

CAROLINA RÜBESAM OURIQUE

**RELATÓRIO DE SUSTENTABILIDADE:
CONSTRUÇÃO DE UM GLOSSÁRIO BILÍNGUE
DE TERMOS E FRASEOLOGIAS**

**PORTO ALEGRE
2018**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA: ESTUDOS DA LINGUAGEM
LINHA DE PESQUISA: LEXICOGRAFIA E TERMINOLOGIA: RELAÇÕES
TEXTUAIS**

**RELATÓRIO DE SUSTENTABILIDADE:
CONSTRUÇÃO DE UM GLOSSÁRIO BILÍNGUE
DE TERMOS E FRASEOLOGIAS**

CAROLINA RÜBESAM OURIQUE

ORIENTADORA: PROFa. DRa. ROZANE RODRIGUES REBECHI

Dissertação de Mestrado em Estudos da Linguagem, apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

**PORTO ALEGRE
2018**

CIP - Catalogação na Publicação

Ourique, Carolina Rübeseam
Relatório de Sustentabilidade: construção de um
glossário bilíngue de termos e fraseologias / Carolina
Rübeseam Ourique. -- 2018.
118 f.
Orientadora: Rozane Rodrigues Rebechi.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2018.

1. Terminologia. 2. Unidades fraseológicas
especializadas. 3. Tradução especializada. 4.
Linguística de Corpus. 5. Relatório de
Sustentabilidade. I. Rebechi, Rozane Rodrigues,
orient. II. Título.

Carolina Rübesam Ourique

**RELATÓRIO DE SUSTENTABILIDADE:
CONSTRUÇÃO DE UM GLOSSÁRIO BILÍNGUE DE TERMOS E FRASEOLOGIAS**

Dissertação de Mestrado em Estudos da Linguagem, apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 19 de dezembro de 2018.

Resultado: Aprovado com nota unânime A.

Banca Examinadora:

Prof^a. Dr^a. Rozane Rodrigues Rebechi (Orientadora)

Programa de Pós-Graduação em Letras

Universidade Federal do Rio Grande do Sul

Prof. Dr^a Anna Maria Becker Maciel

Programa de Pós-Graduação em Letras

Universidade Federal do Rio Grande do Sul

Prof. Dr^a Ilza Maria Tourinho Girardi

Programa de Pós-Graduação em Comunicação

Universidade Federal do Rio Grande do Sul

Prof. Dr^a Marina Leivas Waquil

Universidade Federal do Rio Grande do Sul

Para Olívia, com todo o meu amor.

AGRADECIMENTOS

Aos meus pais, Jane e Clóvis, todo meu carinho pela paciência, pelo apoio, amor e dedicação, pela inestimável ajuda com a Olívia. Este projeto seria inviável sem vocês. À minha mãe Jane e à minha irmã Luciana, obrigada pela inspiração na jornada acadêmica.

À Olívia, amor da minha vida, agradeço pela paciência carinhosa e pelo apoio lúdico.

À minha irmã Angela, obrigada pela paciência e por seres um exemplo para mim e para a Olívia.

À Professora Anna Maria Maciel, um agradecimento especial, por ter aceitado me orientar no início deste projeto. Uma pessoa exemplar, por sua dedicação à academia, com incansável curiosidade e espírito inovador.

Agradeço à professora Rozane Rebechi pela parceria paciente e acolhedora e por ter aceitado esse enorme desafio de assumir a orientação no meio do caminho. Obrigada pela confiança e por transmitir tranquilidade sempre com bom humor.

Às professoras Cleci Bevilacqua, Patrícia Reuillard e Maria José Finatto, obrigada pela dedicação e pelos ensinamentos, que ajudaram a expandir meus horizontes. Aos demais professores que cruzaram meu caminho nesta jornada: professor Félix e professora Sabrina, meu sincero agradecimento pela paciência.

À Maria Lúcia Stein, obrigada pelo apoio e pelo incentivo nesses anos.

Aos meus colegas e amigos que ganhei nessa caminhada, Liana, Mauren e Asafe, obrigada pelos momentos de descontração e pela parceria em meio a tantas superações, vocês tornaram esses últimos dois anos muito mais leves!

A todos amigos e amigas que compreenderam as ausências, meu sincero obrigada!

Finalmente, um agradecimento especial à Universidade Federal do Rio Grande do Sul e ao Programa de Pós-Graduação do Instituto de Letras desta Universidade pela oportunidade de continuar minha formação acadêmica.

RESUMO

No Brasil, desde os anos 1990, o uso dos termos *desenvolvimento sustentável* e *sustentabilidade* vem se consolidando por meio da divulgação dos impactos ambientais provocados pelas atividades de empresas e indústrias na forma de um documento conhecido como Relatório de Sustentabilidade (RS) (GRI, 2016). Diante da tendência de crescimento no número dessas publicações em âmbitos nacional e internacional, percebe-se uma oportunidade para tradutores e redatores devido à necessidade de que esses relatórios sejam traduzidos para o português ou vertidos para o inglês, promovendo melhor entendimento entre as empresas e seus públicos. Além disso, existe uma carência em relação a obras de referência no par de línguas inglês-português produzidas com base em metodologia confiável. Nesse sentido, nosso objetivo nesta pesquisa é realizar o levantamento de termos e fraseologias da área com vistas à construção de um glossário bilíngue com finalidade tradutória, para auxiliar redatores e, principalmente, tradutores em suas tarefas textuais. A pesquisa está fundamentada nos pressupostos teóricos da Teoria Comunicativa da Terminologia (CABRÉ, 1999, 2001), da Tradução de perspectiva funcionalista (NORD, 2013) e com foco nas Unidades Fraseológicas Especializadas (BEVILACQUA, 2004; WAQUIL, 2013). Para esse levantamento, adotamos a abordagem metodológica da Linguística de *Corpus* (BERBER SARDINHA, 2004). O *corpus* se configura como comparável, isto é, formado por textos autênticos escritos no idioma original (e não sua tradução) e foi composto por dois *subcorpora*, com textos em português brasileiro e em inglês estadunidense. Os textos foram coletados e compilados em dois *subcorpora* a partir de determinados critérios, os quais foram, então, submetidos à análise com auxílio das ferramentas computacionais. Foi utilizado o programa *Sketch Engine* (SE) para realizar a extração semiautomática de palavras-chave, que forneceu candidatos a termos e fraseologias, os quais foram organizados para que se procedesse à análise manual a partir das linhas de concordância. A análise incluiu a limpeza manual das listas, isto é, verificação das linhas de concordância de cada candidato em suas ocorrências para que se pudesse selecionar termos e fraseologias estatisticamente mais significativas a fim de compor o modelo para o glossário bilíngue. Esse modelo apresenta a descrição de cada item que comporá as entradas na obra. Ao final, são discutidos os dados encontrados à luz dos pressupostos teóricos, fazendo-se uma avaliação das ferramentas utilizadas.

Palavras-chave: Terminologia, Unidades Fraseológicas Especializadas, Tradução Funcional, Linguística de *Corpus*, Relatório de Sustentabilidade.

ABSTRACT

In Brazil, since the 1990s, the use of terms such as *sustainable development* and *sustainability* has been consolidated in the form of a document known as Sustainability Report (SR) (GRI, 2016). This document is intended to disclose the environmental impacts caused by companies and industries. Despite the growing number of such publications at both national and international levels, there is an opportunity for translators and writers due to the need for these reports to be translated into Portuguese or English, promoting a better understanding between companies and their audiences. In addition to that, there is a gap in reference works in the language pair ‘English-Portuguese’ that can help translators and writers produce these documents. In this regard, this dissertation research aims at carrying out a compilation of the area's terms and phraseologies in order to design a bilingual (English-Portuguese) glossary to help writers, mainly translators, in their writing. This research is based on the theoretical assumptions of the Communicative Theory of Terminology (CABRÉ, 1999, 2001), the Functionalist Approach to Translation (NORD, 2013) and focuses on Specialized Phraseological Units (BEVILACQUA, 2004; WAQUIL, 2013). For this survey, we used *Corpus Linguistics* (BERBER SARDINHA, 2004) as methodological approach. Our *corpus* can be defined as comparable, since it is consisted of authentic texts written in their original language (and not translation) and is comprised of two *subcorpora*, with authentic texts in Brazilian Portuguese and American English. The texts were collected and compiled based on certain criteria. The *subcorpora* were then submitted to analysis using computational tools. The Sketch Engine software was used to perform a semi-automatic extraction of keywords, which provided potential terms and phraseologies that were organized in order for a manual analysis to be conducted based on the concordance lines. This analysis required that the lists were manually ‘cleaned’, that is, the concordance lines of each potential term in their occurrences were checked so that the most statistically significant terms and phraseologies could be selected in order to compose the template for the bilingual glossary. This template provides the description of each item that will comprise the glossary entries. Eventually, the data discovered is discussed in the light of the theoretical assumptions mentioned before, and the tools used in the research are assessed.

Key words: Terminology, Specialized Phraseological Units, Functional Translation, *Corpus Linguistics*, Sustainability Report.

LISTA DE QUADROS

Quadro 1. Esquema sintético do glossário de Prata (2017)	15
Quadro 2. Resumo das denominações, definições e critérios utilizados pelos principais teóricos da fraseologia especializada.....	36

LISTA DE TABELAS

Tabela 1. Composição do <i>corpus</i> em português.....	52
Tabela 2. Composição do <i>corpus</i> em inglês	53

LISTA DE FIGURAS

Figura 1. Exemplo: página de índice.....	23
Figura 2. Exemplo de páginas de texto	24
Figura 3. Trecho da seção descritiva do escopo do “GRI 303: Water and Effluents”	26
Figura 4. Sumário do indicador “GRI 303: Water and Effluents” e os tópicos que serão abordados.....	27
Figura 5. Parte do conteúdo e requisitos do indicador GRI 303-1, que trata da água como recurso compartilhado	28
Figura 6. Esquema da proposta de Nord: a função do texto é atribuída pelo receptor.....	42
Figura 7. <i>Subcorpora</i> Português.....	54
Figura 8. <i>Subcorpora</i> inglês	54
Figura 9. Tela com visualização parcial do SE após aplicação da ferramenta Keywords/Terms (<i>corpus</i> do português).....	57
Figura 10. Tela do SE após levantamento realizado pela ferramenta Keywords/Terms (<i>corpus</i> do inglês).....	57
Figura 11. Item destacado mostrando exemplo de expressão com alta frequência, porém sem representatividade (ocorrência em apenas um relatório).....	59
Figura 12a. Primeira parte da lista de ocorrências, mostrando o texto do qual foram extraídas.....	60
Figura 12b. Segunda parte da lista da ocorrência de “consultoras Natura”, mostrando o texto do qual foram extraídas.....	60
Figura 13. Recorte da lista bruta das palavras simples do <i>corpus</i> de inglês.....	61
Figura 14. Recorte da lista de palavras compostas do <i>corpus</i> de inglês.....	61
Figura 15. Trecho da lista unificada de termos simples e compostos, em ordem alfabética.....	62
Figura 16. Gráfico gerado pela ferramenta <i>Word Sketch</i> para o termo <i>energy</i>	63
Figura 17. Esquema sintético do comportamento do termo <i>energy</i> , gerado pela ferramenta <i>Word Sketch</i>	63
Figura 18. Visualização da tela após a extração das palavras-chave, mostrando as 20 primeiras palavras-chave simples e compostas do <i>corpus</i> RelSustenta_PT (português).....	65

Figura 19. Visualização da tela após a extração das palavras-chave, mostrando as 20 primeiras palavras-chave simples e compostas do corpus RelSustenta_EN (inglês).....	66
Figura 20. Visualização parcial da lista, com os 20 primeiros itens no programa Excel.....	67
Figura 21. Visualização parcial da tela de palavras compostas, destacando “dados complementares”.....	68
Figura 22. Visualização parcial da janela aberta pelo clique na frequência de "dados complementares".....	69
Figura 23. Visualização parcial da lista das palavras agrupadas, em ordem alfabética, após a limpeza.....	70
Figura 24. Visualização parcial dos itens excluídos conforme os critérios de limpeza estabelecidos.....	71
Figura 25. Visualização parcial da janela contextos de "consumo de água", com 31 ocorrências no corpus de português.....	72
Figura 26. Tela das concordâncias com filtro aplicado para o co-ocorrente verbal.....	73
Figura 27. Tela das concordâncias com filtro aplicado para o co-ocorrente não verbal.....	73
Figura 28. Resumo das colocações de uma palavra-chave com a ferramenta <i>Word Sketch</i>	76
Figura 29. Tela da ferramenta <i>Word Sketch</i> para o termo “ <i>water</i> ”.....	79
Figura 30. Modelo 1.....	82
Figura 31. Modelo 2.....	83

LISTA DE SIGLAS

CLE	Combinatórias Léxicas Especializadas
EN	Inglês
GRI	<i>Global Reporting Initiative</i>
LC	Linguística de <i>Corpus</i>
NE	Núcleo Eventivo
NT	Núcleo Terminológico
OANC	<i>Open American National Corpus</i>
PT	Português
RB	Relatório Brundtland
RS	Relatório de Sustentabilidade
SE	Sketch Engine
TCT	Teoria Comunicativa da Terminologia
TGT	Teoria Geral da Terminologia
TBL	<i>Triple Bottom Line</i>
UFE Eventivas	Unidades Fraseológicas Especializadas Eventivas
UN	<i>United Nations</i>

SUMÁRIO

INTRODUÇÃO	13
1 SUSTENTABILIDADE	18
1.1 Contexto histórico: do desenvolvimento sustentável à sustentabilidade	18
1.2. Relatório de Sustentabilidade: apresentação	21
1.2.1 GRI (<i>Global Reporting Initiative</i>): diretrizes e estrutura para o desenvolvimento do Relatório de Sustentabilidade	24
2 BASES TEÓRICAS	29
2.1. Teoria Comunicativa da Terminologia	30
2.1.1 Texto especializado como repositório de terminologia	33
2.1.2. Fraseologia especializada	35
2.2. Tradução especializada: equivalentes funcionais	39
2.3 Linguística de <i>corpus</i> : breve histórico	42
3 METODOLOGIA	45
3.1 <i>Corpus</i> de estudo	49
3.1.1 Descrição do <i>corpus</i> e critérios de construção dos <i>subcorpora</i>	49
3.2 Ferramentas computacionais utilizadas	51
3.2.1 Procedimentos para a seleção dos candidatos a termo	55
3.2.2 Critérios de exclusão de palavras-chave e listas de palavras-chave	58
4 ANÁLISE DOS DADOS	64
4.1 Análise das listas de palavras-chave	64
4.2 Seleção de termos	72
4.3 Tratamento das entradas do glossário	80
4.3.1 Descrição do verbete: entradas, subentradas, equivalentes, exemplos e remissivas	80
4.3.2 Fraseologia e exemplos	81
4.3.3 Apresentação do modelo	82
5 DISCUSSÃO E CONSIDERAÇÕES FINAIS	84
REFERÊNCIAS	88
ANEXOS	93

INTRODUÇÃO

O presente estudo é o resultado da pesquisa terminológica baseada em *corpora* compostos por textos do tipo Relatório de Sustentabilidade (doravante RS), com vistas à elaboração de uma proposta de glossário bilíngue para fins tradutórios.

A proposta de construção de um glossário dessa área de especialidade se justifica por dois motivos principais. O primeiro deles refere-se ao aumento do número de publicações do tipo Relatório de Sustentabilidade: o Brasil ocupa a terceira posição entre os países onde mais se divulgam esses relatos (G1, 2010). Somado a isso, o número de publicações aumentou no mundo todo entre o final da década de 1990 até 2015¹. Segundo o *Governance & Accountability Institute* (G&A) (2017), essa prática cresceu cerca de 62% entre as maiores empresas do mundo² de 2011 a 2016 (SIEW, 2017). Esses dados mostram que há uma tendência de cada vez mais empresas aderirem à prática de publicar o RS, mesmo que essa divulgação ainda não seja obrigatória no Brasil³, pois a publicação vem se mostrando útil na forma de comunicar os resultados e impactos econômicos, ambientais e sociais das grandes empresas, além de promover uma atitude mais transparente em relação aos impactos críticos e relevantes⁴.

O segundo motivo refere-se a questões econômicas: a divulgação globalizada dos RSs torna mais dinâmica a busca por informações, uma vez que investidores e acionistas se apoiam nesses relatórios para a tomada de decisões. A tradução, portanto, mostra-se uma atividade imprescindível na alavancagem dos negócios no atual cenário brasileiro, resultando em uma demanda potencial de trabalho para tradutores e redatores técnicos:

Elaborar relatórios de sustentabilidade que tratam de mensurar e divulgar os impactos socioambientais causados pelas atividades cotidianas das organizações vem sendo uma prática incorporada por empresas localizadas em diversos países. Fato bastante comum na Europa, na América do Norte e no Japão e que se torna cada vez mais relevante nos mercados emergentes. (CAMPOS et al., 2013, documento on-line).

A crescente disponibilização dessas informações, somada ao desenvolvimento tecnológico, facilitou o acesso às publicações por públicos diversos, e não apenas investidores

¹ Disponível em: <https://singep.org.br/5singep/resultado/669.pdf>. Acesso em: 10 nov. 2018.

² Ranking da Standard&Poor. <https://www.investopedia.com/terms/s/sp500.asp>. Acesso em 19/5/2018.

³ Projeto de Lei n. 289/2012. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/106810>. Acesso em: 31 out. 2018.

⁴ Disponível em: <https://exame.abril.com.br/mundo/qual-e-a-importancia-dos-relatorios-de-sustentabilidade>. Acesso em: 28 mai. 2018.

e especialistas. Além disso, a necessidade de traduções tem aumentado no Brasil desde o início da década devido à projeção do país no mercado externo. Exemplo disso são os eventos internacionais sediados no Brasil, como a Copa de 2014 e as Olimpíadas de 2016, que impulsionaram os serviços no setor (LUQUES, 2010; CRESCIMENTO..., 2012).

Além desses fatores, a escolha do tema desta pesquisa resulta da convergência de dois momentos pessoais: o primeiro, de caráter acadêmico, foi um período em que tive a oportunidade de atuar como bolsista de Iniciação Científica (IC) junto ao Termisul⁵ em projetos que utilizam a metodologia da Linguística de *Corpus*. Nessa ocasião, as tarefas realizadas envolviam a limpeza e a etiquetagem dos textos que compunham as bases textuais do Termisul: a Base Legis (textos da legislação ambiental do Brasil, Alemanha, Argentina, Estados Unidos, França, Paraguai e Uruguai) e a Base Gestamb (textos representativos da área de Gestão Ambiental Empresarial distribuídos em manuais, teses e dissertações). Essas tarefas tinham o objetivo de preparar os textos para análise com as ferramentas do acervo.

O outro momento que me conduziu até o tema da sustentabilidade foi posterior e de caráter profissional. Em 2011 e 2012 atuei como gestora em uma empresa de Comunicação, prestando serviços de consultoria para um cliente do setor florestal. Nesse período, houve contato direto com tópicos relacionados à Sustentabilidade, resultando na sensibilização para esse tema devido ao acompanhamento de todo o processo industrial de fabricação do produto final (celulose) desse cliente. A elaboração do RS é um longo processo, pois há um volumoso levantamento de informações. Ao deparar com essa realidade, diversas vezes me questionei sobre a redação desse longo texto e a sua versão para outros idiomas, e dentre os questionamentos, se haveria algum instrumento para auxiliar o tradutor ou mesmo o redator, tendo em vista a extensão desse documento e a diversidade de informações que ele apresenta.

Ao optar por realizar a pesquisa no âmbito da Sustentabilidade, percebi que o meio acadêmico é desprovido de estudos relacionados à descrição da linguagem dos RSs. Um dos estudos encontrados (PRATA, 2017) investigou expressões traduzidas em RSs, buscando identificar as principais técnicas de tradução com base nos estudos de Mona Baker (1993, 1996). O segundo objetivo desse estudo foi identificar os principais termos e expressões nesses tipos de texto (RSs) em português e seus equivalentes em inglês. Embora um dos propósitos do estudo fosse também construir um glossário bilíngue (português-inglês), foram encontradas diferenças na metodologia, na base teórica e na microestrutura. Em relação à metodologia, o estudo foi realizado com a construção de dois *corpora*: um paralelo e outro

⁵ Termisul é o grupo de pesquisa terminológica da Universidade Federal do Rio Grande do Sul. Site: <http://www.ufrgs.br/termisul/>

comparável. O *corpus* paralelo (composto de RSs em português e suas respectivas traduções para o inglês) foi utilizado para levantar os correspondentes tradutórios, analisando alguns aspectos relativos às técnicas de tradução utilizadas. Já o *corpus* comparável (composto por cerca de 730 mil *tokens* [palavras] no português original e mais de 870 mil *tokens* no de inglês original) serviu apenas para referência, pois foi utilizado para validar os correspondentes. Em relação às bases teóricas adotadas por Prata (2017), o estudo não considerou qualquer dos pressupostos teóricos da Fraseologia Especializada, ignorando as fraseologias especializadas equivalentes, a principal lacuna deixada pela autora em seu estudo. O reconhecimento e a identificação das fraseologias especializadas, como será aprofundado no Capítulo 2, não são tarefas fáceis, pois seus limites são de difícil definição, embora sejam importantes porque trazem informações a respeito do comportamento das unidades terminológicas e seu entorno, estabelecendo maior ou menor grau de fixação. São essas as informações necessárias para que o tradutor não precise criar banco de dados ou glossários próprios a cada novo projeto de tradução. Pela ausência dessa fundamentação teórica em relação a estudos fraseológicos no estudo de Prata (2017), a microestrutura da sua proposta para o glossário apresentou uma diferença bastante considerável em relação à nossa pesquisa. A proposta de glossário foi montada em blocos, conforme mostra a estrutura exemplificada e resumida no Quadro 1 a seguir. A coluna A mostra a organização de palavras-chave e termos em português e a coluna B, os de inglês, os quais a autora considerou como os equivalentes.

Quadro 1. Esquema sintético do glossário de Prata (2017).

A	B
Palavra-chave (português - PT) - frequência (F)	Correspondente da palavra-chave (inglês - EN) - (F)
Termo proveniente da palavra-chave (PT) (F)	Termo correspondente (EN) (F) (retirado do <i>corpus</i> paralelo)

Fonte: adaptada de Prata (2017, p. 43).

Nesse formato de blocos, a consulta fica prejudicada, porque as informações estão separadas, e não agrupadas em um único bloco referente à palavra-chave ou ao termo, exigindo mais esforço do consulente.

Assim, a nossa pesquisa pretende preencher uma parte dessa lacuna, propondo um glossário que apresente, de forma clara, os termos e as fraseologias com seus co-ocorrentes,

isto é, elementos que combinam com as fraseologias de forma recorrente e que possuem certo grau de fixação. Por esses motivos, o trabalho realizado na identificação de termos e fraseologias deve considerar seus contextos de uso, com base nas combinações mais frequentes, entre outros aspectos pragmáticos, que possam ser representados na forma de produtos terminográficos de qualidade e que realmente apoiem o tradutor ou o redator.

Para isto, este texto está dividido da seguinte forma: o Capítulo 1 – Sustentabilidade – inicia com uma apresentação sobre o que vem a ser a sustentabilidade, com um breve histórico da evolução do conceito de desenvolvimento sustentável. Na sequência, apresenta-se a estrutura do RS, com breve descrição das características macrotextuais, ilustrando com alguns exemplos. Após, dedicamos uma subseção para explicar o que é a GRI — sigla para *Global Reporting Initiative* —, uma iniciativa internacional responsável por orientar as normas da publicação dos RSs no mundo.

No Capítulo 2, são explicados os conceitos e definições adotados a partir das teorias que embasam a pesquisa. A Teoria Comunicativa da Terminologia, proposta por Cabré (1999), orienta esta pesquisa em toda sua extensão, desde os conceitos sobre o termo, seguindo com o que se entende por linguagem especializada e texto especializado (CABRÉ, 1993, 1999, 2001; CIAPUSCIO, 1998). O capítulo revisa algumas noções de Terminologia bilíngue sob a perspectiva funcionalista da Tradução (NORD, 1997, 2013). A penúltima seção do capítulo tratará das colocações e fraseologias (BEVILACQUA, 2004; L’HOMME, 2000; WAQUIL, 2013) e a seção final traz um breve histórico da Linguística de *Corpus* (BERBER SARDINHA, 2004).

No Capítulo 3, apresenta-se o detalhamento da metodologia adotada, bem como a descrição completa dos *corpora* utilizados. Para isso, são explicados os critérios de seleção dos textos e a escolha do programa Sketch Engine (doravante SE), da Lexical Computing⁶. Nesse capítulo, são descritas as ferramentas que auxiliaram no levantamento dos dados, como *Keywords/Terms* e *Word Sketch*, com base em parâmetros estatísticos. Além disso, todo o detalhamento dos procedimentos adotados para a seleção dos candidatos a termos é apresentado, explicando os critérios de limpeza e a consequente exclusão de palavras-chave. Ao final, são apresentadas as listas das palavras-chave utilizadas para análise dos contextos de ocorrências.

No Capítulo 4, apresenta-se a análise dos dados obtidos pelo levantamento. Alguns comentários em relação a vantagens e desvantagens do programa também foram incluídos. A

⁶ <https://www.sketchengine.eu>

parte final desse capítulo se dedica a apresentar o modelo de verbete do glossário, com exemplos.

Finalmente, no Capítulo 5 são discutidas questões em relação à aplicação das teorias para a proposta do glossário, finalizando com considerações e sugestões para futuros estudos.

1 SUSTENTABILIDADE

O presente trabalho constitui-se em um estudo acerca da linguagem do Relatório de Sustentabilidade (doravante RS). A prática de publicação dos RSs começou a ser adotada por inúmeras empresas no Brasil no final da década de 1990. Anteriormente a isso, a publicação desses relatórios não figurava no cenário nacional, e muito timidamente no âmbito internacional. Devido, por um lado, ao crescente interesse pelas questões ambientais e ao aumento dos índices de poluição ambiental e, por outro lado, por causa das inovações tecnológicas e da preocupação com a manutenção das gerações futuras no planeta – além do fato de as empresas estarem preocupadas com sua imagem no mercado –, o número de publicações desse tipo por grandes empresas vem crescendo no mundo inteiro (RYDLEWSKI, 2008). Para confirmar esses dados, percebe-se a consolidação de rankings promovidos por grandes publicações, como a revista brasileira Exame⁷ e a canadense Corporate Knights⁸, que classificam as empresas mais sustentáveis de acordo com critérios e metodologia próprios, levando em conta as práticas sustentáveis das empresas participantes do *ranking*. A metodologia da GRI, por sua vez, auxilia na produção de RSs de excelente nível, utilizados cada vez mais como fonte de informação e não apenas como mera estratégia de marketing. Esse crescimento no número de publicações não ocorreu de forma aleatória, mas é decorrente de uma série de eventos, como descreveremos a seguir.

1.1 Contexto histórico: do desenvolvimento sustentável à sustentabilidade

No final dos anos 1960, o empresário italiano Aurelio Peccei, presidente honorário da Fiat, e o cientista escocês Alexander King promoveram um encontro em que seriam tratadas as condições humanas futuras no planeta (PENSAMENTO VERDE, 2014). Esse encontro ficou conhecido como o Clube de Roma, cidade que sediou o evento, ocasião que resultou na redação de um projeto com as bases e princípios sobre as questões de desenvolvimento global. Após quatro anos, representantes de 113 nações, de 250 organizações não governamentais e de organismos da ONU voltaram a se reunir na Suécia, para a Conferência das Nações Unidas sobre o Meio Ambiente Humano, conhecida como a Conferência de Estocolmo (BRASIL, 2012). Essa foi a primeira grande conferência sobre questões de nível

⁷ <https://exame.abril.com.br/guia-exame-de-sustentabilidade/>

⁸ <http://www.corporateknights.com/reports/global-100/>

internacional relacionadas ao ambiente, marcada como um momento decisivo no desenvolvimento de políticas ambientais internacionais (UN, 1987).⁹

A Conferência de Estocolmo lançou as bases das ações ambientais em nível internacional, chamando a atenção para as questões relacionadas à preservação e à melhoria do ambiente. Como resultado desse encontro, foi gerada a Declaração sobre o Meio Ambiente Humano, uma carta de princípios chamada de *Our Common Future*, mas também conhecida como Relatório Brundtland (em homenagem à organizadora Gro Harlem Brundtland). *Our Common Future* é “uma carta de princípios de comportamento e responsabilidades que deveriam nortear as decisões sobre políticas ambientais” (BRASIL, 2012). O conteúdo do Relatório Brundtland (doravante RB) previa uma discrepante incompatibilidade entre o desenvolvimento e os padrões de produção e de consumo, já naquela época. O RB já alertava, de forma contundente, que, se as práticas nos meios de produção e de consumo, de empresas e dos próprios indivíduos, não fossem aprimoradas e ajustadas para modos de produção e consumo mais sustentáveis, haveria chances de a economia, o ambiente e as sociedades entrarem em colapso. Esse relatório definiu, pela primeira vez, o conceito de desenvolvimento sustentável, recomendando que houvesse uma conciliação entre o crescimento e o desenvolvimento e os aspectos ambientais e sociais.

A ONU, desde então, realiza eventos importantes que introduzem na agenda mundial uma série de orientações para os países membros. Exemplo disso são a Eco-92 (sediada no Rio de Janeiro em 1992), a Conferência de Quioto (em 1997), a Conferência de Joanesburgo (em 2002) e a Rio+20. Dessa forma, os países são convocados a cooperarem na busca por soluções para os problemas relacionados ao meio ambiente, entre outros assuntos da temática do desenvolvimento sustentável.

Nas circunstâncias da publicação do RB pela ONU, havia uma pressão, principalmente sobre os países desenvolvidos, no sentido de atuarem na proposição de soluções para preservar o meio ambiente ao “definir compromissos mais rígidos para a redução da emissão de gases de efeito estufa, principal causador do aquecimento global” (BRASIL, 2012)¹⁰. O processo de crescimento econômico pressupõe uma relação eficiente entre a produção de bens e a prestação de serviços e o consumo. Significa que aqueles princípios e diretrizes constantes no RB passaram a nortear, resumidamente, a conduta a respeito dessas questões ao redor do mundo.

⁹ No original: “It was the UN's first major conference on international environmental issues, and marked a turning point in the development of international environmental politics.” Todas as traduções neste trabalho são de autoria própria, salvo quando especificado.

¹⁰ <http://www.brasil.gov.br/noticias/meio-ambiente/2012/01/acordos-globais>. Acesso em: 5 out. 2018.

No Brasil, a Eco-92 forneceu alguns princípios e serviu como incentivo para a criação de ferramentas, de maneira que as empresas passassem a incorporar a sustentabilidade na sua gestão (MARIA, 2012). De certa forma, foi um estímulo para que se discutissem alternativas para incluir a sustentabilidade nos negócios. Em 1997, a metodologia do Ibase (Instituto Brasileiro de Análises Sociais e Econômicas), criada pelo sociólogo Herbert de Souza, partia de vários pontos de análise, funcionando como uma metodologia-piloto do que hoje se chama sustentabilidade. Com essas bases e ferramentas, algumas empresas foram encorajadas a começar esse tipo de relato de seus impactos, ainda que de forma bem tímida (MARIA, 2012).

Já no contexto internacional, a prática de publicar os RSs partiu de uma perspectiva econômica. Em Boston, nos Estados Unidos, a busca por novos caminhos que incorporassem a questão da sustentabilidade nos negócios partiu da iniciativa de alguns ativistas ambientais, que atuavam também na área de finanças, em meio ao contexto de acidentes com empresas petrolíferas (MARIA, 2012). O objetivo desse grupo era analisar, além dos resultados econômicos, também as questões ambientais e sociais. A base que havia na época eram os relatórios financeiros. “Assim, surgiu a ideia de inserir as outras dimensões nesse relatório, para que a empresa pudesse então incorporar na sua gestão as questões de sustentabilidade” (MARIA, 2012). Nesse sentido, algumas corporações, principalmente as grandes e de maior impacto, deram início ao desenvolvimento de metodologias próprias para esses relatórios. Nessa mesma época, a GRI começava a surgir, sendo efetivamente estabelecida em 1997, inaugurando sua sede própria em 2002, na cidade de Amsterdam.

