma, ao afirmarmos que palavras como “blindada”, “enjaulado” e “instou” pertencem à categoria de baixa frequência em nosso *corpus*, por terem apenas uma ocorrência, não queremos dizer que elas não sejam importantes ou úteis aos usuários da língua – significa apenas que tais palavras (também conhecidas por *hápax legómenon*⁶) não são frequentes no gênero artigo científico e, portanto, estudantes com fins específicos devem preterir-las em favor de palavras mais frequentes. Em vez do ensino extensivo desse vocabulário, Nation (2001, p. 20) sugere o ensino de estratégias para lidar com ele, tais quais inferir o sentido a partir do contexto, usar morfemas para lembrar do significado da palavra e consultar dicionários.

⁶ “Hápax legómenon”. BAKER, P; HARDIE, A; MCENERY, T. **A Glossary of Corpus Linguistics**. Edinburgh University Press: Edinburgh. 2006.

Na próxima seção, métodos para diferenciar o vocabulário técnico dos demais foram descritos e ilustrados com exemplos do *corpus* deste trabalho.

2.3 Vocabulário técnico

Uma palavra técnica é aquela que requer conhecimento de determinada área para sua plena compreensão. Nation (2001) e Chung e Nation (2003), estabelecem níveis de tecnicidade de uma palavra para classificar seu pertencimento ao vocabulário técnico. Segundo os autores, tais características dependem da frequência com que a palavra ocorre fora do contexto daquela área (qual seu alcance em diferentes áreas) e do sentido em que tais palavras foram utilizadas.

Um dos métodos de se diferenciar o vocabulário técnico dos demais é comparar um *corpus* de uma determinada área com outro de uma área não-relacionada, examinando a lista de palavras mais frequentes (SUTARSYAH *et al.*, 1994). Outro método é observar o alcance da palavra ao longo de diferentes *corpora*, compostos por uma variedade de áreas de conhecimento – quanto menor o alcance de uma palavra, combinado com sua frequência (que em *corpora* específicos costuma ser alta, mas não somente), maior a possibilidade de estarmos lidando com uma palavra pertencente ao vocabulário técnico. Por exemplo, “significado” e “infecção” têm o mesmo número de ocorrências no *corpus* da pesquisa, 417 *tokens*, mas o alcance da primeira palavra é muito maior (todos os nove *subcorpora*), enquanto a segunda palavra restringe-se a três *subcorpora* (Ciências da Saúde, Ciências Biológicas e Ciências Agrárias), caracterizando-a com parte do vocabulário técnico dessas áreas.

Além desses métodos, segundo Chung e Nation (2003), o vocabulário técnico pode ser identificado por alunos ou professores através das seguintes estratégias: intuição de um especialista no assunto, consulta a um dicionário técnico e uso de dicas que o próprio autor do texto fornece ao introduzir um novo termo, como grifos ou notas explicativas.

Nation (2001) propõe quatro níveis de classificação para o vocabulário técnico: a categoria 1 é formada por palavras que raramente ocorrem fora de determinado campo de conhecimento, por exemplo, na área de Ciência da Computação, “wavelet”⁷; na categoria 2, as palavras são utilizadas dentro e fora daquela área, mas com sentidos diferentes, como “hospedar”, que significa naquela área ‘fornecer o equipamento e as condições necessárias para

⁷ Todos os exemplos da área de Ciência da Computação foram retirados do *corpus* desta pesquisa.

manter um site⁸; na categoria 3, as palavras são empregadas com o mesmo significado dentro e fora da área, mas a maior parte de seus usos com sentido técnico restringe-se àquela área, como “memória”, referindo-se ao ‘dispositivo de um computador ou sistema informático que permite o registro, a conservação e a restituição dos dados’⁹; e, por último, na categoria 4, as palavras cujas ocorrências são mais comuns em certa área, porém, não há tecnicidade em seu significado – a única diferença é que peritos no assunto têm uma noção mais aprofundada dos termos, como “programa”, “usuário” e “sistema”.

Das quatro categorias listadas acima, as duas primeiras são claramente palavras técnicas enquanto as duas últimas são menos óbvias, por serem empregadas em diferentes áreas sem perder dramaticamente seu sentido mais básico. Tendo descrito tal classificação, cabe observar que as categorias 2 e 3 servem de exemplo para demonstrar porque *apenas* o uso exclusivo da ferramenta alcance não basta para detectar as palavras técnicas. Nesses casos, o significado da palavra deve ser levado em consideração (o que pode ser feito utilizando ferramentas como as **linhas de concordância**, ver seção 4.3).

Para melhor visualizar tais categorias, seguem alguns trechos retirados do *corpus* da pesquisa:

- (1) As tarefas de um **workflow**, também chamadas de atividades, são as unidades básicas do **workflow**. O sistema de **workflows** desenvolvido neste trabalho possibilita o uso de dois tipos de componentes de software como tarefas de **workflows**: métodos desenvolvidos na linguagem de programação **Java 2** e **serviços Web**. Um componente de software nada mais é do que o encapsulamento de um pedaço de software executável que possua uma **interface** bem definida. (Trecho retirado da área de Ciência da Computação)
- (2) Como se pode observar na Figura 2, o padrão **IEEE 802.11b** possui 14 canais. Cada canal possui um número que representa a frequência central do canal e uma **largura de banda** de 22 MHz. (Trecho retirado da área de Ciência da Computação)

Os itens lexicais em negrito nos exemplos (1) e (2) são avaliados como pertencentes a categoria 1, sendo, portanto, considerados vocabulário técnico na área de Ciência da Computação. É interessante observar que no trecho (1), o termo “workflow” vem seguido de sua definição, exemplificando o fato de que as palavras do vocabulário técnico costumam vir

⁸ “Hospedar”. Dicionário Priberam da Língua Portuguesa. Disponível em: <<http://www.priberam.pt/dlpo/hospedar>>. Acesso em: 30 maio 2016.

⁹ “Memória”. Dicionário Priberam da Língua Portuguesa. Disponível em: <<http://www.priberam.pt/dlpo/memória>>. Acesso em: 30 maio 2016.

seguidas de sua definição, diferente das palavras dos demais vocabulários (com exceção das palavras de baixa frequência em contexto didático de alfabetização).

As palavras com sublinhado duplo exemplificam a categoria 2, em que a forma foi utilizada com sentido técnico. No trecho (2), “canal” refere-se a ‘um meio destinado à transferência de informações [...] pode ser uma ligação física [...] ou consistir numa transmissão electromagnética em uma ou mais frequências’¹⁰. O item em negrito “largura de banda”, é composto pela forma “banda”, que poderia ser entendida como pertencente à categoria 2, porém, é preciso conhecimento técnico da área para saber que o termo é traduzido do inglês *bandwidth* e, portanto, é uma unidade composicional técnica, referindo-se à ‘quantidade de dados que podem ser transferidos através de uma interface num determinado período de tempo’¹¹.

E, por último, as palavras com sublinhado simples ilustram a categoria 4, em que palavras técnicas são utilizadas em outras áreas, mantendo seu sentido, porém, um perito no assunto teria maior conhecimento sobre os termos. No exemplo, a palavra “software”, apesar de técnica da área da Ciência da Computação, ocorre em quase todos os *subcorpora* da pesquisa (em oito dos nove *corpora*, com frequência acima de vinte *tokens*).

Quanto ao ensino desse vocabulário, alguns autores (STREVENS, 1973; CHUNG, NATION, 2003) acreditam que não é papel do professor ensinar palavras técnicas, pois elas são aprendidas através do estudo da própria área de conhecimento. Isso porque o estudante, que está inserido em tal meio, não terá tanta dificuldade ao aprender os termos técnicos ao passo que o professor terá não apenas que aprendê-los, mas também ensiná-los (STREVENS, 1973, p. 228).

Apesar disso, e com base na classificação do vocabulário de Nation (2001), o professor pode auxiliar o aluno com fins específicos naquelas palavras que, mesmo sendo técnicas, co-ocorrem nas listas de palavras mais frequentes (aquelas pertencentes às categorias 3 e 4). É claro, partindo-se do princípio de que o estudante já domina as palavras mais frequentes da língua, assim como o vocabulário acadêmico. Esse tipo de vocabulário foi discutido com maiores detalhes na próxima seção.

¹⁰ “Channel”. In: Dicionário prático de informática. Disponível em: <http://www2.contilnet.com.br/~Curso_Tecnico/Turma133/ingl%EA/Dicion%E1rio%20de%20Inform%Eltica.pdf>. Acesso em: 30 maio 2016.

¹¹ “Bandwidth”. In: Dicionário de termos de informática. Disponível em: <<http://www.faculdadesjt.com.br/tecnico/gestao/arquivosportal/file/Dicion%C3%A1rio%20de%20Informatica.pdf>>. Acesso em: 30 maio 2016.

2.4 Vocabulário acadêmico

Segundo Chung e Nation (2003), o vocabulário acadêmico é constituído por palavras que não são de uso geral, pertencentes às palavras de alta frequência de uma língua, mas são frequentes nos gêneros acadêmicos, independente da área de conhecimento. Em geral, esse vocabulário cobre uma boa parte de textos acadêmicos, cerca de 8,5% a 10% (SUTARSYAH *et al.*, 1994; COXHEAD, 1998). Tal número pode parecer pequeno à primeira vista, porém, basta lembrar que a cada cem palavras (um parágrafo, em média), dez pertencem ao vocabulário acadêmico, o que torna a porcentagem um número bastante relevante.

Não apenas pela quantidade, o vocabulário acadêmico também é importante por ser formado por “itens retóricos”, na terminologia de Baker (1988), que ajudam o leitor a compreender as intenções e avaliações do autor do texto. Martin (1976) enfatiza o papel do vocabulário acadêmico nas três etapas principais do processo acadêmico: pesquisa, análise e avaliação. A **pesquisa** envolve etapas de formulação, investigação, análise, conclusões e relato de resultados. O vocabulário de **análise** inclui formas verbais simples e compostas, que devem ser ensinadas com suas respectivas regências. Por fim, a **avaliação** pressupõe o conhecimento de adjetivos e advérbios, apesar do desejo de objetividade em textos acadêmicos.

De tal forma, o vocabulário acadêmico permite ao usuário da língua fazer generalizações sobre os diferentes processos envolvidos nas atividades científicas, auxiliando pesquisadores a expressarem ações e etapas envolvidas em suas pesquisas, independentemente da área de conhecimento.

Hyland (2006) lista uma série de características do vocabulário acadêmico dignas de nota, tais quais 1) alta densidade lexical, marcada pela elevada proporção de palavras de conteúdo por palavras gramaticais; 2) estilo altamente nominal, em que eventos e ações são expressos como nomes, em vez de verbos, agrupando um fenômeno complexo em uma sentença. Tal nominalização “congela” o evento, reformulando-o como um objeto. Essa transformação (tornando processos em objetos) expressa uma perspectiva científica que busca estabelecer relações entre entidades; 3) construções impessoais, evitando-se o uso da primeira pessoa do singular (ou plural) e de sentimentos. A primeira pessoa é geralmente substituída pela passiva, por processos existenciais (ver seção 3.2.6) e por metáforas metonímicas, em que a agentividade é atribuída a coisas em vez de pessoas.

Além dessas características, Lindqvist (2014), baseado em Deluca (2010), Uccelli *et al.* (2013) e Coxhead e Byrd (2007), acrescenta como próprios do gênero acadêmico 1) itens altamente frequentes na linguagem cotidiana que são substituídos em favor de formas similares

menos frequentes; 2) um vocabulário que favorece palavras precisas, evitando-se termos aproximados como “quase”; 3) marcadores que modalizam a opinião do autor do texto; 4) verbos no presente simples; 5) uso frequente da voz passiva.

Os exemplos (3) a (6), a seguir, ilustram tais características do vocabulário acadêmico:

- (3) A comparação de nossos resultados com outros estudos nacionais *apresentam* duas possíveis limitações. Em primeiro lugar, estudos envolvendo as práticas alimentares de adolescentes brasileiros *são* escassos na literatura científica, principalmente aqueles relacionados ao consumo de alimentos industrializados, ricos em açúcares e/ou gorduras. Após uma revisão bibliográfica, foram localizados onze artigos brasileiros que *avaliaram* o consumo destes alimentos entre adolescentes. (Trecho retirado da área de Nutrição)
- (4) A pintura de aços galvanizados *constitui-se* em uma solução eficaz para a proteção anticorrosiva de superfícies expostas a ambientes corrosivos. *Tem sido comprovado* que, quando materiais como o aço galvanizado *são* expostos a ambientes altamente agressivos ou quando *é* difícil o acesso a esses materiais, requerendo períodos muito longos para a manutenção, a pintura *promove* uma boa proteção nos mesmos. (Trecho retirado da área de Engenharia II)
- (5) O método de análise de covariância foi utilizado para avaliar as diferenças de inclinação e nível entre as curvas de crescimento em diâmetro por idade entre as árvores médias de *Pinus elliottii* e *Pinus taeda*, considerando os níveis de degradação do solo. Na análise de covariância, *utiliza-se*, conjuntamente, os conceitos de análise de variância e regressão e *relacionam-se* as variáveis independentes, em que não *se exerce* um controle exato sobre elas. (Trecho retirado da área de Ciência Agrárias I)
- (6) Os resultados obtidos nesse trabalho *confirmam* a eficiência do PMV como seleção indireta para o PMS. Ainda assim, a necessidade de obtenção de amostras da porcentagem de matéria seca da forragem produzida, utilizadas para as estimativas do PMS, não *está* descartada. (Trecho retirado da área de Biologia I)

O trecho (3) exemplifica as nominalizações em textos acadêmicos, logo na primeira linha com o item “comparação”. Nos exemplos, o presente do indicativo (em itálico) foi utilizado para lidar com fatos e pressuposições teóricas, em contraste com o particípio passado e a voz passiva (sublinhados nos exemplos), utilizados para lidar com ações realizadas pelos pesquisadores, a fim de diminuir o envolvimento explícito dos autores no processo de pesquisa. Para uma comparação entre o uso da voz passiva e/ou sua forma conjugada, ver Tabela 5, na seção 5.2, e o Apêndice A.

Nos trechos (3) e (6) as metáforas metonímicas estão exemplificadas em “onze artigos brasileiros que avaliaram o consumo destes alimentos entre adolescentes” e “os resultados obtidos nesse trabalho confirmam a eficiência do PMV como seleção indireta para o

PMS”, respectivamente. Em ambas sentenças, partes do trabalho de pesquisadores (“artigos” e “resultados”) são tomados como representantes do todo, sendo capazes, portanto, de “avaliar” e “confirmar” algo. Esse é um fenômeno muito frequente em textos acadêmicos, que pudemos comprovar com a análise qualitativa feita nesta pesquisa (ver seção 6.2).

É interessante observar que, de acordo com Cohen *et al.* (1988 *apud* COXHEAD; NATION, 2001), esse vocabulário não é tão bem conhecido quanto o vocabulário técnico, mesmo sendo fundamental aos gêneros acadêmicos. Em seu estudo, os pesquisadores comprovaram que estudantes tinham maior problema em entender o vocabulário acadêmico do que o técnico. As dificuldades não se limitavam a desconhecer determinadas palavras, mas também a não conseguir distinguir usos técnicos de usos não-técnicos de uma palavra (como por exemplo, “correlação”, da área de estatística), assim como não compreender mecanismos de retomada de coesão lexical.

Lindqvist (2014) também discute os problemas de não ensinar especificamente o vocabulário acadêmico. Segundo o autor, o vocabulário acadêmico básico está imerso em usos específicos do gênero, tais quais colocações e arranjos sintáticos, que diferem do uso cotidiano.

Os resultados da pesquisa de Cohen *et al.* (1988) e Lindqvist (2014) contribuem para reforçar a importância do ensino desse vocabulário. Para alunos de LE, aprender o vocabulário acadêmico implica dominar tanto as palavras mais frequentes da língua, quanto as acadêmicas. Esses dois vocabulários combinados cobrem cerca de 90% das palavras de um texto acadêmico (NATION, 2001, p. 197), sendo, por razões quantitativas e qualitativas apresentadas, indispensáveis para a compreensão do texto.

Para alunos de língua materna, dominar esse vocabulário é sinal de que eles já se envolveram em atividades acadêmicas anteriormente, pois a aprendizagem é resultado da experiência. Na maioria das vezes, o texto acadêmico é estranho para aluno que, apesar de ter elaborado diferentes tipos de textos durante sua vida escolar, não domina a particularidade formal e estrutural da escrita acadêmica (NUNES; PERNA, 2015, p. 174). Assim, cabe ao professor de língua materna ensinar ao seu aluno a pragmática que normalmente envolve a escrita acadêmica: demonstrar uma opinião bem fundamentada sem deixar de ser objetivo.

As particularidades semântico-pragmáticas no uso do vocabulário (seja acadêmico ou não), na perspectiva da Linguística Funcional Centrada no Uso, foram discutidas no próximo capítulo, com especial atenção aos verbos (doravante denominados **processos**), pois esse foi o foco desta pesquisa.

3 O SISTEMA DA TRANSITIVIDADE

A LFCU estuda a língua a partir da relação entre a sua estrutura e os variados contextos comunicativos em que é utilizada. Para Thompson (2014, p. 7) e Halliday e Matthiessen (2004, p. 31), a premissa da linguagem é sua função comunicativa, o que implica dizer que a função motiva, em grande parte, a forma linguística.

Dentro dessa perspectiva, a linguagem é entendida como um sistema semiótico em que o sentido se dá através das escolhas linguísticas do falante, de acordo com o contexto cultural e situacional em que ocorreram. Segundo Eggins (2004, p. 3), essa característica permite afirmar o que é ou não apropriado na língua, baseando-se nas escolhas linguísticas que os usuários costumam fazer em determinadas situações.

A ideia de um sistema semiótico nos permite entender tais escolhas como um contraste entre as realizações linguísticas dos usuários (relações sintagmáticas) e as possíveis escolhas que poderiam ter sido feitas, mas não foram (relações paradigmáticas). Esse potencial linguístico, ainda que diversificado e original, tem suas escolhas moldadas pelo contexto em que ocorreram.

Como Eggins (2004, p. 8), Halliday e Matthiessen (2004, p. 27) e Thompson (2014, p. 11) exemplificam, uma prova de como tal afirmação funciona é bastante simples: uma pessoa leiga é capaz de prever o contexto de um texto a partir de sua amostra. A seguir, os itens (7) e (8) evidenciam que bastam algumas linhas para anteciparmos seu contexto.

- (7) No presente trabalho, são descritos diferentes exemplos bem-sucedidos de micro-organismos especialmente desenhados para otimizar a produção de etanol, a produção de plásticos biodegradáveis a partir de recursos renováveis e a biorremediação de metais tóxicos. (Trecho retirado da área de Biotecnologia)
- (8) Era uma vez uma menina encantadora. Todos que batiam os olhos nela a adoravam. E, entre todos, quem mais a amava era sua avó, que estava sempre lhe dando presentes.¹²

Pode-se perceber que o trecho (7), mesmo antes de ter sua referência citada, é um texto técnico. Além dos dêiticos, como “presente trabalho”, grande parte das características listadas na seção 2.4 podem ser encontradas, como alta informatividade e vocabulário menos comum à

¹² GRIMM, Jacob; GRIMM, Wilhelm. Chapeuzinho vermelho. In: MACHADO, Ana Maria. *Contos de Fada* – de Perrault, Grimm, Andersen & outros. Rio de Janeiro: Zahar, 2010.

linguagem cotidiana, como palavras técnicas. Mesmo que um leigo não reconheça a qual área de conhecimento ou a qual gênero acadêmico pertence o texto, é possível caracterizá-lo como um “trabalho” técnico.

Já o trecho (8) pode ser classificado como uma história infantil principalmente pela expressão inicial “era uma vez”, que é frequentemente associada ao gênero. Comparando o trecho inicial do conto “Chapeuzinho vermelho” com o trecho inicial de um artigo científico, fica claro que as duas amostras pertencem a contextos diferentes. Em (8), as escolhas lexicais de palavras de altamente frequentes na língua e o tamanho reduzido da frase contribuem para uma sentença menos densa informativamente.

Ambos exemplos, (7) e (8), servem para demonstrar que linguagem e contexto estão intrinsecamente ligados, sendo formados por diferentes particularidades que permitem sua associação a situações distintas. Halliday e Matthiessen (2004, p. 27) denominam **variação de registro** tais diferenças entre textos, que, quando utilizadas para alcançar propósito comunicativos, configuram o **gênero** (THOMPSON, 2014, p. 42).

Para os autores (HALLIDAY; MATTHIESSEN, 2004, p.28), qualquer texto está inserido em dois contextos sobrepostos: o **contexto de cultura** e o **contexto de situação**. O primeiro determina que os elementos textuais escolhidos pelo falante partem do contexto cultural em que se inserem. Esse contexto associa-se à noção de propósito social da linguagem. Já no contexto de situação estão envolvidas as propriedades que servem ao funcionamento de uma interação, dando “substância às palavras e aos padrões gramaticais que falantes e escritores usam, consciente ou inconscientemente, para construir os diferentes gêneros” (FURTADO DA CUNHA; SOUZA, 2011, p. 25).

A combinação entre os dois contextos resulta em características particulares a um grupo de textos (gênero) que o diferencia dos demais através de três aspectos constitutivos do contexto de situação: campo, relação e modo.

O **campo** do contexto discursivo se refere à ocorrência da interação, seus participantes, processos envolvidos e circunstâncias. A **relação** representa os papéis sociais dos envolvidos, realizados através da relação interpessoal entre eles. E o **modo** do contexto discursivo diz respeito a como a língua é organizada a fim de atingir a finalidade a que se destina (HALLIDAY; MATTHIESSEN, 2004).

Exemplificando o contexto de situação, o gênero artigo científico pode ser entendido quanto ao *campo*: divulgação de pesquisas acadêmicas; quanto à *relação*: pesquisadores e público-alvo, sendo esse, em geral, especializado na área de conhecimento em

que se insere o artigo; e quanto ao *modo*: o texto escrito, veiculado em periódicos de uma determinada área.

Tais aspectos constitutivos do contexto de situação, de acordo com Thompson (2014, p. 28), organizam a linguagem e refletem respectivamente três significados determinadores da forma:

- significado experiencial, através do uso da língua para exprimir a experiência do usuário sobre o mundo, descrevendo eventos e estados através de processos e seus participantes; corresponde à metafunção ideacional.

- significado interpessoal, através do uso da língua para interagir com outros usuários, tendo em mente influenciar seu comportamento, expressar pontos de vista, entre outros; corresponde à metafunção interpessoal.

- significado textual, através do uso da língua para organizar ideias de modo a relacioná-las a outras e ao contexto maior situacional ou ideacional em que se encaixa; corresponde à metafunção textual.

Esses três significados correspondem propositalmente às três metafunções da linguagem (ideacional, interpessoal e textual), que foram discutidas na seção seguinte.

3.1 As metafunções da linguagem

Dentro dessa perspectiva, a gramática não consiste apenas em descrever a estrutura da língua, mas em explicar como tais estruturas linguísticas se relacionam sistematicamente com outras coisas (tanto linguísticas quanto sociais). Esses sistemas são representados por diferentes **metafunções**, cujas estruturas, sob o ponto de vista funcional, determinam a configuração do todo (linguístico e textual).

Os três tipos de significados vistos na introdução anterior servem às diferentes funções da linguagem, contribuindo simultaneamente para a construção do sentido da frase como um todo e sendo representados por diferentes aspectos da forma linguística (THOMPSON, 2014, p. 30).

Todas as funções específicas da linguagem podem ser designadas a três grandes funções, portanto, denominadas metafunções. A primeira é metafunção ideacional, em que a linguagem é utilizada para falar sobre o mundo; a segunda é a metafunção interpessoal, na qual a língua serve à interação com outras pessoas; e a terceira é a metafunção textual, que organiza a linguagem para expressar diferentes conteúdos.

Cada uma dessas metafunções possui um **sistema** próprio de escolhas linguísticas, cujos resultados compõem a **estrutura** da mensagem de cada componente, que determinam a configuração linguística e do texto (THOMPSON, 2014, p. 31). Assim, cada uma das análises feitas sob essas perspectivas corresponde a três tipos diferentes de papéis funcionais que servem aos elementos da sentença simultaneamente. O Quadro 1, a seguir, ilustra a relação das variáveis do contexto de situação associadas verticalmente aos principais sistemas de cada metafunção.

Quadro 1: Relação entre contexto de situação, metafunção e os principais sistemas de cada metafunção.

Contexto de situação	Campo	Relação	Modo
Metafunção	Ideacional	Interpessoal	Textual
Sistema	Transitividade	Modo e modalidade	Tema

A **metafunção ideacional** organiza e representa as ideias dos indivíduos através do sistema de transitividade, que trata da estrutura da oração cujos elementos são participantes, processos e circunstâncias. Os **processos** são ações ou estados; os **participantes** são aqueles que realizam ou são afetados pela ação; e as **circunstâncias** acrescentam informações à ideia expressa pelo processo. No exemplo (9), retirado do *corpus* desta pesquisa, essa metafunção pode ser observada através dos verbos em negrito e dos termos que os cercam, correspondendo a participantes (em itálico) e circunstâncias (sublinhadas).

(9) *Estes autores **argumentam** que crianças surdas sinalizantes não nativas não **participam** da comunicação sobre estados mentais com seus pais e irmãos, pela sua falta de uma linguagem comum, ao passo que *crianças surdas nativas **participam** com maior frequência.* (Trecho retirado da área de Ensino)*

No trecho, “estes autores” refere-se ao participante Dizente do processo verbal “argumentar”, enquanto “que crianças surdas sinalizantes não nativas não participam da comunicação sobre estados mentais com seus pais e irmãos” é a oração projetada desse o processo verbal (para maiores detalhes, ver seção 3.2.5). No aposto se encontra a circunstância do processo “pela sua falta de uma linguagem comum”.

Dentro da oração projetada, há um processo material (seção 3.2.1), “participar”, cujos participantes são um Ator (“crianças surdas sinalizantes não nativas”) e uma Meta (“da

comunicação sobre estados mentais com seus pais e irmãos”), e outro fora da projeção com um participante obrigatório, o Ator, em “crianças surdas nativas”, e uma circunstância, “com maior frequência”.

A **metafunção interpessoal** representa a relação entre os papéis sociais dos participantes através do sistema de modo e modalidade (FURTADO DA CUNHA; SOUZA, 2011, p. 27). No trecho (10), o autor procura dialogar com seu leitor e o faz através de marcas linguísticas, como o uso da primeira pessoa do plural e da forma interrogativa, numa tentativa de levar o leitor ao mesmo raciocínio do autor utilizando estratégias de aproximação, num diálogo explícito e direto entre esses dois participantes.

(10) Mas é possível **conhecermos** quais são essas forças que atribuem sentido ou, ao menos, algumas dessas forças? (Trecho retirado da área de Filosofia)

E a **metafunção textual** associa-se à organização do texto para adequá-lo ao seu contexto por meio de um sistema temático (FURTADO DA CUNHA; SOUZA, 2011, p. 27-28), que se realiza no nível oracional através das escolhas do falante quanto ao fluxo de informação textual. A parte da informação dada, ou já conhecida pelo interlocutor, na sentença é o **tema**, que possui menor grau de dinamismo, comunicando a menor quantidade de informação possível. Essa parte está destacada em negrito no trecho (11). Já a parte que contém informação nova e relevante é o **rema**, sublinhado no exemplo a seguir.

(11) **Uma das tarefas atuais dos pesquisadores** é a busca por novas fontes energéticas, [...] esse fato depende de cada país, de cada região mundial e de cada condição climática. (Trecho retirado da área de Ciências Agrárias I)

Apesar das metafunções terem sido exemplificadas separadamente, elas ocorrem de forma simultânea, pois as formas linguísticas são estudadas a partir dos significados que expressam (THOMPSON, 2014, p. 29). Desse modo, uma análise funcional deve levar em consideração o significado e contexto de uso da língua, pois, dentro da teoria, expressões linguísticas são insignificantes isoladamente.

A LFCU fornece diferentes possibilidades de análise de amostras reais da língua, como as diferentes metafunções e sistemas que permitem seu funcionamento. O foco da

presente pesquisa foi a metafunção ideacional, especificamente os processos realizados pelos verbos, tópico discutido na seção seguinte.

3.2 Metafunção ideacional: o sistema de transitividade

Segundo Halliday e Matthiessen (2004, p. 310), a metafunção ideacional, como explicado anteriormente, refere-se à representação da experiência humana e divide-se em dois componentes: o **experiential** e o **lógico**. O primeiro diz respeito ao conteúdo e às ideias, o segundo refere-se à relação entre tais ideias. O sistema responsável pela realização linguística dessas experiências é o da transitividade. Na LFCU, a transitividade é compreendida “como uma unidade estrutural que serve para expressar uma gama particular de significados ideacionais e cognitivos”, permitindo “identificar as ações e atividades humanas que estão sendo expressas no discurso” (FURTADO DA CUNHA; SOUZA, 2011, p. 68).

A transitividade nesta perspectiva difere do termo na gramática tradicional, sendo uma propriedade da oração e não apenas do verbo; e do termo na abordagem funcionalista norte-americana, em que a transitividade é um fenômeno gradual, manifestando-se por diferentes graus de interação sintático-semântica entre um verbo e os demais elementos da oração (GIVÓN, 1984 *apud* SILVA, 2012, p. 208). Apesar dessas diferenças, tanto a vertente funcionalista de Givón (1984) quanto a de Halliday e Matthiessen (2004) concordam que a transitividade é uma propriedade da oração.

Na abordagem de Halliday e Matthiessen (2004), no sistema de transitividade uma sentença, como exemplificada em (12), é constituída por um processo, participantes envolvidos e possíveis circunstâncias. Em (12), o processo está em negrito (“reconheceu”), seus participantes estão sublinhados (a pessoa envolvida no processo “o historiador Michael Gomez” e o que ele reconheceu “a importância da influência islâmica nas Américas”) e a circunstância está em itálico (“além desses estudiosos” não pertence ao processo, mas ajuda a situá-lo textualmente).

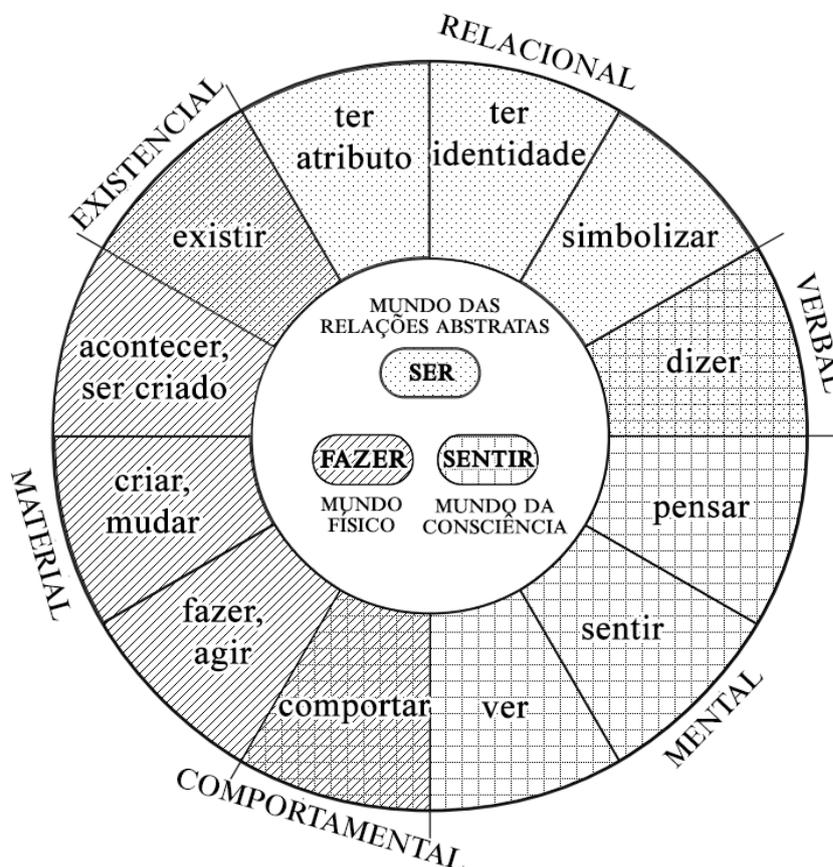
(12) *Além desses estudiosos, o historiador Michael Gomez reconheceu a importância da influência islâmica nas Américas.* (Trecho retirado da área de História)

No sistema de transitividade existem diferentes tipos de processos (materiais, mentais, relacionais, comportamentais, verbais e existenciais) que refletem as experiências do

falante da língua (HALLIDAY; MATTHIESSEN, 2014, p. 214). **Processos materiais** representam a experiência externa enquanto **processos mentais** refletem o mundo exterior através da perspectiva do usuário da língua, que reage, emotivamente ou racionalmente, às ações externas. Um terceiro componente desta teoria é a capacidade de relacionar um fragmento da experiência a outro, numa relação taxonômica de identificação e classificação, denominada **processos relacionais**. Esses três processos são considerados os principais no sistema da transitividade.

Os três outros tipos restantes (comportamentais, verbais e existenciais) são considerados secundários e estão localizados entre os principais, como pode ser observado na Figura 1. Entre os materiais e os mentais estão os **processos comportamentais**, que representam manifestações corporais de estados fisiológicos e emocionais dos seres vivos. Entre os processos mentais e relacionais estão os **processos verbais**, que dizem respeito às relações simbólicas experienciais expressas por meio da língua escrita ou falada. Por último, os **processos existenciais** estão localizados entre os mentais e os relacionais e servem para marcar a existência (ou acontecimento) de coisas.

Figura 1: Tipos de processos

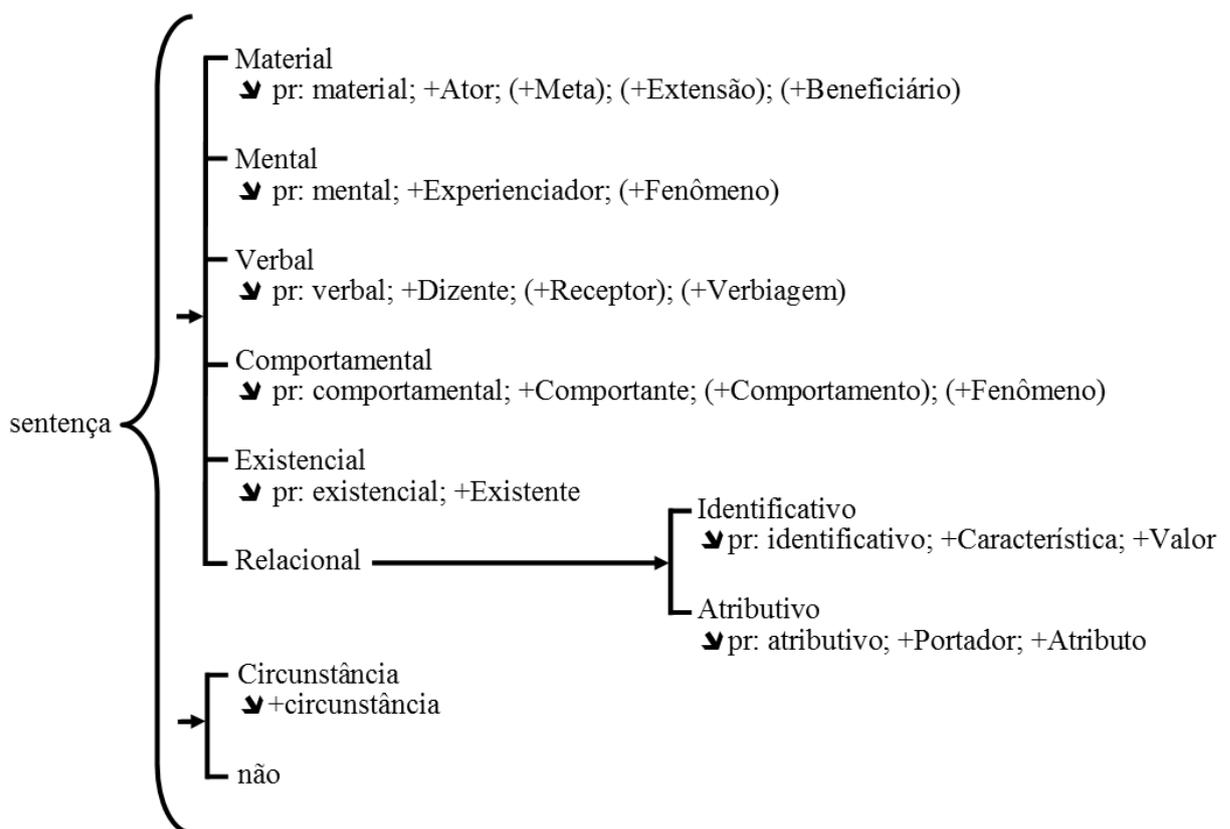


Fonte: Adaptado de Halliday e Matthiessen (2014, p. 216) e Souza (2006, p. 54)

De acordo com Halliday e Matthiessen (2014, p. 215), de onde a Figura 1 foi adaptada, é importante ressaltar que não há prioridade entre os diferentes tipos de processos e, por isso, eles podem ser representados (metaforicamente) em um círculo, em vez de uma linha reta, marcando o sistema de transitividade como um contínuo. Além disso, as principais áreas da figura equivalem aos processos prototípicos de cada tipo, sendo que os limites entre as áreas não são bem definidos. Essa não-clareza é denominada indeterminação sistemática (“*systemic indeterminacy*”) (HALLIDAY; MATTHIESSEN, 2014, p. 217).

A Figura 1 anterior pode ser esquematizada em uma rede sistemática a fim de demonstrar, além dos processos, também seus participantes, como exemplificado na Figura 2.

Figura 2: Sistema de transitividade



Fonte: Adaptado de Eggins (2004, p. 241)

Dentro dessa perspectiva, os processos estão no centro experiencial da sentença, sendo normalmente expressos pelo grupo verbal da oração (ver exemplos 12, anterior, e de 13 a 15, a seguir). Por isso, em análises da metafunção ideacional, os elementos interpessoais (com

sublinhado duplo em 14 e 15) não são levados em consideração (THOMPSON, 2014, p. 92-93).

- (13) As atividades de coleta e seleção de materiais recicláveis **expandem-se** a partir dessas possibilidades. (Trecho retirado da área de Serviço Sociais)
- (14) Para isso os peixes **devem ser anestesiados** para a maior segurança dos operadores (Trecho retirado da área de Zootecnia e Recursos Pesqueiros)
- (15) Nesses casos, pode-se pensar que houve uma grande produção de hidrogênio (Trecho retirado da área de Química)

Normalmente, como observável nos exemplos anteriores, cada processo envolve ao menos um participante (com sublinhado simples), representado por um sintagma nominal. Em dadas ocasiões, um dos participantes pode não estar explícito, mas recuperável pelo contexto, como em (14), em que o Ator (quem *anestesia* os peixes) do processo material não está explícito, mas pode ser inferido como sendo a pessoa que for trabalhar com tais peixes. O mesmo acontece em (15), em que o Experienciador do processo mental “pensar” está omitido (pela partícula “se”).

Nos exemplos de (12) a (15), as partes em itálico representam as circunstâncias dos processos, que costumam ser representadas por sintagmas adverbiais ou preposicionais. Como dito anteriormente, circunstâncias são frequentemente opcionais, porém, como detalhado no estudo de Mendes (2011), que analisou textos de blogs, as circunstâncias contribuem efetivamente para o caráter opinativo do gênero, favorecem à persuasão do leitor às ideias do autor do texto, colabora para atribuir valor de verdade ao que é dito, estabelecem condições para realização dos processos e têm suas funções aproximadas a de um comentário de opinião.

Nas próximas seções, cada um dos tipos de processos (e seus participantes) foram descritos mais detalhadamente e ilustrados com diversos trechos retirados do *corpus* desta pesquisa.

3.2.1 Processos materiais

Os **processos materiais** correspondem àqueles responsáveis por expressar a ação ou a mudança no estado de algo, como acontece em verbos como *correr*, *arranhar* e *cozinhar*. O participante que realiza a ação é o **Ator**, que é o participante obrigatório do processo, mesmo

não estando explícito da sentença. O segundo participante desse processo é a **Meta**, a que a ação do Ator se dirige. Esses são os dois principais e mais frequentes participantes dos processos materiais (THOMPSON, 2014, p. 95). Os trechos (16) a (18) a seguir ilustram esse processo (em negrito) e seus participantes (sublinhados). Em todos os casos, o participante sublinhado foi a Meta do processo.

- (16) Os filmes **foram confeccionados** a partir de dispersões em base aquosa, utilizando método convencional para polímeros termoplásticos (Trecho retirado da área de Ciências Biológicas III)
- (17) Utilizou-se um equipamento convencional para fotografia clínica, provido de filme 35mm, com objetiva macro 100mm, de onde **foi removido o cilindro de foco** (Trecho retirado da área de Odontologia)
- (18) A temperatura do meio reacional **foi mantida** constante através da circulação de etileno glicol pela camisa do reator. O progresso das reações **foi acompanhado** através de titulometria de neutralização ou cromatografia em fase gasosa. (Trecho retirado da área de Química)

Como dito anteriormente, o Ator é obrigatório, mesmo não estando explícito. Em cada um dos exemplos de (16) a (17), o Ator é facilmente retomado pelo contexto como sendo o(s) pesquisador(es) do trabalho.

Além desses dois participantes, Halliday e Matthiessen (2014, p. 236) discutem a existência de outros quatro: Escopo, Beneficiário, Cliente e Atributo. O **Escopo**, exemplificado em (19) com sublinhado duplo, não é afetado de forma alguma pelo processo.

- (19) A pesquisadora **passou o ano de 2010** ^(Escopo) na busca das duas drogarias, estabelecendo contatos telefônicos e construindo redes de relacionamento. (Trecho retirado da área de Farmácia)

Os participantes Beneficiário e Cliente, ilustrados em (20) e (21) respectivamente, são similares, pois ambos se referem aos beneficiados pelo processo material, porém, o **Beneficiário** é aquele a quem algo é dado enquanto o **Cliente** é aquele a quem serviços são prestados (HALLIDAY; MATTHIESSEN, 2014, p. 237).

- (20) Plasma **foi administrado em pacientes** ^(Beneficiário) com hemorragia contínua e TP ou TTP anormal. (Trecho retirado da área de Medicina III)

(21) Posteriormente, a família Porchat ^(Ator) **vendeu a Ilha para o senhor José Fracarolli Sobrinho** ^(Cliente), que construiu uma ponte de madeira facilitando o acesso ao casino. (Trecho retirado da área de Geografia)

Por último, o **Atributo** corresponde a uma característica dada a um dos participantes do processo. Halliday e Matthiessen (2004, p. 194-195) dizem que esse participante claramente pertence aos processos relacionais (ver seção 3.2.3), porém, pode ser associado a processos materiais em alguns casos. Dentre suas propriedades estão a impossibilidade de serem considerados verdadeiros participantes, por não poderem ocupar a posição de sujeito, e o fato de serem opcionais (diferindo-o do participante do processo relacional, que é obrigatório). Como não foram encontrados exemplos desse participante no *corpus* da pesquisa, o trecho (22) traz o exemplo de Halliday e Matthiessen (2004, p. 190):

(22) They stripped her clean ^(Atributo) of every bit of jewellery [[she ever had]].

Uma última observação relevante é que em construções causativas os papéis de Ator e Agente são distintos (ao contrário de sentenças regulares, em que ambos são interpretados pelo mesmo constituinte, embora o termo se restrinja a construções causativas). O trecho (23) exemplifica o participante **Agente**, em que um participante conduz o Ator a realizar o processo material.