A origem dos conceitos de sustentabilidade e desenvolvimento sustentável está diretamente relacionada com o documento *Our Common Future*. O termo **desenvolvimento sustentável**, utilizado pela primeira vez em 1987 no Relatório Brundtland (UN, 1987), foi definido da seguinte forma: “o desenvolvimento que procura satisfazer as necessidades da geração atual, sem comprometer a capacidade das gerações futuras de satisfazerem as suas próprias necessidades.”¹¹ Ou seja, significa promover condições para que as pessoas, no presente momento e no futuro, atinjam “um nível satisfatório de desenvolvimento social e econômico e de realização humana e cultural, fazendo, ao mesmo tempo, um uso razoável dos recursos da terra e preservando as espécies e os *habitats* naturais” (UN, 1987).

¹¹ No original: “Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.” <http://www.un-documents.net/our-common-future.pdf>. Acesso em 20 mar. 2017.

O termo “desenvolvimento sustentável” sofreu algumas interpretações ao longo das últimas décadas, passando a ser mais conhecido pelo nome de “sustentabilidade”. A noção de **sustentabilidade** igualmente sugere que, a partir da percepção da capacidade limitada dos recursos naturais e de seu gradativo esgotamento, é necessária a intervenção do ser humano em processos que preservem o ambiente, de maneira que não coloque em risco os recursos naturais das futuras gerações. Esse conceito é extremamente complexo, por contemplar um conjunto de fatores interdependentes, que devem ser capazes de integrar questões sociais, energéticas, econômicas e ambientais.

Mais recentemente, o termo sustentabilidade se popularizou e passou a ser usado de forma bastante generalizada, principalmente nas áreas de Administração e Marketing, como estratégia de comunicação para agregar valor a um produto, muitas vezes sem vínculo com o conceito original.

Com o surgimento da GRI, aliado às iniciativas das Nações Unidas de promover diretrizes e ações globais para o desenvolvimento sustentável, cada vez mais empresas passaram a aderir à prática de relatar as atividades relacionadas a questões ambientais e sociais na gestão dos negócios.

1.2. Relatório de Sustentabilidade: apresentação

O Relatório de Sustentabilidade (RS), contrariamente a essa onda de popularização do conceito de sustentabilidade, é um documento bastante complexo, porque engloba dados de todos os setores da empresa, com a finalidade de divulgar os impactos provocados pelas atividades das empresas e organizações. Os aspectos considerados importantes para o desempenho das empresas têm como base o *Triple Bottom Line* (TBL), também conhecido como “tripé da sustentabilidade”, pois se refere a **capital humano, capital natural e capital econômico**, necessários para o equilíbrio das organizações (BARBOSA, 2007).

O RS é considerado uma ferramenta importante porque auxilia a empresa ou organização a aprimorar seus processos e comunicar seu desempenho nas esferas econômica, ambiental, social e de governança, definindo objetivos e administrando as mudanças e os impactos de maneira mais eficaz.

Neste trabalho, optamos pela definição da GRI para o termo Relatório de Sustentabilidade:

Um relatório de sustentabilidade é um relato, publicado por uma empresa, sobre os impactos econômicos, sociais e ambientais causados por suas atividades diárias. Um relatório de sustentabilidade também apresenta os valores de uma organização e seu modelo de governança, demonstrando a relação entre sua estratégia e seu compromisso com a economia sustentável global (GRI, 2018)¹².

Em resumo, o RS é uma ferramenta que comunica às partes interessadas (*stakeholders*) os impactos das organizações nos âmbitos econômico, social e ambiental, utilizando uma metodologia que auxilia a transformar temas abstratos em medidas tangíveis e concretas (GRI-G4, 2015). Entende-se por partes interessadas ou *stakeholders* os públicos que se relacionam com a empresa, seja o conselho de administração, funcionários, acionistas, fornecedores, clientes e até mesmo os municípios onde a empresa atua. Em resumo, o RS pode ser considerado uma plataforma voltada para comunicar o desempenho de sustentabilidade e seus impactos, sejam eles positivos ou negativos.

A respeito da macroestrutura do documento, ou seja, da organização interna do texto, o RS é composto por elementos pré-textuais, textuais e pós-textuais. Os elementos pré-textuais se constituem no conjunto de seções anteriores ao texto principal, contendo índice ou sumário, mensagem dos fundadores ou presidente e uma apresentação sobre governança corporativa. A parte dos elementos textuais é sempre a mais longa, contendo textos de caráter descritivo e informativo. A parte final, pós-textual, contém um índice remissivo e informações sobre a produção do relatório (seção chamada de Expediente), sobre a empresa, alguns com mapa da localização da matriz e das sedes. Os RSs variam na dimensão (largura e altura dos formatos impressos, por exemplo), no uso de recursos semióticos (como projeto gráfico e visual, bem como a utilização de gráficos coloridos, figuras e fotos, que servem de apoio à leitura, dando leveza ao visual). Os RSs também variam em relação à extensão: alguns são mais enxutos, outros mais detalhados. Como não são medidos por número de palavras, mas sim pelo conteúdo que apresentam (em relação aos critérios da GRI que são atendidos), esse fator não será considerado nesta pesquisa.

As Figuras 1 e 2 mostram alguns exemplos dos RSs analisados nesta pesquisa.

¹² No original: “A sustainability report is a report published by a company or organization about the economic, environmental and social impacts caused by its everyday activities. A sustainability report also presents the organization's values and governance model, and demonstrates the link between its strategy and its commitment to a sustainable global economy.”. <https://www.globalreporting.org/information/sustainability-reporting/Pages/default.aspx>. Acesso em 27 out. 2018

Figura 1. Exemplo: página de índice.

TABLE OF CONTENTS	
Message from Brent Saunders, Chairman, President and CEO	3
Introduction	4
Sustainability Program and Reporting Scope, Limitations, and Processes	5
Sustainability Structure and Reporting Relationships	6
Allergan and the Precautionary Principle	6
2015–2020 Allergan Strategic Performance Goals and Results	7
Allergan’s 2020 Goals	7
EHS Regulatory Compliance	7
Management Systems	8
Risk Assessment Process	8
Corrective Action Process	8
Allergan’s EHS Audit Program	9
Sustainable Supply Chains	9
Impacts, Materiality, Risks, and Opportunities	10
Innovation	10
Stakeholder Engagement and Collaborations	11
Managing Climate Change	12
Energy Management	13
Waste Management and Recycling	13
Water Management	14
Wastewater Discharge Indicators	15
Air Emissions Indicators	15
Use of Mercurial Preservatives	15
Biodiversity	16
Keeping Employees and Contractors Safe	16
Critical Safety Rules	17
Diversity, Training, and Development	18
Bioethics	19
Corporate Statement on Animal Testing	19
Philanthropy and Cityzenship	20
Grant Limitations	22
Access to Medicines - Our Social Contact	24
The Four Principles	24
“We Care” Grants	25
Product Donations	25
Cost Burden and Health Outcomes	26
Leadership	27
AIF	27
Recognizing Excellence	27
Conclusion	27
Sustainability Performance Summary Table	28
Independent Assurance Statement to Allergan plc	29
Appendix A	30

Fonte: RS da Allergan (2016).

Figura 2. Exemplo de páginas de texto.

64

Relatório
Votorantim
Metais 2015

65

Relatório
Votorantim
Metais 2015

Ambiental

A atividade de mineração e metalurgia provoca impactos importantes no meio ambiente, em seus mais variados aspectos. As operações de extração e beneficiamento de minérios exigem um alto consumo de água e energia, geram grande quantidade de resíduos e, em processos produtivos como o do alumínio, também respondem por uma emissão importante de gases de efeito estufa (GEE).

Essa condição exige um sistema de gestão ambiental robusto e estruturado, que integre a análise permanente de riscos, o monitoramento contínuo das operações, a capacitação profissional e uma série de outros controles que garantam a segurança das atividades minerometalúrgicas.

Na Votorantim Metais, o sistema de gestão ambiental está embasado no PDCA, que tem como premissa a melhoria contínua com base em estudos de benchmark e revisões das práticas e diretrizes. O sistema contempla as normas e legislações ambientais vigentes e abrange os temas mais comuns nas indústrias de mineração e metalurgia, divididos em nove aspectos, a saber: gestão de barragens e depósitos, gestão de resíduos, gestão de águas e efluentes, gestão das emissões, transporte das cargas perigosas, novos projetos, recursos naturais, gestão dos serviços ecossistêmicos e do descomissionamento.

Esses padrões são aplicados em todas as unidades operacionais e consideram o perfil de risco de cada uma delas, de forma que os aspectos mais críticos sejam acompanhados por rotinas preestabelecidas que assegurem a continuidade das operações dentro de parâmetros adequados.

Ao longo dos anos, esse sistema se consolidou como um diferencial relevante para a gestão ambiental da empresa, estabelecendo estruturas e procedimentos próprios. A cada dois anos, ele é revisado e auditado, de modo que possa ser permanentemente atualizado e aprimorado.

Um aspecto importante nesse sentido foi a revisão das metas ambientais previstas no planejamento estratégico, validada em 2015 pela alta liderança da empresa e com prazo de execução estipulado para o ano de 2025. Dessa forma, a Votorantim Metais se compromete a reduzir em 50% a geração específica de resíduos minerometalúrgicos ou o impacto causado por eles, aumentar a eficiência hídrica para 75% e diminuir as emissões de GEE em 5%, todos esses índices em relação aos de 2014.

Trabalho conjunto

Nos últimos cinco anos, o trabalho conjunto das áreas de tecnologia, produção e meio ambiente tem sido fundamental para o desenvolvimento de soluções que minimizam o impacto das atividades da empresa no meio ambiente. Uma busca antiga e permanente dessa parceria é o reaproveitamento de resíduos destinados para as barragens, a fim de encontrar novos usos comerciais para eles e, com isso, evitar a construção de novas estruturas de contenção.

Outros exemplos podem ser encontrados na gestão de recursos hídricos, nas ações para reduzir o consumo de energia e nos investimentos para diminuir o volume de emissões. E o caso de iniciativa implantada na unidade de Três Marias, que permitiu a redução da captação da água do rio São Francisco em 36%. Esse e outros exemplos podem ser encontrados nas próximas páginas.



Investimentos ambientais

Em 2015, a Votorantim Metais teve um aumento significativo nos investimentos ambientais em relação ao ano anterior (veja quadro). Contribuíram para o crescimento os custos com projetos de alinhamento da barragem de rejeitos em Niquelândia, a mudança na fase do projeto Trend Ambrosia em Morro Agudo, além das estações de tratamento de efluentes e o licenciamento de áreas para lava em Miral. [ver mais](#)

Em Cajamarquilha, foram realizadas diversas ações que também contribuíram para o aumento do valor, como a melhoria do sistema de monitoramento e controle de emissões atmosféricas e da qualidade do ar, bem como as remediações de passivos ambientais e os custos com sistemas de tratamento de efluentes.

Em relação aos valores de OPEX (despesas), houve aumento nas despesas de gestão ambiental em Alumínio, resultado do aumento do volume de resíduos que foram enviados para coprocessamento. Em Morro Agudo, houve queda nas despesas de tratamento de efluentes líquidos devido ao projeto de segregação das águas da mina e redução de gastos em Miral, Poços de Caldas, Niquelândia e Milpo.

Total de investimentos e despesas ambientais [ver mais](#)

CAPEX (investimentos) – R\$	2014	2015
Educação ambiental	-	49.000,00
Gestão ambiental	679.903,03	3.254.239,05
Prevenção ambiental	17.080.941,10	-
Preservação, reflorestamento e biodiversidade	226.846,10	511.695,01
Tratamento de emissões atmosféricas	17.545.719,00	10.811.427,00
Tratamento de resíduos	7.146.292,46	89.110.441,11
Tratamento de efluentes líquidos	27.365.292,39	50.317.705,03
DHO – área ambiental	95.000,00	36.500,00
Descomissionamento/remediação de áreas contaminadas (solo e água)	1.135.771,82	1.803.730,32
Seguros ambientais	-	-
Outras despesas ambientais	1.245.247,00	13.032.126,28
TOTAL	72.521.012,90	168.926.863,80

OPEX (despesas)	2014	2015
Educação ambiental	584.011,99	635.716,88
Gestão ambiental	14.200.196,35	8.632.538,74
Prevenção ambiental	2.857.537,46	64.501,06
Preservação, reflorestamento e biodiversidade	10.704.972,72	6.849.508,34
Tratamento de emissões atmosféricas	46.555.020,29	39.727.688,34
Tratamento de resíduos	25.640.285,35	43.834.337,44
Tratamento de efluentes líquidos	36.715.788,63	47.590.080,64
DHO – área ambiental	6.214.383,73	5.237.964,70
Descomissionamento/remediação de áreas contaminadas (solo e água)	8.014.382,18	12.261.108,20
Seguros ambientais	-	-
Outras despesas ambientais	1.905.219,75	1.419.337,94
TOTAL	153.391.798,46	166.252.782,05

Fonte: RS da Votorantim Metais.

No que diz respeito à microestrutura, ou seja, referente a conceitos utilizados, padrões de linguagem em relação ao léxico e a algumas estruturas de natureza morfossintática, os RSs apresentam diversos aspectos em comum. Os RSs escolhidos para esta pesquisa foram produzidos e redigidos seguindo as diretrizes da GRI; portanto, espera-se encontrar semelhanças na terminologia, reservando-se os casos que são próprios de cada idioma pesquisado aqui. Os detalhes em relação aos critérios de seleção e resultados obtidos serão fornecidos nos próximos Capítulos 3 — Metodologia e 4 — Análise.

1.2.1 GRI (Global Reporting Initiative): diretrizes e estrutura para o desenvolvimento de Relatórios de Sustentabilidade

Como visto até aqui, a prática de elaborar e publicar o Relatório de Sustentabilidade vem se consolidando cada vez mais entre as empresas. Sendo um documento tão complexo, faremos uma breve descrição da sua composição, dos objetivos e de especificidades no que diz respeito à forma e ao modo como se produzem esses documentos.

Um relatório de sustentabilidade divulga os impactos – sejam eles positivos ou negativos – de uma organização em relação ao ambiente, à sociedade e à economia. Ao fazê-lo, torna questões abstratas em tangíveis e concretas, portanto auxiliando na compreensão e no gerenciamento dos efeitos do desenvolvimento da sustentabilidade nas estratégias e nas atividades das empresas (GRI, 2018)¹³.

Para tornar concretas essas diversas dimensões abstratas, o RS é construído a partir de parâmetros estabelecidos por entidades globais habilitadas a medir tais impactos. Existem poucas instituições responsáveis por criar metodologias. No Brasil, o instituto Ethos¹⁴ possui um sistema de indicadores, apoiado em normas ISO e alinhado com as normas da GRI. Porém, os indicadores da GRI são os mais utilizados, pois medem o desempenho sustentável de empresas de diversos setores no mundo inteiro. Assim, o modelo mais completo é o da *Global Reporting Initiative* (GRI), pioneira no setor.

A *Global Reporting Initiative* (GRI) é uma instituição global independente e sem fins lucrativos, criada em 1997 para estabelecer uma estrutura que fosse mundialmente aceita com vistas a medir o desempenho sustentável de empresas, repartições públicas, ONGs e outras organizações. Atualmente, mais de 60 países seguem suas diretrizes de desenvolvimento. A GRI conquistou um espaço e ganhou credibilidade em âmbito internacional, grande parte devido ao fato de que muitos acionistas procuram as informações contidas nesses relatórios em busca de consenso entre empresas, sociedade civil, trabalhadores, economistas, acadêmicos e governos em suas deliberações (BITARELLO, 2012). Por esse motivo é considerada hoje a maior referência nesse campo.

Inicialmente, os relatórios eram desenvolvidos de maneira experimental, mas essa prática evoluiu no final dos anos 1990 para o padrão utilizado nas maiores empresas do mundo hoje. De acordo com a GRI, das 250 maiores empresas do mundo, 93% reportam os impactos de suas atividades por meio do RS (GRI, 2018)¹⁵:

No Brasil, 88 das 100 maiores empresas o fazem e estima-se que cerca de 6 mil empresas em todo o mundo o façam também. Os investidores exigem cada vez mais informações sobre o desempenho sustentável das empresas. E as empresas de pesquisa de mercado, como a Bloomberg, oferecem, hoje,

¹³ No original: “A sustainability report conveys disclosures on an organization’s impacts – be they positive or negative – on the environment, society and the economy. In doing so, sustainability reporting makes abstract issues tangible and concrete, thereby assisting in understanding and managing the effects of sustainability developments on the organization’s activities and strategy.”

<https://www.globalreporting.org/information/sustainability-reporting/Pages/default.aspx>. Acesso em 27 out. 2018.

¹⁴ Disponível em: <https://www.ethos.org.br/conteudo/indicadores/#.W9WcZWZ3GuU>. Acesso em 27 out. 2018.

¹⁵ Disponível em: <https://www.globalreporting.org/information/about-gri/Pages/default.aspx>. Acesso em: 02/11/2018.

dados sobre sustentabilidade para mais de 4 mil interessados em seus 300 mil terminais. Os reguladores também estão se tornando mais ativos. Alguns governos, como o da França, tornaram os relatórios de sustentabilidade obrigatórios, e bolsas de valores também se mobilizaram (BITARELLO, 2012).

O modelo de RS da GRI abrange importantes itens sobre o desempenho das empresas, com base no TBL. Para exemplificar, cada uma das dimensões (social, ambiental, econômica e de governança) se divide em subtemas. A dimensão ambiental, por exemplo, está subdividida em energia, água, biodiversidade, emissões, efluentes e resíduos, entre outros. Dentro de cada um desses itens, as empresas devem responder a temas específicos. Por exemplo, no item “água”, a empresa deve responder sobre o uso ou consumo de água e como a empresa gerencia esses impactos. A Figura 3 a seguir mostra um quadro referente ao escopo do item GRI 303: Água e Efluentes¹⁶.

Figura 3. Trecho da seção descritiva do escopo do “GRI 303: Water and Effluents”.

About this Standard	
Responsibility	This Standard is issued by the Global Sustainability Standards Board (GSSB) . Any feedback on the GRI Standards can be submitted to standards@globalreporting.org for the consideration of the GSSB.
Scope	<i>GRI 303: Water and Effluents</i> sets out reporting requirements on the topic of water and effluents. This Standard can be used by an organization of any size, type, sector or geographic location that wants to report on its impacts related to this topic.
Normative references	This Standard is to be used together with the most recent versions of the following documents: GRI 101: Foundation GRI 103: Management Approach GRI Standards Glossary In the text of this Standard, terms defined in the Glossary are <u>underlined</u> .
Effective date	This Standard is effective for reports or other materials published on or after 1 January 2021. Earlier adoption is encouraged.

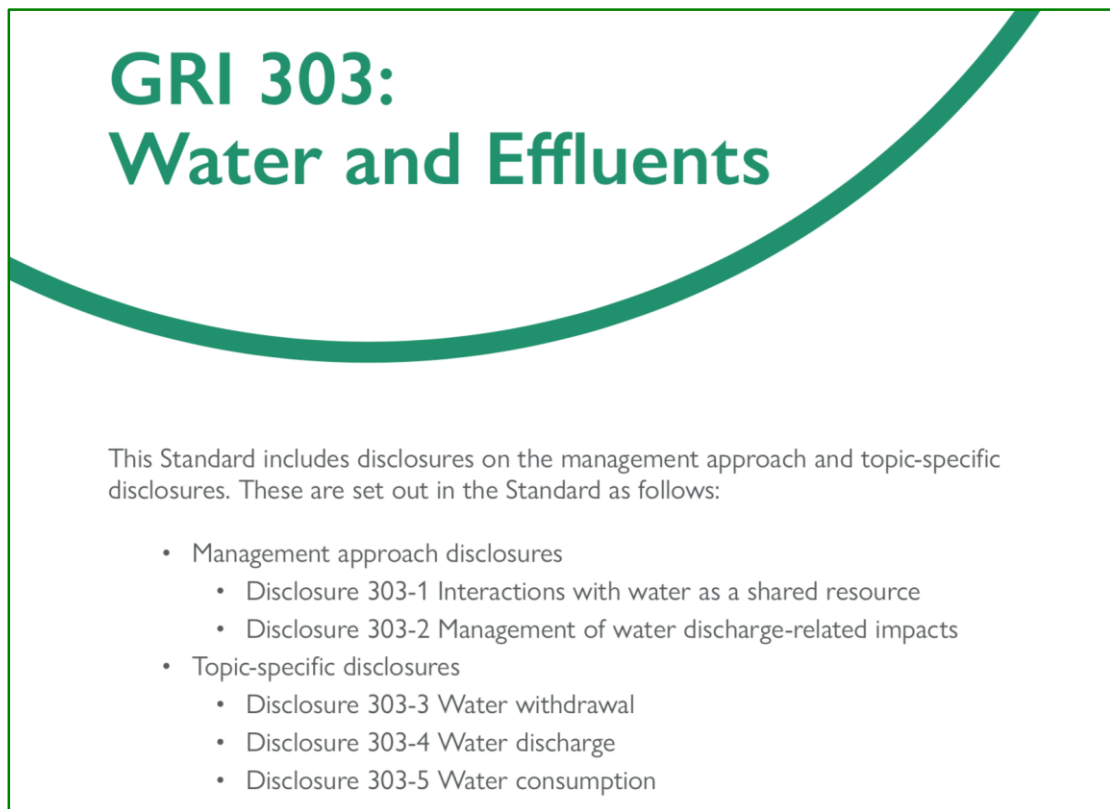
Fonte: GRI Standards (2018).

Assim, cada um desses itens (por exemplo, GRI 303 da Figura 3) é preenchido em forma de tabelas e quadros, de maneira que, ao final, se possa descrever o que foi previsto e o que foi realizado, interpretando e relatando as informações e os dados na forma de um texto, que é o RS propriamente.

¹⁶ O documento completo deste item está disponível no link:
<https://www.globalreporting.org/standards/media/1909/gri-303-water-and-effluents-2018.pdf> .

A Figura 4 apresenta um trecho dessa estrutura de subitens, divididos por temáticas. Assim, no item GRI 303, estão incluídos os tópicos sobre a divulgação de gerenciamento e tópicos específicos da água, cada um mostrando os subitens (Disclosure 303-1, Disclosure 303-2, e assim por diante).

Figura 4. Sumário do indicador “GRI 303: Water and Effluents” e os tópicos que serão abordados.



Fonte: GRI Standards (2018).

A Figura 5, a seguir, apresenta um trecho de como o subitem se divide novamente em uma subestrutura para chegar nos itens a serem avaliados, apresentando os requisitos para o reporte.

Figura 5. Parte do conteúdo e requisitos do indicador GRI 303-1, que trata da água como recurso compartilhado.

Disclosure 303-1
Interactions with water as a shared resource

Reporting requirements

The reporting organization shall report the following information:

- a. A description of how the organization interacts with water, including how and where water is withdrawn, consumed, and discharged, and the water-related impacts caused or contributed to, or directly linked to the organization's activities, products or services by a business relationship (e.g., impacts caused by runoff).
- b. A description of the approach used to identify water-related impacts, including the scope of assessments, their timeframe, and any tools or methodologies used.
- c. A description of how water-related impacts are addressed, including how the organization works with stakeholders to steward water as a shared resource, and how it engages with suppliers or customers with significant water-related impacts.
- d. An explanation of the process for setting any water-related goals and targets that are part of the organization's management approach, and how they relate to public policy and the local context of each area with water stress.

Reporting recommendations

1.2 The reporting organization should report the following additional information:

- 1.2.1 An overview of water use across the organization's value chain;
- 1.2.2 A list of specific catchments where the organization causes significant water-related impacts.

Fonte: GRI Standards (2018).

Assim, considerando os objetivos deste trabalho e o fato de que os RSs se configuram como textos especializados, nos quais o discurso se organiza em torno da sua terminologia prototípica, passaremos ao Capítulo 2, que apresenta as bases teóricas desta pesquisa.

2 BASES TEÓRICAS

O objetivo deste estudo, que é a construção de uma proposta de glossário bilíngue com vistas a auxiliar tradutores e redatores, alinha-se aos princípios teóricos de tendência comunicativa, nos quais nos embasamos para a realização desta pesquisa. Neste capítulo, faremos alguns apontamentos sobre essas teorias e apresentaremos os conceitos adotados para a realização deste estudo.

Partimos da Teoria Comunicativa da Terminologia (TCT), introduzida por Cabré em contraposição à Teoria Geral da Terminologia (TGT), estabelecida desde os anos 1930. Wüster é considerado o principal expoente da TGT, teoria que se tornara insuficiente para atender às necessidades atuais, porque seu caráter é prescritivo, reducionista e idealista, não levando em consideração as variações, naturais à comunicação (ver CABRÉ, 2001, p. 19; a respeito das críticas ao caráter reducionista da TGT, ver também KRIEGER, 2001, p. 50). Novas tecnologias de informação e comunicação passaram a ser cada vez mais utilizadas em diversos processos de trabalho, especialmente no âmbito técnico-científico. Nesse mesmo cenário, outros dois aspectos contribuíram para o estabelecimento de críticas à metodologia vigente: as necessidades terminológicas se tornaram diversificadas, somadas a “uma tendência de adequar as propostas de trabalho e suas aplicações às características de cada aplicação, contexto e situação [...]” (CABRÉ, 2001, p. 20).

Nesse contexto, a TCT passou a nortear diversos estudos e abordagens na década de 1990 em diante. As análises realizadas com os dados obtidos nesta pesquisa alinham-se a esses pressupostos, e da mesma forma as teorias sobre fraseologia e tradução às quais recorreremos. A seguir, serão apresentadas as principais propostas da Teoria Comunicativa da Terminologia (CABRÉ 1999; 2001), considerando o texto especializado tanto como repositório de terminologia (CIAPUSCIO, 1998; CABRÉ, 1999) quanto um ambiente propício à identificação de fraseologias especializadas (BEVILACQUA, 1999, 2004; WAQUIL, 2013). Ao final do capítulo, serão abordadas algumas noções da terminologia bilíngue, sob a perspectiva funcionalista da Tradução, amplamente defendida por Nord (2013), e algumas noções a respeito da Linguística de *Corpus*, que será detalhada no Capítulo 3.

2.1. Teoria Comunicativa da Terminologia

O surgimento de uma proposta de perspectiva comunicativa e mais descritiva do que prescritiva para a Terminologia tem relação direta com o que a linguista Maria Teresa Cabré relata, em sua obra de 1993, a respeito do período pós-industrial. O desenvolvimento da economia de mercado no período de urbanização das sociedades dessa época pós-industrial baseou-se no incentivo à produção e ao consumo. Esse sistema rompeu com a concepção da unidade familiar, uma vez que a concentração da população nos grandes centros urbanos favorecia o individualismo e a busca por bens materiais. Se por um lado essa realidade, assim colocada, parece um tanto negativa e pessimista, por outro confirma que esse novo paradigma contribuiu positivamente para o avanço das ciências e da tecnologia, na mesma medida em que os estudos das disciplinas das mais diversas áreas do conhecimento se desenvolveram exponencialmente. Assim, esse período quebrou paradigmas de maneira significativa: a primeira refere-se à “tecnificação” da sociedade, como salienta Cabré (1993), e a segunda é que a informação passou a ter alto valor para essa nova sociedade pós-industrial. Exemplo disso é o próprio Relatório de Sustentabilidade, um reflexo das novas tendências e fruto do crescimento industrial aliado a uma consciência ecológica. Essas mudanças, incontestavelmente, afetam o modo como se transmite o conhecimento, bem como interferem na linguagem e nas formas de comunicação de modo geral, desencadeando mudanças no comportamento organizacional, nas novas profissões relacionadas à linguagem e, conseqüentemente, nos novos produtos linguísticos.

Para atender a essa nova realidade, despontava uma nova proposta teórica terminológica, a que se denomina Teoria Comunicativa da Terminologia, contrapondo o caráter normativo e que expressava “um apagamento dos aspectos comunicativos e pragmáticos” (KRIEGER, 2001, p. 50). Enquanto disciplina de caráter interdisciplinar e transdisciplinar, a TCT se constitui em um modelo mais aberto, capaz de descrever as unidades terminológicas em toda a sua complexidade. Além disso, fundamenta-se em dois pilares: a perspectiva linguística (a partir da qual se analisam as unidades) e o texto (que é o ambiente natural de ocorrência dessas unidades especializadas em uma comunicação natural) (CABRÉ, 2001, p. 22).

Em consonância com a visão da TCT, entendemos que a Terminologia serve a diversos profissionais e a diferentes finalidades, cumprindo a função de representar e comunicar o conhecimento especializado (CABRÉ, 2001). Destaca-se o pressuposto básico da TCT, “[...] que considera a linguagem especializada como uma **expressão da língua comum**

usada em contexto de especialidade, compartilhando de todas as propriedades da língua natural ao desempenhar as funções comunicativas próprias da área temática” (BEVILACQUA; REUILLARD, 2013, p. 71, grifo nosso). Significa que o léxico especializado não é necessariamente diferente do léxico comum, mas a condição de termo e de especialidade é ativada por um contexto e situação adequados, processo que Cabré chama de **ativação pragmática** (CABRÉ, 2001). Krieger, Santiago e Cabré (2013) resumem com bastante clareza essa ideia:

É ainda interessante observar que Cabré fala de um termo “ativado”. Esse movimento ativador que opera a passagem de um sentido geral para um específico de um campo de conhecimento, ou mesmo do sentido de uma área para outra, está relacionado aos contextos sociocomunicativos. É aí, portanto, que se atualizam as condições de ativação das comunicações especializadas de distintas áreas de conhecimento científico, técnico e tecnológico. (KRIEGER, SANTIAGO, CABRÉ, 2013, p. 332)

Os principais fundamentos da TCT podem ser descritos da seguinte forma, conforme Bevilacqua (2004):

- A terminologia é interdisciplinar (porque se compõe pela soma de uma teoria da linguagem, uma teoria do conhecimento e uma teoria da comunicação) e, ao mesmo tempo, transdisciplinar (pode ser aplicada a vários campos do conhecimento, isto é, as unidades podem ser analisadas sob cada uma dessas perspectivas);
- O objeto de estudo principal são as unidades terminológicas, mas seus estudos abrangem outras unidades transmissoras de conhecimento especializado;
- Considera os textos (ou discursos) especializados como base da comunicação especializada, ou seja, são parte da língua natural mesmo incluindo unidades de outros sistemas simbólicos (por exemplo, fórmulas);
- As unidades terminológicas compartilham das mesmas propriedades da língua (características fonológicas, morfológicas, sintáticas e semânticas), porém, as características pragmáticas são consideradas de acordo com o uso dos termos em âmbitos temáticos ou especializados;
- Os termos e unidades terminológicas são unidades léxicas que adquirem valor especializado (não formam um léxico à parte) pelo uso em contextos e situações específicas;

- A ativação pragmática leva em conta alguns aspectos: temática, tipo de texto, interlocutores, situação comunicativa e função do texto;
- Portanto, as unidades terminológicas não pertencem a um âmbito, mas são utilizadas neles com valor específico (BEVILACQUA, 2004).

Levando em consideração as características mencionadas, tomamos emprestado de Cabré (2001) algumas considerações a respeito do que são *termos* para a TCT:

[...] unidades léxicas especializadas, que aparecem na comunicação natural sobre determinados temas. Essas unidades formam parte do léxico da gramática do falante (usuário), levando em consideração que ele (o falante/usuário) reúne ao mesmo tempo uma condição de falante e de especialista ou profissional em um dos temas específicos. (CABRÉ, 2001, p. 20, destaques nossos)¹⁷.

Significa que, ao mesmo tempo em que um falante utiliza a língua (suas unidades léxicas) para se comunicar em situações comuns, também as utiliza em situações de especialidade na condição de “falante-especialista” ou profissional, nas quais ele domina um tema específico. Cabré explica que os termos:

São *unidades léxicas, ativadas singularmente* por suas condições pragmáticas de adequação a um tipo de comunicação. Compõem-se de forma ou denominação e significado ou conteúdo e possuem valor referencial. (CABRÉ, 2001, p. 24, grifo no original¹⁸).

Entende-se, portanto, que as condições pragmáticas, isto é, temática, usuários e situações de comunicação, é que efetivamente estabelecem a diferença entre os termos e as palavras. É necessário considerar o contexto em que ocorrem tais unidades léxicas de valor terminológico, pois os termos não **pertencem** a um âmbito, mas sim são **utilizados** nesses âmbitos especializados.

A linguagem, de acordo com Cabré (1993), é um sistema complexo e heterogêneo composto por vários subsistemas que se inter-relacionam de maneira complexa em vários níveis: fonológico, morfológico, léxico, sintático e pragmático ou discursivo. Tal lista não é

¹⁷ No original: “[...] *los términos, entendidos como unidades léxicas especializadas, que aparecen en la comunicación natural sobre determinados temas. Estas unidades forman parte del léxico de la gramática de un hablante, teniendo en cuenta que éste reúne al mismo tiempo su condición de hablante y la de profesional o especialista en uno de los temas*”.