(23) A paixão de meu pai ^(Agente) me ^(Ator) **fez renascer diferente**, deu origem a um Edelcio que não se pretendia, nem se pensava, escritor, pois era tão somente padre e professor. (Trecho retirado da área de Filosofia)

3.2.2 Processos mentais

Em oposição aos processos materiais, que refletem o mundo exterior ao sujeito, os **processos mentais** expressam o entendimento humano sobre o mundo através de reações mentais (como percepções sensoriais, desejos, sentimentos e observações) às experiências sociais (THOMPSON, 2014, p. 99). Segundo Halliday e Matthiessen (2004, p. 208), existem quatro tipos de processos mentais: cognitivos, desiderativos, perceptivos e emotivos. Os processos mentais **cognitivos** correspondem a processos do saber, da compreensão e do

entendimento. Os **desiderativos** expressam desejos. Os **perceptivos** expressam reações ao mundo a partir dos sentidos, como *ver* e *ouvir*. E os processos mentais **emotivos** relacionam-se aos sentimentos. Os trechos de (24) a (27) exemplificam cada um dos tipos de processos mentais:

- (24) O Barão do Rio Branco, durante toda sua gestão no Ministério das Relações Exteriores (MRE) (1902-12), cultivou retórica elevada e favorável às boas relações com todas as nações do hemisfério, mas **conhecia** ^(Processo mental cognitivo) os limites da solidariedade sul-americana. (Trecho retirado da área de Ciências Políticas)
- (25) No futuro, Clara **quer** ^(Processo mental desiderativo) se tornar independente através de seu trabalho. Sobre a posição "sonhadora", Clara **lembrou** ^(Processo mental cognitivo) como **queria** ^(Processo mental desiderativo) ser uma desenhista na infância, mas todos sempre a avisavam para sonhar, mas com os pés no chão. Hoje, ela **deseja** ^(Processo mental desiderativo) que seus sonhos se tornem realidade. No futuro, Clara **pretende** ^(Processo mental desiderativo) viver sozinha, mas com a ajuda da mãe e se **vê** ^(Processo mental desiderativo) como uma professora. (Trecho retirado da área de Psicologia)
- (26) A primeira vez que **ouvimos** ^(Processo mental perceptivo) o convite para usar a piscina foi numa roda de conversa na sala de espera do CTA; ela estava convidando a todos os pacientes presentes para irem à piscina no fim de semana. (Trecho retirado da área de Saúde Coletiva)
- (27) Muitas pessoas **sentiam** ^(Processo mental emotivo) náuseas ao percorrer o túnel, especialmente nas horas mais movimentadas. (Trecho retirado da área de História)

Os processos mentais são compostos obrigatoriamente por dois participantes (exceto em casos de projeção, discutidos adiante), o Experienciador e o Fenômeno. Mesmo se um dos participantes não estiver explícito na sentença, ele pode ser recuperado pelo contexto, como acontece com o Ator dos processos materiais. O **Experienciador** é o participante de natureza ativa do processo mental, sendo sempre consciente, mesmo quando inanimado (nesses casos, através de antropomorfização ou de metáforas metonímicas) (EGGINS, 2004, p. 227). Nos exemplos anteriores, todos os Experienciadores do processo estão explícitos na oração e são seres humanos, portanto, conscientes do processo. O trecho (28) seguinte exemplifica a metáfora metonímica, bastante comum no *corpus* da pesquisa, em que “este trabalho” estabelece uma relação metonímica com os pesquisadores que o produziram.

- (28) Este trabalho ^(Experienciador metonímico) **considerou apenas os gastos do Ministério da Saúde** ^(Fenômeno), sendo que o financiamento dos medicamentos do programa é feito com recursos da União e contrapartida dos estados. (Trecho retirado da área de Farmácia)

No exemplo (28) anterior, o outro participante obrigatório dos processos mentais está destacado com sublinhado duplo. O **Fenômeno** é o participante de natureza não-ativa, que expressa algo pensado, sentido ou percebido pelo Experienciador, podendo ser composto tanto por **fatos** (como em 28), quanto por **atos**.

Os processos mentais podem ser expressos tendo o Fenômeno ou o Experienciador como sujeito, o que não acontece com processos materiais (EGGINS, 2004, p. 230). Apesar de nem todos os processos mentais formarem pares compatíveis com essa reversibilidade, a maioria permite.

Como introduzido anteriormente, os processos mentais são capazes de projetar sentenças independentes ou reportar ideias (assim como os processos verbais) (EGGINS, 2004, p. 232). A **projeção** é uma das diferentes relações lógicas que unem sentenças, estando sempre contida naquela com o processo mental que realiza a projeção. Ambas sentenças não obedecem a uma ordem específica (a projeção pode vir antes ou depois do processo) e podem ser dependentes ou independentes da outra. Os itens (29) e (30) exemplificam casos de projeção.

(29) Analisando o sabor, **percebe-se** que não houve muita distinção entre as amostras que foram classificadas em geral como moderadamente doce e moderadamente ácida ^(Projeção). (Trecho retirado da área de Ciência de Alimentos)

(30) Entretanto, **acredita-se** que as causas não infecciosas sejam subestimadas ^(Projeção), pois há grande dificuldade no diagnóstico destas, já que não apresentam lesões histológicas e há, em sua maioria, necessidade de estudo epidemiológico detalhado do animal e da propriedade. (Trecho retirado da área de Zoologia)

3.2.3 Processos relacionais

Os **processos relacionais** são aqueles que estabelecem um relacionamento entre dois conceitos, servindo para sinalizar a existência de uma relação, identificá-la ou classificá-la (THOMPSON, 2014, p. 101). Halliday e Matthiessen (2004, p. 215) dividem os processos relacionais em três tipos (intensivo, possessivo e circunstancial), sendo que cada um ocorre em dois modos distintos (atributivo e identificativo). No tipo **intensivo**, ‘*a é b*’, em que *a* e *b* são grupos nominais, tendo como processo relacional prototípico o verbo *ser*. No tipo **possessivo**, ‘*c pertence a a*’ ou ‘*a tem c*’. E no tipo **circunstancial**, ‘*a (está ou é) em d*’.

Esses três tipos podem ocorrer de dois modos: processos relacionais **atributivos**, que são irreversíveis quanto a seus participantes (denominados Portador e Atributo), ou seja, a

ordem dos participantes não pode ser alterada sem alterar o sentido da sentença ou torná-la agramatical; ou processos relacionais **identificativos**, cujos participantes (denominados Valor e Característica) são reversíveis entre si.

Cada um desses tipos, e seus respectivos modos, foi exemplificado no Quadro 2, a seguir. Vale observar que os participantes dos processos relacionais possessivos podem ser também classificados como **Possuidor** e **Possuído** nos processos atributivos, e **Identificador** e **Identificado** nos processos relacionais identificativos, de acordo com sua função na sentença.

Quadro 2: As principais categorias dos processos relacionais exemplificadas

	Atributivo	Identificativo
<i>Intensivo</i>	Em condições sociais em que <u>os recursos para a aquisição da droga</u> (Portador) são escassos (Atributo), criam-se novos canais através dos quais o consumidor pode se transformar, também, em traficante e se submeter à sujeição criminal. (Trecho retirado da área de Antropologia)	<u>Outliers</u> (Valor) são aqueles valores de um atributo plotado 1,5 vezes a diferença interquartílica acima do terceiro quartil e 1,5 vez a diferença interquartílica abaixo do primeiro quartil (Característica). (Trecho retirado da área de Agronomia)
<i>Possessivo</i>	<u>O texto</u> (Portador: possuído) pertence ao gênero das "memórias" (Atributo: possuidor), portanto, dos escritos que, encomendados pelo governo despótico ilustrado, os homens de ciência luso-brasileiros produziam a fim de fomentar o progresso do império português. (Trecho retirado da área de História)	<u>A Calendula officinalis</u> (Valor: identificado) tem um efeito anti-inflamatório tópico e efeito imunomodulador anti-viral (Característica: identificador), estimulando atividade do sistema imune. (Trecho retirado da área de Biodiversidade)
<i>Circunstancial</i>	Segundo Veloso e Silva, o acúmulo de adiposidade na região abdominal é fator de risco para doenças endócrinas, metabólicas e cardiovasculares, mesmo quando <u>o IMC</u> (Portador) está dentro dos limites	Locais com uma grande densidade de objetos acima do solo podem gerar um MDE com valores de altitude que não representam a altitude do solo, e sim <u>os alvos</u> (Valor) que estão sobre ele

<u>da normalidade</u> ^(Atributo) . (Trecho retirado da área de Nutrição)	^(Característica) . (Trecho retirado da área de Geociências)
---	--

3.2.4 Processos comportamentais

De acordo com Halliday e Matthiessen (2004, p. 248), os **processos comportamentais** estão situados entre os processos mentais e os materiais (conforme observado na Figura 1, na seção 3.2) e, dentre os demais processos, são os mais difíceis de se distinguir por não terem um conjunto bem definido de características, além de serem parte *ação*, parte *mental*.

Desse modo, os processos comportamentais associam-se a processos psicológicos e fisiológicos humanos, tendo como participantes Comportante e Comportamento. O **Comportante** é obrigatório e parecido com Experienciador dos processos mentais, pois ambos requerem um ser dotado de consciência (podendo apenas ser substituído em casos de metáfora metonímica ou antropomorfizações). Já o **Comportamento** é análogo do Escopo, participante dos processos materiais, servindo como complemento do processo (HALLIDAY; MATTHIESSEN, 2014, p.251).

(31) Segundo Lucinéia Contiero, apoiada no depoimento da irmã de Plínio, depois que ele ^(Comportante) **dormiu** algumas noites na rodoviária e de ter sido ajudado por uma mulher (Trecho retirado da área de Artes e Música)

(32) Certa vez, um deles ^(Comportante) **a olhou diretamente** ^(Comportamento) e perguntou: Você é hippie?. A pesquisadora ^(Comportante) **riu** e fez expressão de que não tinha entendido, ele disse que era pelo seu jeito feminino diferente (jovial e descontraído), pelas suas roupas (calça jeans, tênis e blusa de malha), por não usar maquiagem. (Trecho retirado da área de Farmácia)

Os trechos de (31) e (32) exemplificam esse processo. Vale observar que os processos comportamentais não foram frequentes em nosso *corpus*, mesmo em áreas de Ciências da Saúde e Ciências Biológicas. 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.

3.2.5 Processos verbais

Os **processos verbais** são aqueles que expressam a ação de comunicação. Esse processo estaria localizado entre os mentais e os relacionais, representando simbolicamente as operações ocorridas na mente e expressas por meio de linguagem (THOMPSON, 2014, p. 105-106). Processos verbais podem criar narrativas, estabelecendo relações dialógicas, e reportar discursos. Envolvidos nesse processo estão os participantes Dizente, Receptor, Alvo e Verbiagem. O **Dizente** é o participante obrigatório nos processos verbais e é tipicamente humano. O **Receptor** é aquele a quem o dizente está se dirigindo, geralmente expresso por um sintagma preposicionado. O **Alvo** expressa o participante de quem se fala e a **Verbiagem** é a mensagem do Dizente. Os trechos (33), (34) e (35), a seguir, ilustram os processos verbais e seus participantes.

(33) Em 2008, 45 parlamentares ^(Dizente) **discursaram** sobre o tema do aborto ^(Verbiagem)
(Trecho retirado da área de Ciência Política e Relações Internacionais)

(34) Num desses casos, um senhor idoso ^(Dizente) **reclamava do fotógrafo** ^(Alvo) que, mesmo sem ter sido contratado anteriormente, registrou cenas do casamento de sua filha, apresentando a conta ainda durante a cerimônia. (Trecho retirado da área de Antropologia)

(35) Ana Maria Loforte ^(Dizente) (2003), pesquisadora moçambicana, **critica** também os neomarxistas ^(Alvo), principalmente o francês Claude Meillasoux, especialista em antropologia econômica, por tomar as mulheres como uma categoria homogênea, numa análise a-histórica e sem informações etnográficas. (Trecho retirado da área de Sociologia)

Como acontece com os processos comportamentais, distinguir os processos verbais de outros é complicado. No exemplo (36), a seguir, o processo “gritar” poderia ser considerado comportamental, porém, naquele contexto, o processo é verbal, pois seu complemento não descreve a forma como alguém gritou, mas sim *o que* a pessoa (Dizente) gritou.

(36) Naquele dia, no Terreiro do Paço em Lisboa, **gritaram**: "Valorosos Lusitanos, é chegada a hora de acudirmos pela reputação de Portugal, e de comprar com nosso sangue a liberdade da pátria. O Duque de Bragança é nosso legítimo Rei e Senhor natural. Deve-se-lhe a Coroa de direito". (Trecho retirado da área de História)

A citação seguinte ao processo verbal em (36) é bastante característica dos processos verbais, pois eles têm a capacidade de reportar a fala alheia. Esse discurso, seja sob

forma de citações diretas, indiretas ou livres, é chamado de **projeção** e, assim como acontece com os processos mentais, não é considerado um participante do processo. De acordo com Thompson (2014, p. 108-109), a relação entre a sentença projetada (todo conteúdo entre aspas, em 36) e a sentença projetante (“gritaram”) é análoga à relação de uma pintura e sua moldura, respectivamente. Os trechos seguintes trazem outros exemplos dos processos verbais, seus participantes e projeções.

- (37) Dois estudantes ^(Dizente) se apresentaram e, solicitados a darem seu parecer sobre a palestra, disseram ao professor ^(Receptor) *que não tinham qualquer objeção quanto à descrição dos fatos* ^(Projeção). (Trecho retirado da área de Filosofia)
- (38) Aos participantes ^(Receptor) se **perguntou** *como percebiam a crise psíquica e o atendimento dado pelo SAMU à pessoa em crise* ^(Projeção). (Trecho retirado da área de Enfermagem)
- (39) Também cinco foi o número de parlamentares cujo discurso não definia sua posição sobre aborto, sendo três católicos e dois sem religião. Esses últimos discursos ^(Dizente) **defendiam a então candidata Dilma Rousseff** ^(Alvo) dos ataques da imprensa com respeito à sua posição sobre aborto, **criticavam a acusação do deputado Luiz Bassuma** ^(Verbiagem) de que o PT expulsaria os deputados contrários à descriminalização do aborto, e um **acusou de falsa uma lista de parlamentares pretensamente favoráveis ao aborto** ^(Alvo). (Trecho retirado da área de Ciências Políticas)

Em (38), o participante obrigatório, Dizente, não está explícito, mas como ocorre em exemplos anteriores, ele pode ser inferido pelo contexto referindo-se aos pesquisadores envolvidos no trabalho científico. O trecho (39) tem diferentes processos que, apesar de serem verbais, aproximam-se dos processos materiais. Em cada um dos processos, o Dizente (os cinco parlamentares, representados pela metáfora metonímica em “esses últimos discursos”) age verbalmente sobre o Alvo. Segundo Halliday e Matthiessen (2004, p. 256), esses tipos de processos que aceitam Alvo não *projetam* discursos muito bem e, por isso, aproximam-se dos processos materiais. Assim, a motivação do processo (pontilhada no exemplo 39) é construída como circunstância ou oração hipotática (discurso indireto, reportado).

3.2.6 Processos existenciais

Por fim, os **processos existenciais** expressam a existência de um elemento, sem nenhum precedente ou relação com os demais, e se realizam através de verbos como *haver* e *existir*. Esse elemento, que simplesmente existe, chamado **Existente**, é o único participante do

processo – qualquer outra informação sobre o Existente é expressa por circunstâncias, como destacado em itálico nos exemplos (40), (41) e (42):

- (40) *De um lado há o poder do etnógrafo* ^(Existente), que pode levar a interação aos termos necessários à investigação científica (o próprio fato de ser um investigador, com vínculos acadêmicos, expressa uma condição de poder). *De outro há o poder daqueles que conhecem seu cotidiano e que podem falar sobre ele* ^(Existente), permitindo, mais ou menos, o acesso do etnógrafo à sua realidade. (Trecho retirado da área de Educação)
- (41) *Até 1985, existiam 54.000 pesquisadores* ^(Existente), vinculados a 36.000 projetos. (Trecho retirado da área Interdisciplinar)
- (42) *No último século, houve uma diminuição dos índices de pobreza na América Latina* ^(Existente), mesmo assim os modelos que estudam a população em vulnerabilidade continuam sendo instrumentos eficazes de análise da situação dos excluídos latino-americanos (Trecho retirado da área de Planejamento Urbano Regional e Demografia)

Segundo Halliday e Matthiessen (2004, p. 258), entre os processos materiais e os existenciais estão os processos relacionados à meteorologia, que podem se aproximar tanto dos processos materiais quanto dos relacionais, como exemplificado em (43), em que a circunstância “consideravelmente” se aproximaria da Característica do processo relacional atributivo.

- (43) A série temporal correspondente a esse fator (Figura 7c) apresenta máximos em 1996, 2000 e 2006, indicando que *nesses anos choveu consideravelmente no noroeste da Amazônia Oriental* (Trecho retirado da área de Geociências)

Tendo em mente os fundamentos teóricos expostos anteriormente, no próximo capítulo, a metodologia, termos e ferramentas utilizadas nesta pesquisa foram detalhadas.

4 METODOLOGIA

Neste trabalho foram realizadas uma análise quantitativa (ver capítulo 5) e qualitativa (ver capítulo 6) com base nos verbos mais frequentes em artigos científicos de diferentes áreas de conhecimento – quantitativa porque listou os principais verbos acadêmicos no *corpus* e qualitativa porque analisou os principais processos verbais, dentro da perspectiva da LFCU.

Essas duas análises requereram formas distintas de abordar os dados, porém complementares. Primeiro foi realizada a coleta dos artigos científicos, seguindo o modelo teórico de Berber Sardinha (2004), explicitado na seção 4.1. Em seguida, esses textos foram formatados para serem lidos pelo software concordanciador Kitconc© (MOREIRA FILHO, 2008). Com os dados prontos, as abordagens teóricas descritas no capítulo 2 (sobre os diferentes tipos de vocabulário) foram utilizadas na criação de listas de frequência, que permitiram a compilação da lista dos verbos acadêmicos mais frequentes em PB, seguindo métodos detalhados na seção 4.2. A partir desta lista, apresentada no capítulo 5, foi possível observar quais os processos verbais mais frequentes e prosseguir a análise qualitativa das ocorrências.

4.1 Composição do *corpus*

Antes de descrever os procedimentos de coleta, organização e análise dos dados, é importante apresentar o *corpus* e explicar sua composição. Como mencionado no capítulo 1, esta pesquisa é uma expansão do trabalho iniciado na Iniciação Científica e no Trabalho de Conclusão de Curso da autora. Em cada uma dessas atividades acadêmicas, *corpora* formados por artigos científicos foram utilizados.

Na Iniciação Científica, os substantivos acadêmicos mais frequentes foram analisados em um *corpus* de artigos científicos da área de Melhoramento Genético e Aquisição Linguística, com cerca de 300.000 palavras totais. Observou-se que, mesmo quando esses dois *subcorpora* eram unidos, a frequência de palavras técnicas das duas áreas era altíssima. Por isso, no Trabalho de Conclusão de Curso, acrescentou-se ao *corpus* quatro novas áreas de conhecimento (Antropologia, Geografia, Nutrição e Zoologia), mantendo o padrão de 150.000 palavras por área.

O aumento do *corpus* foi motivado pelos resultados observados por Sutarsyah *et al.* (1994), nos quais as palavras acadêmicas tendem a se destacar em um *corpus* composto por textos de diferentes áreas, enquanto o vocabulário técnico tende a se destacar em *corpus* mais especializado (como ocorreu no projeto de Iniciação Científica). Assim, no Trabalho de Conclusão de Curso, foi possível observar os substantivos acadêmicos no *corpus* – aqueles que não eram frequentes na língua, mas muito frequentes no *corpus*.

Contudo, dos trinta substantivos mais frequentes analisados qualitativamente e quantitativamente (ou seja, foram analisadas todas as mais de 35.000 mil ocorrências dos substantivos), apenas seis apresentaram uso acadêmico em números significativos (“análise”, “dado”, “estudo”, “resultado”, “tabela” e “trabalho”). Dentre os trinta itens analisados, mesmo com a variedade de áreas acadêmicas, foi alta a porcentagem do vocabulário técnico (12%, enquanto o acadêmico correspondeu a 23%) na lista, devido à alta especificidade de algumas áreas (como Melhoramento Genético e Aquisição da Linguagem) e a generalidade das demais.

Com base nesses resultados, a presente pesquisa, com o objetivo de listar um vocabulário acadêmico dos verbos mais frequentes, buscou compilar um *corpus* não apenas representativo do gênero, mas também balanceado quanto ao tema. Berber Sardinha (2000) observa seis critérios para compilar um bom *corpus*, e foram nesses itens que nos baseamos: origem, propósito, composição, formatação, representatividade e extensão.

A **origem** diz respeito à necessidade de trabalhar com dados autênticos da língua em uso e, caso os dados sejam de usuários não-nativos da língua, deve ser chamado *corpora* de aprendizes. O **propósito** é a característica que guia a compilação dos textos, pois não adiantaria criar um *corpus* sem ter uma finalidade em mente. Neste trabalho, o foco é o estudo linguístico das formas verbais acadêmicas em artigos científicos, portanto, a coleta se baseou nesse propósito. A **composição** é um aspecto relevante, pois tendo definido o propósito, é fundamental escolher conteúdos criteriosamente para atender à necessidade do pesquisador. Por exemplo, se pretende-se compilar um *corpus* de artigos científicos, deve-se tomar cuidado para não coletar outros gêneros que co-ocorrem no mesmo suporte (periódicos científicos), como cartas do editor, resenhas, resumos, relatos de experiência ou estudos de casos.

A **formatação** corresponde a transformar todos os textos coletados em dados legíveis por computador (em contexto de análises orais, por exemplo, essa característica corresponderia à transcrição do material coletado) e, normalmente, pelos softwares a serem utilizados para auxiliar a análise (como *concordanciadores*). A **representatividade** é outra característica importante, principalmente ao lidar quantitativamente com os dados, pois se

refere a qualidade de um *corpus* ser representativo da variedade amostrada, o que intercala com o último aspecto, a **extensão**, na qual um *corpus* para ser representativo precisa ser vasto.

Como dito anteriormente, esta pesquisa procurou fazer uma análise qualitativa e quantitativa do *corpus*. Portanto, para não repetir os resultados tendenciosos quanto ao vocabulário técnico, encontrados nas pesquisas realizadas na Iniciação Científica e no Trabalho de Conclusão de Curso, procurou-se representar todas as áreas de conhecimento igualmente para que o *corpus* fosse representativo do gênero e também balanceado.

A fim de padronizar tanto a organização dos dados quanto os temas a serem coletados (para que nos artigos da área de Ciências Agrárias I não se sobressaia o tema “Melhoramento Genético”, por exemplo), foi utilizado o sistema de catalogação das diferentes áreas do saber determinadas pela CAPES¹³.

Nesse sistema, a CAPES propõe a distribuição das áreas de conhecimento em três **Colégios** (Ciências da Vida, Ciências Exatas, Tecnológicas e Multidisciplinar e Humanidades) e nove **Grandes Áreas**, que abrangem todas as quarenta e oito **áreas de avaliação**. A tabela disponibilizada pela CAPES¹⁴ se baseia na divisão de Grandes Áreas e áreas de avaliação, que são compostas por inúmeras **subáreas** e **especialidades**.

Este trabalho, por exemplo, organizado segundo a tabela da CAPES obedeceria à catalogação do Quadro 3, que também ilustra a catalogação de um dos artigos do *corpus* da área de Biodiversidade:

Quadro 3: Exemplo do sistema de classificação da CAPES

	Presente Pesquisa	Artigo: “Variação temporal nos recifes de corais de Maragogi, APA Costa dos Corais (2010-2013)”
Colégio	Humanidades	Ciências da Vida
Grande Área	Linguística, Letras e Artes	Ciências Biológicas
Área de Avaliação	Letras e Linguística	Biodiversidade
Subárea	Linguística	Oceanografia

¹³ A página com informações sobre classificação das diferentes áreas de conhecimento pode ser encontrada em <<http://www.capes.gov.br/avaliacao/sobre-as-areas-de-avaliacao>>. Acesso em: 30 maio 2016.

¹⁴ Disponível em <http://www.capes.gov.br/images/stories/download/avaliacao/TabelaAreasConhecimento_072012.pdf>. Acesso em: 30 maio 2016.

Especialidade	Teoria e análise linguística	Inter. entre os organ. marinhos e os parâmetros ambientais

De tal forma, o *corpus* dessa pesquisa, baseado na divisão da CAPES, é formado por quarenta e oito *subcorpora*, que correspondem às diferentes áreas de avaliação. Para construir um *corpus* menos tendencioso para determinados conteúdos, as subáreas de cada área de avaliação foram levadas em consideração na hora de compor os *subcorpora*. A área de avaliação Biodiversidade, mencionada anteriormente, possui quatro subáreas (Ecologia, Oceanografia, Botânica e Zoologia) e cada uma foi representada textualmente no *subcorpus*.

Para coleta dos textos estipulou-se a média de cem mil palavras para cada área de avaliação (assim, no caso anterior, o *corpus* de Biodiversidade teria em média 25% de artigos de cada subárea), totalizando um *corpus* geral de aproximadamente 4,8 milhões de palavras.

A Tabela 3 a seguir especifica a quantidade de palavras e artigos científicos por área de avaliação e Grande Área. Apesar de não estarem apresentados, por questão de espaço, todos os 1.265 artigos da pesquisa foram organizados por periódico e data da publicação, que não excedeu o período de uma década desde o começo da coleta, ou seja, todos os artigos coletados são posteriores a 2005.

Tabela 3: Composição do corpus da pesquisa

	Nº de artigos	Nº de palavras
Ciências exatas e da terra		
Matemática, Probabilidade e Estatística	31	100.212
Ciência da computação	26	101.700
Astronomia e Física	23	101.138
Química	29	99.976
Geociências	25	100.782
Total	134	503.808
Ciências biológicas		
Ciências Biológicas I	39	99.647
Ciências Biológicas II	38	100.416
Ciências Biológicas III	41	102.020
Biodiversidade	27	99.738

	Nº de artigos	Nº de palavras
Total	145	401.821
Engenharias		
Engenharias I	24	100.376
Engenharias II	35	99.504
Engenharias III	19	102.233
Engenharias IV	23	100.248
Total	101	402.361
Ciências da saúde		
Medicina I	37	101.645
Medicina II	37	100.391
Medicina III	47	100.058
Nutrição	32	101.235
Odontologia	40	100.795
Farmácia	25	100.273
Enfermagem	31	101.560
Saúde coletiva	26	95.260
Educação física	32	99.510
Total	307	900.727
Ciências agrárias		
Ciências agrárias I	37	99.763
Zootecnia e Recursos pesqueiros	32	100.116
Medicina Veterinária	43	99.955
Ciência de alimentos	35	100.963
Total	147	400.797
Ciências sociais aplicadas		
Direito	14	99.866
Administração, Ciências Contábeis e Turismo	15	101.827
Economia	17	100.140
Arquitetura e Urbanismo	26	100.789
Planejamentos urbano regional e Demografia	18	101.530
Ciências sociais aplicadas I	20	100.645

	Nº de artigos	Nº de palavras
Serviço social	19	99.798
Total	129	704.595
Ciências humanas		
Filosofia e Teologia	16	99.907
Sociologia	13	100.378
Antropologia e Arqueologia	14	100.991
História	13	100.432
Geografia	21	99.015
Psicologia	18	99.193
Educação	17	100.127
Ciência política e Relações Internacionais	14	100.360
Total	126	800.403
Linguística, letras e artes		
Letras e Linguística	18	100.803
Artes e Música	23	99.541
Total	41	200.344
Multidisciplinar		
Interdisciplinar	22	100.382
Ensino	21	99.212
Materiais	34	100.757
Biotecnologia	35	100.044
Ciências ambientais	23	99.826
Total	135	500.221
Corpus geral	1.265	4.815.077

Na próxima seção, os procedimentos de coleta do *corpus* foram descritos, seguidos pela metodologia de análise preliminar dos dados.

4.2 Procedimentos de coleta

Todos os artigos coletados foram retirados de periódicos disponíveis on-line, sendo a maior parte proveniente do Portal de Periódicos SciELO, que disponibiliza edições completas de periódicos de muitas áreas, seguindo o padrão de organização por Grande Área da CAPES. Porém, algumas áreas de avaliação são mais difíceis de encontrar no portal, pois publicam majoritariamente em língua estrangeira (como Física e Astronomia) ou não possuem periódicos dedicados (como a área de avaliação Interdisciplinar, cujas especialidades relacionam pesquisas multidisciplinares entre as diferentes Grandes Áreas).

De tal forma, os artigos científicos foram escolhidos conforme o tema (se se encaixava na especialidade e subárea de determinada área de avaliação). Para isso, os critérios fundamentaram-se no resumo, nas palavras-chave e, principalmente, na descrição do periódico, que costuma trazer em que área de avaliação suas publicações se inserem.

Quando essa informação não estava disponível, recorreu-se à plataforma Webqualis¹⁵, que oferece listas com o Qualis (sistema de avaliação) de periódicos nacionais e estrangeiros, a fim de se confirmar se determinada revista publica artigos da área de avaliação em questão.

Para melhorar a qualidade dos textos do *corpus*, de cada artigo coletado foram retirados o título, o resumo, as citações em língua estrangeira, as citações longas, os cabeçalhos, tabelas, figuras, subtítulos, agradecimentos e referências, de modo a diminuir a repetição massiva do conteúdo (como nos resumos) e da linguagem menos acadêmica (agradecimentos).

A meta para cada *subcorpus* foi coletar cem mil palavras e cobrir todas as subáreas da área de avaliação. É interessante observar que o número médio de palavras por artigo científico varia de acordo com a área de avaliação. Publicações da área de Ciências Humanas tendem a ser longas, com mais de seis mil palavras (média de 6.352 palavras), enquanto publicações na área de Ciências Agrárias tendem a ser curtas, com pouco mais de duas mil palavras (2.726 palavras, em média). A Tabela 4 a seguir ilustra a média de palavras por artigo científico em cada Grande Área.

Tabela 4: Média de palavras em artigos científicos por Grande Área

Grande Área	Média de palavras por artigo científico
Ciências Exatas e da Terra	3.650
Ciências Biológicas	2.771

¹⁵ Disponível em <<http://qualis.capes.gov.br>>. Acesso em: 30 maio 2016.

Grande Área	Média de palavras por artigo científico
Engenharias	3.983
Ciências da Saúde	2.933
Ciências Agrárias	2.726
Ciências Sociais Aplicadas	5.461
Ciências Humanas	6.352
Linguística, Letras e Artes	4.886
Multidisciplinar	3.705

Vale observar que artigos muito compridos ou muito curtos, que destoavam da média dentro da área de avaliação, não foram incorporados à pesquisa. Além disso, os artigos foram coletados de periódicos variados de uma mesma área, para que as particularidades de publicação de uma revista não influenciassem todo um *subcorpus*. Por esses motivos, a Tabela 4 pode ser considerada representativa das Grandes Áreas, assim como seriam as médias calculadas por área de avaliação com os dados da Tabela 3.

O número total de palavras em cada *corpus* foi calculado com auxílio do programa Kitconc© (MOREIRA FILHO, 2008), que considera apenas palavras¹⁶ para contagem e não qualquer símbolo (como números) entre dois espaços, como faz a maior parte dos editores de texto com esse recurso. Essa ferramenta é bastante útil ao lidar com áreas em que equações fazem parte da sentença, como Matemática, Probabilidade e Estatística. Nessa área de avaliação, que tem 100.212 palavras, um contador de palavras não-sensível a números atribui ao *subcorpus* 106.154 palavras, quase 6 mil a mais (equivalendo a aproximadamente dois artigos daquela área) apenas em números.

Na próxima seção, o programa mencionado anteriormente e as ferramentas utilizadas foram descritas, assim como a metodologia de análise.

4.3 Procedimentos de análise

¹⁶ **Palavra**, nesse sentido, corresponde às unidades escritas linearmente, com espaços em branco em ambos os lados (SAUTCHUK, 2010, p. 5).

O programa Kitconc© (MOREIRA FILHO, 2008), e suas ferramentas, foi utilizado para lidar com os dados coletados de forma quantitativa e qualitativa. Esse aplicativo fornece listas organizadas de acordo com a frequência das palavras, linhas de concordâncias e colocados, permitindo a visualização do contexto de cada *token*. Também calcula a dispersão de palavras no *corpus*, o que é um recurso fundamental para descoberta de palavras pertencentes ao vocabulário acadêmico.

O ponto de partida para identificar os verbos mais frequentes foi a elaboração de **listas de palavras** que mostrassem a frequência de cada *type* no *corpus*. Como as listas de frequências geradas pelo programa apresentam contagem por forma, ou palavra ortográfica, foram somados os números de *tokens* de cada *type* que constituem o lexema verbal a fim de identificar os processos mais frequentes. Assim, por exemplo, a primeira ocorrência do lexema “justificar”, foi a forma “justifica”, que ocupa 2.799ª posição, com 180 *tokens*. Quando somamos as ocorrências de todas as formas do lexema verbal “justificar”, os *tokens* totais do lexema (555) permitiram a inclusão e identificação desse verbo entre os mais frequentes do *corpus* geral, estando na 145ª posição.

Para elaborar uma lista com os verbos mais frequentes, buscou-se todas as formas verbais entre os primeiros cinco mil *types* e, com base nessas formas, uma nova lista foi elaborada, dos lexemas verbais mais frequentes no *corpus*. Calcular os verbos por frequência do lexema é algo importante, pois, como exemplificado anteriormente, certas formas verbais que não estariam entre as mais frequentes da língua (como “justifica”), tornam-se frequentes quando acrescidas das outras formas do seu lexema. A lista com os principais lexemas verbais pode ser encontrada no Apêndice A.

Foram consideradas parte do lexema verbal todas as formas temporais e pessoais conjugadas do verbo e sua forma nominal no infinitivo, gerúndio e particípio, esse último apenas quando acompanhado de verbos auxiliares. Desse modo, itens como “sendo assim”, “haja vista” e “tendo em mente”, não foram considerados parte das ocorrências do lexema, tampouco formas que resultassem em construções como “dar origem” e “estar sujeito” ou expressões como “dar conta”, “levar em conta” e “quer dizer”.

Para facilitar o ensino e a visualização dos dados, cada lexema foi dividido em três partes: as formas conjugadas, nominais (infinitivo e gerúndio) e passivas. Dessa forma, é possível saber que “consistir” é predominantemente usado em sua forma ativa, conjugada, tendo ocorrido nominalmente em apenas 3,8% das ocorrências, nenhuma vez na forma passiva.

Dos cinco mil *types* mais frequentes, houve mais de mil formas verbais que resultaram em aproximadamente trezentos e cinquenta lexemas. Desses, o ponto de corte foi

cem ocorrências, pois abaixo desse número os verbos começaram a se restringir a determinadas Grandes Áreas. Assim, há 281 lexemas verbais que compõem a lista dos verbos mais frequentes no *corpus*, especificada no Apêndice A.

Neste capítulo foram detalhados os procedimentos de compilação do *corpus*, assim como os métodos de organização desses dados. No próximo capítulo, discutiu-se as etapas para a análise quantitativa do *corpus*: a criação de uma lista dos verbos acadêmicos mais frequentes.

5 ANÁLISE QUANTITATIVA: lista dos verbos acadêmicos mais frequentes

Seguindo a metodologia explicitada no capítulo anterior, a lista dos lexemas verbais mais frequentes foi elaborada, conforme observado na tabela do Apêndice A. Com base nesta lista, e seguindo os critérios de identificação do vocabulário acadêmico discutidos no capítulo 2, foram listados os verbos acadêmicos do *corpus*. A seção 5.1 descreveu as etapas da elaboração, e a seção 5.2 apresentou a lista mencionada, assim como observações relevantes sobre o vocabulário acadêmico.

5.1 Identificando o vocabulário acadêmico

Segundo Nation (2001), o vocabulário pode ser dividido em quatro tipos, detalhados no capítulo 2: palavras de alta frequência, vocabulário acadêmico, vocabulário técnico e palavras de baixa frequência. Em nosso *corpus*, trabalhamos majoritariamente com os três primeiros tipos, pois apenas lidamos com as formas mais frequentes, que resultaram nos lexemas verbais mais frequentes.

A tabela no Apêndice A corresponde aos verbos de alta frequência no *corpus*, mas não ao vocabulário acadêmico, pois há verbos que são bastante frequentes em outros contextos além do acadêmico, como “ser”, “fazer” e “mostrar”. Assim, para determinar quais são os verbos acadêmicos (mais frequentes neste *corpus* e menos frequente em outros *corpora*) foi preciso comparar essa lista com os verbos mais frequentes no PB, representativos de diferentes gêneros e do vocabulário de alta frequência.

Berber Sardinha (2004) anexa uma lista das 3.000 palavras mais frequentes ao seu livro, baseadas no Banco do Português, que na época tinha cerca de 223 milhões de palavras. Essa relação foi utilizada como correspondente da GSL em nosso *corpus*, conforme Coxhead (2000) utilizou para elaborar a AWL. Para realizar a comparação, uma outra lista foi criada a partir das formas verbais mais frequentes do *corpus* de Berber Sardinha (2004), porém, como não tivemos acesso ao *corpus* todo, a lista se baseou apenas nas formas lexicais presentes (diferente do nosso *corpus*, em que todas as formas verbais foram contabilizadas, mesmo se não estivessem entre as 5.000 primeiras).

A lista baseada em Berber Sardinha (2004) foi dividida em três partes, organizadas por milhar: primeiras mil palavras (em amarelo no Apêndice A), segundas mil palavras (em

verde) e terceiras mil palavras (em rosa). Por exemplo, há duas formas verbais do lexema “trazer”, uma na 634^o posição, outra na 2.442^o, apesar da diferença, o lexema se encaixaria na primeira categoria, pois sua primeira ocorrência está entre as primeiras mil palavras.

Essa divisão nos permite observar se há verbos no grupo menos frequente (em rosa, abrangendo os colocados entre as posições 2.000 e 3.000) que são frequentes em nosso *corpus*. Esse é um fenômeno esperado, pois os dois *corpora* são compostos por gêneros bastante distintos e é nessa diferença de vocabulário que esta pesquisa se baseia.

Como pode ser observado no Apêndice A, tal divisão serviu para destacar diferentes verbos menos frequentes na lista de Berber Sardinha (2004) que são frequentes neste *corpus*, como “avaliar”, “indicar”, “identificar” e “buscar”.

A divisão também nos permitiu identificar os lexemas mais frequentes em ambos *corpora*, tais quais “considerar”, “realizar” e “apresentar”. Esse último está entre as mil palavras mais frequentes, no entanto, ao organizarmos a lista por frequência de lexema (tal qual nossa lista está organizada), “apresentar” ocupa a 39^a posição dos 192 verbos encontrados enquanto, neste *corpus*, o lexema ocupa a terceira posição, superando o quarto colocado com quase o dobro de ocorrências – ou seja, esse é um lexema claramente relevante para o gênero acadêmico.

Além de comparar os verbos mais frequentes deste trabalho com a lista de Berber Sardinha (2004), cuja tabela organizada por frequência do lexema pode ser encontrada no Apêndice B, comparamos as duas listas com os verbos mais frequentes do *corpus* de Biderman (1998; 2000).

A autora lista apenas os vinte lexemas mais frequentes, em azul nos Apêndices A e B, baseados em um *corpus* de tamanho similar ao deste (cinco milhões de palavras), e pode-se observar, através de comparação, que esses verbos se aproximam mais dos resultados encontrados em Berber Sardinha que dos resultados desta pesquisa. Ilustrados em azul em ambos apêndices, podemos encontrar todos os vinte verbos de Biderman na lista de Sardinha (catorze deles são os primeiros colocados), porém, em nossa lista, eles estão dispersos e um não está presente, o lexema “achar”, que teve apenas 79 ocorrências e, devido ao ponto de corte, não entrou na lista do Apêndice A.

Essa comparação com Biderman é interessante, pois os *corpora* têm aproximadamente o mesmo número de palavras, porém, o da autora é formado por textos de diferentes gêneros (conforme explicado no capítulo 2), o que ajuda a comprovar a existência de um léxico especializado neste *corpus*, que não é observável num *corpus* diversificado.

Essas comparações permitiram perceber características do gênero acadêmico quanto às escolhas lexicais realizadas pelos produtores dos textos, como detrimento de lexemas

bastante frequentes na língua cotidiana, em favor de outros. O verbo mencionado anteriormente, “achar”, é um desses casos. Como pode ser visto na tabela do Apêndice A, lexemas verbais modais são bastante frequentes, porém, “achar”, que também é um modal, sequer teve ocorrências suficiente para entrar na lista. É possível que isso ocorra porque “achar” carrega consigo o lugar-comum de uma palavra frequente, assim como certa subjetividade, desapreciada pelos gêneros acadêmicos. O mesmo ocorre com “querer”, muito frequente nas listas das palavras mais frequentes, mas ocupante da 165ª posição neste *corpus*.

O lexema “dizer” é um caso particularmente interessante. Esse item é altamente frequente na lista baseada em Berber Sardinha (2004), ocupando a 3ª posição, e na de Biderman (2001, p. 342), na 6ª posição. Porém, ocupa a 72ª colocação em nosso *corpus*. Essa discrepância é relevante, pois, diferente de casos como “olhar” (frequente no vocabulário geral, infrequente neste *corpus*), aquele lexema é prototipicamente utilizado para reportar ou citar o discurso alheio e, por isso, esperar-se-ia que fosse frequente nesta pesquisa.

Essa é uma informação especialmente útil para o ensino, por exemplo. Se “dizer” não é tão frequente (com base em dados empíricos), cabe ao professor ou estudante da língua observar quais são os lexemas verbais que são preferidos no lugar daquele. O mesmo serve para outros lexemas, como “querer” ou “achar”. No capítulo 6 analisamos qualitativamente seis dos processos verbais (aqueles que expressam ações comunicativas, como “dizer”) mais frequentes, o que pode fornecer bases para o ensino desse conteúdo.

Por ser bastante produtiva essa comparação entre vocabulário de alta frequência e vocabulário acadêmico, principalmente para fins didáticos, tais listas foram incluídas como apêndices deste trabalho, devidamente legendadas.

Por fim, realizadas tais observações e detalhados os processos de identificação do vocabulário acadêmico, a próxima seção especificou quais foram os verbos acadêmicos encontrados no *corpus*.

5.2 Lista dos verbos acadêmicos mais frequentes

A Tabela 5, a seguir, lista os verbos acadêmicos mais frequentes do *corpus*. Diferente da tabela no Apêndice A, essa exclui todos os lexemas que ocorrem na lista dos verbos mais frequentes em PB (Apêndice B), reduzindo o número total de lexemas encontrados de 281 para 156.

Cabe destacar que a divisão das ocorrências de cada lexema, de acordo com a forma em que ocorreu, pode ser bastante útil para o ensino, pois permite observar quais são as formas de determinado verbo mais utilizadas em artigos científicos. Desse modo, estudantes precisarão apenas focar seus estudos nas formas mais utilizadas.