¹⁸ No original: “*Los términos son unidades léxicas, activadas singularmente por sus condiciones pragmáticas de adecuación a un tipo de comunicación. Se componen de forma o denominación y significado o contenido y tienen valor referencia.*”

exaustiva, uma vez que existem unidades que participam em mais de um nível ao mesmo tempo.

A linguagem especializada, por sua vez, configura-se como uma sublinguagem ou subconjunto da linguagem, e é caracterizada por aspectos pragmáticos, como mencionado anteriormente: temática, usuários e situações de comunicação. A temática especializada se refere ao conhecimento que tenha sido objeto de uma aprendizagem especializada, portanto não faz parte do conhecimento geral da língua. Assim, a variável *usuários* corresponde a esse conjunto de indivíduos que possuem o conhecimento especializado adquirido por aprendizagem. Já as situações comunicativas também condicionam o caráter especializado, devido ao fato de que a comunicação é formal e regulada por certos critérios técnicos ou científicos. Pode-se dizer, portanto, que esses subconjuntos da linguagem são especializados por tais variáveis (temática, usuários e situações comunicativas) e apresentam características particulares de tipo linguístico (unidades, regras) e de tipo textual. Existem outras características relevantes apontadas por Cabré (1993), como o propósito comunicativo, que possibilita a existência de diferentes tipos textuais de acordo com sua função, e o grau de abstração, que depende da própria temática.

A respeito dos âmbitos especializados, interessam-nos, neste trabalho, os textos especializados, que serão brevemente descritos a seguir em relação às suas características.

2.1.1 Texto especializado como repositório de terminologia

A diferença entre linguagem geral e especializada está mais relacionada ao grau do que ao tipo: o grau em que as características fundamentais da linguagem são maximizadas ou minimizadas nas linguagens especiais. As linguagens especializadas são usadas de modo mais consciente do que a língua geral, porque as situações em que são usadas aguçam a preocupação do usuário com a linguagem. É, portanto, no nível do uso que esses critérios de diferenciação são percebidos (SAGER, DUNGWORTH & MC DONALD, 1980, *apud* CABRÉ, 1993, p. 125).

Além de destacarmos essa diferença, é importante definir o que é texto. Recuperando a origem da palavra, chegamos ao verbo “tecer” (do latim *texo*) e ao conceito de texto, que era de “tecido rico e ordenado com sentido” (CIAPUSCIO, 1998, p. 44). Para o propósito desta investigação, consideramos texto como uma forma de discurso que tem por objetivo atender a dada situação comunicativa, decorrente de um processo de comunicação cujos elementos são

específicos, selecionados e combinados entre si (CABRÉ, 1993), ou seja, elementos “tecidos”. Cabré considera o texto como um ambiente que viabiliza tanto a expressão quanto a transmissão cultural, dando oportunidade aos indivíduos de se relacionarem entre si (CABRÉ, 1993). Em consonância com esse pensamento, Bourigault e Slodzian (2004) defendem que a análise da terminologia deve ser feita em “textos reais, que proliferam e circulam em todas as direções, alterando as fronteiras entre as áreas”, ao contrário do que a TGT propõe (BOURIGAULT; SLODZIAN, 2004, p. 107). Segundo Auger e L’Homme (2004, p. 109), “[...] longe de ser tratada como uma unidade isolada, a palavra [termo] é indissociável do texto que a encerra e lhe dá o seu sentido [...]”, ou seja, o contexto.

Ciapuscio (1998) define texto como um objeto de caráter linguístico e ao mesmo tempo comunicativo, e por essa razão é complexo. Sua análise pode se realizar em diferentes níveis: funcional, situacional, temático, estruturação linguística e formulação. Como é um produto da atividade intelectual humana, é passível de análise por diferentes perspectivas, como a do processo (a atividade de produzir um texto) ou a do produto (resultado da atividade). Ciapuscio (1998) prevê que os textos se agrupam em classes textuais conforme aspectos macro e microestruturais. A macroestrutura de um texto, segundo a autora, está relacionada com questões de *como* e *para que* se emprega tal texto. A microestrutura, por sua vez, se refere aos seus padrões léxicos, morfossintáticos e pragmáticos, elementos que se somam para configurar o caráter de especialidade do texto. Os níveis superiores do texto (macroestrutura) determinam os aspectos microestruturais, mas o contrário também é verdadeiro.

O texto de que trata este estudo é, portanto, um produto: um texto especializado, resultante de uma atividade intelectual (a escrita), que apresenta uma macroestrutura característica (isto é, escrito sob as diretrizes da GRI, para reportar os impactos das empresas) e contém características específicas de padrão lexical, morfossintático e, principalmente, pragmático (sua temática é a sustentabilidade, escrito para empresários e especialistas em um contexto de divulgação de informações).

A seguir, descreveremos o conceito adotado neste estudo a respeito das fraseologias, ponto principal desta pesquisa.

2.1.2. Fraseologia especializada

Os textos especializados, como é o caso do Relatório de Sustentabilidade, abrigam uma grande quantidade de termos, cuja identificação, por si só, não garante uma adequada compreensão do texto, assim como não garante que sejam feitas traduções adequadas do ponto de vista funcional. Evidencia-se, assim, a necessidade de mirarmos nossa atenção para as fraseologias: são elas que mostram como os elementos de uma língua se combinam entre si, formando os padrões de uso em âmbitos especializados.

No contexto dos estudos fraseológicos, alguns trabalhos devem ser destacados. Entre os pesquisadores da área, Marie-Claude L’Homme defende que um termo, para o especialista, é a unidade que permite transmitir o conhecimento próprio de seu domínio de especialidade (2004). Segundo L’Homme (2000), os especialistas em geral concordam com o fato de que tanto as colocações (a que a autora entende como combinatórias da língua de modo geral) quanto as combinatórias léxicas especializadas se conformam em um uso convencional entre os seus usuários, sejam os falantes da língua ou os especialistas de certas áreas.

Uma das consequências de se analisar um objeto por diferentes perspectivas é a diversidade de nomes para esse mesmo objeto. Embora Bevilacqua (2004) aponte diferentes denominações, por exemplo, *colocações*, *colocações de linguagens especializadas*, *combinatórias léxicas especializadas*, *entidades fraseológicas*, *fraseologismos* e *unidades fraseológicas especializadas*, o que precisa ser considerado é o ponto de convergência entre elas. Independentemente do nome atribuído, o fato é que muitos pesquisadores vêm se debruçando sobre as características dessas unidades, pois elas revelam um comportamento em relação ao uso do idioma, especialmente importante para os tradutores em âmbito especializado. Entre as características, são apontadas as seguintes:

- Possuem diferentes graus de fixação;
- Apresentam frequências estatisticamente significativas nos textos especializados;
- Têm como ponto de partida o termo, a partir do qual se busca identificar os co-ocorrentes, e não estruturas morfossintáticas preestabelecidas.

A complexidade característica dos estudos em fraseologia ainda permeia os critérios de identificação e análise dessas combinações, ponto em que todas as teorias acabam se tornando difusas, como apontou Waquil (2013). A autora explica que:

Outro limite de difícil precisão é o existente entre uma unidade fraseológica especializada e uma unidade terminológica. Não há consenso estabelecido

para esta distinção e as posturas dos autores em relação a esta problemática são ainda bastante divergentes, já que se constata nos estudos em Terminologia combinatórias de palavras que se caracterizam como termos e não como unidades fraseológicas (WAQUIL, 2013, p. 60).

A respeito dos inúmeros estudos que tomaram as fraseologias como objeto de investigação, a revisão teórica realizada por Waquil (2013) resume as principais correntes teóricas dos estudos de fraseologia de perspectiva lexicológica e da fraseologia especializada, com aderência à proposta de cunho terminológico de Bevilacqua (2004). O Quadro 2, a seguir, apresenta denominações, definições e critérios utilizados pelos principais pesquisadores de fraseologia especializada, resumo realizado nos estudos de Waquil (2013).

Quadro 2. Resumo das denominações, definições e critérios utilizados pelos principais teóricos da fraseologia especializada.

Autor	Denominação	Definição para fraseologia especializada	Crítérios para identificação das fraseologias especializadas
Picht (1990)	<i>Fraseologia LSP / Frase LSP (LSP = Language of Special Purpose)</i>	Estruturas formadas por, pelo menos, dois conceitos: um com característica de verbo e o outro com característica de objeto	a) apresentar combinabilidade conceitual; b) apresentar grau de especialização.
Blais (1993)	<i>Fraseologismo</i>	Estruturas próprias de um domínio de especialidade, formadas por elementos linguísticos (um deles é o termo <i>nódulo</i>) que se ligam sintática e semanticamente e entre os quais há uma restrição paradigmática	a) apresentar dois elementos linguísticos (no mínimo); b) apresentar um termo (no mínimo); c) apresentar um termo que ocupe função central (nódulo); d) apresentar ligações sintáticas e semânticas entre os elementos linguísticos; e) ser uma construção própria da língua de especialidade;

			f) caracterizar-se por limitação na substituição dos elementos constituintes.
Pavel (2003)	<i>Fraseologismos LE ou Fraseologia LE (LE = língua de especialidade)</i>	Estruturas formadas por termos-núcleos e co-ocorrentes, constituindo-se em combinatórias sintagmáticas de termos de uma estrutura conceitual coerente.	a) previsibilidade (relacionado ao nível de conhecimento da LE); b) combinabilidade (capacidade de combinar-se com variados co-ocorrentes); c) comutabilidade (co-ocorrentes intercambiáveis, sinônimos); d) grau de especialização (quanto mais estável e desenvolvido for o domínio, maior o grau de especialização e menor será a comutabilidade de co-ocorrentes); e) frequência expressiva; f) descontinuidade (útil para extração semiautomática); g) grau de fixidez da UF (combinações: fixa/estável, semifixa e livre).
L'Homme (2000)	<i>Combinatórias léxicas especializadas</i>	Combinações especializadas compostas por uma unidade lexical central (geralmente uma unidade terminológica) e um co-ocorrente.	a) convencionalidade (de acordo com a comunidade especialista); b) co-ocorrência lexical livre (liberdade restrita ao campo especializado); c) imprevisíveis para o aprendiz.
Gouadec (1994)	<i>Entidades fraseológicas, Fraseologismo, ou Matrizes fraseológicas</i>	Entidades fraseológicas com <i>pivô</i> terminológico: compostas por um núcleo	a) esterotípiã; b) frequência significativa; c) anomalia ou ruptura (risco);

		(termo, que é a parte fixa) e outros elementos que são o entorno. As matrizes fraseológicas são compostas por elementos invariáveis (matriz, que é fixa) e variáveis (componentes que inserem a fraseologia em um domínio específico). Exemplo: “assumir um compromisso entre [x] e [y]” (‘assumir um compromisso entre’ = parte invariável [x] e [y] = parte variável)	e) cadeia de caracteres significantes; f) recorrência.
Bevilacqua (2004)	Unidades Fraseológicas Especializadas Eventivas (UFEs Eventivas)	São estruturas que se determinam a partir da área em que estão inseridas, compostas por um núcleo eventivo (NE) + núcleo terminológico (NT), partindo das unidades formadas por verbos.	a) caráter sintagmático; b) estabilidade semântica; c) estabilidade sintática; d) grau de fixação; e) inclusão de uma UT; f) semicomposicionalidade ou composicionalidade; g) uso em um âmbito específico; h) frequência significativa.

Fonte: Adaptado de Waquil (2013, p. 58-59).

Desse modo, adotamos como referência para a análise dos dados levantados nesta pesquisa a perspectiva de Bevilacqua (2004). Em sua tese, a autora propõe um deslocamento do foco de atenção exclusivo do termo, e que unidades fraseológicas especializadas (UFEs) passam a ter dois polos ou núcleos de interesse, sendo um o núcleo terminológico (NT), de categoria nominal, com valor referencial e caráter denominativo, e o outro, um núcleo

eventivo (NE), de caráter verbal, deverbal ou participio. Por essa razão, as UFEs passam a se chamar UFEs Eventivas.

Ao observar o Quadro 2 e as propriedades descritas por Bevilacqua (2004), nota-se que a proposta da autora se aproxima das ideias de L'Homme (2000), devido ao fato de que o foco não é apenas o termo, mas é dado o mesmo destaque ao núcleo eventivo ou co-ocorrente. Ao mesmo tempo, há proximidade com a proposta de Picht (1990), uma vez que o autor menciona características de verbo de um dos elementos. Nossa aproximação com a proposta de Bevilacqua (2004) também se deve ao fato de que está alinhada com a TCT, ao relacionar as fraseologias com o uso em situações comunicativas reais, realizadas entre os usuários de um domínio ou âmbito especializado, em que há transmissão do conhecimento especializado por meio das fraseologias.

2.2. Tradução especializada: equivalentes funcionais

Entre as abordagens disponíveis para analisar aspectos relacionados à tradução, a de perspectiva funcional (ou funcionalista) é a que parece mais adequada para esta pesquisa, porque apresenta uma visão mais global e leva em consideração aspectos relacionados à comunicação como um todo, e não apenas os termos.

Katharina Reiss foi o primeiro expoente feminino do funcionalismo, e, juntamente com Hans Vermeer, fez importantes contribuições para a efetiva mudança de paradigma com a sua *Skopostheorie*. Ambos deixaram um legado importante porque colocaram a cultura, a comunicação e a ação no centro de seus estudos, tirando o foco exclusivo dos aspectos linguísticos. Hoje, o funcionalismo tem, na figura de Christiane Nord, sua principal representante.

As perspectivas linguísticas, até então vigentes, tomavam a tradução como uma operação linguística, utilizando técnicas ou metodologias específicas para transpor um texto de uma língua a outra. Porém, no início dos anos 1970, o modelo que privilegiava as equivalências em uma perspectiva linguística começou a ser questionado por Reiss. Para ela, o ideal de tradução deveria ter como objetivo, na língua-alvo, uma equivalência do conteúdo conceitual, da forma linguística e da função comunicativa do texto na língua-fonte. No entanto, a noção básica de equivalência, a qual até então não havia sido questionada, ainda resistiria até meados dos anos 80, pois o foco na perspectiva linguística baseada em equivalência era preservar as características da língua no texto-alvo (NORD, 1997).

A perspectiva funcionalista se consolida com a publicação da obra de Katharina Reiss e Hans J. Vermeer, em 1984, ao proporem uma teoria que romperia com esse paradigma de tradução de perspectiva linguística vigente na época. A teoria do *Skopos* (escopo), como ficou conhecida, tem seu nome derivado da palavra grega *skopos*, que significa “propósito”. Reiss e Vermeer buscaram fornecer o entendimento de que a natureza da tradução se concentra em uma atividade “funcional” (*purposeful*) e, além disso, recorreram a um conceito de cultura para fundamentar sua teoria, deslocando a noção de tradução como atividade estritamente linguística, o que corrobora a adoção do termo “adequação” em vez de “equivalência”.

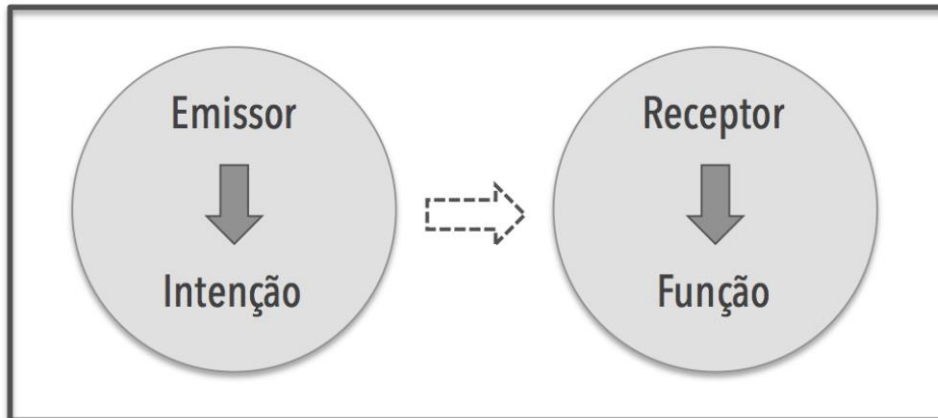
Dessa forma, a teoria do *Skopos* buscou um ponto de partida em teorias comunicativas e culturais, pois elas conseguiriam explicar, em grande parte, a especificidade das situações comunicativas. Vermeer sugeriu que a tradução seria uma forma de “ação translacional” baseada em um texto-fonte, constituído por elementos verbais e não verbais. Qualquer ação tem um objetivo, um propósito, e por esse motivo, a teoria foi denominada de *skopos*. (VERMEER, 1989b, p. 173 apud NORD 1997, p. 25).

Dois méritos dessa teoria são enfatizados por Nord: o primeiro é em relação aos aspectos culturais relevantes e que devem ser considerados na prática tradutória. Vermeer (1987 apud NORD, 2013) propôs uma definição de cultura com base em Göhring (1978 apud NORD, 1995) como sendo “todo o conjunto de normas e convenções que devem ser conhecidas pelos indivíduos para que se tornem parte de uma comunidade e, ao mesmo tempo, sejam diferentes de todos os demais”. Todos esses aspectos culturais Vermeer denominou *culturemas*, explicando que seriam “[...] fenômenos sociais de uma cultura X considerados relevantes por essa comunidade cultural, e, quando comparados com um fenômeno social correspondente em uma cultura Y, são entendidos como sendo específico da cultura X”. (NORD, 2013, p. 3).

O outro mérito da teoria do *Skopos* é em relação à quebra do paradigma da equivalência como objetivo da tradução. De acordo com essa teoria, o primeiro princípio que determina qualquer processo de tradução é o propósito da ação translacional geral, o que se encaixa com o fato de que a intencionalidade é parte de qualquer ação. Além do termo *Skopos*, Vermeer usou também outros sinônimos relacionados para se referir ao escopo: *alvo*, *propósito*, *intenção* e *função*. A fim de evitar confusão, Nord (2005: 2013) propôs uma distinção básica entre intenção e função, porque a primeira se refere ao emissor, e a segunda, ao receptor. O emissor deseja atingir certo objetivo com o texto; porém, é o receptor quem atribui uma função ao texto, de acordo com suas próprias expectativas e necessidades de comunicação. Quando o desejo do emissor coincidir com a necessidade/expectativa do

receptor, pode-se considerar uma situação ideal, em que a intenção e a função serão análogas ou mesmo idênticas (NORD, 2013, p. 2). A Figura 6 esquematiza essa proposta:

Figura 6. Esquema da proposta de Nord: a função do texto é atribuída pelo receptor.



Nord (2013) explica que, no processo translacional, a função, que não é uma qualidade inerente ao texto, corre o risco de ser ignorada, não podendo ser mudada, tampouco preservada. Consequência disso é que a tradução, que para Nord é uma comunicação mediada ou translacional, precisa cruzar uma barreira tanto da linguagem quanto da cultura para chegar ao seu destino. Porém, ao realizar esse movimento, o conteúdo que foi enviado com determinada intenção poderá ser recebido com uma função diferente. Quanto maior a distância cultural (seja ela geográfica, temporal etc.), maiores são as chances de que a intenção e a função não coincidam.

Em suma, na nossa visão, o texto especializado é um repositório de terminologias inserido em determinada cultura, que deve ser considerada para que os aspectos desses textos também sejam adequadamente transpostos para a língua e a cultura de chegada.

2.3 Linguística de *corpus*: breve histórico

As pesquisas em terminologia vêm adotando como procedimento de rotina a utilização de *corpora* especializados e em formato eletrônico para a coleta de termos, fraseologias, entre outros tipos de investigação de âmbito especializado. Além disso, de acordo com Maciel, a utilização de ferramentas computacionais nas pesquisas com textos especializados parece aumentar as garantias e a confiabilidade do ponto de vista científico (MACIEL, 2013).

A importância das ciências da informação (ou simplesmente, informática) é inegável para o desenvolvimento de pesquisas terminológicas, tanto teóricas quanto aplicadas, como é o caso da terminografia. Como demonstra a própria definição: *corpus*, de acordo com Sinclair, “é uma coleção de partes de textos em **formato eletrônico**, selecionados de acordo

com critérios externos para representar, o quanto for possível, uma linguagem ou sua variação como fonte de dados para a pesquisa linguística”¹⁹ (SINCLAIR, 1991, 2004, nosso grifo).

Como observa Berber Sardinha, a respeito da expansão da área de pesquisas utilizando *corpora*:

[...] há um desenvolvimento crescente de centros de pesquisa mantidos por empresas. Estes centros utilizam-se de pesquisas baseadas em *corpus* para várias finalidades comerciais, como o processamento automático de textos, informatização de grandes bases de dados e a montagem de sistemas inteligentes de reconhecimento de voz e gerenciamento de informação. As grandes empresas de telecomunicações investem nestas áreas, reconhecendo o potencial econômico deste campo. (BERBER SARDINHA, 2000, p. 329).

A sofisticação dos equipamentos e dos *softwares* permitiu a execução de tarefas bem mais complexas de modo mais eficaz, principalmente devido ao aprimoramento das mídias (observe-se a evolução desde os cartões perfurados até os documentos eletrônicos em diversas extensões que hoje dispomos, como PDF, TXT e TMX), o que facilitou imensamente a agilidade desse tipo de pesquisa. Com essas inovações, também aumentou a capacidade de armazenamento de textos e, conseqüentemente, de dados, o que permite criar novas possibilidades de estudo, pesquisa e áreas de atuação dos profissionais do texto (BERBER SARDINHA, 2004).

Assim, hoje temos a Linguística de *Corpus*, uma abordagem que trata do uso de *corpora* compilados em formato eletrônico e que possibilita evidenciar dados com a sistematização das informações coletadas empiricamente e processadas pelos *softwares* de análise textual, como o AntConc e o Sketch Engine (ANTHONY, 2014; SKETCH ENGINE, 2017).

Dessa forma, o pesquisador tem condições de analisar a língua com objetividade e observar a frequência relativa das palavras com evidências que sustentam as afirmações sobre o objeto de estudo (REBECHI, 2015). Um *corpus* pode fornecer dados relevantes, desde que tenha sido compilado criteriosamente e seja representativo do contexto a ser examinado. Com base nos aspectos metodológicos relacionados à compilação de um *corpus* para a finalidade de um levantamento terminológico (ver BERBER SARDINHA, 2004), as seguintes etapas foram consideradas para a compilação do *corpus* deste estudo: delimitação do tema, estabelecimento dos critérios para a seleção das fontes; compilação dos textos; conversão de formatos; nomeação dos arquivos; e escolha da ferramenta. No entanto, julgamos importante

¹⁹ No original: “[...] corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.”

fazer um contraponto, a fim de suscitar uma reflexão a respeito do uso de *corpora* em pesquisas:

O *corpus* sempre será um material incompleto, um artefato preparado em função dos critérios preestabelecidos pelo pesquisador em vista de seus propósitos. [...] o *corpus* não é um material inteiramente neutro; reflete, de um lado, a subjetividade de seu compilador; de outro, as opiniões dos autores dos textos e o pensamento de uma dada comunidade em determinada época. Os *corpora* estão sujeitos aos defeitos da passagem do tempo e da evolução das ideias; nenhum deles contém tudo que é necessário para entender a área ou determinar-lhe um corte [...]” (MACIEL, 2013, p. 42).

Isso significa que, por mais específicos que sejam os critérios para a construção de um *corpus*, ou por mais extenso ou representativo que ele seja, um corpus sempre conterá traços subjetivos, como aponta Maciel (2013), seja essa subjetividade atribuída ao próprio pesquisador (uma vez que ele determina os critérios de seleção dos textos para sua pesquisa), seja ela atribuída a quem produziu os textos que compõem o *corpus*.

O Capítulo 3, a seguir, descreve a metodologia utilizada/percorrida nesta pesquisa e outros aspectos importantes a respeito da Linguística de *Corpus*.

3 METODOLOGIA

As pesquisas linguísticas e terminográficas são bastante antigas e datam, no mínimo, do século XVIII, conforme apontam Aluísio e Almeida (2006). Nas últimas décadas, no entanto, houve um aumento na quantidade de estudos baseados em *corpus*, dada a relevância tanto nas pesquisas, especialmente nas áreas de Linguística Computacional e Linguística Aplicada.

O surgimento dos computadores e da internet provoca uma ruptura no paradigma do que se entendia por *corpus*. A introdução do computador proporcionou um aumento no volume de processamento dos dados devido à capacidade de armazenamento desse recurso tecnológico, o que contribuiu significativamente para refinar as pesquisas linguísticas. Ao mesmo tempo, as novas tecnologias inseriram no cenário da Linguística de *Corpus* novos dilemas e questionamentos que, provavelmente, só conseguem ser respondidos devido ao caráter empírico que a pesquisa com *corpus* apresenta.

Berber Sardinha (2000, p. 327) explica com clareza esse panorama:

Com a popularização dos computadores, foi possibilitado o acesso de mais pesquisadores ao processamento de linguagem natural e, concomitantemente, a sofisticação do equipamento permitiu a consecução de tarefas mais complexas, mais eficientemente, sem falar no aumento da capacidade de armazenamento e na introdução de novas mídias (fitas magnéticas, em vez de cartões *hollerith* perfurados, etc.), as quais facilitaram a criação e manutenção de *corpora* em maior número. Com a entrada em cena dos micro-computadores pessoais, nos anos 80, uma nova onda de mudanças aconteceu, como a popularização de *corpora* e de ferramentas de processamento, o que contribuiu decisivamente para o reaparecimento e fortalecimento da pesquisa linguística baseada em *corpus*.

A era digital, portanto, situa a Linguística de *Corpus* em um novo patamar, conforme também aponta Zilio (2010):

[...] a partir do momento em que se pôde armazenar textos em um computador e produzir programas capazes de analisar esses textos de forma automática ou assistida por computador é que os trabalhos com *corpora* se tornaram mais frequentes, criando e dando maior visibilidade à área (ZILIO, 2010, p. 134).

Significa dizer, portanto, que o computador introduz ao cenário de estudos de linguagem uma objetividade de certa forma inédita nas pesquisas. Graças aos computadores surgem programas de análise textual, possibilitando examinar com mais precisão o

comportamento da língua, livre de quaisquer julgamentos subjetivos. Observam-se, dessa forma, aspectos linguísticos de toda ordem: morfológicos, sintáticos, semânticos etc., bem como a produtividade lexical, entre outros, que poderiam passar despercebidos. “Em resumo, por meio de *corpus*, descreve-se a língua de forma objetiva” (ALUÍSIO; ALMEIDA, 2006, p. 158). Ou, nas palavras de Sinclair: “Confie no texto” (SINCLAIR, 2004).

A diferença fundamental entre as definições de *corpus* é justamente seu formato eletrônico. Para a Linguística de *Corpus*, um conjunto grande de textos impressos, como jornais e revistas, não é considerado *corpus*, pois não são passíveis de análise por computador (ALUÍSIO; ALMEIDA, 2006). Seria necessário digitalizar um material impresso para que, então, pudesse submetê-lo a um programa de análise textual. Picht (2004, p. 67) também reforça essa ideia ao mencionar as seguintes vantagens do formato eletrônico: tratamento de grandes quantidades de dados (para fornecer um embasamento mais amplo às pesquisas), processamento rápido e múltiplas finalidades de uso. Essas análises feitas por computador possuem caráter quantitativo e, assim, objetivo, o que vai ao encontro do que argumenta Picht: “um *corpus* não pode oferecer mais do que aquilo que contém” (PICHT, 2004, p. 68).

Nota-se, portanto, que a era digital representa um marco no desenvolvimento da Linguística de *Corpus*, conforme aponta Zilio:

[...] a partir do momento em que se pôde armazenar textos em um computador e produzir programas capazes de analisar esses textos de forma automática ou assistida por computador é que os trabalhos com *corpora* se tornaram mais frequentes, criando e dando maior visibilidade à área (ZILIO, 2010, p. 134).

Assim, utilizarei nesta pesquisa a seguinte definição de *corpus*: “[...] coletânea de textos entendidos num sentido amplo, em formato eletrônico, compilados segundo critérios específicos para o estudo a que se propõem” (TAGNIN, 2013, p. 216).

Essa definição parece preencher todos os requisitos necessários para compor um *corpus* bem construído, trazendo os elementos considerados mais importantes em outras definições (ver BERBER SARDINHA, 2004, p. 16-18). Portanto, levamos em conta os seguintes itens como critérios específicos, ou pré-requisitos de confiabilidade (BERBER SARDINHA, 2004) dos *corpora* desta pesquisa: (I) origem; (II) propósito; (III) composição; (IV) formatação; (V) representatividade; e (VI) extensão. Todos eles serão descritos a seguir.

I) Origem: os textos coletados para a construção do *corpus* são autênticos, ou seja, foram produzidos com uma finalidade real, e não para a pesquisa em si, além de terem sido

escritos por falantes nativos (BERBER SARDINHA, 2000). Todos os textos coletados para esta pesquisa são considerados, portanto, autênticos.

II) Propósito: esse é um importante ponto, pois o *corpus* deve ser construído com a finalidade de servir como objeto de estudo linguístico (BERBER SARDINHA, 2004), em consonância com o item anterior.

III) Composição: esse item refere-se à seleção dos textos, que tem por finalidade atribuir ao *corpus* certas características que atendam às condições de naturalidade e autenticidade (BERBER SARDINHA, 2000). As características desejadas para a realização desta pesquisa estão relacionadas a um tipo de texto, os RSs, e por isso foram estabelecidos critérios de seleção, conforme descrito na próxima seção. Além disso, foi considerada uma “quantidade aceitável” de textos, para que a análise fosse o mais imparcial possível.

IV) Formatação: diz respeito ao *corpus* ser passível de processamento por computador.

Finalmente:

IV) Representatividade: é um dos aspectos mais difusos. Para Berber Sardinha (2004), está relacionada com a extensão do *corpus*, considerando o propósito de que o *corpus* seja representativo de uma língua ou de suas variedades. No entanto, na visão de Koester, “não importa tanto o tamanho do *corpus*, pois tudo depende do que o *corpus* contém e o que está sendo investigado”²⁰ (KOESTER, 2010, p. 68). Por isso, Koester defende que *corpora* especializados não necessitam ser demasiadamente extensos (não tanto quanto os *corpora* de língua geral) para gerarem resultados interessantes, pois é o cuidado durante sua construção que os torna confiáveis, às vezes mais confiáveis do que os de tamanho grande. Além disso, o léxico especializado e suas estruturas são mais propensos a apresentarem padrões regulares e melhor distribuição do que um *corpus* geral (O’Keefe et al., 2007 apud Koester, 2010). A esse respeito, observamos a seguinte explicação, bastante lúcida, fornecida por Koester:

Se um tipo de texto muito específico está sendo investigado, então parece bem simples estabelecer uma representatividade **situacional**, em que todas as amostras coletadas irão representar com precisão o tipo ou gênero. [...] Se todas as amostras se originarem de uma mesma organização, então o corpus será representativo do gênero (ou tipo) conforme é usado na organização, mas não será representativo do gênero como um todo (KOESTER, 2010, p. 69, grifo nosso)²¹.

²⁰ No original: “There is no ideal size for a corpus; it all depends on what the corpus contains and what is being investigated.”

²¹ No original: “If a very specific type of genre is being investigated, then it may be quite straightforward to establish situational representativeness, as all the samples collected will accurately represent that genre. [...]”

Nesta pesquisa a representatividade pode ser verificada pela seleção de apenas um tipo textual, porém de diferentes fontes (isto é, as organizações [empresas]), mais do que pela extensão, visto que o foco não é abranger toda a língua portuguesa ou inglesa, mas sim um tipo de linguagem especializada.

A respeito das linguagens especializadas, atualmente muitas pesquisas que objetivam a construção de dicionários e glossários conciliam Terminologia e Linguística de *Corpus*, conforme apontou Rebechi (2015). As pesquisas de caráter terminológico, desde muitas décadas atrás, já utilizam conjuntos de textos especializados, os quais são uma grande fonte de extração de termos. Segundo Maciel (2001), a linguagem especializada possui o mesmo sistema linguístico que a linguagem geral, porém seu uso é peculiar, pois ela depende de fatores e circunstâncias para sua realização, como a situação e os interlocutores. Se a circunstância for uma comunicação entre especialistas, certamente será diferente da comunicação entre um especialista e um indivíduo leigo:

[...] é ponto pacífico que não se trata de uma língua diferente dos sistemas lingüísticos naturais, mas sim de um uso peculiar da língua, isto é, uma linguagem, cuja configuração depende das múltiplas circunstâncias de sua realização. Tal linguagem não se limita à interação formal entre especialistas que usam um mesmo registro, mas é utilizada por interlocutores de diferentes hierarquias e graus de especialização em diferentes níveis de formalidade (MACIEL, 2001, p. 85-86).