Tabela 5: Lista dos lexemas verbais acadêmicos mais frequentes

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
1	UTILIZAR	1.855	2.158	2.750	6.763
2	OBSERVAR	2.803	982	1.846	5.631
3	VERIFICAR	1.419	787	416	2.622
4	POSSUIR	2.279	235	3	2.517
5	ANALISAR	461	988	719	2.168
6	DEMONSTRAR	1.239	386	222	1.847
7	DETERMINAR	521	618	613	1.752
8	CONSTITUIR	1.090	274	312	1.676
9	DESTACAR	1.041	548	64	1.653
10	RELACIONAR	370	205	1.038	1.613
11	CONTRIBUIR	845	657	110	1.612
12	DESENVOLVER	589	474	491	1.554
13	ESTABELEECER	600	637	264	1.501
14	ENVOLVER	872	492	135	1.499
15	VARIAR	919	440	39	1.398
16	SUGERIR	969	329	56	1.354
17	RESULTAR	790	401	139	1.330
18	DESCREVER	520	279	490	1.289
19	CONTER	396	800	33	1.229
20	COMPREENDER	392	665	163	1.220
21	PERCEBER	645	344	192	1.181
22	CARACTERIZAR	581	276	307	1.164
23	FORMAR	302	568	281	1.151

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
24	COMPARAR	187	670	274	1.131
25	ASSOCIAR	247	95	788	1.130
26	INFLUENCIAR	531	282	302	1.115
27	APLICAR	288	224	580	1.092
28	POSSIBILITAR	707	365	17	1.089
29	EVIDENCIAR	647	249	162	1.058
30	COMPOR	404	153	474	1.031
31	REFERIR	812	141	48	1.001
32	CORRESPONDER	852	127	2	981
33	RESSALTAR	410	524	46	980
34	CALCULAR	161	215	562	938
35	VISAR	383	512	-	895
36	CONSISTIR	834	33	-	867
37	AFETAR	459	220	184	863
38	FORNECER	434	324	102	860
39	RELATAR	604	58	180	842
40	NOTAR	565	226	46	837
41	SUBMETER	67	60	687	814
42	ESTIMAR	211	303	286	800
43	CONCLUIR	485	284	31	800
44	PERMANECER	639	147	13	799
45	ADOTAR	321	201	267	789
46	RECONHECER	307	281	195	783
47	INICIAR	500	185	91	776
48	EXPRESSAR	463	130	181	774
49	CONSTATAR	483	145	124	752
50	ATRIBUIR	265	120	357	742
51	REFLETIR	404	284	22	710
52	PROPORCIONAR	409	286	11	706

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
53	CAUSAR	327	276	87	690
54	ALCANÇAR	218	276	175	669
55	EMPREGAR	167	197	302	666
56	CLASSIFICAR	131	111	412	654
57	FAVORECER	331	261	47	639
58	BASEAR	261	69	308	638
59	CONDUZIR	217	116	300	633
60	TENDER	597	22	-	619
61	ALTERAR	247	234	133	614
62	IMPLICAR	509	64	21	594
63	SELECIONAR	82	125	370	577
64	ABORDAR	261	172	136	569
65	INVESTIGAR	170	263	124	557
66	JUSTIFICAR	290	177	88	555
67	COLETAR	33	165	351	549
68	DIVIDIR	172	86	286	544
69	INSERIR	167	94	256	517
70	NECESSITAR	445	63	1	509
71	FACILITAR	171	296	38	505
72	OCUPAR	307	141	57	505
73	CORROBORAR	262	151	47	460
74	MEDIR	118	159	182	459
75	INCORPORAR	151	156	143	450
76	REQUERER	391	31	27	449
77	EXERCER	270	145	31	446
78	REFORÇAR	248	143	48	439
79	ELABORAR	152	120	156	428
80	ADQUIRIR	178	118	128	424
81	CITAR	192	128	101	421

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
82	DIFERIR	352	63	2	417
83	APROXIMAR	261	124	29	414
84	INTEGRAR	201	142	62	405
85	INTERFERIR	233	166	5	404
86	LOCALIZAR	129	67	203	399
87	EXPLORAR	107	179	107	393
88	ENFATIZAR	217	137	39	393
89	SALIENTAR	196	179	9	384
90	ILUSTRAR	232	76	75	383
91	OPERAR	222	131	30	383
92	ORGANIZAR	130	113	139	382
93	TESTAR	51	162	167	380
94	ORIENTAR	143	136	99	378
95	MINIMIZAR	71	260	40	371
96	DISPOR	228	51	89	368
97	DETECTAR	82	119	166	367
98	AGIR	131	229	3	363
99	ELEVAR	121	93	143	357
100	DIFICULTAR	183	146	25	354
101	EXCLUIR	55	69	218	342
102	SITUAR	197	71	70	338
103	ADICIONAR	60	65	202	327
104	MODIFICAR	118	124	85	327
105	SUPERAR	104	170	51	325
106	OPTAR	272	47	5	324
107	ESTIMULAR	135	142	47	324
108	CONCENTRAR	210	70	41	321
109	COMPROMETER	118	166	37	321
110	PARTIR	143	171	2	316

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
111	LIDAR	73	239	4	316
112	DENOMINAR	116	28	166	310
113	MANIFESTAR	221	74	15	310
114	INTERPRETAR	65	127	115	307
115	QUESTIONAR	96	118	84	298
116	ELIMINAR	65	177	55	297
117	ASSEGURAR	91	170	32	293
118	OBJETIVAR	172	111	3	286
119	CONFERIR	170	99	17	286
120	ABRANGER	191	88	4	283
121	DISTRIBUIR	71	47	159	277
122	RECOMENDAR	180	18	79	277
123	DIFERENCIAR	129	113	32	274
124	SEPARAR	87	83	89	259
125	EMERGIR	201	54	3	258
126	REMETER	220	29	8	257
127	AJUSTAR	77	75	105	257
128	MENCIONAR	125	70	59	254
129	DESEJAR	214	30	10	254
130	ADEQUAR	27	56	169	252
131	ESCLARECER	82	124	27	233
132	MERECER	208	17	2	227
133	TOTALIZAR	39	184	1	224
134	ARTICULAR	97	94	29	220
135	RESIDIR	196	23	-	219
136	AVANÇAR	88	114	15	217
137	PRESERVAR	50	117	32	199
138	VALER	125	69	3	197
139	POR	97	44	53	194

Frequência de ocorrências					
	Lexema verbal	Forma conjugada	Forma nominal	Forma passiva	Total
140	VINCULAR	58	25	105	188
141	DENOTAR	141	38	7	186
142	ADAPTAR	49	87	50	186
143	PRESSUPOR	144	31	6	181
144	FORMULAR	60	68	52	180
145	DECORRER	158	20	1	179
146	INDUZIR	52	90	37	179
147	DESTINAR	76	13	84	173
148	PUBLICAR	76	23	69	168
149	DERIVAR	112	23	33	168
150	COMPLETAR	79	59	29	167
151	PRATICAR	93	43	31	167
152	PREDOMINAR	133	27	-	160
153	DESCARTAR	21	40	99	160
154	RETOMAR	64	66	25	155
155	AGREGAR	71	64	18	153
156	VALIDAR	25	76	51	152

Este capítulo apresentou a lista dos verbos acadêmicos mais frequentes no *corpus*, assim como parte dos procedimentos de identificação desse vocabulário, distinguindo-o das palavras de uso geral. Foram observadas particularidades, baseadas nos dados quantitativos, que demonstraram relevância para o ensino do gênero. O capítulo seguinte descreveu a análise qualitativa realizada, que focou nos processos verbais mais frequentes do *corpus*, baseados na lista da Tabela 5.

6 ANÁLISE QUALITATIVA: processos verbais em destaque

Este capítulo relatou a análise qualitativa feita dos principais processos verbais acadêmicos. O procedimento metodológico precedente a esta etapa da pesquisa foi descrito na seção 4.3, logo após a explicação da parte quantitativa.

A próxima seção apresentou os processos verbais acadêmicos panoramicamente, analisando quais formas linguísticas realizaram tais processos e com que frequência foram utilizados. Já na seção 6.2, foram observadas as relações desses processos com os participantes da oração, Dizente, Verbiagem e Receptor, assim como os casos de orações projetadas.

6.1 Panorama geral dos processos verbais mais frequentes

O uso de processos verbais em gêneros acadêmicos está associado a argumentação, trazendo vozes alheias para o discurso, a fim de corroborar e validar o que está sendo apresentado no texto (PENHA, 2012). Isso difere de outros gêneros, como o narrativo, por exemplo, em que os processos verbais servem para introduzir o discurso alheio, dos personagens, com objetivo de narrar diálogos em uma história.

Para uma análise qualitativa dos processos verbais acadêmicos mais frequentes, ou seja, daqueles que não pertencem aos mais frequentes na língua, de acordo com a metodologia utilizada neste trabalho, foi realizado um levantamento das formas linguísticas de cada processo (ilustrada na Tabela 6). Isso permite verificarmos as particularidades de cada processo antes de partirmos para observação de suas linhas de concordância.

Tabela 6: Formas linguísticas dos processos verbais mais frequentes analisados

DEMONSTRAR		DETERMINAR		DESTACAR	
DEMONSTRA	331	DETERMINA	212	DESTACA	516
DEMONSTRARAM	297	DETERMINAM	108	DESTACAM	312
DEMONSTRAM	275	DETERMINOU	96	DESTACOU	84
DEMONSTROU	271	DETERMINARAM	28	DESTACARAM	70
DEMONSTRAVA	13	DETERMINAVA	17	DESTACAMOS	34
DEMONSTRAVAM	12	DETERMINE	16	DESTACAVAM	6

DEMONSTRAR		DETERMINAR		DESTACAR	
DEMONSTRE	10	DETERMINARÁ	9	DESTACAVA	5
DEMONSTRAMOS	8	DETERMINAREM	7	DESTACARÁ	3
DEMONSTREM	13	DETERMINAMOS	6	DESTACO	3
DEMONSTRARIA	3	DETERMINARIAM	5	DESTACAREM	2
DEMONSTRARIAM	2	DETERMINEM	5	DESTACARMOS	2
DEMONSTRARÁ	2	DETERMINARA	3		
DEMONSTRAREM	2	DETERMINARÃO	3		
		DETERMINARIA	3		
		DETERMINAVAM	2		
Total conjugado	1.239	Total conjugado	520	Total conjugado	1.037
Forma nominal	386	Forma nominal	618	Forma nominal	548
Forma passiva	222	Forma passiva	613	Forma passiva	64
Total de ocorrências	1.847	Total de ocorrências	1.752	Total de ocorrências	1.649
ESTABELECEER		SUGERIR		DESCREVER	
ESTABELECE	279	SUGERE	483	DESCREVE	261
ESTABELECEM	119	SUGEREM	349	DESCREVEM	109
ESTABELECEU	106	SUGERIRAM	44	DESCREVERAM	48
ESTABELECERAM	41	SUGERIU	35	DESCREVEU	35
ESTABELEÇA	17	SUGERIMOS	23	DESCREVEMOS	22
ESTABELECEREM	16	SUGERIA	12	DESCREVEREMOS	9
ESTABELECIA	13	SUGERIREM	6	DESCREVIA	9
ESTABELECIAM	9	SUGIRO	6	DESCREVEREM	8
ESTABELEÇAM	8	SUGIRAM	3	DESCREVA	6
ESTABELECEMOS	6	SUGERI	2	DESCREVAM	4
ESTABELECERIA	3	SUGERIAM	2	DESCREVIAM	4
ESTABELECERMOS	3	SUGIRA	2	DESCREVERIA	3
ESTABELECERÁ	2			DESCREVERÁ	2
Total conjugado	622	Total conjugado	967	Total conjugado	520
Forma nominal	637	Forma nominal	329	Forma nominal	279
Forma passiva	264	Forma passiva	56	Forma passiva	490
Total de ocorrências	1.523	Total de ocorrências	1.352	Total de ocorrências	1.289

Em todos os processos há predominância do uso de formas na terceira pessoa do presente do indicativo, sempre com o dobro de ocorrências das formas correspondentes no pretérito. Isso permite que uma ação no passado seja trazida para o presente do leitor do texto, causando um “efeito de permanência do conteúdo” (FUZER, 2012). Os trechos (44) a (47), a seguir, referem-se a estudos realizados no passado, cujos resultados introduzidos por processos no presente representam certa atemporalidade de tais estudos. Já nos trechos de (48) e (49), os processos no tempo presente representam a atualidade das normas descritas, que continuam estando em vigor (no momento da produção do artigo).

- (44) Tuan (2005), entretanto, **descreve** a seca como algo mais abrangente que a simples ausência de chuva: "é um fenômeno invisível, exceto pela devastação que provoca: colheitas reduzidas, animais mortos e pessoas mal nutridas". (Trecho retirado da área de Ciências Ambientais)
- (45) Acentua-se que pesquisas **destacam** a candidose bucal como a lesão mais prevalente em portadores do HIV/Aids nas suas diferentes apresentações clínicas. (Trecho retirado da área de Odontologia)
- (46) Os resultados de outros estudos **demonstram** algumas diferenças: geralmente os profissionais de enfermagem mais novos são menos positivos em suas avaliações da cultura de segurança. (Trecho retirado da área de Enfermagem)
- (47) Este resultado, na safra 2004/2005, **sugere** que, para algumas parcelas da área estudada a redução do índice de vegetação não foi seguida na proporção prevista de aumento da LST, possivelmente, por causa da predominância do efeito de sobreaquecimento generalizado da superfície vegetada por uma onda de calor e não diretamente relacionada com a forçante imposta pelas condições prolongadas de deficiência hídrica. (Trecho retirado da área de Agronomia)
- (48) Segundo Gomes (2006), a norma brasileira de lixiviação, a NBR 10005 (ABNT, 2004a), assim como o método americano 1311 TCLP (UNITED..., 1992), tem o propósito de classificar um resíduo de forma simples, reprodutível e com baixo custo. Ambos os procedimentos **estabelecem** que 100 g de amostra com granulometria menor que 9,5 mm permaneçam durante 18 h em contato com uma solução ácida, sob agitação, em uma proporção de amostra/solução lixiviante de 1:20. (Trecho retirado da área de Engenharia I)
- (49) A anisakiase não está listada na portaria nº 5, porém, ela pode ser enquadrada no Art. 2º, desta portaria, que **determina** que: «A ocorrência de agravo inusitado, caracterizado como a ocorrência de casos ou óbitos de doenças de origem desconhecida ou alteração no padrão epidemiológico de doença desconhecida, independente de constar na lista nacional de doenças e agravos de notificação compulsória, deverá também ser notificada às autoridades sanitárias». (Trecho retirado da área de Ciências Biológicas III)

Nos exemplos de (44) a (49), os processos verbais foram utilizados a fim de trazer discursos de outros pesquisadores, o que fortalece a argumentação do texto acadêmico, dando-lhe maior credibilidade. Ao mesmo tempo, o uso da terceira pessoa do presente pode

caracterizar o oposto, correspondendo à omissão ou à indeterminação do sujeito, que é comum ao gênero. A impessoalização tira o foco do autor do texto (ou do Dizente), distanciando-o de seu objeto de estudo, e favorece a mensagem, como exemplificado nos trechos (50), (51) e (52).

- (50) Desta forma, **sugere-se** que a resolução espaço-temporal dos dados de TSM influencia de forma pouco significativa na representação da ZCAS pelo modelo BRAMS, tanto para eventos ocorridos durante o El Niño, quando a ZCAS é mais oceânica, quanto para os ocorridos durante a La Niña, quando a ZCAS é mais continental. (Trecho retirado da área de Geociências)
- (51) Entre os diferentes fatores que influem sobre o desempenho da levedura no processo de produção de etanol, **destaca-se** a contaminação do meio de fermentação por bactérias, responsável por perdas importantes na produtividade das destilarias. (Trecho retirado da área de Biotecnologia)
- (52) Na justificativa do projeto, **demonstra-se** que a solução para a 'precarização' é o concurso público com contrato pela Consolidação das Leis do Trabalho (CLT). (Trecho retirado da área de Saúde Coletiva)

As formas passivas são bastante frequentes, tanto aquelas que explicitam o Agente, como em (53), quanto as que o omitem, conforme ilustrado nos exemplos (54) a (56), em que todos o(s) autor(es) do texto (e Agentes) foram omitidos, inclusive nos demais processos (destacados em *itálico*).

- (53) Um modelo computacional de analogia visual para a resolução de problemas foi **descrito** por Davies, Nersessian e Goel (2005), exemplificando-se sua utilização para representar um pouco do raciocínio de Maxwell sobre o eletromagnetismo. (Trecho retirado da área de Ensino)
- (54) *Aplicou-se* o teste de Kolgomorov-Smirnov para avaliar a simetria das distribuições das variáveis contínuas e foram *calculadas* as medidas de tendência central e de dispersão para cada período de observação. O consumo de nutrientes foi **descrito** para os dados ajustados pelo consumo de energia total segundo o método de resíduos. *Testou-se* a associação entre variáveis categóricas e o peso ao nascer classificado em adequado e inadequado pelo teste do Qui-quadrado. (Trecho retirado da área de Nutrição)
- (55) Logo, ficou **demonstrado** um aumento do espaço aéreo bucofaringeano e nasofaringeano e uma redução da resistência nasal após a realização da ERM, acontecimento já relatado anteriormente em trabalhos existentes na literatura. (Trecho retirado da área de Odontologia)
- (56) O teste de Qui-quadrado foi *utilizado* para verificar heterogeneidade entre as proporções. O nível de significância adotado *foi* de $\alpha = 5\%$. Foram **sugeridos** pontos de corte para as variáveis antropométricas por meio das curvas ROC (Receiver Operating Characteristic). (Trecho retirado da área de Educação Física)

A passiva permite a explicitação do Dizente ou seu apagamento e, portanto, relaciona-se à continuidade do tópico discutido no texto e à objetividade acadêmica (FUZER, 2012) – nos casos de apagamento do Dizente, seja para omitir a participação efetiva do autor, ou para parafrasear outros autores. Esse último caso, como observável em (57) e (58), torna mais difícil delimitar qual parte é paráfrase e qual parte é atribuída ao autor do texto. Em (57), a autoria está entre parênteses no final da sentença, “CARR; SUTER; BARBIERI, 2005”, e, no exemplo (58), a referência externa está em “[27]”, que associa o trecho à referência bibliográfica do artigo.

(57) Em 2002, a Population Environment Research Network (PERN), rede voltada para discussões sobre população e ambiente, partiu do artigo produzido por Lambin e Geist para discutir e disseminar os avanços que o trabalho proporcionou do ponto de vista do papel da população. A complexidade do fenômeno, seu caráter espacial e temporal e, sobretudo, a rejeição de um fator causal único foram **destacados** (CARR; SUTER; BARBIERI, 2005). (Trecho retirado da área de Planejamento Urbano Regional e Demografia)

(58) Tem sido **sugerido** que alterações na funcionalidade mitocondrial podem estar relacionadas ao fenômeno de estresse oxidativo, visto que as mitocôndrias constituem o principal ambiente celular de geração de radicais livres ^[27]. (Trecho retirado da área de Materiais)

De tal forma, a passiva, em conjunto com o uso da terceira pessoa e da passiva sintética, compõe uma das diferentes formas de tornar o discurso mais objetivo e impessoal, sendo, por isso, bastante frequente na análise, como apresentado na Tabela 6.

Na próxima seção explicitou-se a análise de linhas de concordância dos seis processos verbais acadêmicos, focando em seus participantes e nos modos de realização da mensagem.

6.2 A transitividade em processos verbais acadêmicos

Cada processo contribui de uma forma distinta para a construção da experiência, baseado na expressão linguística de cada gênero. Assim, apesar da mescla de processos predominantes em um texto, é possível perceber tendências e relevâncias distintas dentro de um gênero específico. Por exemplo, processos mentais são típicos de conversações casuais,

processos verbais são típicos de notícias e processos existenciais e relacionais são característicos de textos narrativos (SOUZA, 2006, p. 107).

No caso dos gêneros acadêmicos, como artigos científicos, Halliday e Matthiessen (2004, p. 253), discutem sobre como processos verbais são fundamentais para esse gênero, possibilitando citar e reportar discursos de diferentes estudiosos ao mesmo tempo em que indica a posição do autor do texto, ao fazer uso de verbos como “afirmar”, “negar”, “sugerir” e “declarar”.

Os processos verbais analisados neste capítulo foram frequentes em todas as Grandes Áreas, conforme ilustrado na Tabela 7, com algumas particularidades expostas a seguir.

Tabela 7: Número total de ocorrências de cada processo por Grande Área

	Demonstrar	Determinar	Destacar	Estabelecer	Sugerir	Descrever
C. Saúde	536	291	239	254	362	230
C. Humanas	225	134	277	320	198	173
C. Sociais	183	159	334	286	138	111
Aplicadas						
C. Exatas	139	313	177	109	96	248
Multidisciplinar	189	146	175	158	137	104
C. Biológicas	194	177	150	65	133	72
C. Agrárias	245	241	102	79	132	190
Engenharias	73	268	132	163	86	104
Linguística, Letras e Artes	63	22	63	89	70	57
Total	1.847	1.751	1.649	1.523	1.352	1.289

É interessante observar os números totais de ocorrência por Grande Área de cada processo, que nos permite verificar quais são os mais frequentes naquela área (“demonstrar” em Ciências da Saúde se destaca dos demais) e quais apenas foram inseridos na lista por serem frequentes como um todo, mas não particularmente relevante (“determinar” em Letras e Artes).

Os destaques em negrito feitos nas ocorrências mais frequentes por área, assim como o agrupamento de *subcorpora* com o mesmo número de *tokens* (delimitado por linhas

pontilhadas), possibilita a visualização das médias dentro da Grande Área e quais os processos que se destacam por serem mais, ou menos, frequentes.

Na área de Ciências da Saúde, destacam-se os processos “demonstrar” e “sugerir” entre os mais frequentes. Em Ciências Humanas o processo verbal mais relevante é “destacar”, cabendo ressaltar que “determinar” fica bastante abaixo da média (221 ocorrências por processo), diferente da área de Ciências Exatas, em que o processo é o mais frequente, representando quase o dobro da média da área (165 ocorrências por processo em Ciências Exatas e Multidisciplinar, já que ambas compartilham o mesmo total de *tokens*).

Já a Grande Área Multidisciplinar é marcada pelo equilíbrio, nenhum processo se destacou acentuadamente dos demais, provavelmente porque há nela uma convergência de áreas de avaliação e subáreas. As áreas de Ciências Agrárias e Biológicas compartilham o mesmo processo mais frequente (“demonstrar”), porém, junto com Engenharias, os processos menos frequentes devem ser destacados (“descrever” e “estabelecer” em Ciências Biológicas, “estabelecer” em Ciências Agrárias e “demonstrar” em Engenharias), pois estão bastante abaixo da média das áreas (144 ocorrências por processo), correspondendo a quase metade do número.

Essas observações contribuem para a ideia de que o vocabulário acadêmico deve ser estudado e ensinado. Caberia aqui o argumento de que esse vocabulário deveria ser ensinado por área de conhecimento, afinal, conhecer os muitos empregos de “estabelecer” pode não ser tão proveitoso para o estudante de Ciências Agrárias quanto dominar os usos de “demonstrar”. Porém, essa diferença entre número total de ocorrências é interessante para demonstrar que existe variação entre as áreas e que o ensino do gênero pode ser feito focando em alternativas lexicais aos itens mais frequentes.

Para análise qualitativa, foram selecionadas aleatoriamente 70 linhas de concordância de cada um desses processos, levando em consideração seus participantes e o contexto em que ocorreram.

Com base nessas amostras, outras observações gerais puderam ser feitas sobre o comportamento dos processos verbais. Dentre as 420 ocorrências, a maior parte dos Dizentes expressos nas orações correspondeu a outras pesquisas ou a metáforas metonímicas, como exemplificado em (59) e (60). Quando não estava expresso, o Dizente podia ser recuperado através do contexto, ou facilmente deduzido, conforme (61), geralmente referindo-se ao próprio autor do texto.

- (59) A CIF foi aprovada em 2001 e antecipa o principal desafio político da definição de deficiência proposta pela Convenção sobre os Direitos das Pessoas com Deficiência: o documento ^(Dizente) **estabelece** critérios para mensurar as barreiras e a restrição de participação social. (Trecho retirado da área de Direito)
- (60) Esses resultados ^(Dizente) **sugerem** a necessidade de maior tamanho ótimo de parcela (Xo) para avaliar MVR, quando comparado ao Xo necessário para avaliar MVT e MVPA, o que evidenciaria variabilidade do Xo entre caracteres. (Trecho retirado da área de Agronomia)
- (61) Desta forma a imagem gravada não estava sincronizada com os sinais elétricos. No intuito de sincronizar estes, a associação das imagens gravadas com os sinais adquiridos pelo sistema foi feita em três etapas: inicialmente definiu-se a velocidade média de soldagem, na sequência **determinou**-se a distância do início da filmagem à origem da abertura do arco (através da régua posicionada a frente da chapa de metal) e por fim, através destas **determinou**-se o tempo do início de abertura do arco até o início da filmagem. (Trecho retirado da área de Engenharia II)

As mensagens comunicadas pelo processo verbal podem ser a Verbiagem ou uma oração projetada, como visto anteriormente (seção 3.2.5). A Verbiagem normalmente é associada a um sintagma nominal, que pode assumir feições oracionais, enquanto a projeção (que não é considerada um participante) representa uma oração completa, subordinada ao processo (SOUZA, 2006, p. 157-158).

O uso de orações com discurso direto ou indireto relaciona-se ao teor argumentativo do gênero, pois a inserção de vozes externas na narrativa tende a funcionar como argumento de autoridade (em vez de relatos de diálogos, como acontece em textos narrativos) (SOUZA, 2006, p. 158). Em nossos dados, vale observar que houve apenas 10 ocorrências de citações diretas (2,4% do total), a maior parte das mensagens ocorreram sob a forma de sintagmas nominais ou citações indiretas. Outra observação válida é a ausência do participante Receptor (optativo nos processos verbais) nas linhas de concordância analisadas.

De tal modo, nos próximos subitens, cada um dos seis processos verbais mais frequentes foi detalhado e exemplificado com trechos analisados.

6.2.1 DEMONSTRAR

Esse foi o processo protipicamente verbal mais frequente no *corpus*, se mantendo na média em todas as Grandes Áreas, com exceção de Engenharias, em que foi o menos frequente dos seis analisados. A forma linguística mais utilizada foi a conjugada, com 1.239 ocorrências totais, sobre 386 formas nominais (infinitivo e gerúndio) e 222 formas passivas.

- (62) Estudos clínicos ^(Dizente) **demonstram** que a maturação fetal pode ser iniciada de forma precoce ^(Oração projetada), em situações nas quais o feto é submetido a estresse crônico, como o desenvolvimento de lesões placentárias e episódios de hipóxia (Rossdale et al., 1991; Mays et al., 2002). (Trecho retirado da área de Medicina Veterinária)
- (63) Os casos de botulismo ocorreram em todos os meses do ano acometendo bovinos de praticamente todas as faixas etárias, embora a maioria dos surtos tenha ocorrido entre os meses de setembro e março acometendo animais com idade acima de 19 meses. Esses resultados ^(Dizente) **demonstram** que embora a maioria dos surtos diagnosticados esteja relacionada à osteofagia (53 surtos), outras fontes de intoxicação como água ou alimento (3 surtos) também foram relatadas ^(Oração projetada). (Trecho retirado da área de Zoologia)
- (64) Em um estudo de intervenção randomizado realizado na Índia os autores ^(Dizente) **demonstraram** melhorias importantes no nível de conhecimento dos pacientes ^(Verbiagem) que receberam aconselhamento quando comparados a pacientes que não receberam a intervenção (Ponnusankar et al., 2004) (Trecho retirado da área de Ciências Biológicas II)
- (65) Além disso, foi **demonstrado** que algumas espécies de fungos podem acumular altas concentrações de metais em seus corpos de frutificação ^(Oração projetada) (TURNAU, 1991), ou em outros tecidos fúngicos, na parede celular e em compartimentos internos (COLPAERT; VAN ASSCHE, 1987; GADD, 1993; BLAUDEZ et al., 2000; GRAZZIOTTI et al., 2001; BELLION et al., 2006). (Trecho retirado da área de Ciências Biológicas III)
- (66) Neste trabalho, pudemos **demonstrar** que a ocorrência de eventos adversos relacionados à Enfermagem levou a um risco de morte significativamente maior ^(Oração projetada) ($\chi^2=10,49$ p:0,001). (Trecho retirado da área de Enfermagem)
- (67) Estudos em ostracodes coletados na Turquia (Fig. 9) mostram que, apesar da ocorrência de diferentes espécies antes e após o limite Permiano-Triássico, nesta região, os grupos de ostracodes típicos do Paleozóico puderam sobreviver por mais tempo à extinção em massa (CRASQUIN-SOLEAU et al., 2002). Desta forma, **demonstra-se** que as mudanças na biosfera não foram homogêneas em escala global ^(Oração projetada). (Trecho retirado da área de Planejamento Urbano Regional e Demografia)

Nos trechos anteriores, o processo tem a função de reportar os resultados obtidos por pesquisadores, sendo, portanto, característico dos processos verbais. Em (63) e (64), ambos Dizentes são sintagmas nominais que representam tais resultados através de metonímia. Em (63), o Dizente refere-se a trabalhos alheios, fato evidenciado pela referência bibliográfica no final do exemplo, “(Rossdale et al., 1991; Mays et al., 2002)”, enquanto em (64), o Dizente associa-se aos próprios resultados apresentados no artigo, comprovado tanto pela ausência de referências quanto pela seção do artigo em que se encontra o trecho – a discussão.

O mesmo comportamento do Dizente em (63) acontece em (65), em que há referência ao trabalho dos autores mencionados, ao final da sentença, e também uma breve

explicação da pesquisa na circunstância “um estudo de intervenção randomizado realizado na Índia”.

Em (66), o Dizente é omitido da oração, mas facilmente recuperável pelo mesmo mecanismo de explicitação das fontes de (63) e (65). Tal facilidade (e necessidade) ao se recuperar o Dizente, ocorreu em boa parte das passivas analisadas em todos os seis processos, seja por ter o Dizente expresso posteriormente ou por se referir ao próprio autor do texto, não-expresso linguisticamente, mas subentendido pelo contexto.

No trecho (67), o autor se insere no texto, reforçando os resultados de seu trabalho através da desinência de terceira pessoa do plural. Por outro lado, em (68), o Dizente omitido serve para centralizar o foco da frase na mensagem (“as mudanças na biosfera não foram homogêneas em escala global”), cuja veracidade é comprovada por fontes explicitadas na frase anterior, inserindo o discurso alheio através das referências.

6.2.2 DETERMINAR

Esse é o segundo processo mais frequente, sendo mais frequente em Ciências Exatas e menos frequente em Ciências Humanas e Linguística, Letras e Artes. Suas formas linguísticas foram bastante equilibradas entre conjugada (520 ocorrências), infinitivo e gerúndio (548 ocorrências) e passiva (613 ocorrências), resultando em 1.751 *tokens* totais.

(68) Ambas as contas referências 1985 e 2000 diferenciam entre as despesas de consumo final das administrações públicas voltadas "(...) para benefício da comunidade como um todo (consumo coletivo)" e as voltadas "para benefício das famílias individuais" (IBGE 2008b, p.42). Ambas ^(Dizente) **determinam**, ademais, que "o valor do consumo final das administrações públicas em bens e serviços de caráter individual" deve ser "computado como transferências sociais em espécie" ^(Oração projetada) (ibid)[9] (Trecho retirado da área de Economia)

(69) A decisão sobre a realização de laparoscopia ou laparotomia como abordagem cirúrgica inicial foi baseada na avaliação pré-operatória realizada pelos cirurgiões. O protocolo do estudo ^(Dizente) **não determinava qual abordagem a ser realizada** ^(Oração projetada). As cirurgias foram realizadas por um especialista em oncologia ginecológica assistido por um ou mais médicos residentes ou pós-graduandos. (Trecho retirado da área de Medicina III)

(70) Aqui não faz muito sentido a afirmação apodítica "Todo caso é diferente", visto que são as próprias regras jurídicas ^(Dizente) que **determinam quais aspectos de cada caso real são relevantes para a decisão** ^(Oração projetada) (NEUMANN, 2001, p. 244 s.; RÖHL; RÖHL, 2008, p. 604 s.). (Trecho retirado da área de Direito)

- (71) Ao contrário do que **determinam as normas técnicas do Ministério da Saúde** ^(Dizente), a verdade do estupro para o acesso ao aborto legal não se resume a uma narrativa íntima e com presunção de veracidade, mas é uma construção moral e discursiva produzida pela submissão da mulher aos regimes periciais dos serviços. (Trecho retirado da área de Medicina I)
- (72) (RESULTADOS E DISCUSSÃO) A análise de variância mostrou que não houve diferença significativa entre as três populações (efeito fixo), com $Pr > F = 0,94$. Isto indica que as gerações nos três cruzamentos têm um comportamento semelhante para resistência à doença. Considerando-se os efeitos aleatórios (Avaliação e Geração) e suas estimativas de valores genotípicos (BLUP's), para as gerações e suas interações com as populações, bem como entre as avaliações ao longo do tempo, pôde ser **determinado que as gerações segregantes dos três cruzamentos possuem comportamentos semelhantes** ^(Oração projetada). (Trecho retirado da área de Ciências Biológicas II)
- (73) As denúncias tiveram efeito, pois no próprio dia 19 de janeiro, o Chefe de Polícia enviou um ofício ^(Dizente) ao 1º Delegado Auxiliar **determinando que respondesse** "com toda brevidade possível, as referências que lhe são feitas na representação pública hoje pelo Jornal do Commercio".[27] (Trecho retirado da área de História)

Em (69), (70), (71) e (72) os Dizentes estão explícitos, referindo-se a documentos oficiais (em 69, o numeral refere-se a “as contas referências 1984 e 2000”; em 71, a “as regras jurídicas”; e, em 72, a “as normas técnicas do Ministério da Saúde”) ou a “protocolo de estudo”, em (70), que são veículos da informação contida nas orações projetadas. Nesses casos, as projeções introduzem discursos alheios de modo indireto em (69), como resumo de seu conteúdo (em 70 e 71) e como uma forma de discordar do que foi dito (72), usando informações contidas no Dizente como argumentos contrários ao posicionamento do autor.

Em (71), cabe ainda ressaltar que a menção às regras jurídicas, assim como as referências de diferentes autores no final da sentença, serve para legitimar a proposta de que a afirmação “todo caso é diferente” seria incorreta dentro daquele contexto.

No exemplo (73), o Dizente não está explícito no trecho, mas pode ser inferido como referindo-se aos próprios pesquisadores, em função do contexto – tanto da seção em que se encontra, “resultados e discussão”, quanto da sequência de frases cujos Agentes foram omitidos, todos fazendo referência aos produtores do texto.

No trecho (74), o Dizente associa-se ao ofício enviado pelo Chefe de Polícia, pedindo uma resposta. Nesse caso, o discurso indireto presente na sentença é uma oração projetada por outro processo verbal, “respondesse”, que faz parte da oração projetada por “determinar”.

Nos exemplos acima, o processo adquiriu uma qualidade verbal (em oposição a ocorrências materiais e relacionais em casos não analisados detalhadamente por esta pesquisa), correspondendo a ‘prescrever’, nos itens (69) a (72), ‘indicar’, em (73) e ‘demandar’, em (74).

6.2.3 DESTACAR

Dos 1.649 *tokens* do processo “destacar”, a maioria das ocorrências foram em forma conjugada, 1.041. Metade desse valor ocorreram no gerúndio e infinitivo, 548, havendo apenas 64 passivas. O lexema foi frequente em todas as Grandes Áreas, mantendo-se na média, com exceção de Ciências Sociais Aplicadas, em que foi o mais relevante.

- (74) EL DAMIR (2006) ^(Dizente) **destacou** que o tipo de substrato é importante para a produção massal de fungos ^(Oração projetada), uma vez que está diretamente relacionado com os custos e a qualidade dos propágulos produzidos. (Trecho retirado da área de Ciências Biológicas III)
- (75) Estudos prévios ^(Dizente) já **destacavam** que a elevada idade entre os cuidadores principais [7, 15, 22, 24] pode estar associada ao fato de que muitas vezes este cuidador é um cônjuge [24] ou mesmo um filho(a) ^(Oração projetada). (Trecho retirado da área de Enfermagem)
- (76) O mesmo entendimento se aplica aos Sistemas Especialistas, cujo recurso que o diferencia dos demais é a base de dados de conhecimento específico que trabalha com inferência na localização da informação que possivelmente ajudará a solucionar determinado problema, o que caracteriza uma decisão a respeito de uma dada situação. Por último, **destaca-se** a Inteligência de Negócios ^(Verbiagem) (Business Intelligence - BI) que tem nos Sistemas de Informação Baseados em Computador (SIBCs) sua base de ação. (Trecho retirado da área de Ciências Sociais)
- (77) Entre as muitas obras que demarcaram rumos interpretativos à educação como investigações teóricas exemplares acerca do que significa educar, **destacam-se**: Ética a Nicômaco (cerca de 334 a. C.) de Aristóteles; Cartas a Lucílio (65) de Sêneca; De magistro (389), de Agostinho; Da afeição dos pais pelos filhos, em Ensaio (1580 e 1588), de Montaigne; Alguns pensamentos sobre educação (1693), de Locke; Emílio ou da educação (1762), de Rousseau; Cartas sobre a educação estética da humanidade (1795), de Schiller e Sobre a pedagogia (1803), de Kant ^(Verbiagem). Das discussões desses pensadores, surgiram os fundamentos filosóficos da educação moderna. (Trecho retirado da área de Educação)
- (78) Foi introduzida em açudes do nordeste no início da década de 1950, pelo DNOCS - Departamento Nacional de Obras Contra a Seca, porém, somente a partir da década de 1970 que P. squamosissimus ^(Verbiagem) **destacou-se** na produtividade pesqueira (Fontenele & Peixoto, 1978; Sato & Godinho, 1999). (Trecho retirado da área de Zoologia)

(79) À luz deste fato, esta passou a ser a frequência considerada como foco dos testes, e, portanto, a impedância paralela dominante proposta foi um filtro harmônico sintonizado para a frequência de 180 Hz, cujos parâmetros ^(Verbiagem) são **destacados** na Tabela 6. (Trecho retirado da área de Engenharia IV)

Diferente do processo anterior, “determinar”, todas as ocorrências encontradas de “destacar” corresponderam a processos verbais. Nos exemplos (75) e (76), o Dizente está explícito, sendo que em (76), há inclusive referências a tais estudos, marcadas pela numeração entre colchetes, “[7, 15, 22, 24]”.

Já nos trechos de (77) a (79) o Dizente não está expresso, mas refere-se ao autor do texto. Em (77), isso pode ser percebido pela indeterminação do sujeito no processo analisado e no processo inicial da sentença anterior (“aplica”), tendo como Verbiagem “a Inteligência de Negócios”. Em (78), reocorre a indeterminação, porém, pela falta de menções a outros estudos (através de referências dentro do texto ou ao final da sentença), entende-se que o destaque das obras foi feito pelo autor – diferente do trecho (79), em que, apesar do pronome reflexivo, a autoria da informação é atribuída às fontes mencionadas no final da sentença, “Fontenele & Peixoto, 1978; Sato & Godinho, 1999”. Esse mecanismo de omissão do Dizente, ou sua referência posterior, foi bastante comum nos trechos analisados do processo.

Novamente, em (80), houve a omissão do Dizente, mas pode ser compreendido como o autor do texto através da circunstância “na Tabela 6”, que se refere a tabela publicada no artigo, cuja fonte diz ser os próprios autores.

Em todos os trechos analisados, o sentido básico do processo foi ‘fazer sobressair’ a Verbiagem ou projeção.

6.2.4 ESTABELEECER

Esse processo, nem sempre verbal, teve 1.523 *tokens*, sendo 622 ocorrências conjugadas, 637 nominais e 264 passivas. No *corpus*, “estabelecer” foi mais frequente na Grande Área de Ciências Humanas e menos utilizado em Ciências Biológicas e Ciências Agrárias, estando bastante abaixo da média em ambos casos.

(80) A principal alteração na política migratória deste período e que basicamente pauta as alterações subsequentes foi a adoção da lei Turco-Napolitano ^(Dizente), de 1998, que **estabelece os Centros de Permanência Temporária para o confinamento dos imigrantes irregulares** ^(Verbiagem) que não podem ser imediatamente expulsos,

mediante o dispositivo da detenção administrativa 14. (Trecho retirado da área de Planejamento Urbano Regional e Demografia)

- (81) A ABNT NBR 10004 ^(Dizente) **estabelece** que a concentração do referido elemento, expressa como fluoreto (F-), não deve ultrapassar 150 mg L-1 no extrato lixiviado ^(Oração projetada). (Trecho retirado da área de Química)
- (82) D'Ambrósio também é o líder internacional e o disseminador mundial das ideias que envolvem a etnomatemática e suas aplicações em Educação Matemática (ROSA; OREY, 2005). Em seus estudos, na área sociopolítica, D'Ambrosio (2004) ^(Dizente) **estabeleceu um relacionamento amplo entre a matemática, a antropologia e a sociedade** ^(Verbiagem). (Trecho retirado da área de Educação)
- (83) Com o objetivo de encontrar um padrão de normalidade nas proporções verticais da face, vários autores [4, 5, 6, 14, 17, 18, 20, 30] ^(Dizente) **estabeleceram valores médios considerados normais para essas proporções** ^(Verbiagem), observando que os pacientes que apresentavam tais medidas, ou que se encontravam próximos delas, possuíam faces mais equilibradas e harmoniosas. (Trecho retirado da área de Odontologia)
- (84) Esta equação será útil mais tarde, na Seção 4, ao **estabelecermos a equivalência entre o MCU e o MHS** ^(Verbiagem) empregando os dados de Galilei. (Trecho retirado da área de Astronomia e Física)
- (85) [MÉTODO] O estudo consistiu em uma pesquisa descritiva, transversal e individualizada. Foram *avaliadas* 50 crianças, sendo 25 asmáticas e 25 não-asmáticas na faixa etária de 7 a 12 anos, vinculadas aos Ambulatórios de Alergologia e Imunologia e/ou ao Ambulatório de Pediatria Geral do Hospital das Clínicas da Universidade Federal de Pernambuco - Brasil. A coleta foi *realizada* no período de agosto de 2008 a julho de 2009. Esta faixa etária ^(Verbiagem) foi **estabelecida** por ser este período considerado de transição no desenvolvimento do sistema respiratório e marcar o fim do crescimento rápido e das modificações estruturais das unidades respiratórias periféricas (12). (Trecho retirado da área de Medicina III)

Em (81) e (82), os Dizentes referem-se a leis e normas que regem instituições (em 81) e regras (82). Esse tipo de Dizente foi bastante frequente nos processos verbais de “estabelecer” (assim como foram nos processos verbais de “determinar”). Nesses casos, o processo funciona atribuindo às mensagens um valor de ‘ordem’.

Nos trechos (83) e (84), o processo refere-se a discursos de outros pesquisadores, os Dizentes, que determinaram através de suas pesquisas o conteúdo da Verbiagem em ambos casos. Já em (85), o Dizente está associado ao próprio autor do texto, marcado pela desinência número-pessoal do processo, cujo emprego está associado ao sentido de ‘demonstrar’ a Verbiagem destacada.

Em (86), o Dizente não está explícito, mas pode-se perceber pela seção em que se encontra o trecho (“método”), pela referência na primeira sentença, “o estudo”, referindo-se ao próprio artigo, e pela sequência de passivas que omitem o Agente (destacadas em itálico), que

o Dizente se refere ao autor do texto. Nesse exemplo, o uso do processo se aproxima do sentido de ‘determinar’ algo, no caso, a Verbiagem “esta faixa etária”.

Diferente de outros processos verbais, como “destacar”, há nos trechos (81) e (82), certa imperatividade, que Halliday e Matthiessen (2004, p. 255) chamam de semiose imperativa, associada ao trabalho alheio (como leis, normas e estudos que estabelecem padrões) – o que não há quando o Dizente referia-se aos próprios autores do texto.

Além disso, pode-se perceber nesse processo como os limites entre os diferentes tipos não são claros, mas complementares, conforme ilustrado na Figura 1, na seção 3.2. Nos exemplos abaixo, foi ilustrada a ocorrência (frequente) de processos materiais de “estabelecer”, em (87), com sentido de ‘criar’, e, em (88), com sentido de ‘surgir’ – ambos tipicamente materiais.

(86) O Tratado criou o Comitê Executivo Permanente, com uma Secretaria, único órgão administrativo conjunto, que permanecerá no segundo Tratado de Montevidéu em 1980. De certo modo, serviu de modelo para a secretaria do Mercosul, também **estabelecida** em Montevidéu a partir de 1991, igualmente com características técnicas e destituída de força política. (Trecho retirado da área de Ciências Políticas)

(87) A púrpura traumática (Bateman) não se associou aos níveis de densidade óssea, apesar de descrita em outros estudos, o que deve se justificar pela maior sensibilidade das medidas da espessura cutânea na estimativa da variação da densidade óssea, já que a púrpura só se **estabelece** em quadros mais avançados de atrofia cutânea. (Trecho retirado da área de Medicina I)

6.2.5 SUGERIR

O processo prototipicamente verbal “sugerir” obteve 1.352 *tokens* no *corpus*, sendo 967 na forma conjugada, 329 na forma nominal e apenas 56 na forma passiva. Com exceção de Ciências Exatas e Engenharias, em que teve a menor quantidade de ocorrências, o processo se manteve constante nas demais Grandes Áreas.