Isso significa que não são as palavras usadas que se diferenciam, mas sim o contexto em que elas são utilizadas e o público a que se dirigem os textos, o que leva à ativação de novos significados. O léxico de uma língua é o mesmo utilizado tanto na linguagem comum (em textos não especializados) quanto na linguagem especializada. Nesta pesquisa não foram investigadas novas palavras ou termos, mas sim as ocorrências em novas situações de uso, com o propósito de auxiliar o trabalho dos profissionais do texto, sejam tradutores ou redatores. A seguir, farei o detalhamento da construção do *corpus* de estudo utilizado neste trabalho, à medida em que apresentarei as ferramentas computacionais utilizadas e as configurações adotadas para o processamento dos textos.

3.1 *Corpus* de estudo

A Linguística de *Corpus*, conforme apontado anteriormente, é uma metodologia bastante útil para investigação linguística, porque oportuniza a análise de textos autênticos similares com a utilização do computador (TAGNIN, 2007).

Assim, os *corpora*, de modo geral, são classificados de acordo com sua composição em alguns tipos, dos quais dois destaco aqui: *corpus* comparável e *corpus* paralelo. **Corpus comparável** se refere ao conjunto de textos de um mesmo gênero ou tipo, escritos originalmente em idiomas diferentes. Considera-se como **corpus paralelo** o conjunto composto por textos originais e suas respectivas traduções. **Corpus comparável** se difere de **corpus paralelo** na composição e nos objetivos, conforme aponta Tagnin (2007):²²

[...] um *corpus* comparável, composto de textos originalmente escritos nas duas línguas, inglês e português no caso em questão, **fornecerá o termo efetivamente usado por aquela comunidade**. Em outras palavras, o *corpus* fornecerá o termo mais provável de ocorrer no contexto que o tradutor estiver pesquisando, ou seja, o termo mais recorrente. Nem sempre, é óbvio, o tradutor deseja usar o termo mais recorrente, mas, mesmo nesse caso, o *corpus* lhe será de grande valia justamente para identificar o termo que *não* pretende usar (TAGNIN, 2007, p. 2, grifo meu).

Kenning também observou que uma coleção de textos (ou seja, os *corpora*) pode ser considerada comparável quando esses textos forem reunidos de acordo com os mesmos critérios, como tamanho, assunto e período (KENNING, 2010).

3.1.1 Descrição do *corpus* e critérios de construção dos *subcorpora*

O *corpus* de estudo desta pesquisa foi composto por dois *subcorpora*: um construído com textos autênticos e originais em português brasileiro, o qual denominamos RelSustenta_PT, e o outro por textos autênticos e originais em inglês estadunidense, denominado RelSustenta_EN, ambos compostos apenas por Relatórios de Sustentabilidade (RS). Os *corpora* de estudo são classificados, portanto, como **comparáveis**, pelas características de sua composição. O objetivo principal deste trabalho, conforme explicado anteriormente, é realizar o levantamento dos termos e fraseologias no *corpus* em português, buscando seus equivalentes funcionais em língua inglesa e analisando os contextos das ocorrências com a finalidade de apresentar uma proposta de glossário bilíngue.

²² http://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/Stella_Abrapui%202007_artigo.pdf

Outro tipo de *corpus* que deve ser destacado aqui é o ***corpus de referência***, que serve de comparação com o *corpus* de estudo, funcionando como um padrão de referência da língua. Os *corpora* de referência utilizados nesta pesquisa serão descritos na próxima seção, na qual serão também descritas as ferramentas utilizadas na análise.

Para compor os dois *subcorpora* desta pesquisa, os Relatórios de Sustentabilidade (RSs) foram selecionados a partir de duas publicações, da seguinte forma: os textos originalmente escritos em português foram extraídos da edição 2016 do Guia de Sustentabilidade da Revista Exame, o mais reconhecido do Brasil, e os textos escritos originalmente em inglês estadunidense foram retirados do *ranking* da Corporate Knights 2017, reconhecida publicação anual em que são eleitas as 100 empresas mais sustentáveis do mundo.²³ Cada uma dessas publicações possui sua própria metodologia para avaliar as empresas participantes do *ranking*²⁴, embora seus critérios tenham semelhanças, pois ambos se baseiam nas diretrizes da GRI.

Da lista das empresas constantes no *ranking* do Guia Exame de Sustentabilidade, foram escolhidos os RSs da Duratex, da Fibria, da Klabin, da Natura e da Votorantim Metais, eleitas as mais sustentáveis em seus respectivos setores, cujos relatórios estavam disponíveis para *download* nos sites institucionais das empresas, no formato PDF. Para os textos em inglês, foram escolhidos apenas os RSs de empresas norte-americanas, para evitar entrar no mérito da variação linguística, que não é o foco desta pesquisa. A partir do *ranking* da Corporate Knights²⁵, foram selecionadas as empresas americanas com maior pontuação, privilegiando setores diferentes, conforme o critério de representatividade. O armazenamento dos textos seguiu as etapas detalhadas a seguir.

Primeiro, realizou-se o *download* de cada arquivo em PDF, seguido do salvamento e nomeação do arquivo de forma padronizada, conforme detalhado a seguir. Após essa padronização, todos os PDFs foram convertidos para o formato **.txt**, que é o mais aceito por diversos *softwares* e plataformas de análise textual (embora o programa utilizado para analisar os textos desta pesquisa, SE²⁶, consiga fazer a leitura de arquivos no formato PDF, outros como AntConc e WordSmith²⁷ não o fazem. Com vistas a disponibilizar os *corpora* para

²³ Para conhecer melhor os rankings, acesse os sites: <http://www.corporateknights.com/reports/2017-global-100/2017-global-100-results-14846083> e <https://exame.abril.com.br/edicoes/guia-de-sustentabilidade-2016/>.

²⁴ Os questionários do Guia Exame de Sustentabilidade estão disponíveis no link: <http://sustentabilidade.exame2.com.br/>

²⁵ O ranking está disponível no site: <http://www.corporateknights.com/magazines/2018-global-100-issue/2018-global-100-results-15166618/> e a metodologia adotada pela publicação está disponível em: https://www.corporateknights.com/wp-content/uploads/2018/10/2019-Global-100_Methodology-Final.pdf.

²⁶ O Sketch Engine será detalhado a seguir.

²⁷ O WordSmith Tools encontra-se disponível no link: <https://www.lexically.net/wordsmith>.

futuras pesquisas, optamos por fazer a conversão para a extensão .txt). Os nomes dos arquivos foram padronizados da seguinte forma: primeiro, identificou-se o idioma com o código “PT” para o português e “EN” para o inglês; na sequência, o código para o nome da empresa com três letras (DTX para Duratex, FBR para Fibria, KLB para Klabin, NTR para Natura, VTM para Votorantim Metais, ALL para Allergan, CGP para Colgate-Palmolive, CIS para Cisco, INT para Intel, e, finalmente, JNJ para Johnson & Johnson). Ao final, a identificação do texto e o ano do *ranking*, seguidos da extensão .txt; a seguir, apresenta-se um exemplo do nome completo de um dos arquivos: PT_VTM_RS2015.txt. Após realizada a conversão para o formato txt, procedeu-se à escolha do programa de análise para realizar os procedimentos.

3.2 Ferramentas computacionais utilizadas

A análise pela frequência de *tokens* nos fornece o número total de ocorrências de uma unidade — que é considerada, do ponto de vista da Linguística de *Corpus*, a sequência de letras com espaço antes e depois, ou uma palavra no texto corrido, não importa quantas vezes se repitam. Substantivos e verbos, para exemplificar, podem ocorrer, respectivamente, na forma singular (ação) e na forma infinitiva (consumir), mas também aparecem na forma plural (ações) e conjugadas (consumiram, consumiu, consome etc.). Contudo, correspondem a um mesmo lema (por exemplo, ação, consumir). O SE faz a lematização e considera cada ‘lema’ como um *type*. Portanto, a frequência de *types* corresponde às diferentes palavras, que seriam as palavras lematizadas. Exemplo disso são verbos no infinitivo, substantivos e adjetivos na forma singular e no gênero masculino (LEAL; BISOL, 2017; EVISON, 2010). Como as ferramentas disponíveis no SE apresentaram algumas divergências na lematização das palavras²⁸, fizemos um cálculo aproximado dos *types*.

O cálculo do número de *tokens* é efetuado automaticamente pelo programa a partir da compilação dos *corpora*. Para compilá-los, basta selecionar os textos desejados (salvos no computador) e fazer o seu carregamento (upload). É possível inserir novos textos após a compilação.

As Tabelas 1 e 2 apresentam o levantamento de *tokens* e *types* em português e em inglês, respectivamente, fornecidos pelo SE, para o qual *token* está definido como:

[...] a menor unidade em que o *corpus* é dividido. Tipicamente, cada forma de palavra (ou simplesmente palavra) e pontuação (vírgula, ponto, etc.) é um

²⁸ Essas questões serão discutidas mais adiante.

diferente token (exceto no inglês, palavras como *don't* consistem em dois tokens. Os corpora contêm mais tokens do que palavras e os espaços entre as palavras não são tokens (SKETCH ENGINE, 2018)²⁹.

O SE não denomina *types* o número de formas lematizadas de cada palavra (*token*). O cálculo para palavras lematizadas foi encontrado na aba de informações sobre cada texto, no item “Lemma” das estatísticas do texto (para o português) e no item “lemma_lc” (para os textos em inglês, sendo “lc” a silga para “*lower-case*”, pois no inglês ele faz essa distinção). A razão forma/item (*type/token ratio*) é normalmente expressa em percentagem e indica a riqueza lexical de um texto. “Quanto maior o seu valor, mais palavras diferentes o texto conterà. Em contraposição, um valor baixo indicará um número alto de repetições, o que pode indicar um texto menos rico ou variado, do ponto de vista de seu vocabulário” (BERBER SARDINHA, 2004, p. 94).

A primeira coluna apresenta o nome de cada arquivo do *corpus* de português, a segunda coluna apresenta o respectivo número de *tokens* e a terceira coluna, o número aproximado de *types*.

Tabela 1. Composição do *corpus* em português

Nome do arquivo	Número de <i>tokens</i>	Número de <i>types</i>
PT_DTX_RS2016.txt	33.382	4.198
PT_FBR_RS2016.txt	20.648	3.508
PT_KLB_RS2016.txt	8.065	1.873
PT_NTR_RS2016.txt	55.769	5.607
PT_VTM_RS2016.txt	41.416	4.281
Total	159.280	19.467

O *corpus* de inglês ficou constituído da seguinte forma (Tabela 2):

²⁹ No original: *Token is the smallest unit that each corpus divides to. Typically, each word form and punctuation (comma, dot, ...) is a separate token (but don't in English consists of 2 tokens). Therefore, corpora contain more tokens than words. Spaces between words are not tokens. A text is divided into tokens by a tool called a tokenizer which is often specific for each language.*

Tabela 2. Composição do *corpus* em inglês

Nome do arquivo	Número de <i>tokens</i>	Número de <i>types</i>
EN_ALL_RS2017.txt	12.938	1.901
EN_CGP_RS2016.txt	45.687	3.521
EN_CIS_RS2016.txt	76.803	5.699
EN_INT_RS2016.txt	51.348	4.411
EN_JNJ_RS2016.txt	46.446	7.021
TOTAL	235.334	22.553

Ao observarmos essas duas tabelas, é possível perceber que o número de palavras do *corpus* em inglês supera o *corpus* em português em cerca de 50%. Essa diferença se explica, possivelmente, devido a fatores culturais, corroborando as conclusões de Katan (1999), como explica Rebechi (2015, p. 195):

Retomando o conceito de ‘contexto’ (*contexting*) do antropólogo Edward Hall, Katan (1999, p. 177) sugere que “[...] indivíduos, grupos e culturas (e em diferentes momentos) possuem prioridades divergentes no que tange à quantidade de informação (texto) necessária a ser explicitada para que a comunicação seja bem-sucedida”. Denomina *high context culture* a cultura que durante a comunicação apoia-se no conhecimento compartilhado pelo grupo; já em uma *low context culture*, supõe-se que o conhecimento comum seja insuficiente para garantir a comunicação demandando maior explicação.

O SE é o programa de escolha para realizar a análise dos *subcorpora*. Trata-se de um *software* on-line que possibilita a criação de *corpus* personalizado pelo usuário. O programa pode ser utilizado por meio do cadastro do usuário e do pagamento, para o qual são oferecidas várias modalidades; para a realização desta pesquisa, foi utilizado o plano de pagamento mensal na versão estudante, a um custo de cerca de US\$ 8 (oito dólares americanos)³⁰.

As Figuras 7 e 8 apresentam a tela do SE mostrando o resumo da compilação de cada *subcorpus*.

³⁰ A licença de uso do Sketch Engine pode ser adquirida por meio do site: the.sketchengine.co.uk.

Figura 7. *Subcorpora* Português.

Sketch Engine

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Keywords/terms
Corpus info
Manage corpus
My jobs
User guide [↗](#)

Word list
Corpus: RelSustenta
Total number of items: 5
Total frequency: 5

file.filename	document frequency
PT_VTM_RS2015.txt	1
PT_NTR_RS2016.txt	1
PT_KLB_RS2015.txt	1
PT_FBR_RS2015.txt	1
PT_DTX_RS2016.txt	1

Figura 8. *Subcorpora* inglês

Sketch Engine

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Keywords/terms
Corpus info
Manage corpus
My jobs
User guide [↗](#)

Word list
Corpus: RelSustenta_EN
Total number of items: 5
Total frequency: 5

file.filename	document frequency
EN_JNJ_RS2016.txt	1
EN_INT_RS2016.txt	1
EN_CIS_RS2016.txt	1
EN_CGP_RS2016.txt	1
EN_ALL_RS2017.txt	1

Além disso, a ferramenta disponibiliza diversos *corpora* de língua geral, por exemplo, ptTenTen e OANC, mencionados anteriormente, que foram utilizados como *corpora* de referência. Assim, foram compilados os corpora de português e de inglês e, após, foram geradas as listas de palavras-chave nos respectivos idiomas.

3.2.1 Procedimentos para a seleção de candidatos a termo

Para a realização desta pesquisa, foram selecionados dois *corpora* de referência dentre os 14 que o SE oferece, por se adequarem aos seguintes critérios: *corpus* composto por textos escritos (e não orais) e por textos em português brasileiro (uma vez que vários *corpora* disponibilizados pelo programa em português contêm textos mistos com português europeu). Assim, selecionamos os *corpora* ptTenTen 11, com quase 4 bilhões de palavras do português, sendo composto por cerca de 80% com textos do português brasileiro, e *Open American National Corpus - written* (doravante OANC), com aproximadamente 13 bilhões de palavras, composto por textos de artigos, jornais, guias de viagem, textos de ficção e não ficção, entre outros.

Para extrair as palavras-chave, cada *subcorpus* do *corpus* de estudo foi comparado ao *corpus* de referência do respectivo idioma, pela ferramenta *Keyword/Terms*, da seguinte forma:

- *Subcorpus* RelSustenta_PT (pt) x ptTenTen11
- *Subcorpus* RelSustenta_EN x OANC (*written*)

Dessa comparação, foram geradas duas listas em cada idioma, contendo as *keywords* (ou palavras-chave) necessárias para eleger os candidatos a termo. As palavras-chave serão detalhadas a seguir.

As palavras-chave são o "[...] resultado da comparação entre o *corpus* de estudo e o *corpus* de referência. Essa comparação elimina palavras com frequência relativa similar nos dois *corpora*, de modo que restem as palavras cuja frequência é estatisticamente significativa" (TAGNIN, 2013, p. 218). Assim, é possível afirmar que as palavras-chave são as palavras (simples ou compostas) cuja frequência no *corpus* de estudo é significativamente maior do que no *corpus* de referência, isto é, se uma dada palavra possui alta frequência em ambos os *corpora*, ela será desconsiderada, pois o objetivo é identificar as que são apenas significativamente mais frequentes no *corpus* de estudo. As palavras-chave exercem, portanto, uma função bastante importante, uma vez que se trata das palavras relativamente mais frequentes no *corpus* de estudo, objeto de análise textual.

Desse modo, o levantamento dos candidatos a termos por meio da lista de palavras-chave foi realizado aplicando-se os seguintes parâmetros à ferramenta *Keywords/Terms*:

- Frequência mínima de cinco ocorrências para o termo/palavra-chave;
- Máximo de palavras-chave de 1000;

- Valor *simple math* de 1000 (inclui uma variável que permite ao usuário focar nas palavras de maior ou menor frequência; quanto maior o número, serão selecionadas as palavras de frequência mais alta).

Os parâmetros aplicados podem ser ajustados individualmente na ferramenta, e os testes realizados para ajustar esses parâmetros mostraram que seus resultados estão interligados. Assim, para determinar uma frequência mínima de ocorrências adequada à extensão do *corpus*, foi realizado um teste-piloto com a ferramenta antes de se determinar o valor de cinco ocorrências. Configurou-se a frequência mínima de três ocorrências para os termos ou palavras-chave e o parâmetro *simple math* foi configurado com o valor mínimo (1). Como se esperava, esse levantamento teve pouco aproveitamento, pois a lista obtida exibiu com muitos ruídos nas palavras com a frequência mínima de três ocorrências (por exemplo, expressões sem sentido, como “lealdade cns”, “cuidado com”, “inclui capão”, “rodovia br”). No teste-piloto, o parâmetro *simple math* foi testado com outros valores, mantendo-se o valor de frequência mínima de três ocorrências. Como os dados se mantiveram semelhantes ao do primeiro teste, os valores foram reajustados. Assim, os testes seguintes somente apresentaram melhora na qualidade dos dados com os seguintes parâmetros: *simple math* 1000, frequência mínima de cinco ocorrências e máximo de palavras-chave no valor de 1000.

Após aplicar os parâmetros, a ferramenta gerou duas listas, uma de palavras simples e outra de palavras compostas, conforme mostram as Figuras 9 e 10 a seguir, respectivamente, candidatos a termos em português e em inglês.

Figura 9. Tela com visualização parcial do SE após aplicação da ferramenta *Keywords/Terms* (corpus do português).

RelSustenta: Extracted keywords / terms ?

[Change extraction options](#) Download singlewords: [TBX CSV](#). Download multiwords: [TBX CSV](#).

Singlewords and multiwords are ordered by [keyness score](#). The score and corpus frequency (leading to the respective concordance) are displayed in parentheses. **Highlighted** words were used as seeds in a previous WebBootCaT run within this corpus.

[<< Back to corpus files](#) Use WebBootCaT with selected words

Single-word	Score	F	RefF	Multi-word	Score	F	RefF
<input type="checkbox"/> button	W 907.26	999	13,498	<input type="checkbox"/> votorantim metais	W 438.21	287	560
<input type="checkbox"/> metais	W 379.94	320	6,007	<input type="checkbox"/> relatório votorantim	W 112.44	71	0
<input type="checkbox"/> votorantim	W 335.35	376	14,117	<input type="checkbox"/> relatório votorantim metais	W 104.59	66	0
<input type="checkbox"/> natura	W 263.50	442	30,287	<input type="checkbox"/> asseguração externa	W 79.48	50	0
<input type="checkbox"/> fibria	W 250.19	194	4,091	<input type="checkbox"/> dados complementares	W 79.36	51	396
<input type="checkbox"/> gri	W 130.36	91	1,912	<input type="checkbox"/> operações internacionais	W 77.35	49	138
<input type="checkbox"/> asseguração	W 123.37	81	715	<input type="checkbox"/> forma de gestão	W 64.38	44	1,635
<input type="checkbox"/> klabin	W 108.02	86	4,790	<input type="checkbox"/> consultoras natura	W 60.52	38	41
<input type="checkbox"/> duratex	W 101.38	69	1,446	<input type="checkbox"/> vide página	W 47.95	30	52
<input type="checkbox"/> cajamar	W 88.25	65	3,097	<input type="checkbox"/> instituições fortes	W 45.83	29	282
<input type="checkbox"/> ods	W 85.25	55	451	<input type="checkbox"/> número total	W 45.83	48	12,321
<input type="checkbox"/> dma	W 73.76	51	1,829	<input type="checkbox"/> cadeia de valor	W 45.33	35	4,331
<input type="checkbox"/> gee	W 72.90	58	4,859	<input type="checkbox"/> asseguração externa vide	W 44.95	28	0
<input type="checkbox"/> deca	W 66.64	49	3,130	<input type="checkbox"/> emprego digno	W 42.57	27	356
<input type="checkbox"/> emissões	W 64.67	49	3,789	<input type="checkbox"/> índice remissivo	W 42.22	27	512

Figura 10. Tela do SE após levantamento realizado pela ferramenta *Keywords/Terms* (corpus do inglês)

RelSustenta_EN: Extracted keywords / terms ?

[Change extraction options](#) Download singlewords: [TBX CSV](#). Download multiwords: [TBX CSV](#).

Singlewords and multiwords are ordered by [keyness score](#). The score and corpus frequency (leading to the respective concordance) are displayed in parentheses. **Highlighted** words were used as seeds in a previous WebBootCaT run within this corpus.

[<< Back to corpus files](#) Use WebBootCaT with selected words

Single-word	Score	F	RefF	Multi-word	Score	F	RefF
<input type="checkbox"/> GRI	W 757.37	712	0	<input type="checkbox"/> supply chain	W 221.96	208	0
<input type="checkbox"/> Colgate	W 577.55	733	19	<input type="checkbox"/> management approach	W 71.25	76	0
<input type="checkbox"/> Cisco	W 568.56	692	16	<input type="checkbox"/> supplier diversity	W 56.24	52	0
<input type="checkbox"/> Governance	W 312.36	396	19	<input type="checkbox"/> corporate responsibility	W 56.24	52	0
<input type="checkbox"/> GHG	W 271.89	255	0	<input type="checkbox"/> energy efficiency	W 53.47	67	0
<input type="checkbox"/> Sustainability	W 252.62	250	3	<input type="checkbox"/> metric tonne	W 47.74	44	0
<input type="checkbox"/> Supply	W 229.69	279	16	<input type="checkbox"/> water conservation	W 46.68	43	0
<input type="checkbox"/> CSR	W 226.93	374	41	<input type="checkbox"/> renewable energy	W 45.60	57	0
<input type="checkbox"/> FY17	W 216.65	203	0	<input type="checkbox"/> palm oil	W 44.55	41	0
<input type="checkbox"/> Intel	W 207.71	670	132	<input type="checkbox"/> greenhouse gas	W 44.12	49	0
<input type="checkbox"/> Environment	W 195.04	416	69	<input type="checkbox"/> water use	W 42.43	39	0
<input type="checkbox"/> REPORT	W 192.23	300	36	<input type="checkbox"/> material topic	W 37.12	34	0
<input type="checkbox"/> Allergan	W 191.87	183	1	<input type="checkbox"/> green power	W 32.87	30	0
<input type="checkbox"/> Supplier	W 180.02	181	4	<input type="checkbox"/> energy consumption	W 32.37	43	0
<input type="checkbox"/> sustainability	W 173.86	220	19	<input type="checkbox"/> environmental sustainability	W 31.81	29	0

Cada lista das palavras simples (coluna *single word* do *corpus* do português e do inglês) e das palavras compostas (coluna *multi-word*, nos dois idiomas) foi armazenada no programa Excel. Elas foram organizadas em abas diferentes, da seguinte forma: simples_PT, composta_PT, simples_EN e composta_EN. Dessa forma, foi possível classificá-las tanto em ordem alfabética quanto em ordem de frequência ou de chavicidade. A propósito do índice de chavicidade, ele é um parâmetro fornecido pelos programas de análise textual. No SE, calcula-se a chavicidade, ou seja, a frequência relativa, de uma palavra-chave ou termo pela seguinte equação:

$$\frac{fpm_{focus} + n}{fpm_{ref} + n}$$

Onde:

- fpm_{focus} = frequência da palavra no *corpus* de estudo
- fpm_{ref} = frequência da palavra no *corpus* de referência
- n = número padrão (1) usado como “*smoothing parameter*” para regularizar (ou suavizar) o cálculo (em estatística, esse valor é utilizado com função de aproximação dos valores, para tentar capturar padrões relevantes nos dados) (SIMONOFF, 1998).

Assim, se a frequência da palavra for muito alta no *corpus* de estudo e muito baixa no *corpus* de referência, significa que o índice de chavicidade é alto, indicando que a palavra ou expressão tem ocorrência estatisticamente relevante, configurando em candidato a termo.

3.2.2 Critérios de exclusão de palavras-chave

As listas geradas pelo programa, por meio da ferramenta *Keywords/Terms*, foram salvas no programa Excel, para melhor manuseio dos dados, sendo agrupadas em apenas uma lista. Esse agrupamento foi seguido de uma tarefa de “limpeza”, eliminando-se candidatos que apresentassem algum dos seguintes critérios:

- Nomes próprios (pessoas, organizações ou lugares. Por exemplo: Fibria, Ekos, Deca, Cajamar, Niquelândia, Cisco, Ushahidi, Protex etc.);
- Números, elementos químicos e símbolos do sistema métrico (por exemplo: m³, CO₂, kW, hectares, Zn, etc.);

- Abreviaturas (por exemplo: fem., masc., un., FY16, CEO, etc., com exceção daquelas que representam termos da área, como ODS – objetivos do desenvolvimento sustentável);
- Ruídos (por exemplo: button, nd, title, number etc., pois são indicações de página, cabeçalhos, rodapés, informações que constam na configuração dos arquivos PDF convertidos para acesso on-line, alguns itens de linguagem de programação, como códigos de programação que a ferramenta acabou considerando por causa da ausência de *stop-list*);
- Candidatos que ocorreram em apenas um dos textos foram desconsiderados por falta de representatividade, pois poderia indicar alguma particularidade de um texto de determinada empresa (como é o caso da ocorrência da expressão “consultoras natura”, que ocorre 38 vezes, mas em apenas um texto (Relatório da Natura) (ver Figuras 11 e 12). Essa informação só pôde ser verificada ao clicarmos na sua frequência.

Figura 11. Item destacado mostrando exemplo de expressão com alta frequência, porém sem representatividade (ocorrência em apenas um relatório).

Multi-word	Score	F	RefF
<input type="checkbox"/> votorantim metais	W 578.29	287	560
<input type="checkbox"/> são paulo	W 153.21	75	434
<input type="checkbox"/> relatório votorantim	W 149.59	71	0
<input type="checkbox"/> relatório votorantim metais	W 139.12	66	0
<input type="checkbox"/> asseguração externa	W 105.64	50	0
<input type="checkbox"/> dados complementares	W 103.72	51	539
<input type="checkbox"/> operações internacionais	W 103.55	49	0
<input type="checkbox"/> forma de gestão	W 83.31	44	1,631
<input type="checkbox"/> consultoras natura	W 80.31	38	41
<input type="checkbox"/> vide página	W 63.53	30	57
<input type="checkbox"/> instituições fortes	W 60.46	29	282
<input type="checkbox"/> asseguração externa vide	W 59.60	28	0
<input type="checkbox"/> relatório anual	W 57.50	27	0

Figura 12a. Primeira parte da lista de ocorrências, mostrando o texto do qual foram extraídas.

Term 38 (238.57 per million) ⓘ

Page 1 of 2 Go Next Last

file652187... 709 360 119 Lucro retido 24 154 178 **Consultoras Natura** 4.122 4.421 4.430 Colaboradores 1.075 1.245

file652187... : Operações Internacionais; CNs: **Consultoras Natura** ; CNOs: Consultoras Natura Orientadoras.

file652187... ; CNs: Consultoras Natura; CNOs: **Consultoras Natura** Orientadoras. NOTAS: 1. fonte: Bloomberg. 2.

file652187... . Seguimos ampliando nossa rede de **Consultoras Natura** , que cresceu 7,5% e se aproximou da marca de 550

file652187... estreitos e de qualidade com as **Consultoras Natura** . Implantação da estratégia Revisamos

file652187... - Brasil 2012 2013 2014 2015 2016 **Consultoras Natura** no Brasil (milhares) 1.268,5 1.290 1.319 1.377

file652187...) 1.268,5 1.290 1.319 1.377 1.256 **Consultoras Natura** Orientadoras 12.125 11.957 11.328 9.500 8.310

file652187... (CNs)1 4.2902 4.161 4.028 **Consultoras Natura** Orientadoras (CNOs)2 14.306 17.614 18.428 1.

file652187... no estreitamento da relação com as **Consultoras Natura** (CNs) em nossas operações na América Latina. O

file652187... na qualidade do relacionamento. **Consultoras Natura** (CN) - Operações Internacionais, por país 2016

file652187... Peru 94,0 Colômbia 87,8 França 0 **Consultoras Natura** (CN) - Operações Internacionais 2015 2016

file652187... 2015 2016 Total 505,1 543,0 **Consultoras Natura** Orientadoras (CNO) - Operações

file652187... CNO na Argentina e no México. **Consultoras Natura** Orientadoras (CNO) - Operações

file652187... / 102-44 Índice de lealdade de **Consultoras Natura** Operações Internacionais (%)1 2 3 2014 2015

file652187... , hoje formada por 1,8 milhão de **Consultoras Natura** (CNs). Para isso, estruturamos uma

file652187...). Ele acompanha a situação das **Consultoras Natura** (CNs) de acordo com três dimensões: saúde,

file652187... o desenvolvimento social das **Consultoras Natura** (CNs) e de toda a sociedade. A partir de 2014, a

file652187... de beleza do Brasil. Nossas **Consultoras Natura** Digitais (CNDs), que atuam nessa plataforma,

file652187... de consumidoras cadastradas. As **Consultoras Natura** (CNs) que participam do modelo online tiveram

file652187... às Gerentes de Relacionamento e às **Consultoras Natura** Orientadoras (CNOs) para que suas vendas no

Page 1 of 2 Go Next Last

Figura 12b. Segunda parte da lista da ocorrência de “consultoras Natura”, mostrando o texto do qual foram extraídas.

Term 38 (238.57 per million) ⓘ

Page 2 of 2 Go

file652187... com a formação digital das **Consultoras Natura** . Além do impacto positivo sobre seus ganhos,

file652187... de catadores e o envolvimento de **Consultoras Natura** na coleta de materiais. A experiência serviu

file652187... 2016 Atingimento da meta **Consultoras Natura** (CNs) Despertar o interesse pelo aprendizado

file652187... de catadores e o envolvimento de **Consultoras Natura** na coleta de materiais. Após a captação,

file652187... Ver. 2. Média da porcentagem de **Consultoras Natura** que compraram algum item de Crer para Ver,

file652187... Crer para Ver. 2. Porcentagem de **Consultoras Natura** envolvidas com compras da linha Crer para Ver,

file652187... de colaboradores e **Consultoras Natura** (CNs), mudanças climáticas, resíduos,

file652187... da melhoria da educação pública. • **Consultoras Natura** • Colaboradores • Instituto Natura • Sociedade

file652187... prioritários: colaboradores, **Consultoras Natura** , fornecedores e comunidades fornecedoras e

file652187... e engajamento do consumidor e **consultoras Natura** , trazendo o impacto social e ambiental que sua

file652187... de catadores e o envolvimento de **Consultoras Natura** na coleta de materiais. A experiência serviu

file652187... • Beleza livre de estereótipos **Consultoras Natura** Aumentar significativamente a renda média

file652187... de alguns grupos de trabalho. **Consultoras Natura** (CN) Produtividade 203-2 - Renda média anual

file652187... (CNs)1 4.290 4.161 4.028 **Consultoras Natura** Orientadoras (CNOs)2 14.306 17.614 18.428 1.

file652187... e Reconhecimento Participação de **Consultoras Natura** em treinamentos - Brasil (milhares) 2014 2015

file652187... - Brasil (milhares) 2014 2015 2016 **Consultoras Natura** iniciantes 496 543 554 Treinamento inicial 378

file652187... iniciativas. treinamentos **Consultoras Natura** - Operações Internacionais 1 2 2014 2015 2016

file652187... manifestações de consumidores e **Consultoras Natura** oriundas dos canais de atendimento com esses

First | Previous Page 2 of 2 Go

As Figuras 13 e 14, a seguir, mostram, como exemplo, uma parte da listagem das palavras-chave em inglês no programa Excel: lista de palavras simples, lista de palavras compostas e a união dessas duas listas em uma só. Usamos a palavra “energy” para

exemplificar o procedimento realizado. A Figura 13 apresenta um recorte da tabela do Excel, mostrando a palavra na posição 556 da lista de palavras simples, na etapa anterior ao agrupamento das listas. Nesse momento, a lista estava organizada por ordem de chavicidade (coluna B), tal como foi copiada do SE.

Figura 13. Recorte da lista bruta das palavras simples do *corpus* de inglês.

	A	B	C	D
1	Single-word	Score	F	RefF
551	applicable	9.00	44,00	414,51
552	videoconference	8.99	7,00	14,69
553	hackathon	8.96	6,00	4,10
554	external	8.96	11,00	58,42
555	development	8.94	98,00	1,004,152
556	energy	8.94	442,00	4,743,996
557	conserve	8.93	22,00	178,63
558	labor	8.92	85,00	864,97
559	encouraging	8.88	6,00	4,76

A Figura 14 apresenta um recorte semelhante ao da Figura 13 anterior, porém da lista de palavras compostas (*multi-word*), copiadas para uma tabela do Excel, mostrando algumas expressões com a palavra *energy* em sua composição (ver destaques na Figura). A lista, da mesma forma, estava ordenada pela chavicidade (coluna B), tal como foi copiada do SE.

Figura 14. Recorte da lista de palavras compostas do *corpus* de inglês.

	A	B	C	D
1	Multi-word	Score	F	RefF
4	supplier diversity	56.24	52	0
5	corporate responsibility	56.24	52	0
6	energy efficiency	53.47	67	0
7	metric tonne	47.74	44	0
8	water conservation	46.68	43	0
9	renewable energy	45.60	57	0
10	palm oil	44.55	41	0
11	greenhouse gas	44.12	49	0
12	water use	42.43	39	0
13	material topic	37.12	34	0
14	green power	32.87	30	0
15	energy consumption	32.37	43	0
16	environmental sustainability	31.81	29	0
17	energy use	31.61	31	0
18	oral care	29.68	27	0
19	environmental impact	28.24	32	0
20	oral health	28.18	29	0
21	sustainability profile	26.50	24	0
22	climate change	24.94	42	0
23	energy reduction	22.25	20	0
24	carbon footprint	20.99	21	0

As palavras destacadas nas Figuras 13 e 14 (*energy*, *energy efficiency*, *renewable energy*, e assim por diante) aparecem juntas na Figura 15. Essa figura mostra um recorte de uma nova lista, gerada com o agrupamento de todas as palavras (simples e compostas).

Figura 15. Trecho da lista unificada de termos simples e compostos, em ordem alfabética.

	A	C	D	E
1	WORD	Score	Freq	Freq/ref
177	energy	8.99	659,00	4,522,637
178	energy conservation	20.66	9,00	20,422
179	energy consumption	47.61	43,00	64,955
180	energy efficiency	43.34	67,00	127,016
181	energy intensity	52.04	13,00	1,819
182	energy management	43.03	17,00	15,933
183	energy reduction	77.14	20,00	2,594
184	energy usage	12.65	5,00	17,244
185	energy use	47.36	31,00	40,955
187	engagement	17.71	111,00	583,458
189	engages	11.59	12,00	79,195
190	entrepreneurship	89.77	56,00	37,738
191	environment	153.14	416,00	239,622

Nessa lista agrupada, ordenaram-se as palavras alfabeticamente. Como mostra o exemplo da Figura 15, o agrupamento ordenado alfabeticamente possibilitou visualizar a palavra-chave simples (*energy*) e parte das palavras compostas, mas isso só ocorreu quando as expressões iniciavam com a mesma letra do termo (no caso, letra E). As demais ocorrências, como *renewable energy*, foram identificadas ao longo da limpeza dessa lista. Outras informações importantes apresentadas na Figura 15: a coluna C mostra o *score* (índice de chavicidade), a coluna D representa a frequência no *corpus* de estudo e a coluna E apresenta a frequência no *corpus* de referência. Foi importante manter essas informações à vista, uma vez que esses valores auxiliaram na validação dos dados.

Para que nenhuma expressão se perdesse nesse procedimento, fizemos uma verificação utilizando a ferramenta *Word Sketch*. Essa ferramenta mostra, de maneira sintética, um resumo de todas as expressões que ocorreram combinadas com a palavra-chave principal. O *Word Sketch* possibilita visualizar todas as ocorrências na forma de um gráfico ou de uma lista. As Figuras 16 e 17 mostram exemplos dessas duas apresentações para o termo *energy*.

4 ANÁLISE

Conforme explicado no Capítulo 3 - Metodologia -, os candidatos a termos foram selecionados a partir das listas de palavras-chave obtidas com a ferramenta *Keywords/Terms* do SE, que fornece, de acordo com as configurações desejadas pelo usuário, duas listas de palavras: uma de palavras simples e outra de combinações de palavras. As opções de extração de palavras-chave são ajustáveis e permitem que o usuário defina alguns campos, como:

- *Corpus* de referência;
- Atributo (classificação por palavra ou lema);
- Frequência mínima;
- Número máximo de palavras-chave e de termos.

No capítulo anterior, foram explicadas essas configurações. Deixamos, no entanto, o item do atributo para ser descrito aqui no Capítulo da Análise, pois se trata de um item que não interfere no cálculo estatístico. O atributo é um parâmetro que possibilita extrair as palavras-chave na forma de palavra ou por lema. A diferença entre eles consiste no fato de que, quando classificado por palavra, exhibe as ocorrências nas formas flexionadas (em gênero e número ou conjugações verbais). A classificação lematizada junta as ocorrências de cada palavra na sua forma canônica. Porém, considerando que a ferramenta apresentou alguns erros, a lematização não foi totalmente aplicada. Houve casos de plurais figurando na lista, ou seja, que não foram devidamente lematizados ou que escaparam da métrica utilizada pela ferramenta. Isso nos leva a cogitar a possibilidade de que a ferramenta priorize a frequência nesses casos. Mesmo assim, esses cálculos não são muito bem esclarecidos, mas os dados fornecidos puderam ser utilizados, sem prejuízo para a análise. A seguir, serão explicados os procedimentos para análise das listas de palavras-chave.

4.1 Análise das listas de palavras-chave

As análises das listas obtidas pela ferramenta *Keyword/Terms* teria sido menos trabalhosa se outras configurações estivessem disponíveis no SE no par de idiomas deste trabalho. É o caso da ferramenta *stop-list*, que seria extremamente útil, por exemplo, na limpeza de palavras gramaticais nas tabelas das palavras-chave, como será detalhado adiante neste capítulo. Mesmo sem contar com esse recurso, partimos para a verificação dos contextos.

As duas listas, de palavras simples e palavras compostas, são apresentadas lado a lado após a extração, como mostram as Figuras 18 e 19 (português e inglês, respectivamente), e podem ser analisadas sob dois aspectos: pelo *score* (índice de chavicidade) ou pela frequência no *corpus* de estudo (F). No programa SE, a seleção da ferramenta *Keywords/Terms* fornece automaticamente as palavras por ordem decrescente do índice de chavicidade; na coluna imediatamente à direita da coluna *score*, apresenta-se a frequência da palavra no *corpus* de estudo, indicado por F, e a coluna da extrema direita (RefF) apresenta a frequência do *corpus* de referência (Figuras 18 e 19).

Figura 18. Visualização da tela após a extração das palavras-chave, mostrando as 20 primeiras palavras-chave simples e compostas do *corpus* RelSustenta_PT (português).

Single-word	Score	F	RefF	Multi-word	Score	F	RefF
<input type="checkbox"/> button	W 1,209.71	999	214	<input type="checkbox"/> votorantim metais	W 352.83	287	560
<input type="checkbox"/> votorantim	W 473.04	376	1	<input type="checkbox"/> relatório votorantim	W 90.15	71	0
<input type="checkbox"/> metais	W 402.81	320	0	<input type="checkbox"/> relatório votorantim metais	W 83.87	66	0
<input type="checkbox"/> fibria	W 244.60	194	0	<input type="checkbox"/> dados complementares	W 63.95	51	396
<input type="checkbox"/> natura	W 235.30	442	7,724	<input type="checkbox"/> asseguração externa	W 63.78	50	0
<input type="checkbox"/> projetos	W 225.76	179	0	<input type="checkbox"/> operações internacionais	W 62.17	49	138
<input type="checkbox"/> atividades	W 152.93	121	0	<input type="checkbox"/> forma de gestão	W 52.54	44	1,635
<input type="checkbox"/> gri	W 114.80	91	23	<input type="checkbox"/> consultoras natura	W 48.64	38	41
<input type="checkbox"/> klabin	W 108.97	86	1	<input type="checkbox"/> número total	W 39.98	48	12,321
<input type="checkbox"/> asseguração	W 102.31	81	22	<input type="checkbox"/> vide página	W 38.58	30	52
<input type="checkbox"/> duratex	W 87.41	69	15	<input type="checkbox"/> cadeia de valor	W 37.86	35	4,331
<input type="checkbox"/> renovável	W 83.64	85	1,632	<input type="checkbox"/> instituições fortes	W 36.97	29	282
<input type="checkbox"/> cajamar	W 82.62	65	0	<input type="checkbox"/> asseguração externa vide	W 36.16	28	0
<input type="checkbox"/> na	W 81.36	64	0	<input type="checkbox"/> gases de efeito	W 35.08	41	11,475
<input type="checkbox"/> sustentabilidade	W 76.33	219	14,823	<input type="checkbox"/> emprego digno	W 34.38	27	356
<input type="checkbox"/> governança	W 75.86	108	4,538	<input type="checkbox"/> índice remissivo	W 34.15	27	512
<input type="checkbox"/> gee	W 73.70	58	10	<input type="checkbox"/> asseguração externa vide página	W 33.65	26	0
<input type="checkbox"/> ods	W 69.91	55	12	<input type="checkbox"/> gases de efeito estufa	W 33.13	36	9,137
<input type="checkbox"/> objetivos	W 68.81	54	0	<input type="checkbox"/> visão de sustentabilidade	W 32.08	25	223
<input type="checkbox"/> mulheres	W 68.81	54	0	<input type="checkbox"/> relatório anual	W 31.37	27	2,607

Figura 19. Visualização da tela após a extração das palavras-chave, mostrando as 20 primeiras palavras-chave simples e compostas do *corpus* RelSustenta_EN (inglês).

Single-word				Multi-word			
	Score	F	RefF		Score	F	RefF
<input type="checkbox"/> S	W 1,367.85	965	0	<input type="checkbox"/> management approach	W 98.32	76	75
<input type="checkbox"/> Colgate	W 1,039.24	733	0	<input type="checkbox"/> supply chain	W 75.98	208	2,061
<input type="checkbox"/> GRI	W 1,009.50	712	0	<input type="checkbox"/> supplier diversity	W 72.25	52	24
<input type="checkbox"/> Cisco	W 971.25	685	0	<input type="checkbox"/> metric tonne	W 62.88	44	5
<input type="checkbox"/> Intel	W 961.34	678	0	<input type="checkbox"/> corporate responsibility	W 60.91	52	161
<input type="checkbox"/> Report	W 862.19	608	0	<input type="checkbox"/> material topic	W 49.16	34	0
<input type="checkbox"/> Johnson	W 723.38	510	0	<input type="checkbox"/> water conservation	W 48.24	43	202
<input type="checkbox"/> Environment	W 590.23	416	0	<input type="checkbox"/> palm oil	W 46.09	41	201
<input type="checkbox"/> CSR	W 530.74	374	0	<input type="checkbox"/> green power	W 40.22	30	58
<input type="checkbox"/> Governance	W 515.16	363	0	<input type="checkbox"/> water use	W 39.27	39	308
<input type="checkbox"/> Index	W 474.09	334	0	<input type="checkbox"/> oral care	W 37.61	27	31
<input type="checkbox"/> Society	W 452.84	319	0	<input type="checkbox"/> sustainability profile	W 34.99	24	0
<input type="checkbox"/> Assessment	W 404.68	285	0	<input type="checkbox"/> oral health	W 31.42	29	242
<input type="checkbox"/> Supply	W 386.27	272	0	<input type="checkbox"/> energy consumption	W 30.63	43	728
<input type="checkbox"/> Business	W 384.85	271	0	<input type="checkbox"/> environmental sustainability	W 30.52	29	270
<input type="checkbox"/> Chain	W 373.52	263	0	<input type="checkbox"/> energy efficiency	W 29.66	67	1,592
<input type="checkbox"/> GHG	W 362.19	255	0	<input type="checkbox"/> energy reduction	W 27.92	20	36
<input type="checkbox"/> Appendix	W 357.94	252	0	<input type="checkbox"/> fy07 baseline	W 26.50	18	0
<input type="checkbox"/> Water	W 345.19	243	0	<input type="checkbox"/> greenhouse gas	W 26.02	49	1,216
<input type="checkbox"/> Approach	W 328.19	231	0	<input type="checkbox"/> s business	W 25.88	18	17

Em ambas as Figuras (18 e 19), nota-se que, embora as primeiras palavras da lista simples estejam relacionadas na ordem decrescente do *score* (chavicidade), elas não parecem apresentar dados muito relevantes, já que nessa primeira parte da lista a maior parte das ocorrências representa nomes das empresas (o que era esperado, uma vez que os textos são oriundos de cada organização), nomes próprios (por exemplo, Cajamar, Amazônia), entre outros que constam nos critérios de exclusão (ver Capítulo Metodologia). Portanto, a lista fornecida pelo programa não é capaz de determinar, por si, quais elementos são, de fato, termos. É preciso que se analisem, manualmente, os contextos dessas palavras-chave para classificá-las como tal. Para prosseguir com essa análise, foram unificadas as duas listas em apenas uma, de modo que todos os candidatos (total de 1.660 da lista bruta de português e total de 1.320 da lista de inglês) fossem agrupados em ordem alfabética, a fim de simplificar a busca e auxiliar a distinguir entre termos simples e compostos e a identificar os núcleos das combinações de palavras.

A Figura 20, a seguir, mostra a primeira parte da lista bruta dos candidatos extraídos pela ferramenta *Keywords/Terms* do SE, ordenada alfabeticamente no Excel, antes da análise manual. A lista bruta parece não ter sido corretamente etiquetada, visto que foram computados termos compostos formados por artigo + substantivo, como mostram as linhas 2 a 13 da Figura 20.

Figura 20. Visualização parcial da lista, com os 20 primeiros itens no programa Excel.

	A	B	C	D
1	WORD	Score	F	RefF
2	a empresa	7.61	7,00	6,64
3	a fibria	13.53	10,00	40,00
4	a forma	17.39	14,00	1,59
5	a forma de gestão	18.58	14,00	2,00
6	a gestão	9.17	7,00	1,55
7	a klabin	12.29	9,00	14,00
8	a natura	17.26	13,00	91,00
9	a sustentabilidade	10.92	8,00	261,00
10	a sustentabilidade na	11.05	8,00	2,00
11	a sustentabilidade na klabin	11.05	8,00	0,00
12	a votorantim	31.10	24,00	26,00
13	a votorantim metais	31.14	24,00	3,00
14	abastecer	7.46	13,00	7,50
15	abastecimento	4.29	20,00	28,81
16	abnt	7.91	6,00	449,00
17	abnt nbr	6.65	5,00	2,19
18	abranger	4.11	20,00	30,30
19	absenteísmo	9.10	8,00	1,21
20	acelerar	4.97	14,00	15,51

Além desse aparente problema de etiquetagem da ferramenta no português, houve casos em que palavras foram lematizadas equivocadamente, como se fossem verbos. Por exemplo, "pela" (contração da preposição per + a) foi lematizada para o verbo "pelar"; "agosto" foi lematizado para "agostar", "correto" para "corretar". No inglês ocorreu outro problema: algumas palavras não foram corretamente etiquetadas, pois a ferramenta etiquetou alguns verbos (por exemplo, *source*) como substantivo. Esses casos foram considerados como um erro da ferramenta, indicando que o programa necessita de aprimoramentos, especialmente no processamento dos textos em língua portuguesa. A ferramenta não esclarece a forma de lematização das palavras (no nível de programação computacional, até porque esse tipo de informação seria pouco útil ao pesquisador linguista), tampouco explica como corrigir esses erros durante o processamento. Levando-se em conta os critérios de seleção das palavras que nos interessam, realizamos a limpeza de elementos como os das ocorrências da linha 3 "a fibria" e da linha 10 "a sustentabilidade na" (Figura 20). Esses itens foram reconhecidos pelo programa como palavras compostas, porém as palavras gramaticais ("a", "de") em posição inicial deveriam ter sido ignoradas pela ferramenta, uma vez que não iniciam termos compostos. Nesses casos, excluíram-se os itens, pois foram considerados como erro de parâmetro nas configurações utilizadas para a extração das palavras-chave. Mesmo fazendo ajustes e modificando os parâmetros, os resultados não se mostraram diferentes, o que leva a concluir que são limitações do próprio SE.

Para realizar a limpeza propriamente dita, compararam-se manualmente todos os itens pela verificação dos seus contextos. Valendo-se da lista do Excel, que ajudou a organizar alfabeticamente os candidatos, o contexto só pôde ser verificado no ambiente do SE.

A Figura 21, a seguir, é um recorte da tela do SE, mostrando o item "dados complementares". Percorrendo-se a lista de cima para baixo, esse item está colocado como sendo o primeiro item que não corresponde a "nomes próprios" (nomes próprios é um dos critérios de exclusão).

Figura 21. Visualização parcial da tela de palavras compostas, destacando “dados complementares”

Multi-word	Score	F	RefF
<input type="checkbox"/> votorantim metais	W 352.83	287	560
<input type="checkbox"/> relatório votorantim	W 90.15	71	0
<input type="checkbox"/> relatório votorantim metais	W 83.87	66	0
<input checked="" type="checkbox"/> dados complementares	W 63.95	51	396
<input type="checkbox"/> asseguração externa	W 63.78	50	0
<input type="checkbox"/> operações internacionais	W 62.17	49	138
<input type="checkbox"/> forma de gestão	W 52.54	44	1,635
<input type="checkbox"/> consultoras natura	W 48.64	38	41
<input type="checkbox"/> número total	W 39.98	48	12,321
<input type="checkbox"/> vide página	W 38.58	30	52
<input type="checkbox"/> cadeia de valor	W 37.86	35	4,331
<input type="checkbox"/> instituições fortes	W 36.97	29	282
<input type="checkbox"/> asseguração externa vide	W 36.16	28	0
<input type="checkbox"/> gases de efeito	W 35.08	41	11,475
<input type="checkbox"/> emprego digno	W 34.38	27	356
<input type="checkbox"/> índice remissivo	W 34.15	27	512
<input type="checkbox"/> asseguração externa vide página	W 33.65	26	0
<input type="checkbox"/> gases de efeito estufa	W 33.13	36	9,137
<input type="checkbox"/> visão de sustentabilidade	W 32.08	25	223
<input type="checkbox"/> relatório anual	W 31.37	27	2,607

A ferramenta computou 51 ocorrências dessa expressão no *corpus*. Então, seguindo-se o procedimento, ao clicar sobre esse número da frequência, o programa abre uma janela contendo todas essas ocorrências e seus contextos (Figura 22), para verificar se a expressão atende aos demais critérios.

Figura 22. Visualização parcial da janela aberta pelo clique na frequência de "dados complementares".

The screenshot shows the Sketch Engine search results page. The search term is "Term 51 (320.19 per million)". The results are displayed in a table with columns for file ID, description, and page information. The first column contains file IDs starting with "file511907...". The second column contains descriptions of various GRI indicators and their related data. The third column indicates the page number for each result, with "(dados complementares e/ou DMA) Asseguração externa" appearing in parentheses next to the page number. The interface includes a search bar, navigation buttons like "Next" and "Last", and a sidebar with various tools and options.

File ID	Description	Page Info
file511907...	Índice remissivo Descrição GRI	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	para provedores de capital e governos	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	em unidades operacionais importantes	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	infraestrutura e serviços oferecidos	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	em unidades operacionais importantes	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	EN1 - Materiais usados por peso e volume	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	provenientes da reciclagem	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	de energia dentro da organização	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	EN8 - Total de retirada de água por fonte	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	total de água reciclada e reutilizada	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	. Não aplicável. ASPECTO EMISSÕES	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	de emissões de gases de efeito estufa	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	, por tipo e método de disposição	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	por categoria de produtos	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	com leis e regulamentos ambientais	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	como do transporte de seus empregados	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	com proteção ambiental, por tipo	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	critérios ambientais	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	por meio de mecanismo formal	Página (dados complementares e/ou DMA) Asseguração externa
file511907...	por faixa etária, gênero e região	Página (dados complementares e/ou DMA) Asseguração externa

Ao analisar as ocorrências da expressão "dados complementares", notou-se que ela ocorreu em apenas um dos textos do *corpus* (o texto está identificado com o código **file511907**, na Figura 18). Como esse era um dos critérios de exclusão mencionados no Capítulo 3, de que o item deveria ocorrer no mínimo em dois textos, o item foi excluído da planilha. Esse procedimento de abrir cada ocorrência clicando no número da frequência foi realizado em todos os itens, verificando se atendiam aos critérios de seleção. Quando necessária, a exclusão foi realizada na planilha do programa Excel, por ser editável.

A Figura 23, a seguir, mostra uma parte da planilha de Excel ordenada alfabeticamente, com a limpeza já realizada. Referente à lematização, é possível observar no exemplo da linha 6 que o item "ações" deveria ter sido lematizado para "ação", e isso contradiz o fato de que o programa lematiza todas as palavras. É possível supor, portanto, que certas palavras ocorrem significativamente com maior frequência no plural do que na sua forma singular e, por esse motivo, foram computadas na forma plural.

Figura 23. Visualização parcial da lista das palavras agrupadas, em ordem alfabética, após a limpeza.

	A	B	C	D
1	WORD	score	F	REF_F
2	abastecer	7.46	13,00	7,50
3	abranger	4.11	20,00	30,30
4	acelerar	4.97	14,00	15,51
5	acionistas	42.44	33,00	0,00
6	ações	6.03	193,00	222,92
7	ações de mitigação	7.13	5,00	482,00
8	adicionar	3.94	18,00	28,30
9	adotadas	13.56	10,00	0,00
10	adotar	11.05	8,00	0,00
11	afastamento	9.84	28,00	15,16
12	agilidade	6.02	10,00	7,10
13	água	6.35	269,00	296,44
14	alcançar	9.23	49,00	32,71
15	alinhar	19.78	30,00	5,41
16	alterações climáticas	10.82	13,00	13,91
17	ambiental	14.86	258,00	118,24
18	ambiente de trabalho	7.24	13,00	32,20
19	ampliar	7.03	68,00	64,00
20	anticorrupção	24.75	20,00	312,00

Para a limpeza dessa lista organizada no Excel, conforme os critérios estabelecidos, comparou-se a lista alfabética (organizada na planilha do Excel) com a lista ordenada pelo *score* (índice de chavicidade), fornecida automaticamente pelo programa. A Figura 24, a seguir, apresenta o ordenamento pelo *score*, no SE; os itens tachados correspondem aos que foram excluídos na limpeza.

Figura 24. Visualização parcial dos itens excluídos conforme os critérios de limpeza estabelecidos.

Multi-word	Score	F	RefF
<input type="checkbox"/> votorantim metais	352.83	287	560
<input type="checkbox"/> relatório votorantim	90.15	71	0
<input type="checkbox"/> relatório votorantim metais	83.87	66	0
<input type="checkbox"/> dados complementares	63.95	51	396
<input type="checkbox"/> asseguração externa	63.78	50	0
<input type="checkbox"/> operações internacionais	62.17	49	138
<input type="checkbox"/> forma de gestão	52.54	44	1,635
<input type="checkbox"/> consultoras natura	48.64	38	41
<input type="checkbox"/> número total	39.98	48	12,324
<input type="checkbox"/> vide página	38.58	30	52
<input checked="" type="checkbox"/> cadeia de valor	37.86	35	4,331
<input type="checkbox"/> instituições fortes	36.97	29	282
<input type="checkbox"/> asseguração externa vide	36.16	28	0
<input type="checkbox"/> gases de efeito	35.08	41	11,475
<input type="checkbox"/> emprego digno	34.38	27	356
<input type="checkbox"/> índice remissivo	34.15	27	512
<input type="checkbox"/> asseguração externa vide página	33.65	26	0
<input checked="" type="checkbox"/> gases de efeito estufa	33.13	36	9,137
<input type="checkbox"/> visão de sustentabilidade	32.08	25	223
<input type="checkbox"/> relatório anual	31.37	27	2,607
<input type="checkbox"/> a votorantim metais	31.14	24	3
<input type="checkbox"/> a votorantim	31.10	24	26
<input type="checkbox"/> temas materiais	29.87	23	8
<input type="checkbox"/> três lagoas	28.65	39	17,216
<input checked="" type="checkbox"/> consumo de água	28.44	31	9,339
<input type="checkbox"/> igualdade de gênero	27.43	24	3,131
<input type="checkbox"/> avaliação de fornecedores	27.17	21	169

O ordenamento pelo *score* é importante porque indica os itens estatisticamente significativos. Porém, conforme dito anteriormente, somente pela análise manual é possível confirmar a validade do dado obtido, já que não é possível incluir os critérios desta pesquisa nas ferramentas do SE. O procedimento de limpeza dos dados ordenados pelo *score* apenas na planilha do Excel poderia ter sido uma solução mais rápida. No entanto, quando se filtraram os dados por esse parâmetro, os valores não foram lidos como números, e sim como texto. Isso causou um ordenamento que começou em 99.99 até 9, seguido de 89.99 até 8, e assim sucessivamente. Por esse motivo, optamos por acompanhar o *score* diretamente no SE.

A Figura 25 mostra a nova janela, referente à ferramenta Concordanciador, aberta pelo clique na frequência do candidato "consumo de água". Essa ferramenta lista todos os

exemplos da palavra de busca (ou palavras-chave, como foi o caso), sejam palavras ou frase do *corpus*, acompanhadas de seu contexto à direita e à esquerda da palavra ou frase buscada.

Figura 25. Visualização parcial da janela dos contextos de "consumo de água", com 31 ocorrências no *corpus* de português.

The screenshot shows the Sketch Engine interface. On the left is a navigation menu with options like Home, Search, Word list, Word sketch, Thesaurus, Sketch diff, Keywords/terms, Corpus info, Manage corpus, My jobs, and User guide. Below this is a section for Save, Make subcorpus, View options (KWIC, Sentence), Sort (Left, Right, Node, References, Shuffle), and a Shuffle button. The main area displays search results for the term 'consumo de água'. At the top, it shows 'Term 31 > Sort Left 31 (194.63 per million)'. Below that, there are navigation controls for page 1 of 2, and buttons for Next, Last, and Concordance sorting. The results are a list of concordance lines, each starting with a file identifier (e.g., file652188...) and followed by a snippet of text where the search term is highlighted in red. The text snippets are aligned to the right of the search term. At the bottom, there are more navigation controls for page 1 of 2.

4.2 Seleção de termos

Conforme mencionado anteriormente, o procedimento de abrir cada janela das concordâncias foi realizado para confirmar o item como candidato ou se ele deveria ser excluído. Na análise dos contextos, verificaram-se os seguintes aspectos: origem da ocorrência (em qual texto do *corpus* a expressão foi encontrada) e contexto, por meio das linhas de concordância. A palavra de busca foi alinhada de três formas: à esquerda, para que pudessem ser analisadas as palavras à esquerda da expressão em ordem alfabética; à direita, para analisar as palavras à direita da expressão em ordem alfabética comparando os resultados; e também centralizada, conforme ilustrado na Figura 25.

O alinhamento da palavra de busca à esquerda trouxe resultados significativos quanto às colocações. A primeira observação é em relação à origem dessas ocorrências: a expressão foi identificada em pelo menos três textos (identificados por “file652188”, “file652187” e “file511907” na coluna em cor azul à esquerda, na Figura 25), mostrando que o candidato “consumo de água” está apto a figurar na próxima etapa da análise. Assim, ao observar as

linhas de concordância, verificou-se que, das 31 ocorrências, 11 apresentaram-se combinadas da seguinte forma:

- 6 ocorrências com co-ocorrente verbal (“reduzir o consumo de água”) (Figura 26);
- 5 ocorrências com co-ocorrente nominal (“redução do consumo de água”) (Figura 27).

Figura 26. Tela das concordâncias com filtro aplicado para o co-ocorrente verbal

Query **consumo, de, água** 21 > Positive filter (excluding KWIC) **reduzir** 6 (37.67 per million) ⓘ

file652188... desenvolvemos têm como objetivo: • **Reduzir** o **consumo de água** , ampliar seu reaproveitamento e diminuir a
file652188... , na própria operações para **reduzir** o **consumo de água** e energia nas fábricas. Fizemos um mapeamento
file652188... mais sugerem melhorias e ideias para **reduzir** o **consumo de água** , energia e de insumos." Na divisão Madeira, os
file652188... os níveis de produção desejáveis • **Reduzir** o **consumo de água** por hectare de plantio na irrigação em 50% •
file652188... da Deca e Hydra-Corona, que permitem **reduzir** o **consumo de água** e energia durante seu uso quando comparados aos
file511907... dos demais empregados. Além de **reduzirem** o **consumo de água** , as medidas possibilitaram o desligamento

Figura 27. Tela das concordâncias com filtro aplicado para o co-ocorrente não verbal

Query **consumo, de, água** 21 > Positive filter (excluding KWIC) **redução** 5 (31.39 per million) ⓘ

file652188... 1 m3 de produto. Para promover a **redução** do **consumo de água** No último ano, na Unidade Taquari (RS), concluí
file652188... , as principais iniciativas para a **redução** do **consumo de água** estiveram relacionadas ao aprimoramento dos
file652188... para reafirmar que nossas ações de **redução** do **consumo de água** estão no caminho certo. Para dar transparência
file652188... 56 Não temos como medir o impacto da **redução** de **consumo de água** e energia de nossos produtos economizadores,
file652187... monitorados. Como esperado, houve **redução** do **consumo de água** e energia elétrica, reflexo da crise hídrica e

A busca pelos equivalentes procurou estabelecer uma correlação funcional por meio da comparação entre as concordâncias geradas pela ferramenta. Foram analisados os termos e fraseologias do português e os termos e fraseologias do inglês que desempenhassem a mesma função nos respectivos textos. Essa análise foi realizada com o uso concomitante das ferramentas Concordanciador e *Word Sketch* a partir da lista de palavras-chave compostas (geradas pela ferramenta *Keywords/Terms*). Tomando-se como exemplo a fraseologia “redução do consumo de água”, a fim de identificar possíveis equivalentes funcionais, percorremos a lista de palavras-chave à procura de itens que contivessem como núcleo terminológico (NT) o termo “*water*”. Foram encontrados os seguintes itens, listados a seguir com as respectivas frequências no *corpus* de estudo:

- “*water conservation*” (F = 43);
- “*water use*” (F = 39);
- “*water risk*” (F = 18);
- “*water stress*” (F = 17);

- “*water consumption*” (F = 17);
- “*water conservation awareness*” (F = 13);
- “*water usage*” (F = 12);
- “*water scarcity*” (F = 9);
- “*water stewardship program*” (F = 6);
- “*(manufacturing) water intensity*” (F = 6);
- “*reducing water*” (F = 6).

Nessa lista, observam-se as ocorrências de “*water use*”, “*water consumption*” e “*water usage*”. Para validar qual delas seria o equivalente de “consumo de água” mais adequado funcionalmente, tomamos emprestada a classificação de Bevilacqua (2004), que categorizou as unidades fraseológicas especializadas eventivas (UFEs Eventivas) (ou co-ocorrentes) em relação aos processos de aproveitamento de recursos naturais. A pesquisa de Bevilacqua, deu enfoque exclusivo ao termo “energia”. Partindo da premissa de que “água” é um recurso natural, assim como “energia”, comparamos essa classificação nas ocorrências levantadas na nossa investigação.

Das ocorrências listadas acima, “*consumption*” e “*use*” seriam de uma mesma categoria, conforme a proposta de Bevilacqua (2004, p. 77). A priori, ambos poderiam ser apresentados no verbete “água” como equivalentes de “consumo de água”. No entanto, no âmbito da Sustentabilidade, distinguem-se “*water use*” de “*water consumption*”: “*water use*” descreve o total de água retirada de sua fonte para ser utilizada nos processos industriais. “*Water consumption*” é a quantidade de água usada que não retorna à fonte original depois de ser retirada. O consumo ocorre quando há perda de água por evaporação ou pela incorporação a produtos (por exemplo, plantas), se tornando indisponível para reúso (REIG, 2013). Com base nessa distinção, o equivalente mais adequado para “consumo de água”, portanto, é “*water consumption*”.

Entretanto, a aplicação dessa classificação no *corpus* de estudo mostrou algumas diferenças em relação aos NTs. De modo geral, esses núcleos terminológicos identificados não correspondem a recursos naturais como a classificação de Bevilacqua propõe. A maioria deles se refere a substantivos abstratos, característicos do âmbito empresarial. Os exemplos a seguir partem de NEs e são seguidos de exemplos de NTs identificados no *corpus* de estudo:

- *Captar* (NE): investimentos (NT); capital (NT); tendências (NT).
- *Acumular* (NE): inflação (NT); negócios (NT); valor (NT); lucro (NT).