(88) Como não temos o hábito de comer soja como os orientais - consumimos apenas de 2% a 3% de nossa produção -, a então ministra do Meio Ambiente, Marina Silva ^(Dizente), **sugeriu** que toda a soja transgênica fosse exportada ^(Oração projetada). (Trecho retirado da área de Sociologia)

(89) O apagamento da necessidade de formação de professores-mentores, de sua inserção nas experiências das IES e da inserção de professores formadores no

cotidiano das escolas é um dos mecanismos que opera para seu caráter regulador e, neste aspecto, de permanência das práticas. Quando falo na tensão entre potencial transformador e potencial regulador o faço ciente de que, como **sugerem Furlong et al. (2000)** ^(Dizente), políticas educacionais não são ingenuamente interpretadas e postas em práticas ^(Oração projetada). (Trecho retirado da área de Ensino)

- (90) O problema inteiro-misto foi resolvido com a utilização do solver CPLEX para um conjunto de instâncias hipotéticas e os resultados ^(Dizente) **sugeriram que o estudo pode contribuir para a tomada de decisões eficiente em situações de desastres** ^(Oração projetada). (Trecho retirado da área de Engenharia III)
- (91) Apesar de Du Bois examinar a linha de cor e sua conseqüente dupla consciência em referência específica a tensões raciais entre brancura e negritude nos Estados Unidos, esses conceitos ^(Dizente) **sugerem implicações significativas quanto à forma pelas quais as outras etnias são vistas e recebidas na cultura americana** ^(Oração projetada). (Trecho retirado da área de Linguística e Letras)
- (92) Assim, sendo de 2003, as RDC nº 359 e nº 360 foram baseadas na regulamentação da OMS lançada em 2003,14 e não na Estratégia Global para Promoção da Alimentação Saudável, Atividade Física e Saúde, lançada em maio de 2004 ^[16]. Em 2010, a regulamentação da publicidade de alimentos no Brasil ^(Dizente) **sugeria que "alimento com quantidade elevada de gordura trans era aquele que possuía em sua composição uma quantidade igual ou superior a 0,6 g para 100 g na forma como estava exposto à venda"** ^(Oração projetada). (Trecho retirado da área de Nutrição)
- (93) Desta forma, mesmo com o grande recuo da linha de costa nesse setor do atual território do Estado de São Paulo, impulsionado pela redução do nível do mar durante a última glaciação, altos índices de umidade, oriunda do Oceano Atlântico e/ou da intensificação das massas de ar de origem polar, continuaram a penetrar nessa região ^(Oração projetada), como foi **sugerido por De Oliveira (1992)** ^(Dizente). (Trecho retirado da área de Ciências Ambientais)

Nos trechos (89) e (90), os Dizentes fazem referência a discursos alheios, em (89) à fala de Marina Silva, e em (90) ao trabalho de outros pesquisadores, devidamente citados. A construção (90), “como [Dizente] [processo]” seguida ou precedida por Verbiagem ou oração projetada foi bastante frequente nos trechos analisados.

Em (91) e (92), os Dizentes se referem a termos não-específicos como resultados de pesquisas (em 91), conceitos explicitados por outros pesquisadores (“Du Bois”, em 92) e regulamentos (em 93). Esse tipo de referente metonímico, em que “dados” ou “resultados” ou alguma outra parte do trabalho alheio *sugere* algo, foi igualmente frequente nos trechos.

O exemplo (94) corresponde à forma menos utilizada nas linhas de concordância analisadas, em que o processo verbal se encontra na passiva com o Dizente expresso – na maior parte dos casos, a passiva foi preferida por tirar o foco do Dizente, que era o próprio autor do texto, e ressaltar a mensagem.

6.2.6 DESCREVER

Em “descrever”, as formas conjugada e passiva foram as mais frequentes, com 520 e 490 ocorrências respectivamente. Acrescidas das 279 formas nominais, compõem os 1.289 *tokens* do processo verbal. Em geral, o lexema foi bastante regular em todas as Grandes Áreas, com exceção de Ciências Biológicas em que ficou abaixo da média.

- (94) Alguns produtores ^(Dizente) **descrevem**, ainda, que os bovinos "dão pulos repentinos", caem ao solo e morrem em poucos minutos ^(Oração projetada). Há relato de aproveitamento para o consumo quando os bovinos são imediatamente sangrados. (Trecho retirado da área de Zoologia)
- (95) O trabalho está estruturado da seguinte forma: a Seção 1 faz uma revisão da utilização da metodologia DEA proposta em processos licitatórios; a Seção 2 ^(Dizente) **descreve os princípios de DEA, os modelos clássicos e o método da fronteira invertida ^(Verbiagem) e a Seção 3 apresenta o estudo de caso, descrevendo a modelagem e a análise e discussão dos resultados. (Trecho retirado da área de Engenharia I)**
- (96) No entanto, a literatura ^(Dizente) **descreve que a presença de antioxidante é essencial para a maturação de oócitos ^(Oração projetada), particularmente para a maturação citoplasmática, necessária para o desenvolvimento embrionária (Furnus et al., 2008; Silva, et al., 2011). (Trecho retirado da área de Medicina Veterinária)**
- (97) A ZCAS ^(Verbiagem) é **descrita** como região com alta variabilidade da atividade convectiva no leste da América do Sul durante o verão austral (Carvalho et al., 2004). (Trecho retirado da área de Geociências)
- (98) O artigo está estruturado em outras 5 seções, além desta introdução. Nas duas próximas seções, apresentam-se uma discussão teórica que fundamenta a pesquisa. Na quarta, **descreve-se a metodologia e os dados utilizados** ^(Verbiagem). (Trecho retirado da área de Economia)
- (99) [DISCUSSÃO] Neste estudo **descreveram-se as cinco fases do processo de adaptação transcultural e contributos para validação para o português brasileiro do Questionário sobre as dimensões de qualidade e a satisfação de pacientes com os cuidados primários de saúde** ^(Verbiagem). Produziu-se um instrumento equivalente à versão portuguesa para uso no Brasil, garantindo sua equivalência idiomática e semântica. (Trecho retirado da área de Enfermagem)

Nos trechos de (95) a (97) o Dizente está expresso, associando-se a um Ator consciente, em (95), “alguns produtores”, e a um Dizente metonímico, em (96), no qual “a Seção 2” se refere a uma parte do trabalho realizado pelo autor do texto. O mesmo ocorre em (97), em que o Dizente “a literatura” está relacionado às pesquisas referenciadas no final do trecho, “(Furnus et al., 2008; Silva, et al., 2011)”.

No exemplo (98), o Dizente não está explícito no texto, mas marcado no final da sentença, sob a forma de citação. Essa construção foi frequente nas formas passivas analisadas.

Em (99) e (100), o sujeito indeterminado omite o Dizente, que é o autor dos textos, em ambos os casos. Em (99), o Dizente pode ser identificado por ser usado para detalhar as seções do artigo, cuja autoria só pode ser o próprio pesquisador; e, em (100), a presença da circunstância “nesse estudo” permite ao leitor chegar à semelhante conclusão.

Junto com “sugerir” e “destacar”, todas as ocorrências analisadas de “descrever” foram processos verbais – diferente dos demais processos. Tal fato serve para ilustrar que a divisão entre eles nem sempre é clara, sendo fundamental uma análise qualitativa de suas linhas de concordância.

Dentre os trechos analisados, os processos verbais foram utilizados a fim de reportar o discurso alheio trazendo para o texto não apenas novas informações, mas também servindo para corroborar com o posicionamento do autor; além disso, foram utilizados a fim de apresentar resultados/dados observados e fomentar diálogos com outras pesquisas.

Tais propósitos cumprem funções específicas no texto, atribuindo a ele credibilidade. Reportar a fala do outro, seja para legitimar posições teóricas ou para desconstruí-las, passa uma imagem para o leitor da qualidade do estudo realizado – quanto mais autores foram lidos, quanto mais teorias e pesquisas foram trazidas para discussão, maior seria a compreensão do autor sobre o assunto. Assim, esses discursos de autoridades servem não apenas para fundamentar trabalhos, mas também para tornar os autores “fiadores de veracidade de um certo ponto de vista” (PLATÃO; FIORIN, 2001, p. 202), caso sejam bem empregados – pertinentes ao tema e bem discutidos.

Já a introdução do nome de algum pesquisador, ao qual é tipicamente atribuída uma teoria, explicitada ou não, pode funcionar a favor ou contra o posicionamento dos autores. A fala alheia, representativa de uma perspectiva teórica, tende a complementar a argumentação, sendo, portanto, desvalorizada ou valorizada de acordo com a opinião dos autores. Nos exemplos (101) e (102) abaixo, esse aspecto fica claro. Em (101), todas as referências feitas (“2, 14-16”) atribuem ao autor maior credibilidade em sua pesquisa, pois teriam sido feitas leituras suficientes para chegar àquele resultado; enquanto, em (102), a fala de “(TRAUGOTT, 2002)”, representante da teoria funcionalista, foi empregada para complementar o posicionamento teórico no texto.

(100) Diversos autores já demonstraram que a ASG e suas versões parecem detectar, com maior precisão, os pacientes desnutridos hospitalizados, comparativamente

a outros métodos de avaliação nutricional (2,14-16). (Trecho retirado da área de Medicina II)

- (101) Neste trabalho, embora priorizemos a concepção de mudança como motivada por práticas discursivas e sociais que levam em conta a interação entre falante e ouvinte (TRAUGOTT, 2002), aspectos relacionados à fonte lexical também são contemplados em nossa análise. (Trecho retirado da área de Letras e Linguística)

De tal forma, os processos verbais fazem parte de uma estratégia argumentativa capaz de levar o leitor a concordar com as ideias expressas, contribuindo para justificar e explicar escolhas e posicionamentos teóricos no texto. Para tanto, muitas vezes foram feitos usos de paráfrases ou citações diretas, introduzindo a fala alheia na íntegra ou apenas sua essência – aquela que interessa para a argumentação, instanciando um argumento de autoridade. Os processos verbais também contribuem para atribuir credibilidade ao autor, pois todos os argumentos e resultados expostos teriam sido fruto de vasta leitura crítica sobre as teorias/pesquisas naquela área, ou para distanciá-lo de seu objeto de estudo, através da omissão do Dizente, mantendo a impessoalidade do texto e permitindo que a atenção do leitor recaia na mensagem expressa pelo processo.

7 CONSIDERAÇÕES FINAIS

Esta pesquisa objetivou elaborar uma lista dos lexemas verbais acadêmicos, de acordo com a divisão do vocabulário da língua proposta por Nation (2001), e analisar qualitativamente os processos verbais acadêmicos mais frequentes. Para tanto, foi compilado um *corpus* composto por artigos científicos, representativo do gênero, de todas as áreas de avaliação da CAPES, a fim de observar a existência de diferenças nas escolhas lexicais de cada Grande Área; de particularidades do gênero, que pudessem ser percebidas em nossos dados em escala qualitativa; e de diferenças entre os usos dos processos verbais mais frequentes.

A primeira pergunta desta pesquisa procurou descobrir quais foram os verbos mais frequentes em artigos científicos de diferentes áreas; como resposta/resultado apresentamos o Apêndice A, que descreve os verbos mais frequentes, assim como a quantidade de ocorrências nas formas conjugada, nominal e passiva. Esse modo de visualização dos dados foi pensado a fim de facilitar o uso da lista para o ensino, permitindo constatar, por exemplo, que “consistir” é predominante em sua forma ativa, conjugada, tendo ocorrido nominalmente em apenas 3,8% das ocorrências e nenhuma vez na forma passiva.

Além disso, o Apêndice A está entrelaçado com os resultados da segunda e terceira perguntas, destacando verbos de uso geral (segundo BERBER SARDINHA, 2004 e BIDERMAN, 1998) que foram muito utilizados neste *corpus*, em maior ou menor frequência – quanto menos frequentes em *corpora* não-acadêmicos e mais frequentes neste *corpus*, mais relevante para o ensino, como as formas “avaliar”, “indicar”, “identificar”, “buscar” e “gerar”, que foram menos frequentes no *corpora* representantes do vocabulário geral, porém, estão entre os cinquenta primeiros verbos mais frequentes que encontramos.

A segunda pergunta se relacionou à capacidade de, a partir dos resultados obtidos na primeira, listar um vocabulário de verbos acadêmicos. Conforme demonstrado na Tabela 5, na seção 5.2, a elaboração de tal lista se mostrou possível, tendo como base a terceira pergunta, que indagou sobre as particularidades do *corpus* deste trabalho, comparado a outros *corpora*. Fundamentada nessa comparação, pudemos chegar aos resultados da segunda pergunta. Ao definirmos dois *corpora* compostos por diferentes gêneros, não apenas acadêmicos, tais quais o de Berber Sardinha (2004) e o de Biderman (1998; 2001), e os compararmos com o resultado da pergunta número um, chegamos à lista dos verbos acadêmicos, que poderá orientar o ensino do vocabulário acadêmico do PB, independente da área de avaliação em que se insiram os alunos.

Conforme pode ser observado no Apêndice A, com base no Apêndice B, nem todos os verbos mais frequentes da língua foram encontrados ou foram frequentes neste *corpus*. Verbos dos *corpora* não-acadêmicos altamente frequentes como “dizer”, “falar” e “achar” tiveram pouquíssimas ocorrências no *corpus* de artigos acadêmicos, sendo que o último sequer teve *tokens* suficientes para entrar na lista dos verbos mais frequentes (Apêndice A). Essa particularidade parece estar relacionada à característica do gênero científico em evitar itens lexicais que carreguem o lugar-comum da palavra frequente, assim como certa subjetividade no caso de “achar”, ambas desapreciadas nos gêneros acadêmicos.

As duas últimas perguntas deste trabalho relacionaram-se aos aspectos da LCFU, a fim de identificar quais foram os principais processos verbais do *corpus*, assim como possíveis semelhanças e diferenças entre seus usos. Para responder a essas questões, foram analisados qualitativamente seis processos verbais mais frequentes. Percebemos que metade foram prototipicamente processos verbais, “destacar”, “sugerir” e “descrever”, enquanto na outra parte, “demonstrar”, “determinar” e “estabelecer”, outros tipos de processos também ocorreram. Apesar dessa diferença, todos os processos verbais foram empregados de forma similar ao longo dos trechos analisados.

Como procuramos demonstrar ao longo do trabalho, o domínio dos gêneros acadêmico é importante tanto para o aluno de língua materna quanto de PB enquanto LE. Assim, o ensino/aprendizagem deve ser feito com base nas palavras (e nos usos) mais frequentes do gênero, o léxico acadêmico, e não apenas no vocabulário geral da língua. Além da frequência, justifica-se o ensino desse vocabulário porque corresponde às palavras que ajudam o leitor a compreender intenções e avaliações do autor do texto, seu ensino alcança a mais alunos em turmas heterogêneas e, diferente do vocabulário técnico, para este não há glossário ou notas explicativas durante a leitura.

Através da análise dos lexemas verbais mais frequentes, pudemos confirmar com dados reais algumas noções (como as de LINDQVIST, 2014, explicitadas na seção 2.4) sobre o gênero. Os itens altamente frequentes na linguagem cotidiana não o são em gêneros acadêmicos, bastando conferir os itens destacados em azul no Apêndice A – verbos comuns como “achar”, “falar”, “querer” e “saber” são pouco utilizados, em favor de outros que cumprem funções semelhantes, como o verbo modal “deve”, outros processos que reportam discursos, como “demonstrar” e “determinar”, e processos mentais alternativos, como “verificar” e “considerar”. A lista proposta por este trabalho ajuda a indicar quais formas são pouco utilizadas, ao mesmo tempo em que fornece as alternativas que foram preferidas e a frequência de suas formas linguísticas.

Além disso, com a análise qualitativa dos processos verbais, pudemos perceber a predominância de formas na terceira pessoa do presente do indicativo, que tem a capacidade de fazer com que uma ação no passado seja trazida para o presente do leitor, causando um efeito de atemporalidade (FUZER, 2012), ou omitindo-se o sujeito da oração através da indeterminação, o que distancia o autor de seu objeto de estudo. Esse mesmo objetivo pode ser alcançado com o uso da passiva, a fim de omitir a participação efetiva do autor ou de destacar a mensagem expressa pelo processo verbal. Tais mecanismos contribuem para tornar o discurso científico mais objetivo e impessoal.

Apesar de ser um participante obrigatório dos processos verbais, o Dizente muitas vezes não estava presente, devendo ser inferido pelo contexto. Essa característica foi bastante utilizada para omitir a participação do autor, distanciando-o de seu objeto de estudo, de forma a manter a impessoalidade do texto acadêmico e permitir que o foco da frase esteja na informação comunicada pelo processo – na Verbiagem ou na oração projetada, que estiveram expressas em grande parte dos casos analisados.

Essa mensagem que reporta discursos alheios, majoritariamente por meio de citações diretas ou de paráfrase, confere maior credibilidade ao autor ao trazer a fala de outras pessoas, ou trabalhos já publicados, validando o posicionamento teórico do texto através de argumentos de autoridade.

Em geral, os resultados que obtivemos permitiram perceber que há diferenças entre as Grandes Áreas, inclusive nas palavras acadêmicas preferidas por cada uma. Porém, as diferenças não foram suficientes para invalidar a utilidade do ensino do vocabulário acadêmico em sala de aula – serviram para demonstrar que as particularidades existem, mas não são radicais, pois não interferiram nos dados quantitativamente, e apenas puderam ser observadas quando comparadas entre si, como a preferência por certos processos em detrimento de outros (apresentada na seção 6.2).

Assim, cabe aqui sugestões para trabalhos futuros, tais quais utilizar métodos alternativos para identificar o vocabulário acadêmico (GARDNER; DAVIES, 2014), a fim de aumentar a precisão de uma possível lista de itens acadêmicos; analisar mais a fundo o léxico de cada Grande Área através de comparação e refinamento dos dados de cada *subcorpus*; observar mais linhas de concordância dos processos mais frequentes para determinar quais os tipos mais frequentes de cada área. Essas são apenas algumas ideias que podem ser realizadas aprofundando estudos nessa área.

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.
- BERBER SARDINHA, Tony. **Linguística de Corpus: Histórico e Problemática**. 2000. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005>. Acesso 22 jul. 2016.
- BERBER SARDINHA, Tony. **Linguística de Corpus**. Barueri: Manole, 2004.
- BIDERMAN, Maria Tereza. A face quantitativa da linguagem: um dicionário de frequências do português. **Alfa: revista de linguística**, v. 42(n.esp.), p. 161-181. 1998.
- _____. **Teoria linguística: teoria lexical e linguística computacional**. São Paulo: Martins Fontes, 2001.
- CAMPOY, Mari Carmen Campoy Cubillo. Verb types in chemistry research articles: a semantic perspective. **The ESP**, v. 23, n. 1, p. 19-35. 2000.
- CARROLL, J. B., DAVIES, P., & RICHMAN, B. **The American Heritage Word Frequency Book**. New York: American Heritage, 1971.
- CHUNG, Teresa Mihwa; NATION, Paul;. Technical vocabulary in specialized texts. **Reading in a Foreign Language**, v. 15, n. 2. 2003.
- COHEN, A. D; GLASMAN, H. ROSENBAUM-COHEN, P; FERRARA, J. FINE, J. Reading English for specialized purposes: discourse analysis and the use of student informants. IN: CARRELL, P; DEVINE, J; ESKEY, E. **Interactive approaches to second language reading**. Cambridge: Cambridge University Press, 1988.
- COXHEAD, A. A New Academic Word List. **TESOL Quarterly**, v. 34, n. 2, p. 213-238. 2000.
- COXHEAD, A.; BYRD, P. Preparing writing teachers to teach the vocabulary and grammar of academic prose. **Journal of Second Language Writing**, v. 16, n. 3, p. 129–147. 2007.
- COXHEAD, A.; NATION, P. The specialised vocabulary of English for academic purposes. IN.: FLOWERDEW, J; PEACOCK, M. **Research Perspectives on English for Academic Purposes**. Cambridge: Cambridge University Press, 2001.
- CUNHA, Celso; CINTRA, Luís F. Lindley. **Nova gramática do português contemporâneo**. 5.ed. Rio de Janeiro: Lexikon, 2008.

DAVIES, M. 2012. **Corpus of Contemporary American English** (1990–2012). Disponível em: <<http://corpus.byu.edu/coca/>>. Acesso em 22 jul. de 2016.

DELL'ISOLA, Regina Lúcia Peret. **O sentido das palavras na interação leitor ⇔ texto**. Belo Horizonte: Faculdade de Letras da UFMG, 2005.

DELUCA, E. Unlocking Academic Vocabulary. **Science Teacher**, v. 77, n. 3, p. 27–32. 2010

EGGINS, S. **An introduction to systemic functional linguistics**. 2nd ed. Londres: Continuum, 2004.

FURTADO DA CUNHA, Maria Angélica Furtado da; SOUZA, Maria Medianeira de. **Transitividade e seus contextos de uso**. Rio de Janeiro: Lucerna, 2011.

FUZER, C. realização de processos verbais em textos científicos da área de engenharia civil. **DELTA**, v. 28, n. especial, 2012.

GARDNER, D.; DAVIES, M. A new academic vocabulary list. **Applied Linguistics**, 35(3), 2014. p. 305-327. Disponível em: <<http://appliedjournals.org/content/35/3/305.full>> Acesso 22 de julho de 2016.

GIVÓN, Talmy. **Syntax I: a functional-typological introduction**. Amsterdam: John Benjamins, 1984.

HALLIDAY, Michael. **An Introduction to Functional Grammar**. 3 ed. Revised by Christian M. I. M. Matthiessen. London, Edward Arnold, 2004.

HALLIDAY, Michael. **An Introduction to Functional Grammar**. 4 ed. Revised by Christian M. I. M. Matthiessen. London, Edward Arnold, 2014.

HYLAND, K. **English for Academic Purposes: An Advanced Resource Book**. London: Routledge, 2006.

HWANG, K; NATION, I. S. P. Where would general service vocabulary stop and special purposes vocabulary begin? **System**, v. 1, n. 23, 1995. p. 35-41.

LINDQVIST, F. **English for Academic Purposes Vocabulary**. 2015. Disponível em <https://gupea.ub.gu.se/bitstream/2077/38442/1/gupea_2077_38442_1.pdf>. Acesso 22 jul. 2016.

MARTIN, Anne V. Teaching Academic Vocabulary to Foreign Graduate Students. **TESOL Quaterly**, vol. 10, n. 1. p. 91-98. 1976.

MARTÍNEZ, Iliana A. *et al.* Academic vocabulary in agriculture research articles: a corpus-based study. **English for Specific Purposes**, v. 28, p. 183-198. 2009.