- *Concentrar* (NE): esforços (NT); portfólio (NT); atuação (NT); investimentos (NT).
- *Gerar* (NE): valor (NT); impacto (NT); renda (NT); riqueza (NT); benefícios (NT).

A aplicação dessa classificação foi adaptada ao presente trabalho para que abrangesse uma parcela do *corpus* de estudo, de modo que pudesse abranger processos empresariais, já que Bevilacqua (2004) referiu-se apenas a processos relativos a recursos naturais.

Dando continuidade à análise, a ferramenta *Word Sketch* disponibilizada pelo SE possibilitou a visualização das colocações ou *clusters* (agrupamentos) mais frequentes. Essa ferramenta complementa a análise por apresentar um resumo bastante organizado do comportamento gramatical e colocacional das palavras. No *Word Sketch*, as colocações são separadas de acordo com as relações gramaticais que apresentam (THOMAS, 2015), ordenadas pela frequência no *corpus* de estudo junto com o valor do *score*. A Figura 28 mostra como são organizadas as informações que essa ferramenta fornece.

Figura 28. Resumo das colocações de uma palavra-chave com a ferramenta *Word Sketch*.

água (<i>noun</i>)			1
RelSustenta freq = <u>269</u> (1,688.84 per million)			
<u>n_modifier</u>	2		22.30
limpo	14	3	12.51
6 . Água limpa e saneamento			
novo	8		11.79
reutilizado	7		11.72
subterrâneo	7		11.58
potável	3		10.60
pluvial	3		10.58
superficial	3		10.52
modifies			
			38.66
consumo	31		12.13
consumo de água		4	
volume <u>10</u>	13		10.69
tratamento <u>3</u>			
retirada	7		11.01
total <u>7</u>	11		10.04
percentual <u>4</u>			
recirculação	4		10.21
fluxo	3		9.72

Na Figura 28, os seguintes aspectos são observados:

1) **Cabeçalho principal:** a linha 1 apresenta os seguintes dados: lema (água), classificação morfológica (*noun* — substantivo), classificação morfológica alternativa (*alternative PoS* — quando a palavra apresenta classificação em outra categoria, por exemplo, verbo e adjetivo, mas nesse caso, não há); a linha 2 apresenta: nome do *corpus*, frequência do lema.

2) Cabeçalho de categoria gramatical: esse cabeçalho apresenta o nome da categoria: observando a figura, nota-se que a ocorrência de “água” apresenta duas categorias, *n-modifier* e *modifies*.

“*N-modifier*” apresenta os casos em que a palavra de busca (“água”, que é um *noun* = substantivo) está sendo modificada por alguma outra palavra, por exemplo, um adjetivo (por exemplo, “limpo”, “novo” etc.). Com a lematização, as palavras são apresentadas nas formas canônicas, por isso aparece “limpo”, e não “limpa”. No entanto, ao clicar na frequência, visualiza-se a forma flexionada da ocorrência no texto (por exemplo, “água limpa”). Os casos observáveis das fraseologias recorrentes do glossário são oriundos desse campo (como será explicado na subseção 4.3 - *Modelo para tratamento das entradas no glossário*).

“*Modifies*” são os casos em que a palavra de busca (neste caso, “água”) modifica outras, por exemplo, outros substantivos. Na figura, onde mostra o número 4, é a palavra “consumo” que está sendo modificada por “água”, formando “consumo de água”. Ao clicar no valor da frequência (por exemplo, 31 para a palavra “consumo”), são exibidas as concordâncias em uma nova tela, assim é possível visualizar o contexto. As fraseologias identificadas para compor o glossário são oriundas desse campo.

3) Frequência de cada colocação: pode ser clicada para exibir as concordâncias; ao lado da frequência, é exibido o valor do *score*.

4) Uso mais típico da colocação: as partes em cinza referem-se às combinações mais longas e comuns, mostrando a realização mais frequente, o que ajuda a entender o uso mais típico da colocação. Cada quadro corresponde a uma função sintática; por exemplo, *n_modifier* refere-se a qualquer palavra que modifica o n (*noun*, substantivo), sendo um adjetivo ou outro substantivo. As palavras em negrito se referem às principais ocorrências e também foram lematizadas, o que foi confirmado clicando-se na frequência.

Na Figura 28, observa-se o resultado da aplicação da ferramenta *Word Sketch* para a palavra-chave “água”. As frases em cor cinza (sem negrito) mostram exemplos de como a palavra de busca combina suas colocações: no quadro visualizado na parte superior da Figura, são apresentadas as combinações em que a palavra “água” funciona como um substantivo sendo modificado (indicado por *n_modifier*, em que n é *noun* — substantivo — e *modifier* são os adjetivos combinados). O quadro logo abaixo apresenta as ocorrências em que a palavra água está modificando outras palavras (substantivos), indicadas por “*modifies*”. A ferramenta é bastante intuitiva, porque faz uma etiquetagem gramatical ao compilar o *corpus*, facilitando esse tipo de análise. Eventualmente, ocorreram erros, conforme já mencionado anteriormente.

Voltando brevemente à Figura 23, a palavra-chave "água" encontra-se na linha 13, entre os 20 primeiros itens da planilha, após realizada a limpeza. No entanto, como os demais itens que estavam posicionados alfabeticamente perto dela foram excluídos por se encaixarem em algum dos critérios de limpeza estabelecidos, só seria possível encontrar outro candidato a termo percorrendo-se a lista até a letra C, em que consta a expressão "consumo de água". Com a ferramenta *Word Sketch*, foi possível detectar essa ocorrência com mais agilidade, principalmente devido às diferenças entre os dois idiomas, porque a construção morfossintática do inglês é diferente do português. Assim, foi possível observar que as colocações mais frequentes com a palavra água foram "consumo de água" e "retirada de água". Ao examinar esses itens, nota-se que "consumo de água" consta na lista unificada do Excel (com 31 ocorrências), mas está agrupado com outras formas da palavra "consumo", sendo esta a mais frequente, seguida de "consumo de energia", com 25 ocorrências. Os critérios da estruturação da proposta do glossário serão detalhados mais adiante.

Por essa razão, a utilização da ferramenta *Word Sketch* foi útil. Outro dado interessante que esse exemplo nos traz é que o *score* da categoria gramatical "*modifies*" é maior do que o *score* de "água" como substantivo modificado por adjetivos. Esse dado se comprova no exame das concordâncias, que mostraram que ocorrências como "água limpa" e "água reutilizada", constantes no quadro superior da Figura 28, não foram estatisticamente relevantes.

A ferramenta *Word Sketch* foi igualmente aplicada ao *corpus* de inglês, para o equivalente de água, *water*. A diferença é bastante considerável: no *corpus* de língua portuguesa, a frequência da palavra água foi de 269 ocorrências, e no de língua inglesa, 737 ocorrências, conforme mostra a Figura 29, a seguir.

Figura 29. Tela da ferramenta *Word Sketch* para o termo "water".

Sketch Engine RelSustenta_EN

Home
Search
Word list
Word sketch
Thesaurus
Sketch diff
Keywords/terms
Corpus info
Manage corpus
My jobs
User guide

Save
Change options
Cluster
Sort by freq
Hide gramrels
More data
Less data
Sketch grammar
Translate
- Arabic
- Bulgarian
- Czech
- Dutch
- French
- German
- Italian
- Polish
- Russian
- Spanish

water (noun) Alternative PoS: verb (freq: 1)
RelSustenta_EN freq = 737 (3,131.71 per million)

modifiers of "water"	17.91	nouns and verbs modified by "water"	66.49	verbs with "water" as object	14.25	"water" and/or ...	11.67
clean	15 11.62	use	51 11.13	save	10 11.23	energy	14 11.45
clean water		water use		save water		water and energy	
save	6 10.33	withdrawal	16 9.99	withdraw	7 11	sanitation	9 11.37
fresh	4 9.87	water withdrawals		conserve	6 10.66	system	6 10.35
less	4 9.77	consumption	16 9.69	use	24 10.64	air	3 10.04
safe	4 9.70	water consumption		water used		waste	4 9.91
potable	3 9.50	stress	13 9.69	bring	6 10.27	food	3 9.81
surface	3 9.48	water stress		replenish	3 9.81	change	3 9.67
recycled	3 9.04	awareness	12 9.44	clean	3 9.76	climate	2 9.23
cdp	3 9.04	water conservation awareness to		reuse	3 9.72	service	2 8.48
ultrapure	2 8.93	conservation	11 9.43	need	3 9.43	management	2 8.46
replenish	2 8.93	water conservation		recycle	3 9.35		
drinking	2 8.92	intensity	11 9.35	require	4 9.34	prepositional phrases	
accessible	2 8.90	water intensity		reclaim	2 9.26	... of "water"	43 5.83
award-winning	2 8.89	scarcity	10 9.34	air	2 9.19	... to "water"	19 2.58
hot	2 8.86	water scarcity		reduce	8 9.12	"water" for ...	16 2.17
food	2 8.62	risk	13 9.32	treat	2 9.08	"water" per ...	7 0.95
change	2 8.52	water risks		associate	2 8.60	"water" from ...	5 0.68
sustainable	2 8.33	usage	10 9.26	support	2 7.69	"water" in ...	5 0.68
waste	2 8.19	water usage				... with "water"	4 0.54
n	3 8.13	source	10 9.13	verbs with "water" as subject		... for "water"	3 0.41
climate	2 8.05	day	9 9.01	message	2 11	"water" to ...	2 0.27
water	3 7.30	system	12 8.99	use	5 10.13	"water" with ...	2 0.27
energy	3 7.26	biodiversity	7 8.73	be	10 7.39	... by "water"	2 0.27
management	2 7.12	project	9 8.65				

Após a limpeza, os itens que permaneceram na lista tiveram seus contextos analisados e serão apresentados na forma de um modelo de verbete de glossário. Como a lista é extensa, apresentaremos, a seguir, um modelo do tratamento das entradas para a proposta de glossário, em que são contemplados os seguintes itens descritos em sequência.

4.3 Tratamento das entradas no glossário

A proposta deste modelo se alinha à finalidade desta pesquisa, que é apresentar uma ferramenta funcional que auxilie o tradutor especializado nas suas tarefas, bem como revisores e redatores. Para isso, o modelo está baseado na proposta de Rebecchi (2015) e Bevilacqua (2004).

4.3.1 Descrição do verbete: entradas, subentradas, equivalentes, exemplos e remissivas

Os verbetes serão compostos por entrada, subentrada, fraseologia e seu equivalente fraseológico, exemplo de uso da fraseologia extraído do *corpus* de estudo, co-ocorrente mais frequente da fraseologia, equivalente para o co-ocorrente, exemplo de uso e remissivas.

Para a seleção das entradas, foi considerado o núcleo terminológico das palavras-chave compostas, como no exemplo utilizado "consumo de água": o que determinou que essa unidade fraseológica tivesse entrada em "água" e não em "consumo" foi a natureza do núcleo: consumo diz respeito a um processo, e água se refere a um conceito. Embora "consumo" possa também ser considerado um conceito (visto que o diferenciamos de "uso"), entende-se que aqui ele deriva de consumir, portanto, configura-se mais como um processo do que um conceito nominalizado. Assim, cada entrada foi selecionada a partir das palavras-chave que permaneceram nas listas após a limpeza e o exame dos contextos de suas ocorrências. Neste modelo, estão sendo consideradas como entradas apenas os termos simples, uma vez que o objetivo é fornecer termos e fraseologias equivalentes para o tradutor. As palavras-chave constantes na lista de *multi-words* do SE (palavras compostas) serão consideradas no glossário como fraseologias. A decisão sobre a escolha da entrada foi apoiada com o uso da ferramenta *Word Sketch*, que auxiliou a visualizar essas ocorrências de modo simultâneo. As entradas são apresentadas em letras maiúsculas, na forma canônica, em negrito e sem distinção quanto ao gênero.

As subentradas são previstas para casos específicos de subcategorização do termo principal, por apresentarem conceitos diferentes desse termo (ver REBECHI, 2015). Por exemplo, a entrada "água" poderia apresentar como subentradas "água reutilizada", "água potável", entre outras. Para o exemplo que apresentaremos a seguir, no modelo do glossário, esses casos específicos não apresentaram a frequência mínima estabelecida (que era de cinco ocorrências). Nos casos em que houve frequência mínima, não houve contexto para identificar

as fraseologias (pois eram itens de listas, tabelas, títulos etc.) ou ocorreram em apenas um texto, o que determinou sua exclusão (conforme os critérios apresentados no Capítulo 3 Metodologia). No entanto, como se trata de um modelo, iremos incluir algumas dessas subentradas como forma de consolidar a proposta, especialmente porque essas subentradas se mostraram mais abundantes no inglês do que no português (ver Figuras 28 e 29, que mostram o *Word Sketch* para “água” e “water”).

Tanto as entradas quanto as subentradas são imediatamente seguidas por seus equivalentes, abreviados por “**Eq.**” em negrito. Os equivalentes foram selecionados a partir da comparação das listas de palavras-chave e termos, bem como dos resultados apresentados com a ferramenta *Word Sketch*, conforme descrito na seção 4.2.

4.3.2 Fraseologias e exemplos

Essas unidades são as mais representativas, resultado de toda a análise dos contextos propriamente. De maneira objetiva, as fraseologias, ou UFEs — unidades fraseológicas especializadas — são exibidas sob a abreviatura de “**Fras.**” no glossário, para que o consulente encontre imediatamente a colocação adequada à sua busca. Serão tratadas como fraseologias, de acordo com Bevilacqua (2004): água seria um termo ou unidade terminológica porque “possui um referente específico”, ao passo que “consumo de água” é uma UFE porque denota um processo da área.

Como os significados se dão pelas relações das palavras umas com as outras, e não de forma isolada, a busca por padrões de ocorrência nos ajudou a identificar os co-ocorrentes mais frequentes dos termos identificados. A utilização de uma metodologia apoiada nas ferramentas computacionais, como o Concordanciador e o *Word Sketch*, facilitou essa identificação, embora as ocorrências encontradas não sejam limitadas ao nosso *corpus* de estudo.

Dessa forma, a função das fraseologias é mostrar ao consulente as principais combinações dos termos, fornecendo os equivalentes fraseológicos conforme a análise das linhas de concordância. Para isso, são apresentados os principais co-ocorrentes, identificados no modelo com a sigla **COFras**. Os equivalentes fraseológicos serão identificados pela sigla **Eq,COFras**.

A respeito dos exemplos, conforme aponta Rebechi (2015), em obras terminográficas construídas com base em *corpora*, as evidências do uso fornecem exemplos reais, ao contrário

das tradicionais obras lexicográficas que criam exemplos apenas para encaixar nas definições. “De fato, a extração de exemplos contextualizados em textos autênticos de áreas especializadas certamente resulta em extratos significativos que, em geral, contribuem para ampliar a compreensão do consulente” (REBECHI, 2015, p. 263). Como a proposta desse glossário não inclui definições, os exemplos escolhidos mostram as colocações recorrentes, cujos contextos fossem o mais extenso possível para exibir as fraseologias.

Dessa forma, foram extraídos exemplos nos dois idiomas a partir das linhas de concordância dos termos e fraseologias neste estudo, mostrando o contexto de uso. Ao final de cada exemplo, são indicados de quais textos se originaram.

As remissivas, quando possível, foram incluídas, pois elas podem auxiliar o consulente ao estabelecer relações sintáticas e semânticas, norteando a busca por uma melhor compreensão das informações (REBECHI, 2015). Podem ser remissivas por semelhança ou por diferença, exibidas no glossário como “*Ver também*” ou “*Comparar*”. São o último item do verbete, sem negrito.

A seguir, apresentaremos o modelo, conforme descrito.

4.3.3 Apresentação do modelo

A seguir, são apresentados dois exemplos do modelo proposto, um com o termo **água** e outro com o termo **energia**, seguindo os critérios descritos anteriormente.

Figura 30. Modelo 1.

ÁGUA. Eq. WATER. Fras. consumo de água. **Eq. Fras. water consumption. Ex.** “*We have made great improvements in our ability to track water consumption, but this continues to be a challenge [...]*” (EN_CIS_RS2016.txt). **COFras.** redução do consumo de água. **Eq.COFras. reduce water consumption. Ex.** “*We have reduced water consumption through production efficiency improvements [...]*” (EN_ALL_RS2017.txt). **COFras.** reduzir o consumo de água. **Eq.COFras. reduce water consumption. Ex.** “[...] *we understand the importance of reducing water consumption as much as we can in our operations [...]*” (EN_CIS_RS2016.txt). Ver também: consumo de **energia**.

Figura 31. Modelo 2.

ENERGIA. Eq. ENERGY Fras. eficiência energética. **Eq. Fras. energy efficiency**
Ex. "[...] Cisco and our partners are boosting *energy efficiency* while reducing the total carbon footprint [...] (EN_CIS_RS2016.txt). **COFras.** aumentar a eficiência energética **Eq.COFRas. (1) increase energy efficiency.** **Ex.** "[...] and work with our supply chain partners to reduce their impact and *increase their energy efficiency.*" (EN_CIS_RS2016.txt). **Eq.COFRas. (2) improve energy efficiency** **Ex.** "[...] We continuously strive to *improve energy efficiency*, reduce emissions, and reduce our footprint throughout our operations. (EN_INT_RS2016.txt) **Fras.** consumo de energia **Eq. Fras. energy consumption.** (EN_CIS_RS2016.txt). **Ex.** "[...] we are scaling, or adjusting, the energy consumed by ASICs to achieve performance standards and minimize *energy consumption.*" (EN_CIS_RS2016.txt) **COFras.** reduzir o consumo de energia. **Eq. COFRas. reduce energy consumption.** **Ex.** "[...] The systems not only reduce *energy consumption* through daylight harvesting and occupancy sensing, but also create a platform [...]" (EN_INT_RS2016.txt).

No exemplo com o termo “energy”, a fraseologia “energy efficiency” foi identificada como equivalente para “eficiência energética”, demonstrando as singularidades nos padrões sintáticos de cada idioma: no português, usa-se um adjetivo (“energética”) para modificar o substantivo “energia”, ao passo que no inglês essa modificação é realizada com outro substantivo. Embora nosso foco não seja a descrição morfossintática dessas estruturas, são aspectos relevantes para as escolhas tradutórias, reafirmando a importância da adequação tanto da forma quanto do sentido.

5 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Ao chegar nesse ponto, é necessário refletir sobre os diversos aspectos que permearam o desenvolvimento desta pesquisa. Alguns deles são bastante positivos, mas é preciso ser crítico e reconhecer eventuais divergências e dificuldades, inerentes aos processos de aquisição do conhecimento e aprimoramento de técnicas de pesquisa. Portanto, nesta seção, faremos algumas reflexões acerca da metodologia e dos resultados, à luz das teorias que nos apoiam, e, finalmente, algumas considerações a respeito da pesquisa como um todo.

Com relação à metodologia empregada, não há dúvidas de que a Linguística de *Corpus* exerce um papel essencial junto aos estudos das linguagens. Conhecer suas inúmeras possibilidades de trabalho é fundamental, porém requer tempo. Há muitas ferramentas disponíveis na forma de *softwares* cada vez mais bem elaborados, com importante apoio das áreas computacionais. De outro lado, esse é um dos motivos que nos faz, como terminólogos e pesquisadores, construir *corpora* bastante criteriosos, conforme os princípios que mencionamos neste trabalho, no Capítulo 3. A tarefa não é fácil, pois existem, ainda, divergências consideráveis a respeito do que são critérios razoáveis para construir um *corpus*.

Entre as questões mais complexas nesse sentido está a da representatividade, um dos critérios de construção do *corpus*. Alguns autores defendem, como é o caso de Berber Sardinha (2004), que a representatividade está atrelada à extensão: “[...] quanto maior o número de palavras, maior será a chance de o *corpus* conter palavras de baixa frequência” (p. 25). Isso porque as palavras de baixa frequência compõem a maioria das palavras de uma língua. Porém, em uma investigação que busca observar os padrões de combinação entre as palavras, esse critério de extensão acaba se neutralizando diante de outros mais relevantes. Como apontou Koester (2010), que defende os *corpora* ‘pequenos’: “o propósito de um *corpus* informatizado é permitir o armazenamento eletrônico e a análise automática de grandes quantidades de texto, o que não seria possível manualmente”³¹ (KOESTER, 2010, p. 66). Portanto, podemos considerar o *corpus* deste estudo pequeno (cerca de 319 mil palavras) em relação aos imensos *corpora* de língua geral existentes (por exemplo, *British National Corpus*, com 96 milhões de palavras, ou o *English Web 2015*, com 15 bilhões). No entanto, o autor faz uma ressalva importante, com base nos comentários de Sinclair (2004): ainda que seja possível obter resultados válidos de um *corpus* pequeno, muito provavelmente tais resultados serão limitados em certos aspectos, como questões envolvendo o léxico

³¹ No original: “[...] the point of a computer-based corpus is to allow the electronic storage and machine analysis of huge amounts of text which could not be handled manually.”

(quantidade). Nesse quesito, é preciso reconhecer que o desenho do nosso *corpus* de estudo acabou demonstrando certas limitações. Isso foi possível de perceber quando se estabeleceram os critérios para extrair o máximo de relevância quanto à frequência de ocorrências para palavras-chave, sem conseguir estabelecer uma frequência mínima de fraseologias que se desejava encontrar. A inclusão desse parâmetro nos obrigaria a aumentar o *corpus* ou mudar o tipo de composição (*corpus* paralelo), algo a ser pensado para futuros estudos.

Ainda assim, em estudos com *corpora* pequenos, todas as ocorrências dos itens de alta frequência podem ser examinadas, o que se mostra vantajoso; e o fato de que nas pesquisas com *corpora* pequenos fica mais evidente a relação entre o *corpus* e os contextos em que se inserem os textos (KOESTER, 2010). Os textos selecionados compreendem um recorte nesse âmbito de especialidade da Sustentabilidade, cujo propósito não foi incluir toda a gama de variações possíveis, mas sim oferecer um caminho viável e objetivo para analisar os padrões da linguagem desses Relatórios de Sustentabilidade.

De modo geral, o levantamento realizado com o programa Sketch Engine foi considerado bastante eficaz. No entanto, durante a etapa de verificação dos contextos, o *software* sofreu uma atualização e não houve tempo hábil de se estudar as novas funções, tampouco se as dificuldades com alguns parâmetros poderiam ter sido resolvidas nessa nova versão. Portanto, toda a conferência dos contextos e concordâncias foi realizada na versão anterior, não sendo possível, por ora, detectar se houve prejuízo nas análises. Além disso, levou-se muito tempo na etapa de limpeza das listas, fato que poderia ter sido evitado caso o programa tivesse funcionado com as *stop-lists*.

As análises do *corpus* de estudo trouxeram dados bastante promissores, principalmente com relação às fraseologias especializadas e ao potencial de criação de uma base terminológica com foco nessa linguagem. De fato, os textos dos Relatórios de Sustentabilidade configuram-se como textos especializados, especialmente pelos aspectos pragmáticos de uso e situação comunicativa. Seu caráter evidentemente transdisciplinar reforça os pressupostos da TCT de que tais estruturas não pertencem à linguagem especializada, mas sim adquirem o caráter de especialidade pelas condições em que se realizam.

Exemplo disso é a expressão analisada “consumo de água”, que poderia ser utilizada em âmbitos não especializados, porém sem carregar o sentido implícito de que a água, aqui, se refere às águas de rios utilizadas em processos industriais. Além do termo “água”, o sentido atribuído a “consumo”, principal co-ocorrente do termo nesse âmbito, corrobora a classificação das unidades fraseológicas especializadas proposta por Bevilacqua (2004), para

quem “consumo” se refere aos processos finais de emprego de recursos naturais em processos produtivos relacionados a meio ambiente.

Conforme visto no Capítulo 2, a tradução de abordagem funcionalista entende o texto como uma ação comunicativa, em que há pelo menos dois interlocutores: de um lado, um deles com intenção de comunicar algo e, de outro, um interlocutor atribuindo uma função para o que é comunicado. A intenção dos Relatórios de Sustentabilidade, como um ato comunicativo, é disponibilizar informações a respeito dos impactos de uma empresa em relação a diversos âmbitos. Um leitor que não esteja situado em um mesmo contexto cultural, ao se deparar com uma tradução, deve encontrar condições de realizar a leitura do texto de chegada de maneira fluida e natural. Essas condições são fornecidas por diversos aspectos intratextuais, como as relações estabelecidas entre os elementos de um texto. As fraseologias e seus equivalentes são exemplos disso. Por isso, as equivalências que compõem a proposta do glossário partem da observação em contexto autêntico de realização: textos com características similares e construídos sobre as mesmas bases metodológicas. Nesta investigação, foi possível encontrar equivalentes adequados ao propósito comunicativo utilizando as ferramentas do SE, que possibilitaram analisar os contextos.

Conforme observado nos exemplos de verbetes, por meio da comparação dos contextos e da análise das listas das palavras-chave, com auxílio da classificação proposta por Bevilacqua (2004), foi possível encontrar equivalentes que cumpriam a mesma função nos textos. No entanto, o segundo exemplo analisado fornece indícios de que a variação terminológica deva ser considerada para futuros estudos, uma vez que o *corpus* apresentou duas possibilidades de equivalência: “*increase energy consumption*” ou “*improve energy consumption*”, por meio de metodologia a ser discutida em futuros estudos. Além disso, a aplicação da classificação de Bevilacqua (2004) no nosso *corpus* de estudo mostrou que é necessário criar uma nova proposta de classificação, de modo que se englobem não apenas os processos relativos a recursos naturais, mas uma proposta que possa abarcar outras áreas do conhecimento e que utilizam como referência os processos descritos por Bevilacqua (2004). Isso mostra, mais uma vez, o caráter transdisciplinar dos RSs.

Finalmente, algumas considerações para futuros estudos: a GRI, ao estabelecer uma metodologia para a produção dos RSs, acaba estabelecendo alguns termos. Embora se tenha analisado apenas superficialmente o glossário oferecido no site³² por não ser o foco deste estudo, haveria possibilidade de integrar um glossário de fraseologias aos termos já inseridos

³² <https://www.globalreporting.org/standards/gri-standards-download-center/gri-standards-glossary/>

na base da GRI e, principalmente, ampliá-la com uma base metodológica mais adequada aos propósitos linguísticos e comunicativos. Além disso, em consonância com todos esses direcionamentos globais de preservação do ambiente, será inevitável ampliar essas bases de dados por meio de novas investigações que envolvam outros idiomas.

REFERÊNCIAS

ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa lingüística. *Calidoscópico*, v. 4, n. 3, p. 156-178. São Leopoldo: Unisinos, 2006.

ARAÚJO, J. S. P. Direitos Humanos, meio ambiente e sustentabilidade. *Revista Direito Ambiental e sociedade*, v. 3, n. 1, p. 289-317. Caxias do Sul: UCS, 2013. Disponível em: <<http://www.ucs.br/etc/revistas/index.php/direitoambiental/article/viewFile/3627/2077>>. Acesso em: 20 out. 2018.

BARBOSA, P. R. A. *Índice de sustentabilidade empresarial da bolsa de valores de São Paulo (ISE-BOVESPA): exame da adequação como referência para aperfeiçoamento da gestão sustentável das empresas e para formação de carteiras de investimento orientadas por princípios de sustentabilidade corporativa*. 2007. Dissertação (Mestrado em Administração) – Universidade Federal do Rio de Janeiro, Instituto COPPEAD de Administração, 2007.

BERBER SARDINHA, T. Retrospectiva/retrospective lingüística de corpus: histórico e problemática. *D.E.L.T.A.*, v. 16, n. 2. São Paulo: LAEL-PUC, 2000, p. 323-367. Disponível em: <<http://www.scielo.br/pdf/%0D/delta/v16n2/a05v16n2.pdf>>. Acesso em: 8 abr 2017.

BERBER SARDINHA, T. *Lingüística de Corpus*. Barueri, SP: Manole, 2004.

BEVILACQUA, C. R. *Unidades fraseológicas especializadas eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Tese (Doutorado em Linguística Aplicada – Léxico) - Instituto Universitario de Linguística Aplicada (IULA), Universidad Pompeu Fabra, Barcelona, 2004. 243 p.

BEVILACQUA, C. R.; REUILLARD, P. C. Gouadec e Gémard: suas contribuições para o projeto Combinatórias Léxicas Especializadas. *Synergies Brésil*, nº 11, p. 67-77, 2013.

BITARELLO, M. *Ernst Ligteringen: A importância dos relatórios de sustentabilidade e sua presença na Rio+20*. Disponível em: <<https://exame.abril.com.br/mundo/qual-e-a-importancia-dos-relatorios-de-sustentabilidade/>>. Acesso em: 20 mar. 2017.

BOURIGAULT, D.; SLODZIAN, M. Por uma terminologia textual. *Cadernos de Tradução*, n. 17, Porto Alegre: IL/UFRGS, 2004, p. 101-108.

BRASIL. Ministério Meio Ambiente. *Acordos globais*. Disponível em: <<http://www.brasil.gov.br/noticias/meio-ambiente/2012/01/acordos-globais>>. Acesso em: 26 out. 2018.

CABRÉ, M. T. *La terminología. Teoría, metología, aplicaciones*. Barcelona: Editorial Antártida/Empuries, 1993.

CABRÉ, M. T. *Terminología: representación y comunicación: una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 1999.

CABRÉ, M. T. Sumario de principios que configuran la nueva propuesta teórica y consecuencias metodológicas. In: CABRÉ, M. T.; FELIU, J. (Ed.). *La terminología científicotécnica: reconocimiento, análisis y extracción de información formal y semântica: Informe DGES PB9 9690293*. Barcelona: Universitat Pompeu Fabra; Institut Universitari de Lingüística Aplicada, 2001.

CAMPOS et al. Relatório de sustentabilidade: per l das organizações brasileiras e estrangeiras segundo o padrão da Global Reporting Initiative. *Gest. Prod.*, São Carlos, v. 20, n. 4, p. 913-926, 2013. Disponível em: <http://www.scielo.br/pdf/gp/v20n4/aop_gp0954_ao.pdf>. Acesso em: 19 mai. 2018.

CIAPUSCIO, G. E. La Terminología desde el punto de vista textual: selección, tratamiento y variación. *Organon*, v. 12, n. 26, 1998, p. 43-66.

CRESCIMENTO... O crescimento do mercado de trabalho para tradutores e intérpretes. Gaúcha ZH. 17/11/2012. Disponível em: <<https://gauchazh.clicrbs.com.br/educacao-e-emprego/noticia/2012/11/o-crescimento-do-mercado-de-trabalho-para-tradutores-e-interpretres-3953441.html>>. Acesso em: 26 out. 2018.

EVISON, J. What are the basics of analysing a corpus? In: O’Keeffe, A.; McCarthy, M. *The Routledge Handbook of Corpus Linguistic*. 2010, p. 124

GRI. GRI Standards Download Center. 2018. Disponível em: <<https://www.globalreporting.org/standards/gri-standards-download-center/>>. Acesso em: 16 jul. 2018.

KENNING, M. M. What are parallel and comparable corpora and how can we use them? In: O’Keeffe, A.; McCarthy, M. *The Routledge Handbook of Corpus Linguistic*. 2010, p. 487-500.

KOESTER, A. Building small specialised corpora. In: O’Keeffe, A.; McCarthy, M. *The Routledge Handbook of Corpus Linguistics*. 2010. p. 66-79.

KRIEGER, M.G. Terminologia revisitada. In: KRIEGER, M. G.; MACIEL, A. M. B. (org) *Temas de Terminologia*. Porto Alegre/São Paulo: Ed. Universidade/UFRGS/Humanitas/USP, 2001.

KRIEGER, M. G.; MACIEL, A. M. B.; FINATTO, M. J. B. Terminografia das leis do meio ambiente: princípios teórico-metodológicos. *TradTerm*, n. 6. São Paulo: USP, 2000. Disponível em: <<http://www.revistas.usp.br/tradterm/article/view/49520/53595>>. Acesso em: 20 out. 2018.

KRIEGER, M. G.; SANTIAGO, M. S.; CABRÉ, M. T. *Terminologia em foco: uma entrevista comentada com Maria Teresa Cabré*. Calidoscópio, v. 11, n. 3, p. 328-332, set/dez, 2013.

LEAL, E. G.; BISOL, L. Frequência de uso: Tokens e Types na harmonia vocálica. *ReVEL*, n. 14, 2017. Disponível em: <www.revel.inf.br>. Acesso em: 26 set. 2018

L’HOMME, M. C.; BERTRAND, C. Specialized Lexical Combinations: Should they be described as collocations or in terms of selectional restrictions? *Proceedings of Euralex 2000*.