MENDES, J. M. de O. **Processos verbais em artigos científicos de Administração de Empresas e Engenharia: perspectivas sistêmico-funcionais**. 2012. 112 f. Dissertação (Mestrado em Linguística Aplicada a Estudos da Linguagem), PUC-SP, São Paulo, 2013.

MIRANDA, M. V; OLIVEIRA, A. A. Substantivos em artigos científicos: o vocabulário técnico e acadêmico na perspectiva da Linguística de Corpus. **Revista Signótica**, v. 28, n.2, 2016.

MOREIRA FILHO, J. L. **Kitconc 4.0**. 2008. Disponível em <<http://www.fflch.usp.br/dl/li/x/?p=394>> Acesso em: 8 ago. 2013.

NATION, Paul. **Learning vocabulary in another language**. Cambridge: Cambridge University Press, 2001.

NUNES, S; PERNA, C. Uma Análise dos Substantivos como Marcadores de Posicionamento em Artigos Acadêmicos em Língua Portuguesa. **BELT: Brazilian English Language Teaching Journal**, v. 6, n. 2, p. 172-195, 2015.

PENHA, R. F. A transitividade dos Processos Verbais Dizer e Afirmar em artigos científicos de graduandos em Letras. **Revista Ao Pé da Letra**. v. 14, n. 1, 2012.

PLATÃO, F; FIORIN, J. L. **Manual do candidato: português**. Brasília: Fundação Alexandre de Gusmão, 2001.

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.

SANTOS, Vanderlei dos Santos. **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.

SILVA, E. C. Previsões em horóscopos: o sistema de transitividade nas representações de futuro. IN: SOUZA, Medianeira (org). **Sintaxe em foco**. Recife: PPGL/UFPE, 2012.

SOUZA, Medianeira. **Transitividade e construção de sentido no gênero editorial**. Recife: UFPE, Programa de Pós-graduação em Letras, 2006. (Tese de doutoramento)

STREVENS, P. Technical, technological, and scientific English'. **ELT Journal**, v.27, p. 223-234. 1973.

SUTARSYAH, Cucu; NATION, Paul; KENNEDY, Graeme. How useful is EAP vocabulary for ESP? A corpus based study. **RELC Journal**, v. 25, n. 2, p. 34-50. 1994.

THOMPSON, G. **Introducing Functional Grammar**. 3 ed. Nova Iorque: Routledge, 2014.

UCCELLI, P; DOBBS, C. L; SCOTT, J. Mastering academic language: organization and stance in the persuasive writing of high school students. **Written Communication**, v. 30, n. 1, p. 36-62. 2013.

WEST, Michael. **A general service list of English words**. London: Longman, Green & Co. 1953.

VIVAN, E. G. S. **Principais usos de processos verbais e metáforas interpessoais em artigos de Linguística Aplicada**. São Paulo: PUC-SP, 2010. (Tese de doutoramento)

XUE, Guoyi; NATION, Paul. A university word list. **Language Learning and Communication**, v. 3, n. 215, 1984.

APÊNDICE A: Lista dos lexemas verbais mais frequentes no *corpus*, contrastados com os lexemas verbais pertencentes ao vocabulário de alta frequência (BERBER SARDINHA, 2004; BIDERMAN, 2000)

Legenda

Entre os 1.000 mais frequentes
Entre as 2.000 mais frequentes
Entre as 3.000 mais frequentes
Pertencem aos lexemas verbais mais frequentes de Biderman (2000)

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
1	SER	38.420	6.406	476	45.302
2	PODER	17.057	673	2	17.732
3	APRESENTAR	8.834	1.209	1.349	11.392
4	HAVER	6.228	596	57	6.881
5	UTILIZAR	1.855	2.158	2.750	6.763
6	CONSIDERAR	1.450	2.367	2.275	6.092
7	OBSERVAR	2.803	982	1.846	5.631
8	DEVER	4.956	118	133	5.207
9	REALIZAR	1.022	805	3.342	5.169
10	TER	4.039	958	71	5.068
11	FAZER	1.585	1.373	1.585	4.543
12	MOSTRAR	3.197	436	273	3.906
13	ENCONTRAR	2.329	475	885	3.689
14	OCORRER	2.886	535	132	3.553
15	OBTER	922	912	1.160	2.994
16	PERMITIR	2.284	615	80	2.979
17	AVALIAR	658	1.381	857	2.896

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
18	VERIFICAR	1.419	787	416	2.622
19	TORNAR	1.861	695	61	2.617
20	REPRESENTAR	1.706	476	344	2.526
21	POSSUIR	2.279	235	3	2.517
22	ESTAR	2.269	238	2	2.509
23	INDICAR	1.519	700	113	2.332
24	DAR	1.142	619	501	2.262
25	ANALISAR	461	988	719	2.168
26	PASSAR	1.729	326	46	2.101
27	LEVAR	990	864	207	2.061
28	IDENTIFICAR	588	971	492	2.051
29	TRATAR	1.212	454	258	1.924
30	EXISTIR	1.666	195	2	1.863
31	DEMONSTRAR	1.239	386	222	1.847
32	AUMENTAR	967	758	99	1.824
33	DEFINIR	713	434	657	1.804
34	DETERMINAR	520	618	613	1.751
35	INCLUIR	708	673	317	1.698
36	MANTER	750	634	306	1.690
37	CONSTITUIR	1.090	274	312	1.676
38	DESTACAR	1.037	548	64	1.649
39	RELACIONAR	370	205	1.038	1.613
40	CONTRIBUIR	845	657	110	1.612
41	DESENVOLVER	589	474	491	1.554
42	ESTABELEECER	622	637	264	1.523
43	ENVOLVER	872	492	135	1.499
44	BUSCAR	905	490	40	1.435
45	AFIRMAR	991	406	23	1.420
46	GERAR	562	615	237	1.414
47	VER	570	392	436	1.398
48	VARIAR	919	440	39	1.398

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
49	USAR	269	597	517	1.383
50	SUGERIR	967	329	56	1.352
51	PRODUZIR	646	451	243	1.340
52	RECEBER	914	315	106	1.335
53	RESULTAR	790	401	139	1.330
54	FICAR	1.060	219	17	1.296
55	REDUZIR	458	636	198	1.292
56	DESCREVER	520	279	490	1.289
57	PARECER	1.166	65	10	1.241
58	CONTER	396	800	33	1.229
59	COMPREENDER	392	665	163	1.220
60	PERCEBER	645	344	192	1.181
61	APONTAR	828	197	149	1.174
62	CARACTERIZAR	581	276	307	1.164
63	EXPLICAR	393	413	349	1.155
64	FORMAR	302	568	281	1.151
65	SEGUIR	628	407	109	1.144
66	COMPARAR	187	670	274	1.131
67	ASSOCIAR	247	95	788	1.130
68	CRIAR	394	466	261	1.121
69	ENTENDER	451	442	227	1.120
70	INFLUENCIAR	531	282	302	1.115
71	REVELAR	842	209	55	1.106
72	DIZER	443	587	66	1.096
73	APLICAR	288	224	580	1.092
74	POSSIBILITAR	707	365	17	1.089
75	EVIDENCIAR	647	249	162	1.058
76	COMPOR	404	153	474	1.031
77	PROMOVER	456	541	30	1.027
78	REFERIR	812	141	48	1.001
79	CORRESPONDER	852	127	2	981

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
80	RESSALTAR	410	524	46	980
81	CONSTRUIR	269	358	352	979
82	ATINGIR	436	418	110	964
83	TOMAR	360	378	216	954
84	CALCULAR	161	215	562	938
85	VISAR	383	512	-	895
86	TRABALHAR	442	369	75	886
87	COLOCAR	402	235	249	886
88	ASSUMIR	537	291	57	885
89	GARANTIR	264	574	43	881
90	SABER	573	257	43	873
91	CONSISTIR	834	33	-	867
92	AFETAR	459	220	184	863
93	FORNECER	434	324	102	860
94	DIMINUIR	426	383	39	848
95	RELATAR	604	58	180	842
96	NOTAR	565	226	46	837
97	PRECISAR	794	37	1	832
98	CHEGAR	514	305	11	830
99	ATUAR	522	293	9	824
100	SUBMETER	67	60	687	814
101	PENSAR	230	490	92	812
102	ESTIMAR	211	303	286	800
103	CONCLUIR	485	284	31	800
104	CONHECER	195	361	243	799
105	PERMANECER	639	147	13	799
106	ADOTAR	321	201	267	789
107	RECONHECER	307	281	195	783
108	ESTUDAR	281	342	158	781
109	INICIAR	500	185	91	776
110	EXPRESSAR	463	130	181	774

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
111	CHAMAR	455	140	176	771
112	PARTICIPAR	415	328	26	769
113	CONSEGUIR	583	145	38	766
114	ATENDER	298	352	114	764
115	EVITAR	117	588	53	758
116	CONSTATAR	483	145	124	752
117	ATRIBUIR	265	120	357	742
118	RESPONDER	383	283	47	713
119	REFLETIR	404	284	22	710
120	PROPORCIONAR	409	286	11	706
121	SOFRER	507	159	38	704
122	CAUSAR	327	276	87	690
123	ALCANÇAR	218	276	175	669
124	EMPREGAR	167	197	302	666
125	REGISTRAR	284	94	284	662
126	OFERECER	336	259	65	660
127	CLASSIFICAR	131	111	412	654
128	FAVORECER	331	261	47	639
129	BASEAR	261	69	308	638
130	CONDUZIR	217	116	300	633
131	TENDER	597	22	-	619
132	EXIGIR	483	107	26	616
133	VIVER	360	229	27	616
134	ALTERAR	247	234	133	614
135	DISCUTIR	246	192	169	607
136	SERVIR	385	199	22	606
137	APARECER	498	89	10	597
138	IMPLICAR	509	64	21	594
139	CONFIRMAR	289	183	119	591
140	ACONTECER	464	107	10	581
141	SELECIONAR	82	125	370	577

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
142	ABORDAR	261	172	136	569
143	PROPOR	183	170	205	558
144	INVESTIGAR	170	263	124	557
145	JUSTIFICAR	290	177	88	555
146	COLETAR	33	165	351	549
147	MELHORAR	124	397	25	546
148	DIVIDIR	172	86	286	544
149	TRANSFORMAR	248	213	78	539
150	ACREDITAR	460	76	-	536
151	COMEÇAR	438	90	8	536
152	PROVOCAR	304	189	42	535
153	CABER	490	37	-	527
154	CONTINUAR	411	103	8	522
155	IR	391	128	1	520
156	INSERIR	167	94	256	517
157	NECESSITAR	445	63	1	509
158	ESPERAR	259	125	122	506
159	FACILITAR	171	296	38	505
160	OCUPAR	307	141	57	505
161	RESOLVER	131	281	89	501
162	ACOMPANHAR	185	164	152	501
163	FALAR	225	261	7	493
164	QUERER	433	47	-	480
165	PRETENDER	441	20	1	462
166	CORROBORAR	262	151	47	460
167	MEDIR	118	159	182	459
168	AMPLIAR	150	259	45	454
169	INCORPORAR	151	156	143	450
170	REQUERER	391	31	27	449
171	ESCOLHER	133	115	199	447
172	EXERCER	270	145	31	446

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
173	GANHAR	320	105	20	445
174	AJUDAR	286	146	7	439
175	REFORÇAR	248	143	48	439
176	ELABORAR	152	120	156	428
177	TENTAR	248	164	14	426
178	ADQUIRIR	178	118	128	424
179	CITAR	192	128	101	421
180	DEFENDER	287	101	31	419
181	ACABAR	393	20	4	417
182	DIFERIR	352	63	2	417
183	APROXIMAR	261	124	29	414
184	ABRIR	235	125	52	412
185	TRAZER	148	239	22	409
186	VOLTAR	244	99	66	409
187	INTEGRAR	201	142	62	405
188	INTERFERIR	233	166	5	404
189	LOCALIZAR	129	67	203	399
190	EXPLORAR	107	179	107	393
191	ENFATIZAR	217	137	39	393
192	ACEITAR	186	120	87	393
193	ESCREVER	136	161	88	385
194	SALIENTAR	196	179	9	384
195	ILUSTRAR	232	76	75	383
196	OPERAR	222	131	30	383
197	ORGANIZAR	130	113	139	382
198	TESTAR	51	162	167	380
199	MUDAR	192	168	19	379
200	PERTENCER	320	58	1	379
201	ORIENTAR	143	136	99	378
202	PERDER	195	131	52	378
203	MINIMIZAR	71	260	40	371

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
204	DISPOR	228	51	89	368
205	DETECTAR	82	119	166	367
206	AGIR	131	229	3	363
207	ELEVAR	121	93	143	357
208	DIFICULTAR	183	146	25	354
209	FUNCIONAR	242	104	4	350
210	APRENDER	98	235	16	349
211	MARCAR	165	56	123	344
212	LIGAR	77	29	238	344
213	IMPEDIR	184	138	21	343
214	EXCLUIR	55	69	218	342
215	SITUAR	197	71	70	338
216	LEMBRAR	127	188	19	334
217	CONTROLAR	85	164	83	332
218	ADICIONAR	60	65	202	327
219	MODIFICAR	118	124	85	327
220	SENTIR	159	116	51	326
221	SUPERAR	104	170	51	325
222	OPTAR	272	47	5	324
223	ESTIMULAR	135	142	47	324
224	RETIRAR	67	69	188	324
225	CONCENTRAR	210	70	41	321
226	COMPROMETER	118	166	37	321
227	PARTIR	143	171	2	316
228	LIDAR	73	239	4	316
229	PREVER	166	101	44	311
230	DENOMINAR	116	28	166	310
231	MANIFESTAR	221	74	15	310
232	INTERPRETAR	65	127	115	307
233	NASCER	132	164	9	305
234	PREPARAR	46	46	208	300

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
235	QUESTIONAR	96	118	84	298
236	ELIMINAR	65	177	55	297
237	APROVAR	35	13	247	295
238	ASSEGURAR	91	170	32	293
239	OBJETIVAR	172	111	3	286
240	CONFERIR	170	99	17	286
241	ENFRENTAR	107	137	39	283
242	ABRANGER	191	88	4	283
243	DISTRIBUIR	71	47	159	277
244	RECOMENDAR	180	18	79	277
245	DIFERENCIAR	129	113	32	274
246	SEPARAR	87	83	89	259
247	EMERGIR	201	54	3	258
248	REMETER	220	29	8	257
249	AJUSTAR	77	75	105	257
250	PROCURAR	32	192	32	256
251	MENCIONAR	125	70	59	254
252	DESEJAR	214	30	10	254
253	ADEQUAR	27	56	169	252
254	ESCLARECER	82	124	27	233
255	MERECER	208	17	2	227
256	TOTALIZAR	39	184	1	224
257	ARTICULAR	97	94	29	220
258	RESIDIR	196	23	-	219
259	AVANÇAR	88	114	15	217
260	PRESERVAR	50	117	32	199
261	VALER	125	69	3	197
262	POR	97	44	53	194
263	RECUPERAR	52	102	37	191
264	VINCULAR	58	25	105	188
265	DENOTAR	141	38	7	186

	Lexema Verbal	Frequência			Total
		Forma conjugada	Forma nominal	Forma passiva	
266	ADAPTAR	49	87	50	186
267	PRESSUPOR	144	31	6	181
268	FORMULAR	60	68	52	180
269	DECORRER	158	20	1	179
270	INDUZIR	52	90	37	179
271	DESTINAR	76	13	84	173
272	PUBLICAR	76	23	69	168
273	DERIVAR	112	23	33	168
274	COMPLETAR	79	59	29	167
275	PRATICAR	93	43	31	167
276	PREDOMINAR	133	27	-	160
277	DESCARTAR	21	40	99	160
278	RETOMAR	64	66	25	155
279	AGREGAR	71	64	18	153
280	VALIDAR	25	76	51	152
281	OLHAR	34	112	3	149

APÊNDICE B: Lista dos lexemas verbais mais frequentes na lista de palavras mais frequentes do Banco do Português (BERBER SARDINHA, 2004), contrastado com os mais frequentes em Biderman (2000)

Legenda

Pertencem aos lexemas verbais mais frequentes de Biderman (2000)

Lexema verbal		Total de ocorrências	Lexema verbal		Total de ocorrências
1	SER	2.969.006	21	RECEBER	63.363
2	ESTAR	530.215	22	DEIXAR	63.334
3	DIZER	478.062	23	ACHAR	54.010
4	TER	455.619	24	LER	53.376
5	IR	442.535	25	LEVAR	51.903
6	PODER	311.646	26	AFIRMAR	51.308
7	FAZER	298.647	27	COMEÇAR	51.284
8	HAYER	288.503	28	GANHAR	46.898
9	DEVER	195.427	29	TENTAR	46.452
10	FICAR	146.657	30	CONSEGUIR	45.613
11	QUERER	134.553	31	SAIR	45.412
12	DAR	109.120	32	PEDIR	43.503
13	SABER	98.070	33	ACONTECER	43.388
14	CHEGAR	93.946	34	FALAR	40.448
15	PASSAR	90.104	35	ACABAR	39.939
16	VIR	74.841	36	CONTINUAR	38.578
17	VOLTAR	72.656	37	MANTER	36.568
18	VER	70.197	38	PARTICIPAR	36.153
19	MOSTRAR	69.244	39	APRESENTAR	35.651
20	CONTAR	64.085	40	PERDER	35.623
			41	PAGAR	35.558

	Lexema verbal	Total de ocorrências		Lexema verbal	Total de ocorrências
42	EXISTIR	35.299	74	PERMITIR	19.916
43	TRABALHAR	34.865	75	ENTRAR	19.620
44	MORRER	33.225	76	VIRAR	19.422
45	CRIAR	32.957	77	ENFRENTAR	19.147
46	CAIR	32.696	78	ATINGIR	19.059
47	PARECER	30.404	79	INFORMAR	18.754
48	PRECISAR	30.036	80	PROCURAR	18.154
49	ABRIR	29.398	81	CHAMAR	18.007
50	ENCONTRAR	29.185	82	FECHAR	17.847
51	MUDAR	29.025	83	ANUNCIAR	17.779
52	AUMENTAR	28.875	84	EXPLICAR	17.743
53	JOGAR	28.583	85	INCLUIR	17.330
54	TORNAR	27.669	86	TOMAR	17.278
55	VENCER	27.331	87	PREVER	17.236
56	OCORRER	27.248	88	GOSTAR	16.600
57	TRAZER	26.983	89	OFERECER	16.414
58	DECIDIR	25.236	90	NEGAR	16.282
59	LANCAR	24.993	91	GARANTIR	16.272
60	TRATAR	24.307	92	LEMBRAR	15.571
61	ENTRAR	23.459	93	PENSAR	15.426
62	CRESCER	23.311	94	OBTER	14.950
63	ACREDITAR	22.988	95	MATAR	14.848
64	VIVER	22.942	96	CUSTAR	14.198
65	USAR	22.188	97	RESOLVER	13.422
66	DEFENDER	22.053	98	FUNCIONAR	13.357
67	EVITAR	22.011	99	CONHECER	13.229
68	PRETENDER	21.993	100	SOFRER	13.031
69	VENDER	21.976	101	MARCAR	12.964
70	COLOCAR	21.516	102	SIGNIFICAR	12.794
71	COMPRAR	20.480	103	TERMINAR	12.371
72	SUBIR	20.398	104	REDUZIR	12.351
73	ESCREVER	20.314	105	BATER	12.333

	Lexema verbal	Total de ocorrências		Lexema verbal	Total de ocorrências
106	ESTUDAR	12.320	138	REGISTRAR	7.290
107	ACEITAR	11.973	139	PROMETER	7.173
108	ANDAR	11.878	140	PRODUZIR	7.144
109	PEGAR	11.502	141	NASCER	6.877
110	ASSUMIR	11.416	142	IMPEDIR	6.729
111	LIGAR	11.305	143	NEGOCIAR	6.680
112	CONSIDERAR	10.744	144	PREPARAR	6.677
113	REUNIR	10.606	145	INDICAR	6.430
114	TOCAR	10.601	146	CONSTRUIR	6.380
115	RESPONDER	10.365	147	APONTAR	6.271
116	DISCUTIR	10.220	148	APURAR	6.199
117	DEFINIR	10.167	149	VIOLAR	6.103
118	AJUDAR	9.570	150	ESCOLHER	6.081
119	REPRESENTAR	9.547	151	OLHAR	6.056
120	TIRAR	8.853	152	SERVIR	6.046
121	ENTENDER	8.779	153	BUSCAR	6.040
122	DECLARAR	8.619	154	APROVAR	5.981
123	QUEBRAR	8.592	155	CUMPRIR	5.963
124	DEPENDER	8.584	156	CONFIRMAR	5.929
125	EXIGIR	8.484	157	TRANSFORMA	5.826
126	VOTAR	8.439	158	ADMITIR	5.821
127	REVELAR	8.429	159	GUIAR	5.729
128	REALIZAR	8.342	160	DIMINUIR	5.664
129	APARECER	8.234	161	SEGUIR	5.636
130	ESPERAR	8.176	162	ASSISTIR	5.608
131	INVESTIR	7.943	163	CORRER	5.594
132	BASTAR	7.905	164	EXIBIR	5.564
133	OUVIR	7.891	165	CRITICAR	5.555
134	ATENDER	7.607	166	ACOMPANHAR	5.537
135	MELHORAR	7.441	167	PROMOVER	5.516
136	PARAR	7.419	168	CABER	5.448
137	DISPUTAR	7.391	169	PERTENCER	5.402

	Lexema verbal	Total de ocorrências		Lexema verbal	Total de ocorrências
170	PROVOCAR	5.327	182	ATRAIR	4.859
171	SURGIR	5.227	183	TROCAR	4.852
172	AVALIAR	5.192	184	PREFERIR	4.820
173	ATUAR	5.145	185	COMER	4.800
174	RECUPERAR	5.122	186	RECORRER	4.784
175	PROPOR	5.111	187	FUGIR	4.774
176	MONTAR	5.062	188	TEMER	4.709
177	COLABORAR	5.029	189	IDENTIFICAR	4.705
178	SENTIR	5.029	190	GERAR	4.699
179	COSTUMAR	4.943	191	AMPLIAR	4.664
180	RETIRAR	4.864	192	CONTROLAR	4.653
181	APRENDER	4.863			