Disponível em: <<https://euralex.org/publications/specialized-lexical-combinations-should-they-be-described-as-collocations-or-in-terms-of-selectional-restrictions/>>. Acesso em: 15 out. 2018

L'HOMME, M. C. *La terminologie: principes et techniques*. Montréal: Presses de l'Université de Montréal, 2004.

LUQUES, I. *Mercado para tradutor e intérprete está aquecido*. O Globo. Economia. 2010. Disponível em: <<https://oglobo.globo.com/economia/emprego/mercado-para-tradutor-intérprete-esta-aquecido-2956048>>. Acesso em: 19 maio 2018.

MACIEL, A. M. B. *Para o reconhecimento da especificidade do termo jurídico*. 2001. Tese (Doutorado em Letras: Teorias do Texto e do Discurso). Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre.

MACIEL, A. M. B. Terminologia e Corpus. *Corpora na Terminologia*. São Paulo: HUB Editorial, 2013, p. 29-45.

MARIA, A. *Relatórios de sustentabilidade: de onde vêm e para onde vão?* Entrevista com Glaucia Terreo, representante da GRI no Brasil. *Envolverde Jornalismo & Sustentabilidade*. 27 nov. 2012. Disponível em: <<http://envolverde.cartacapital.com.br/relatorios-de-sustentabilidade-de-onde-vem-e-para-onde-va-entrevista-com-glaucia-terreo-representante-da-gri-no-brasil/>>. Acesso em: 18 out. 2018.

NEVES, F. Relatórios de sustentabilidade obrigatórios? *VER Valores, Ética, Responsabilidade*. 10 Jul. 2014. Disponível em: <<http://www.ver.pt/relatorios-de-sustentabilidade-obrigatorios/>>. Acesso em: 26 set. 2018

NORD, C. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. London: Routledge, 1997.

NORD, C. *Text analysis in translation: theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi, 2005.

NORD, C. Functional approaches to translation. In: CHAPPELLE, C. *The Encyclopedia of Applied Linguistics*. New Jersey: Blackwell Publishing, 2013.

UN Documents. *Report of the World Commission on Environment and Development: Our Common Future*. 1987. Disponível em: <<http://www.un-documents.net/our-common-future.pdf>>. Acesso em: 20 mar. 2017.

PENSAMENTO VERDE. *Clube de Roma e o relatório "Os limites do crescimento" (1972)*. *Sustentabilidade*, 2. Jan. 2014. Disponível em: <<https://www.pensamentoverde.com.br/sustentabilidade/clube-roma-relatorio-limites-crescimento-1972/>>. Acesso em 20 out. 2018.

PICHT, H. Corpora como ponto de partida para a extração de dados terminológicos. *Cadernos de Tradução*, n. 17, out-dez. Porto Alegre: Instituto de Letras - UFRGS, 2004, p. 67-78.

PRATA, V. *Uma análise da linguagem da tradução presente em relatórios de sustentabilidade à luz dos estudos da tradução baseados em Corpus*. 2017. Dissertação (Mestrado em Estudos da Tradução). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8160/tde-07022018-131526/pt-br.php>>. Acesso em: 18 mai. 2018.

REBECHI, R. R. A tradução da culinária típica brasileira para o inglês: um estudo sob o enfoque da Linguística de Corpus. 2015. Tese (Doutorado em Letras - Estudos Linguísticos e Literários em Inglês). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8147/tde-29092015-162829/pt-br.php>>. Acesso em: 14 mar. 2018.

REIG, P. What's the Difference Between Water Use and Water Consumption? *World Resources Institute*, 12 mar. 2013. Disponível em: <<https://www.wri.org/blog/2013/03/what-s-difference-between-water-use-and-water-consumption>>. Acesso em: 20 nov. 2018.

REISS, K.; VERMEER, H. J. *Towards a general theory of translational action: Skopos theory explained*. London: Routledge, 2014.

RIBEIRO, Fabiana. Relatórios de sustentabilidade obrigatórios para empresas públicas e privadas. *O Globo*. 14/06/2012. Disponível em: <<https://oglobo.globo.com/economia/rio20/relatorios-de-sustentabilidade-obrigatorios-para-empresas-publicas-privadas-5209914>>. Acesso em: 23 mar. 2018.

RYDLEWSKI, C. Pesquisa aponta melhores relatórios de sustentabilidade do Brasil. *Revista Época Negócios*. 2008. Disponível em: <<http://colunas.revistaepocanegocios.globo.com/empresaverde/2008/12/09/pesquisa-aponta-melhores-relatorios-de-sustentabilidade-do-brasil/>>. Acesso em: 26 out. 2018.

SIEW, R. *Sustainability: we know it matters, but how do we measure it?* World Economic Forum, 06/09/2017. Disponível em: <<https://www.weforum.org/agenda/2017/09/struggling-to-measure-sustainability/>>. Acesso em: 24 set 2018.

SIMONOFF, J. *Smoothing methods in statistics*. New York: Springer, 1998.

SINCLAIR, J. *Trust the text*. London/New York: Routledge, 2004.

SINCLAIR, J. Corpus and Text: Basic Principles. In: WYNNE, M. *Developing Linguistic Corpora: a Guide to Good Practice*. ADHS, Tuscan Word Centre, 2004. Disponível em: <<http://ota.ox.ac.uk/documents/creating/dlc/>>. Acesso em: 18 ago. 2017.

SKETCH ENGINE. 2018. Disponível em: <<https://www.sketchengine.eu/>>. Acesso em: 10 ago. 2017.

TAGNIN, S. E. O. A identificação de equivalentes tradutórios em corpora comparáveis. I Congresso Internacional da ABRAPUI, 2007. *Anais...*, 2007. Disponível em: <http://comet.fllch.usp.br/sites/comet.fllch.usp.br/files/u30/Stella_Abrapui%202007_artigo.pdf>. Acesso em: 20 out. 2018.

TAGNIN, S. O. Glossário de Linguística de Corpus. In: TAGNIN, S. O.; BEVILACQUA, C. R. (org.) *Corpora na Terminologia*. São Paulo: HUB Editorial, p. 215-219, 2013.

THOMAS, J. E. *Discovering English with Sketch Engine (DESKE)*, Word Sketches, p. 161–176, 2015.

WAQUIL, M. L. *Tradução de textos especializados: unidades fraseológicas especializadas e técnicas tradutórias*. 2013. Dissertação (Mestrado em Letras - Estudos Linguísticos do Léxico) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: <<https://lume.ufrgs.br/handle/10183/72737> 2013>. Acesso em: 23 jul. 2018.

ZILIO, L. Terminologia textual e Linguística de *Corpus*: estudo em parceria. *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa* Porto Alegre: EDIPUCRS, 2010. p. 128-151.

ANEXO I Lista de palavras-chave – *corpus* português

Word	Score	Freq	Freq/ref
abastecer	7.46	13,00	7,50
abranger	4.11	20,00	30,30
acelerar	4.97	14,00	15,51
acionistas	42.44	33,00	0,00
ações	6.03	193,00	222,92
ações de mitigação	7.13	5,00	482,00
adicionar	3.94	18,00	28,30
adotar	11.05	8,00	0,00
afastamento	9.84	28,00	15,16
agilidade	6.02	10,00	7,10
água	6.35	269,00	296,44
alcançar	9.23	49,00	32,71
alinhar	19.78	30,00	5,41
alterações climáticas	10.82	13,00	13,91
ambiente de trabalho	7.24	13,00	32,20
ampliar	7.03	68,00	64,00
anticorrupção	24.75	20,00	312,00
aperfeiçoar	6.24	9,00	5,50
apoiar	5.76	52,00	59,50
aprendiz	4.70	12,00	13,72
aprendizes	7.28	5,00	0,00
aprimoramento	9.44	22,00	11,52
aprimorar	15.54	31,00	8,90
apuração	4.81	12,00	13,28
área comercial	5.36	5,00	8,28
áreas corporativas	8.49	6,00	131,00
áreas de conservação	8.13	6,00	1,14
áreas florestais	7.86	6,00	1,98
arrendamento	5.08	6,00	3,86
aspecto	3.64	141,00	271,43
aspectos ambientais	6.55	5,00	2,56
aspectos de direitos humanos	8.53	6,00	13,00
aspectos materiais	9.62	7,00	408,00
aspiração	12.57	20,00	6,11
assédio	4.78	6,00	4,44
asseguração	102.31	81,00	22,00
asseguração externa	63.78	50,00	0,00
assembleia	4.56	7,00	6,51
assessoramento	5.90	6,00	2,53
associação	4.00	46,00	77,61
ater	13.35	16,00	3,29

aterro	13.85	<u>22,00</u>	<u>6,05</u>
atingimento	20.27	<u>18,00</u>	<u>930,00</u>
atividades	152.93	<u>121,00</u>	<u>0,00</u>
ativos	33.65	<u>26,00</u>	<u>0,00</u>
atmosférico	7.28	<u>7,00</u>	<u>1,95</u>
atores	13.56	<u>10,00</u>	<u>0,00</u>
atua	19.83	<u>15,00</u>	<u>0,00</u>
atuação	6.57	<u>91,00</u>	<u>93,76</u>
auditado	9.54	<u>7,00</u>	<u>149,00</u>
auditor	5.83	<u>9,00</u>	<u>6,29</u>
auditores independentes	8.19	<u>6,00</u>	<u>985,00</u>
auditoria	19.66	<u>62,00</u>	<u>17,07</u>
auditoria externa	6.84	<u>5,00</u>	<u>1,50</u>
auditoria interna	6.81	<u>5,00</u>	<u>1,59</u>
avaliação	4.55	<u>160,00</u>	<u>245,76</u>
avaliação ambiental	9.32	<u>7,00</u>	<u>1,17</u>
avaliação de desempenho	11.41	<u>14,00</u>	<u>14,52</u>
avaliação de fornecedores	27.17	<u>21,00</u>	<u>169,00</u>
avançar	8.48	<u>20,00</u>	<u>11,78</u>
barragem	30.27	<u>59,00</u>	<u>8,39</u>
beneficiamento	10.90	<u>15,00</u>	<u>4,65</u>
beneficiar	5.17	<u>31,00</u>	<u>38,08</u>
beneficiário	4.75	<u>5,00</u>	<u>3,02</u>
benefícios ambientais	6.85	<u>5,00</u>	<u>1,45</u>
biodiversidade	29.78	<u>61,00</u>	<u>9,10</u>
biodiversidade brasileira	7.98	<u>6,00</u>	<u>1,60</u>
biomassa	9.40	<u>20,00</u>	<u>10,08</u>
cadeia de fornecedores	17.91	<u>14,00</u>	<u>868,00</u>
cadeia de fornecimento	12.91	<u>10,00</u>	<u>1,16</u>
cadeia de valor	37.86	<u>35,00</u>	<u>4,33</u>
cadeia produtiva	7.15	<u>13,00</u>	<u>32,96</u>
cadeias produtivas	20.40	<u>21,00</u>	<u>7,91</u>
capacidade produtiva	6.12	<u>6,00</u>	<u>9,15</u>
capacitação	10.49	<u>46,00</u>	<u>26,09</u>
capacitar	6.91	<u>14,00</u>	<u>9,57</u>
capital humano	6.24	<u>6,00</u>	<u>8,50</u>
capital natural	8.53	<u>6,00</u>	<u>26,00</u>
capital social	4.21	<u>5,00</u>	<u>16,89</u>
captação	7.21	<u>20,00</u>	<u>14,86</u>
captar	6.78	<u>25,00</u>	<u>21,41</u>
carbono	15.17	<u>59,00</u>	<u>22,38</u>
cargos de liderança	9.41	<u>7,00</u>	<u>923,00</u>
casos de discriminação	8.42	<u>6,00</u>	<u>320,00</u>
categoria funcional	10.54	<u>8,00</u>	<u>1,12</u>
categorias passíveis	7.28	<u>5,00</u>	<u>5,00</u>

categorias passíveis de avaliação	7.28	<u>5,00</u>	<u>0,00</u>
celulose	7.27	<u>5,00</u>	<u>22,00</u>
cenário	4.89	<u>40,00</u>	<u>53,75</u>
centro de distribuição	9.32	<u>8,00</u>	<u>4,28</u>
centro de estudos	3.03	<u>5,00</u>	<u>32,44</u>
certificação	15.05	<u>25,00</u>	<u>6,53</u>
certificar	12.26	<u>18,00</u>	<u>5,24</u>
ciclo de vida	4.97	<u>6,00</u>	<u>16,59</u>
cliente	6.40	<u>102,00</u>	<u>108,67</u>
co2 equivalente	8.40	<u>6,00</u>	<u>377,00</u>
código	5.15	<u>39,00</u>	<u>49,37</u>
código de conduta	22.41	<u>22,00</u>	<u>6,42</u>
código de ética	5.92	<u>7,00</u>	<u>15,12</u>
colaborador	42.80	<u>212,00</u>	<u>29,71</u>
colaboradores	34.90	<u>27,00</u>	<u>0,00</u>
combustíveis fósseis	8.37	<u>9,00</u>	<u>10,88</u>
combustível	13.99	<u>43,00</u>	<u>16,61</u>
começar	4.64	<u>26,00</u>	<u>35,41</u>
comercializar	4.53	<u>15,00</u>	<u>19,12</u>
comitê	28.95	<u>69,00</u>	<u>11,49</u>
comitê de auditoria	15.55	<u>12,00</u>	<u>770,00</u>
comitê de diversidade	7.27	<u>5,00</u>	<u>25,00</u>
comitê executivo	14.76	<u>13,00</u>	<u>4,01</u>
comitês de assessoramento	7.23	<u>5,00</u>	<u>148,00</u>
commodities	8.17	<u>10,00</u>	<u>3,74</u>
companhia	10.64	<u>170,00</u>	<u>108,60</u>
compartilhar	4.43	<u>26,00</u>	<u>37,40</u>
compensação	3.65	<u>9,00</u>	<u>13,42</u>
competência	4.22	<u>33,00</u>	<u>51,32</u>
competitividade	5.70	<u>22,00</u>	<u>22,81</u>
compilação	5.81	<u>6,00</u>	<u>2,66</u>
complementar	5.15	<u>58,00</u>	<u>75,53</u>
compor	3.99	<u>36,00</u>	<u>60,02</u>
compostagem	8.59	<u>8,00</u>	<u>1,62</u>
compra	4.98	<u>59,00</u>	<u>79,74</u>
compromisso	7.72	<u>78,00</u>	<u>66,93</u>
comunidade	8.92	<u>258,00</u>	<u>200,72</u>
comunidade local	4.32	<u>5,00</u>	<u>15,81</u>
comunidades de base	8.31	<u>6,00</u>	<u>636,00</u>
comunidades locais	17.62	<u>18,00</u>	<u>7,85</u>
comunidades sustentáveis	7.22	<u>5,00</u>	<u>186,00</u>
comunidades vizinhas	15.20	<u>12,00</u>	<u>1,32</u>
concorrência desleal	5.76	<u>5,00</u>	<u>6,12</u>
condições de trabalho	2.38	<u>5,00</u>	<u>47,64</u>
conduta	5.92	<u>34,00</u>	<u>36,12</u>

confidencialidade	5.93	<u>5,00</u>	<u>1,29</u>
conselho	3.76	<u>92,00</u>	<u>170,14</u>
conselho de administração	21.23	<u>48,00</u>	<u>43,64</u>
consequência	5.16	<u>12,00</u>	<u>11,99</u>
conservação	8.25	<u>45,00</u>	<u>33,85</u>
consolidar	7.29	<u>34,00</u>	<u>28,29</u>
consultor	29.46	<u>96,00</u>	<u>17,72</u>
consumidor	6.27	<u>91,00</u>	<u>98,51</u>
consumidores	13.56	<u>10,00</u>	<u>0,00</u>
consumir	5.69	<u>29,00</u>	<u>31,57</u>
consumo	9.35	<u>188,00</u>	<u>138,02</u>
consumo absoluto	7.27	<u>5,00</u>	<u>20,00</u>
consumo de água	28.44	<u>31,00</u>	<u>9,34</u>
consumo de combustíveis	6.88	<u>5,00</u>	<u>1,35</u>
consumo de energia	15.69	<u>25,00</u>	<u>24,61</u>
consumo total de energia	8.40	<u>6,00</u>	<u>366,00</u>
contábeis	3.65	<u>5,00</u>	<u>5,63</u>
contabilizar	15.81	<u>28,00</u>	<u>7,29</u>
contatos	11.05	<u>8,00</u>	<u>0,00</u>
contemplar	5.98	<u>35,00</u>	<u>36,95</u>
contratação	13.85	<u>50,00</u>	<u>20,43</u>
contratações	4.00	<u>5,00</u>	<u>4,65</u>
contratar	4.43	<u>26,00</u>	<u>37,38</u>
contribuir	5.65	<u>100,00</u>	<u>121,25</u>
controles	7.28	<u>5,00</u>	<u>0,00</u>
convenção	3.77	<u>8,00</u>	<u>10,92</u>
coprocessamento	10.79	<u>8,00</u>	<u>132,00</u>
corrupção	8.02	<u>39,00</u>	<u>29,65</u>
créditos de carbono	6.78	<u>6,00</u>	<u>5,96</u>
crença	4.75	<u>16,00</u>	<u>19,49</u>
crescimento econômico	16.66	<u>35,00</u>	<u>39,28</u>
crise hídrica	8.52	<u>6,00</u>	<u>38,00</u>
critérios ambientais	17.98	<u>14,00</u>	<u>775,00</u>
critérios socioambientais	8.42	<u>6,00</u>	<u>314,00</u>
cultura corporativa	8.26	<u>6,00</u>	<u>765,00</u>
decisão estratégica	6.89	<u>5,00</u>	<u>1,32</u>
deficiência	3.75	<u>24,00</u>	<u>41,43</u>
demanda	4.74	<u>65,00</u>	<u>93,11</u>
demandar	4.04	<u>13,00</u>	<u>18,61</u>
demonstrações financeiras	7.00	<u>6,00</u>	<u>5,06</u>
denúncia	7.69	<u>30,00</u>	<u>22,84</u>
depreciação	5.83	<u>6,00</u>	<u>2,63</u>
desaceleração	6.52	<u>7,00</u>	<u>2,84</u>
descarte	12.57	<u>23,00</u>	<u>7,81</u>
descomissionamento	23.40	<u>18,00</u>	<u>49,00</u>

descumprimento	6.86	<u>8,00</u>	<u>3,46</u>
desempenho	8.17	<u>178,00</u>	<u>150,15</u>
desempenho econômico	11.30	<u>9,00</u>	<u>2,05</u>
desenvolvimento	4.28	<u>324,00</u>	<u>533,85</u>
desenvolvimento de cadeias	7.25	<u>5,00</u>	<u>106,00</u>
desenvolvimento de produtos	7.98	<u>8,00</u>	<u>8,91</u>
desenvolvimento humano	6.49	<u>7,00</u>	<u>11,74</u>
desenvolvimento local	15.42	<u>14,00</u>	<u>4,74</u>
desenvolvimento sustentável	11.92	<u>18,00</u>	<u>22,67</u>
desenvolvimento territorial	11.55	<u>9,00</u>	<u>1,51</u>
desligamento	17.60	<u>21,00</u>	<u>3,14</u>
destinação	12.93	<u>22,00</u>	<u>6,88</u>
diálogo	10.25	<u>65,00</u>	<u>40,02</u>
diálogo aberto	10.22	<u>8,00</u>	<u>1,87</u>
direcionadores	9.79	<u>7,00</u>	<u>0,00</u>
direcionar	4.29	<u>5,00</u>	<u>3,95</u>
direitos humanos	12.10	<u>58,00</u>	<u>117,96</u>
diretoria	6.32	<u>43,00</u>	<u>43,65</u>
diretrizes	65.04	<u>51,00</u>	<u>0,00</u>
discriminação	5.67	<u>22,00</u>	<u>22,92</u>
discriminar	4.62	<u>11,00</u>	<u>12,49</u>
disponibilização	5.59	<u>6,00</u>	<u>2,98</u>
disponibilizar	3.62	<u>12,00</u>	<u>19,51</u>
disposição de resíduos	8.26	<u>6,00</u>	<u>771,00</u>
disseminar	3.88	<u>7,00</u>	<u>8,62</u>
distorções relevantes	8.52	<u>6,00</u>	<u>44,00</u>
distribuir	4.52	<u>20,00</u>	<u>27,06</u>
diversidade	5.92	<u>47,00</u>	<u>51,75</u>
diversificação	3.90	<u>7,00</u>	<u>8,58</u>
diversificar	4.32	<u>7,00</u>	<u>7,18</u>
divulgação	9.65	<u>59,00</u>	<u>38,42</u>
doenças ocupacionais	11.03	<u>9,00</u>	<u>2,66</u>
dragagem	6.52	<u>5,00</u>	<u>663,00</u>
ecoeficiência	8.39	<u>6,00</u>	<u>95,00</u>
economia de baixo carbono	9.29	<u>7,00</u>	<u>1,24</u>
economia local	5.83	<u>6,00</u>	<u>10,73</u>
efeito estufa	23.31	<u>39,00</u>	<u>26,45</u>
eficiência energética	6.83	<u>9,00</u>	<u>18,54</u>
efluente	16.92	<u>40,00</u>	<u>11,49</u>
emissão	32.45	<u>138,00</u>	<u>24,77</u>
emissões atmosféricas	9.43	<u>7,00</u>	<u>870,00</u>
emissões de gee	20.14	<u>16,00</u>	<u>1,10</u>
emissões diretas de gee	7.28	<u>5,00</u>	<u>3,00</u>
emissões indiretas	8.53	<u>6,00</u>	<u>1,00</u>
emissões relativas	8.52	<u>6,00</u>	<u>53,00</u>

empoderamento	17.84	<u>16,00</u>	<u>1,03</u>
empoderamento feminino	7.26	<u>5,00</u>	<u>49,00</u>
empreendedorismo	6.67	<u>8,00</u>	<u>3,72</u>
empregado	11.89	<u>131,00</u>	<u>73,22</u>
emprego digno	34.38	<u>27,00</u>	<u>356,00</u>
emprego saúde	7.28	<u>5,00</u>	<u>11,00</u>
empresa	7.61	<u>7,00</u>	<u>6,64</u>
energético	17.85	<u>39,00</u>	<u>10,20</u>
energia	6.97	<u>162,00</u>	<u>160,58</u>
energia elétrica	7.26	<u>27,00</u>	<u>88,05</u>
energia solar	5.52	<u>7,00</u>	<u>17,91</u>
engajamento	32.98	<u>60,00</u>	<u>7,45</u>
engajar	11.78	<u>19,00</u>	<u>6,29</u>
entornar	17.41	<u>34,00</u>	<u>8,56</u>
entrega	3.63	<u>19,00</u>	<u>33,14</u>
equidade	7.11	<u>7,00</u>	<u>2,14</u>
equidade de gênero	6.98	<u>5,00</u>	<u>996,00</u>
equipar	15.61	<u>49,00</u>	<u>17,04</u>
erradicação	4.85	<u>7,00</u>	<u>5,77</u>
erradicar	10.59	<u>14,00</u>	<u>4,28</u>
escritório	12.17	<u>39,00</u>	<u>17,61</u>
estabilidade operacional	11.01	<u>8,00</u>	<u>71,00</u>
estruturar	11.15	<u>38,00</u>	<u>19,09</u>
estufar	4.85	<u>6,00</u>	<u>4,31</u>
evolução	5.36	<u>64,00</u>	<u>80,43</u>
evoluir	6.19	<u>20,00</u>	<u>18,22</u>
expansão	3.70	<u>37,00</u>	<u>67,04</u>
exploração	3.64	<u>24,00</u>	<u>42,80</u>
exploração mineral	10.36	<u>8,00</u>	<u>1,53</u>
exportação	5.08	<u>21,00</u>	<u>24,86</u>
externalidades	7.28	<u>5,00</u>	<u>0,00</u>
extração	3.83	<u>15,00</u>	<u>23,70</u>
fauna	5.75	<u>13,00</u>	<u>11,41</u>
firmar	4.62	<u>17,00</u>	<u>21,76</u>
focar	5.35	<u>18,00</u>	<u>19,32</u>
foco	9.34	<u>60,00</u>	<u>40,65</u>
fomentar	7.99	<u>15,00</u>	<u>8,40</u>
fomento	10.87	<u>21,00</u>	<u>8,61</u>
fonte renovável	8.21	<u>6,00</u>	<u>912,00</u>
fontes de captação	7.20	<u>5,00</u>	<u>242,00</u>
fontes renováveis	13.63	<u>13,00</u>	<u>6,28</u>
forçar	12.99	<u>19,00</u>	<u>5,18</u>
forma de gestão	18.58	<u>14,00</u>	<u>2,00</u>
forma de medição	8.49	<u>6,00</u>	<u>108,00</u>
forma sustentável	7.61	<u>7,00</u>	<u>6,63</u>

fornecedor	60.15	<u>358,00</u>	<u>36,78</u>
fornecedores locais	16.67	<u>13,00</u>	<u>912,00</u>
fornecedores terceiros	11.04	<u>8,00</u>	<u>13,00</u>
fornecimento	14.27	<u>54,00</u>	<u>21,65</u>
fornecimento externo	11.01	<u>8,00</u>	<u>69,00</u>
fornecimento interno	7.27	<u>5,00</u>	<u>26,00</u>
fortalecer	11.73	<u>47,00</u>	<u>23,34</u>
fortalecimento	5.39	<u>18,00</u>	<u>19,15</u>
fóssil	15.31	<u>14,00</u>	<u>1,21</u>
garantir	3.89	<u>83,00</u>	<u>147,78</u>
gás	11.24	<u>59,00</u>	<u>32,19</u>
gás natural	3.88	<u>7,00</u>	<u>35,24</u>
gases de efeito estufa	33.13	<u>36,00</u>	<u>9,14</u>
geração	8.30	<u>92,00</u>	<u>73,87</u>
geração de energia	7.63	<u>10,00</u>	<u>17,95</u>
geração de renda	6.94	<u>9,00</u>	<u>17,86</u>
geração de resíduos	9.13	<u>7,00</u>	<u>1,68</u>
geração de valor	23.70	<u>19,00</u>	<u>1,13</u>
gerar	5.82	<u>153,00</u>	<u>182,28</u>
gerência	15.91	<u>33,00</u>	<u>9,45</u>
gerenciamento de riscos	18.52	<u>15,00</u>	<u>1,64</u>
gerenciamento de riscos	7.04	<u>5,00</u>	<u>774,00</u>
gerentes de relacionamento	8.50	<u>6,00</u>	<u>99,00</u>
gerir	3.68	<u>6,00</u>	<u>7,47</u>
gestão	21.80	<u>373,00</u>	<u>116,37</u>
gestão	9.17	<u>7,00</u>	<u>1,55</u>
gestão ambiental	15.73	<u>19,00</u>	<u>13,42</u>
gestão de barragens	7.27	<u>5,00</u>	<u>18,00</u>
gestão de pessoas	11.37	<u>12,00</u>	<u>9,56</u>
gestão de resíduos	5.98	<u>5,00</u>	<u>5,02</u>
gestão de riscos	26.83	<u>23,00</u>	<u>2,63</u>
gestor	7.41	<u>36,00</u>	<u>29,69</u>
<i>ghg protocol</i>	19.61	<u>15,00</u>	<u>263,00</u>
governança	75.86	<u>108,00</u>	<u>4,54</u>
governança corporativa	16.00	<u>15,00</u>	<u>5,55</u>
grupo de trabalho	3.25	<u>7,00</u>	<u>46,53</u>
hidrelétricas	5.61	<u>6,00</u>	<u>2,95</u>
homologação	13.93	<u>15,00</u>	<u>2,40</u>
igualdade	7.82	<u>40,00</u>	<u>31,46</u>
igualdade de gênero	27.43	<u>24,00</u>	<u>3,13</u>
igualdade de oportunidades	6.10	<u>6,00</u>	<u>9,23</u>
igualdade de remuneração	7.22	<u>5,00</u>	<u>186,00</u>
impactar	26.92	<u>34,00</u>	<u>3,53</u>
impacto	17.35	<u>240,00</u>	<u>93,10</u>
impacto ambiental	8.72	<u>13,00</u>	<u>22,83</u>

impacto positivo	14.96	14,00	5,60
impactos ambientais	17.79	22,00	14,08
impactos negativos	10.14	9,00	4,92
impactos negativos significativos	8.53	6,00	11,00
impactos sociais	8.80	7,00	2,59
impactos socioambientais	8.20	6,00	938,00
implementação	4.63	36,00	50,88
implementar	6.30	46,00	47,15
impulsionar	6.18	15,00	12,51
incentivar	4.74	24,00	31,53
incineração	6.30	5,00	879,00
incluir	5.33	134,00	174,29
incorporar	3.69	26,00	46,00
indicador relativo	7.25	5,00	103,00
indicadores de saúde	6.76	5,00	1,78
indústria	4.22	58,00	93,57
informação	7.97	179,00	154,89
infraestrutura	53.74	42,00	0,00
iniciativa	11.76	160,00	91,61
inovação	25.44	116,00	27,00
inovação aberta	7.11	5,00	544,00
inovação tecnológica	4.24	5,00	16,56
inovador	6.95	22,00	17,68
innovar	5.53	8,00	5,65
instituições fortes	36.97	29,00	282,00
insumo	19.38	61,00	17,02
integridade	6.40	18,00	15,22
intensidade de emissões	7.27	5,00	36,00
intensificar	5.92	17,00	15,71
inventário	14.45	25,00	7,04
inventário de emissões	7.23	5,00	146,00
investidores	17.32	13,00	0,00
investimento	7.39	160,00	149,22
investir	6.35	61,00	63,55
justiça	4.53	33,00	47,48
lançamento	4.03	22,00	34,61
lançar	10.93	38,00	19,60
legislação	4.33	43,00	66,33
legislação ambiental	4.83	5,00	11,76
licença	10.05	27,00	14,02
licenciamento	4.66	7,00	6,25
liderança	13.83	75,00	33,34
logística	17.49	38,00	10,12
logística reversa	13.69	11,00	1,91
logístico	20.43	25,00	3,32

lucro líquido	5.92	<u>7,00</u>	<u>15,14</u>
manejo	6.15	<u>43,00</u>	<u>45,03</u>
manejo florestal	15.61	<u>14,00</u>	<u>4,40</u>
manejo sustentável	9.93	<u>8,00</u>	<u>2,61</u>
manifestação	5.26	<u>30,00</u>	<u>35,98</u>
mapeamento	7.09	<u>17,00</u>	<u>12,20</u>
mapeamento de riscos	9.74	<u>7,00</u>	<u>114,00</u>
materialidade	13.78	<u>18,00</u>	<u>4,04</u>
matriz de materialidade	8.53	<u>6,00</u>	<u>20,00</u>
matriz energética	13.56	<u>14,00</u>	<u>8,57</u>
medição	4.07	<u>15,00</u>	<u>21,92</u>
meio ambiente	2.25	<u>20,00</u>	<u>244,79</u>
mercado externo	10.26	<u>13,00</u>	<u>15,92</u>
mercado interno	5.36	<u>11,00</u>	<u>40,80</u>
meritocracia	8.15	<u>6,00</u>	<u>268,00</u>
mitigação	24.54	<u>21,00</u>	<u>654,00</u>
mitigação de riscos	7.20	<u>5,00</u>	<u>242,00</u>
mitigar	22.52	<u>21,00</u>	<u>1,22</u>
modelo comercial	7.12	<u>5,00</u>	<u>503,00</u>
modelo de gestão	7.33	<u>9,00</u>	<u>15,71</u>
modelo de governança	9.53	<u>7,00</u>	<u>634,00</u>
modelo de negócio	6.72	<u>6,00</u>	<u>6,27</u>
modelos de negócio	6.49	<u>5,00</u>	<u>2,82</u>
monetário	14.48	<u>17,00</u>	<u>3,08</u>
monitoramento	13.77	<u>51,00</u>	<u>21,10</u>
monitorar	6.24	<u>9,00</u>	<u>5,51</u>
mudança	4.77	<u>82,00</u>	<u>117,95</u>
mudanças climáticas	16.14	<u>32,00</u>	<u>35,88</u>
negócio	34.38	<u>245,00</u>	<u>45,21</u>
níveis de produção	6.93	<u>5,00</u>	<u>1,16</u>
objetivos de desenvolvimento sustentável	11.05	<u>8,00</u>	<u>0,00</u>
objetivos estratégicos	7.56	<u>6,00</u>	<u>2,98</u>
ocorrência de trabalho	10.99	<u>8,00</u>	<u>107,00</u>
ocorrência de trabalho infantil	7.27	<u>5,00</u>	<u>19,00</u>
omissão	10.69	<u>23,00</u>	<u>10,18</u>
ongs	9.60	<u>7,00</u>	<u>113,00</u>
operação	21.66	<u>261,00</u>	<u>80,34</u>
operacional	16.98	<u>126,00</u>	<u>47,46</u>
operações	6.81	<u>61,00</u>	<u>58,95</u>
operações florestais	9.71	<u>7,00</u>	<u>181,00</u>
operações industriais	8.39	<u>6,00</u>	<u>402,00</u>
operações internacionais	62.17	<u>49,00</u>	<u>138,00</u>
oportunidade	4.68	<u>79,00</u>	<u>115,67</u>
organização	3.70	<u>100,00</u>	<u>188,09</u>

órgão de governança	12.30	<u>9,00</u>	<u>12,00</u>
órgãos reguladores	7.70	<u>6,00</u>	<u>2,51</u>
origem renovável	10.90	<u>8,00</u>	<u>311,00</u>
ouvidoria	44.90	<u>40,00</u>	<u>798,00</u>
pacto	5.84	<u>16,00</u>	<u>14,78</u>
pacto global	12.95	<u>10,00</u>	<u>1,09</u>
parceiro	7.00	<u>42,00</u>	<u>37,84</u>
parceria	11.44	<u>94,00</u>	<u>53,32</u>
pegada hídrica	13.47	<u>10,00</u>	<u>156,00</u>
perenidade	7.52	<u>6,00</u>	<u>768,00</u>
pesquisa de engajamento	9.79	<u>7,00</u>	<u>13,00</u>
pilar	6.67	<u>19,00</u>	<u>15,46</u>
planejamento	3.84	<u>71,00</u>	<u>127,37</u>
planejamento estratégico	19.66	<u>29,00</u>	<u>20,88</u>
plano de desenvolvimento	7.39	<u>8,00</u>	<u>11,44</u>
planos de ação	19.89	<u>18,00</u>	<u>4,31</u>
plantar	5.85	<u>31,00</u>	<u>33,02</u>
plástico	4.30	<u>13,00</u>	<u>17,17</u>
plataforma	9.50	<u>31,00</u>	<u>18,15</u>
poder público	1.96	<u>5,00</u>	<u>62,74</u>
políticas corporativas	7.22	<u>5,00</u>	<u>181,00</u>
políticas públicas	2.19	<u>10,00</u>	<u>120,37</u>
potencializar	7.31	<u>13,00</u>	<u>7,77</u>
povos indígenas	4.21	<u>6,00</u>	<u>23,79</u>
prática	3.94	<u>116,00</u>	<u>205,18</u>
práticas de gestão	7.80	<u>7,00</u>	<u>5,91</u>
práticas de manejo	7.94	<u>6,00</u>	<u>1,72</u>
práticas trabalhistas	22.24	<u>17,00</u>	<u>112,00</u>
precificação	6.77	<u>5,00</u>	<u>427,00</u>
precificação de carbono	7.28	<u>5,00</u>	<u>2,00</u>
preservação	3.75	<u>19,00</u>	<u>31,88</u>
prestador	3.65	<u>7,00</u>	<u>9,55</u>
prevenção	3.88	<u>24,00</u>	<u>39,82</u>
priorizamos	7.08	<u>5,00</u>	<u>155,00</u>
processar	3.93	<u>22,00</u>	<u>35,60</u>
processo de avaliação	4.76	<u>5,00</u>	<u>12,24</u>
processo de gestão	7.07	<u>6,00</u>	<u>4,79</u>
processo de homologação	10.79	<u>8,00</u>	<u>558,00</u>
processo de revisão	6.17	<u>5,00</u>	<u>4,17</u>
processo produtivo	15.88	<u>19,00</u>	<u>13,09</u>
processos industriais	6.15	<u>5,00</u>	<u>4,24</u>
processos internos	7.47	<u>6,00</u>	<u>3,31</u>
processos produtivos	10.80	<u>10,00</u>	<u>5,91</u>
produção de celulose	8.04	<u>6,00</u>	<u>1,43</u>
produção responsáveis	16.06	<u>12,00</u>	<u>12,00</u>

produção responsáveis	11.05	<u>8,00</u>	<u>0,00</u>
produtividade	3.84	<u>36,00</u>	<u>62,58</u>
produtivo	8.25	<u>98,00</u>	<u>79,59</u>
produto	6.33	<u>318,00</u>	<u>352,92</u>
programa carbono	9.75	<u>7,00</u>	<u>104,00</u>
programa carbono neutro	9.76	<u>7,00</u>	<u>72,00</u>
programa de desenvolvimento	5.86	<u>6,00</u>	<u>10,54</u>
programa de educação	4.73	<u>5,00</u>	<u>12,48</u>
programa de reconhecimento	7.21	<u>5,00</u>	<u>211,00</u>
programa de voluntariado	12.80	<u>10,00</u>	<u>1,37</u>
projetos	225.76	<u>179,00</u>	<u>0,00</u>
projetos sociais	7.76	<u>10,00</u>	<u>17,30</u>
promover	4.16	<u>83,00</u>	<u>137,76</u>
proposta de valor	12.92	<u>10,00</u>	<u>1,14</u>
prosperidade	6.24	<u>10,00</u>	<u>6,64</u>
protagonismo	8.34	<u>8,00</u>	<u>1,84</u>
público interno	9.40	<u>8,00</u>	<u>4,04</u>
públicos de interesse	7.04	<u>5,00</u>	<u>785,00</u>
públicos de relacionamento	17.19	<u>13,00</u>	<u>189,00</u>
qualidade de vida	3.79	<u>27,00</u>	<u>189,80</u>
queixas	7.28	<u>5,00</u>	<u>0,00</u>
rastreabilidade	16.31	<u>14,00</u>	<u>789,00</u>
reabilitar	5.32	<u>5,00</u>	<u>2,09</u>
reaproveitamento	11.68	<u>10,00</u>	<u>913,00</u>
reaproveitar	9.52	<u>7,00</u>	<u>160,00</u>
reassentamento	9.05	<u>7,00</u>	<u>466,00</u>
receita	4.08	<u>55,00</u>	<u>91,55</u>
receita líquida	27.12	<u>28,00</u>	<u>7,72</u>
reciclagem	11.32	<u>24,00</u>	<u>9,92</u>
reciclar	27.15	<u>50,00</u>	<u>7,65</u>
recirculação	12.64	<u>10,00</u>	<u>413,00</u>
reclamação	32.76	<u>50,00</u>	<u>5,37</u>
recomendação	6.09	<u>15,00</u>	<u>12,79</u>
recursos	17.32	<u>13,00</u>	<u>0,00</u>
recursos hídricos	3.09	<u>5,00</u>	<u>31,39</u>
recursos naturais	6.09	<u>16,00</u>	<u>56,95</u>
rede de relações	6.86	<u>5,00</u>	<u>1,42</u>
redefinição	4.02	<u>5,00</u>	<u>4,60</u>
redução	6.57	<u>160,00</u>	<u>168,50</u>
redução de consumo	6.83	<u>5,00</u>	<u>1,52</u>
redução de custos	7.44	<u>10,00</u>	<u>19,04</u>
reduzir	4.79	<u>106,00</u>	<u>153,14</u>
reduzir o consumo	7.26	<u>5,00</u>	<u>68,00</u>
refinar	4.21	<u>5,00</u>	<u>4,14</u>
refil	7.28	<u>5,00</u>	<u>0,00</u>

reflorestamento	6.46	<u>8,00</u>	<u>4,03</u>
reforçar	10.09	<u>26,00</u>	<u>13,23</u>
regulatórios	7.28	<u>6,00</u>	<u>974,00</u>
rejeito	13.32	<u>15,00</u>	<u>2,77</u>
relacionamento	12.53	<u>104,00</u>	<u>53,83</u>
relatório	28.38	<u>254,00</u>	<u>58,22</u>
relatório anual	31.37	<u>27,00</u>	<u>2,61</u>
relatório de sustentabilidade	20.07	<u>16,00</u>	<u>1,18</u>
remediação	6.72	<u>5,00</u>	<u>471,00</u>
remediação de áreas	7.23	<u>5,00</u>	<u>150,00</u>
remuneração	14.36	<u>56,00</u>	<u>22,47</u>
repartição	4.81	<u>6,00</u>	<u>4,38</u>
repartição de benefícios	7.05	<u>5,00</u>	<u>745,00</u>
reportar	13.60	<u>30,00</u>	<u>10,44</u>
reputação	8.98	<u>13,00</u>	<u>5,26</u>
resíduo	54.78	<u>146,00</u>	<u>13,40</u>
resíduos sólidos	3.61	<u>5,00</u>	<u>23,54</u>
resiliência	5.14	<u>5,00</u>	<u>2,36</u>
responsabilidade social	5.97	<u>8,00</u>	<u>19,69</u>
restaurar	4.69	<u>8,00</u>	<u>7,68</u>
resultados	14.81	<u>11,00</u>	<u>0,00</u>
retenção	3.74	<u>11,00</u>	<u>16,81</u>
retirada	5.66	<u>12,00</u>	<u>10,44</u>
retirada de água	9.68	<u>7,00</u>	<u>271,00</u>
retração	5.35	<u>9,00</u>	<u>7,36</u>
reúso	12.09	<u>9,00</u>	<u>100,00</u>
reutilização	19.79	<u>20,00</u>	<u>1,81</u>
reverso	7.20	<u>12,00</u>	<u>6,97</u>
revisar	10.59	<u>25,00</u>	<u>11,67</u>
revitalização	9.03	<u>11,00</u>	<u>3,63</u>
risco de ocorrência	9.51	<u>7,00</u>	<u>676,00</u>
risco	42.44	<u>33,00</u>	<u>0,00</u>
riscos sociais	8.28	<u>6,00</u>	<u>701,00</u>
riscos socioambientais	8.49	<u>6,00</u>	<u>121,00</u>
rotatividade	9.79	<u>12,00</u>	<u>3,64</u>
rotulagem	10.59	<u>10,00</u>	<u>1,59</u>
rotulagem de produtos	8.42	<u>6,00</u>	<u>301,00</u>
sanção	9.08	<u>13,00</u>	<u>5,14</u>
satisfação	6.07	<u>35,00</u>	<u>36,30</u>
saúde e segurança	7.25	<u>5,00</u>	<u>94,00</u>
segmentação	6.20	<u>8,00</u>	<u>4,42</u>
segurança	7.92	<u>146,00</u>	<u>126,18</u>
sensibilização	4.79	<u>6,00</u>	<u>4,44</u>
sensibilizar	3.78	<u>6,00</u>	<u>7,14</u>
serviço	5.36	<u>115,00</u>	<u>148,05</u>

serviços ecossistêmicos	7.16	5,00	371,00
setor de mineração	6.96	5,00	1,06
setores	22.35	17,00	0,00
sinergia	7.40	8,00	2,79
sistema de gestão	18.82	30,00	24,41
sistema de verificação	9.60	7,00	445,00
sociedade civil	1.68	7,00	111,94
socioambiental	24.37	23,00	1,28
sociobiodiversidade	37.33	29,00	13,00
solidez	5.22	5,00	2,24
solucionar	5.83	17,00	16,06
soluções sustentáveis	8.18	6,00	1,01
sondagem	4.84	7,00	5,80
stakeholders	28.37	31,00	2,31
startups	13.53	10,00	10,00
suborno	5.00	5,00	2,58
subsídio	5.99	10,00	7,16
suprimento	4.15	11,00	14,57
suprimentos	12.30	9,00	0,00
suspensão	5.02	24,00	29,50
suspensão temporária	22.49	19,00	2,44
sustentabilidade	76.33	219,00	14,82
sustentabilidade empresarial	7.06	5,00	724,00
sustentável	33.25	122,00	20,61
tema material	22.35	17,00	2,00
tomada de decisão	8.37	11,00	17,80
trabalhista	10.39	47,00	27,08
trabalho de asseguuração	17.19	13,00	184,00
trabalho decente	10.01	8,00	2,40
trabalho escravo	4.72	6,00	18,72
trabalho infantil	18.92	25,00	16,48
traçar	5.05	9,00	8,13
transparência	10.09	27,00	13,93
transporte de madeira	10.85	8,00	413,00
tratamento de efluentes	10.80	9,00	3,22
tratamento de emissões	8.52	6,00	39,00
tratamento de resíduos	8.42	7,00	3,77
treinamento	9.12	96,00	69,84
unidades industriais	20.78	18,00	3,14
unidades operacionais	21.14	17,00	1,32
unidades produtivas	7.78	6,00	2,25
uso de materiais	6.53	5,00	2,66
uso de produtos	7.51	6,00	3,14
uso sustentável	7.57	7,00	6,79
valor econômico	4.31	5,00	15,95

valor monetário	9.13	<u>7,00</u>	<u>1,67</u>
valoração	7.17	<u>7,00</u>	<u>2,07</u>
valores	21.09	<u>16,00</u>	<u>0,00</u>
valorização	3.86	<u>19,00</u>	<u>30,83</u>
varejo	8.27	<u>18,00</u>	<u>10,51</u>
vazante	14.30	<u>12,00</u>	<u>699,00</u>
verificação	11.10	<u>40,00</u>	<u>20,48</u>
verificação externa	11.05	<u>8,00</u>	<u>3,00</u>
vida útil	7.99	<u>12,00</u>	<u>23,42</u>
violação	4.88	<u>8,00</u>	<u>7,16</u>
voluntariado	18.04	<u>18,00</u>	<u>1,75</u>
voluntário	15.71	<u>29,00</u>	<u>7,83</u>

ANEXO II Lista de palavras-chave – *corpus* inglês

Word	Score	Freq	Freq/ref
academia	8.73	8,00	68,302
accordance	10.16	42,00	378,476
accountability	13.86	10,00	48,572
acquisitions	14.35	20,00	113,36
action plan	7.19	5,00	47,607
additional information	7.09	13,00	157,492
advancing	11.47	30,00	147,496
advocacy	10.46	8,00	53,269
ambitious goal	22.65	6,00	3,843
animal testing	27.74	8,00	5,947
annual	16.02	68,00	388,407
annual basis	13.33	6,00	22,452
annual report	9.29	6,00	42,04
antitrust claims	12.57	7,00	32,831
applicable	9.74	43,00	405,925
approximately	12.40	15,00	95,872
artificial intelligence	28.75	13,00	21,718
assessment	10.83	450,00	808,618
assessment process	31.52	11,00	11,683
assurance	17.35	61,00	206,716
attract	149.75	52,00	10,94
audit	32.75	161,00	260,934
audit findings		15,00	
audit process	22.71	6,00	3,787
audit program	33.12	8	1,278
auditor	14.74	10,00	44,286
awareness	8.51	86,00	955,871
baseline	29.24	60,00	139,787
benchmarking	11.52	6,00	29,501
benefit		17,00	
biodiversity	24.87	31,00	98,502
boundary	84.78	36,00	18,525
breakthroughs	10.90	7,00	41,32
broad range	9.38	11,00	92,87
business	16.84	278,00	1,572,439
business continuity	12.26	5,00	18,48
business performance	15.94	6,00	15,039
business strategy	15.47	9,00	34,918
business success	12.43	6,00	25,712
calendar year	8.95	8,00	66,075
carbon	8.92	86,00	683,418

carbon footprint	27.65	21,00	51,419
carbon monoxide	9.76	6,00	38,947
career development	19.88	10,00	26,972
cdp	299.27	101,00	9,923
certification	13.37	91,00	455,559
certify	10.58	7,00	43,284
chain	230.49	537,00	85,413
supply chain	116.34	208,00	150,039
chairman	33.15	142,00	391,39
challenge (n)	12.04	40,00	299,746
charitable	11.88	7,00	36,081
child labor	36.64	14,00	14,776
circular economy	60.75	15,00	1,492
citizenship	8.66	49,00	158,304
clean water	17.38	13,00	50,801
climate change	10.54	42,00	364,22
climate strategy	21.88	5,00	383
closure	14.63	5,00	11,816
code	26.49	143,00	499,22
collaborate	15.75	53,00	138,102
collaboration	10.97	161,00	595,553
collective bargaining	17.97	10,00	32,255
commitment	11.47	195,00	1,047,702
commit	10.77	135,00	1,027,950
communities	10.07	172,00	1,455,889
community engagement	12.22	5,00	18,627
commute	8.71	11,00	46,381
company	8.28	148,00	1,203,957
compensation	21.33	22,00	77,907
compliance	18.23	192,00	513,435
comply	8.98	11,00	33,537
concern	11.77	7,00	36,632
conduct	18.47	111,00	15,081
conformance	20.18	7,00	11,873
conservancy	17.16	10,00	34,852
conservation	17.22	76,00	404,693
energy conservation			
water conservation	104.92	43	17,059
conserve	8.49	11,00	105,062
consortium	10.14	8,00	55,642
consumer	18.64	63,00	227,089
consumer use	20.47	5,00	1,96
consumption	13.55	115,00	712,635
content	10.56	37,00	317,566
continue	11.10	28,00	48,91

continuous	17.55	11,00	39,053
continuous improvement	35.92	19,00	28,959
contributions	9.02	49,00	504,094
cooling	10.59	36,00	307,43
corporate	55.75	164,00	261,627
corporate governance	20.36	11,00	30,559
corporate responsibility	150.29	52,00	10,825
corrective	35.77	44,00	61,553
corrective action(s)	20.77	18,00	15,563
criteria	8.98	56,00	581,954
csr	564.43	374,00	41,27
customer satisfaction	6.79	7,00	80,121
cybersecurity	23.27	9,00	15,581
data	10.27	74,00	608,964
data center	20.45	11,00	30,31
declaration	10.17	13,00	102,919
deepen	9.33	8,00	62,445
define	10.00	5,00	27,798
deliver	11.61	14,00	29,095
derivatives	21.42	22,00	77,448
design	9.38	6,00	41,449
devaluations	20.80	5,00	1,564
developing	11.60	15,00	104,055
development	9.03	98,00	1,026,994
diligence	9.95	19,00	115,423
due diligence	14.71	15,00	77,289
directors	14.73	40,00	240,879
disadvantaged	9.27	9,00	73,461
disaster relief	15.34	6,00	16,513
discharges	13.16	7,00	30,355
disclosure	9.70	37,00	218,383
disruptions	8.79	5,00	34,751
diverse supplier	22.10	5,00	146
diverse talent	53.99	13,00	949
diversity	24.54	364,00	469,957
divested	30.86	9,00	6,158
divestiture	40.89	11,00	3,792
divestitures	38.76	10,00	2,764
drinking water	2.58	5,00	173,019
driving	10.12	17,00	141,698
droughts	11.46	5,00	21,351
earnings	9.94	5,00	28,102
earthwards	68.94	16,00	4
economic growth	4.05	8,00	173,658
ecosystem	13.43	5,00	14,906

efficiencies	8.32	8,00	72,834
efficiency	37.28	31,00	58,139
efficiency	16.87	149,00	831,354
efficient	8.67	5,00	35,533
effluents	241.75	57,00	124
efforts	8.48	164,00	1,847,327
electricity	10.78	87,00	668,838
electricity consumption	29.03	9,00	7,99
electricity use	39.86	11,00	4,494
electronic	9.24	19,00	178,133
electronics industry	17.37	5,00	6,377
embedding	69.73	19,00	3,899
emerging	10.63	9,00	61,105
emission	16.56	516,00	101,067
emissions	77.07	495,00	425,946
ghg emissions	98.68	23,00	0
employee	15.51	894,00	974,255
employee engagement	39.26	16,00	17,192
employee health	30.81	9,00	6,227
employment	8.92	17,00	163,741
empower	14.63	37,00	130,603
empowerment	21.08	10,00	24,146
encourage	8.85	7,00	56,16
energy	8.99	659,00	4,522,637
energy conservation	20.66	9,00	20,422
energy consumption	47.61	43,00	64,955
energy efficiency	43.34	67,00	127,016
energy intensity	52.04	13,00	1,819
energy management	43.03	17,00	15,933
energy reduction	77.14	20,00	2,594
energy usage	12.65	5,00	17,244
energy use	47.36	31,00	40,955
engagement	17.71	111,00	583,458
engage	11.59	12,00	79,195
entrepreneurship	89.77	56,00	37,738
environment	153.14	416,00	239,622
environmental	21.89	286,00	1,239,308
environmental footprint	33.59	10,00	6,686
environmental impact	38.99	52,00	57,072
environmental performance	52.83	18,00	10,604
environmental policy	15.47	5,00	9,945
environmental sustainability	70.83	29,00	17,123
environmental sustainability section	26.48	6,00	0
equal	13.86	11,00	55,535
equal opportunity	11.66	5,00	20,618

equal pay pledge			
equality	10.74	8,00	51,317
ergonomics	13.68	6,00	21,272
ethical	11.99	60,00	309,143
ethical business	19.18	5,00	3,627
ethical conduct	19.17	5,00	3,643
ethics	12.51	108,00	249,094
ethnicity	10.79	18,00	68,849
evaluates	11.82	7,00	36,371
evaluation	30.15	38,00	99,663
excellence	10.44	18,00	145,852
exclude	35.56	30,00	40,343
expand	15.79	8,00	27,6
explanation	72.26	56,00	24,337
extend	22.76	9,00	16,433
external	8.37	81,00	786,891
facilities	8.93	128,00	1,362,795
factory	8.25	14,00	143,8
fair pay	20.68	5,00	1,719
fairness	8.48	12,00	116,58
fatalities	9.97	8,00	56,992
feedback	20.97	32,00	820,506
filings	10.12	8,00	55,786
finding	49.87	28,00	31,92
footprint	31.22	58,00	157,316
forced	20.87	10,00	24,612
freshwater	10.28	9,00	63,968
fuel combustion	20.94	5,00	1,413
full potential	7.42	6,00	58,368
gap	9.41	35,00	205,291
gas	16.58	38,00	199,853
generate	31.16	50,00	8,977
generation	9.27	15,00	135,948
geographies	24.69	9,00	13,379
global	14.07	304,00	2,064,756
global electricity	30.01	7,00	549
global energy	16.24	5,00	8,405
global health	12.51	5,00	17,683
global initiative	20.32	5,00	2,158
global manufacturing	23.83	6,00	2,533
global problem	56.02	15,00	3,541
global public health	33.37	8,00	1,113
global supply	41.54	13,00	8,032
global supply chain	40.99	11,00	3,748
global water	23.33	6,00	3,082

global workforce	20.23	5,00	2,26
globally	21.46	51,00	207,718
governance	20.45	61,00	266,372
governance body	67.93	16,00	354
green building	18.47	9,00	25,54
green chemistry	48.16	12,00	1,81
green power	101.83	30,00	5,942
greenhouse gas	48.72	49,00	74,825
gri	2,372.46	712,00	6,249
hackathon	24.73	6,00	1,613
harassment	28.45	10,00	11,992
hazardous	13.08	48,00	163,54
hazardous waste	36.83	16,00	19,84
non-hazardous waste		16,00	
healthy	11.60	25,00	187,215
healthy activity	25.00	6,00	1,353
healthy movement	21.93	5,00	323
healthy work environment	21.32	5,00	985
helping	11.28	12,00	81,96
high risk	8.57	11,00	103,78
hot water	2.93	6,00	182,658
human health	17.47	12,00	44,901
human right	13.03	5,00	16,059
human trafficking	21.37	12,00	32,545
identify	9.49	6,00	40,664
impact	9.68	350,00	2,403,615
environmental impact	38.99	32,00	57,072
economic impact	20.24	11,00	30,862
impact assessment	24.46	8,00	9,789
impact	8.51	8,00	70,667
implement	13.11	7,00	30,537
improvement	10.14	69,00	93,734
improve	13.42	11,00	58,097
incidents	10.36	30,00	259,067
include	11.45	139,00	105,662
inclusion	218.37	109,00	25,556
inclusion	29.37	78,00	234,392
inclusion strategy	21.93	5,00	327
inclusive	25.08	53,00	182,117
inclusive leadership	21.93	5,00	319
inclusive work	21.86	5,00	407
inclusive workforce	26.27	6,00	185
increase	12.72	33,00	115,596
industry	12.95	49,00	344,161
initiative	18.97	199,00	187,072

injury	10.57	11,00	79,887
innovate	15.41	12,00	53,902
innovation	12.85	155,00	677,742
innovator	26.42	18,00	43,896
inquiry	10.21	16,00	130,775
intensity	10.25	39,00	346,85
internal audit	15.97	5,00	8,927
investment	9.11	107,00	541,274
investor	41.56	28,00	42,851
invest	8.65	5,00	35,696
irrigation	9.57	14,00	120,914
issue	8.76	152,00	145,199
keep	22.33	51,00	198,785
kpi	43.77	14,00	8,662
labeling	8.87	15,00	88,638
labor	9.60	78,00	763,836
labs	20.94	29,00	112,017
land fill	27.20	52,00	70,37
land use	166.43	39,00	27
leader	9.13	14,00	127,849
leadership	24.94	81,00	291,662
learn	14.88	5,00	11,244
learnings	17.62	6,00	11,436
led lighting	12.06	5,00	19,194
lead	28.36	25,00	63,163
life cycle	24.37	21,00	61,391
limitations	11.38	5,00	21,676
limited assurance	26.30	6,00	175
location	10.57	31,00	262,553
logistics	13.52	40,00	171,76
loop	12.05	10,00	59,251
management	16.88	205,00	1,150,580
manufacturer	9.56	8,00	60,437
manufacture	24.09	278,00	787,531
market	8.51	11,00	104,735
match	19.49	10,00	27,973
material topic	145.39	34,00	10
materiality	47.88	114,00	12,02
material	55.34	93,00	139,886
matter	11.98	44,00	67,806
measurement	11.11	6,00	31,43
medicine	14.65	7,00	24,939
membership	9.52	7,00	50,596
mentoring	12.96	7,00	31,178
microprocessors	27.41	8,00	6,274

minerals	8.64	46,00	270,215
minimize	9.27	31,00	302,608
minorities	8.80	31,00	111,413
mitigate	10.90	13,00	94,476
natural gas	3.81	9,00	211,15
networking equipment	20.05	5,00	2,48
new product	3.68	5,00	114,558
next generation	7.05	13,00	158,58
ngos	15.54	20,00	102,956
non-hazardous waste	67.53	16,00	491
nonprofit organization	10.53	6,00	34,438
nonprofits	25.00	20,00	55,409
notable	12.23	6,00	26,478
notice	19.87	8,00	17,278
nutrition	16.50	27,00	136,56
operating company	30.18	8,00	3,626
operation	16.02	347,00	1,304,464
opportunity	16.43	20,00	96,123
optimize	10.39	15,00	63,034
organization	8.72	193,00	1,729,684
outcomes	8.81	5,00	34,613
oversees	9.70	9,00	69,155
oversight	9.08	24,00	160,459
package material	22.12	5,00	136
packaging	39.95	185,00	347,51
paperboard	19.62	5,00	3,028
participates	8.95	7,00	55,282
partnering	38.42	28,00	15,553
partners	11.05	121,00	1,036,212
partnership	10.44	66,00	164,281
performance	32.34	108,00	300,338
performance management	10.83	5,00	23,946
personal information	2.54	6,00	213,777
personal leadership	23.62	6,00	2,771
pledge	13.43	18,00	29,286
policy	20.07	131,00	517,02
positive impact	11.70	10,00	61,718
power consumption	8.78	5,00	34,861
power efficiency	34.25	9,00	3,317
power purchase	23.21	6,00	3,213
practices	11.40	317,00	1,283,658
preferred valuable feedstock	22.23	5,00	0
presence	8.88	7,00	55,884
previous year	6.16	6,00	74,992
priority	9.07	95,00	299,047

prioritization	14.60	5,00	11,894
prioritize	16.70	20,00	59,567
privacy	26.87	55,00	175,729
process	13.43	33,00	216,172
procurement	16.45	75,00	142,977
procure	9.27	5,00	31,789
products	8.31	683,00	5,520,545
professional development	6.68	9,00	110,805
program	10.66	132,00	1,175,158
programs	8.68	23,00	235,693
progress	131.16	178,00	108,44
promote	16.12	11,00	44,545
promoting	26.55	17,00	39,925
promoting transparency	30.01	7,00	547
protecting	9.33	6,00	41,77
protocol	21.07	27,00	102,016
providing	11.25	15,00	108,035
purchase	32.42	12,00	13,697
race	16.26	8,00	26,148
rainwater	13.83	9,00	41,748
real estate portfolio	28.72	7,00	1,613
receive	9.15	5,00	32,476
recognition	21.72	18,00	58,315
recognizing	12.39	8,00	41,425
recordable	38.18	17,00	617
recyclability	92.20	23,00	1,596
recycle	17.86	124,00	92,066
recycler	21.87	17,00	4,788
recycling	34.85	128,00	268,746
reduce	9.38	275,00	2,027,340
reduction	15.24	198,00	814,592
refer	16.92	12,00	47,081
refiner	43.20	13,00	6,844
refrigerant	11.78	10,00	20,179
regulatory	10.29	60,00	401,517
relations	13.15	27,00	177,256
reliability	12.28	6,00	26,303
remote collaboration	25.99	6,00	446
replenishment	16.67	6,00	13,37
report	53.75	317,00	546,937
reporting	15.56	306,00	754,393
represent	22.98	7,00	7,669
reprocessed	30.04	8,00	3,737
requesting organization	22.10	5,00	153
resale	10.71	9,00	60,537

resilience	16.78	6,00	13,135
resiliency	21.34	9,00	19,043
resource	9.83	43,00	401,649
respect	49.41	28,00	5,083
responsibility	12.90	256,00	1,250,847
responsibly	32.13	25,00	53,079
restructure	54.92	18,00	9,323
retention	12.50	30,00	210,674
reuse	28.67	52,00	85,783
reward	25.04	7,00	5,158
right	34.02	164,00	443,205
risk	20.60	175,00	208,006
roadmap	144.74	56,00	14,775
roundtable	16.33	8,00	25,952
routing	13.57	5,00	14,519
safety	24.89	132,00	490,186
scarcity	12.42	10,00	56,805
scope	319.01	194,00	36,042
scorecard	31.15	14,00	21,376
segments	11.59	29,00	220,822
selection process	9.04	6	43,839
senior management	15.61	11	46,767
shared value	21.05	5	1,282
shipments	9.07	10,00	86,142
significant progress	13.23	5	15,476
skills	15.97	30,00	160,039
smelter	55.56	15,00	3,738
society	33.01	319,00	910,794
socioeconomic	9.83	59,00	38,488
solvent	10.45	8,00	53,343
solve	13.89	6,00	20,595
source	11.88	139,00	84,824
specification	13.63	6,00	21,436
spills	10.54	9,00	61,838
sponsorships	10.60	5,00	24,924
stakeholder	51.95	171,00	55,749
standard	8.87	92,00	371,762
statement	24.87	55,00	164,484
stewardship program	25.56	6,00	837
stockholder	13.38	6,00	22,25
strategy	30.54	76,00	218,191
subcontractors	9.70	5,00	29,393
subsidiaries	18.45	14,00	51,761
substances	12.88	5,00	16,51
supplier	313.13	181,00	33,138

supplier	53.14	184,00	311,915
supply chain	116.34	208,00	150,039
support	10.07	9,00	65,774
survey	8.35	27,00	292,036
sustainability	210.89	470,00	91,802
system power efficiency	34.97	8,00	16
technology industry	26.31	9,00	11,182
technology sector	16.73	5,00	7,482
third-party auditor	34.66	8,00	219
total waste	23.62	7,00	6,855
traceability	36.92	14,00	14,49
tracked	10.53	16,00	126,054
trade association	13.48	5,00	14,751
trafficking	9.29	25,00	125,131
training	52.62	15,00	5,22
transparency	15.07	61,00	222,126
trash	18.05	9,00	26,653
treasure	20.08	17,00	60,115
underserved	49.93	31,00	37,658
usage	12.84	9,00	46,685
use electricity	25.94	6,00	485
use of renewable energy	22.19	5,00	54
validate	11.00	26,00	75,832
valuable feedstock	22.23	5,00	16
value chain	29.88	14,00	23,255
value	25.39	24,00	69,424
venture capital	7.15	5,00	47,968
verification	11.99	35,00	164,305
verify	9.64	16,00	139,767
videoconferencing	21.06	7,00	10,434
violation	20.32	17,00	11,645
volunteer	9.40	18,00	92,621
volunteerism	31.53	12,00	14,728
waste	12.78	157,00	1,164,630
water	30.32	255,00	770,695
webinars	10.25	7,00	45,38
wind power	10.54	7,00	43,528
withdrawals	19.27	14,00	48,585
workforce	17.75	93,00	321,172
work	8.70	44,00	467,865
workplace design	21.57	5,00	726