

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

JOEL SOSSAI COLETI

**BASE DE DADOS MORFOLÓGICOS DE
TERMINOLOGIAS DO PORTUGUÊS DO BRASIL.
DESCRIÇÃO E ANÁLISE MORFOLÓGICA COM VISTAS À
DISPONIBILIZAÇÃO ON-LINE**

SÃO CARLOS
2012

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

JOEL SOSSAI COLETI

**BASE DE DADOS MORFOLÓGICOS DE
TERMINOLOGIAS DO PORTUGUÊS DO BRASIL.
DESCRIÇÃO E ANÁLISE MORFOLÓGICA COM VISTAS À
DISPONIBILIZAÇÃO ON-LINE**

Dissertação apresentada ao Programa de
Pós-Graduação em Linguística, para
obtenção do título de mestre em
Linguística.

Orientação: Prof^a. Dr^a Gladis Maria de Barcellos Almeida

SÃO CARLOS
2012

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C694bd Coleti, Joel Sossai.
Base de dados morfológicos de terminologias do português do Brasil : descrição e análise morfológica com vistas à disponibilização on-line / Joel Sossai Coleti. -- São Carlos : UFSCar, 2012.
146 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2012.

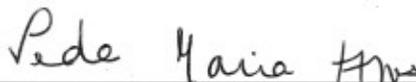
1. Linguística. 2. Morfologia. 3. Terminologia. 4. Nanociência. 5. Nanotecnologia. 6. Energia da biomassa. I. Título.

CDD: 410 (20^a)

**BANCA EXAMINADORA DA DISSERTAÇÃO DE MESTRADO DE
JOEL SOSSAI COLETI**



Prof^a. Dr^a. Gladis Maria de Barcellos Almeida
Orientadora e Presidente
UFSCar/São Carlos



Prof^a. Dr^a. Ieda Maria Alves
Membro titular
USP/São Paulo



Prof. Dr. Oto Araujo Vale
Membro titular
UFSCar/São Carlos

Submetida a defesa pública em sessão realizada em: 12/abril/2012.
Homologada na 4^a reunião da CPGL, realizada em 03/05/2012



Prof. Dr. Oto Araújo Vale
Coordenador do PPGL

Dedico esta dissertação a Wilson Roberto Coleti e Elisa Sossai Coleti. Obrigado por serem meus pais, minha maior fonte de inspiração, força, alegria, fé e aprendizado diário.

AGRADECIMENTOS

A Deus e a Nossa Senhora.

Aos meus pais, Wilson e Elisa, que amo mais que tudo, por simplesmente tudo.

À Marineile, minha "amora", por ter me compreendido, me incentivado, me ajudado e me amado. Agradeço por ter compartilhado comigo todos os momentos felizes, mas também os mais difíceis dessa e de outras caminhadas.

À minha família e amigos, pelo apoio em todos os momentos e pelo incentivo constante.

À Gladis, minha orientadora, por toda sabedoria, dedicação e, acima de tudo, competência incomum.

À Margarita, minha coorientadora, por sua ajuda, avaliação e sábias ideias.

A todos do GETerm e do NILC, pela amizade e ensinamentos. Agradeço especialmente a Daniela, minha grande parceira e amiga, desde a I.C., pelos muitos bons momentos compartilhados, pelas conversas, convites, ajudas e, principalmente, pelo companheirismo.

Aos professores do Departamento de Letras da UFSCar que, ao longo de toda minha trajetória nesta universidade, sempre me ofereceram mais que ensinamentos, dando-me a oportunidade de sempre contar com apoio e amizade.

Aos professores Oto Araújo Vale e Ieda Maria Alves, pela contribuição que deram a este trabalho com suas leituras e colaborações, por ocasião das bancas.

À Nani e ao Leonardo da secretaria do PPGL.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, pela concessão de bolsa de estudos para a realização desta pesquisa.

Algumas palavras acharão neste livro que nos dicionários portugueses se não encontram; mas as línguas vivas se enriquecem com o progresso da civilização, e das ciências, e uma nova ideia pede um novo termo.

Gonçalves Magalhães
Suspiros Poéticos e Saudades, 1836

Resumo

As pesquisas nas áreas de *Linguística e Letras* são extremamente ricas, constantes e eficientes no Brasil, entretanto, ainda são pouco expressivas no cenário brasileiro as pesquisas envolvendo questões morfológicas em Terminologia. Diante deste cenário, esta pesquisa pretende descrever a estrutura de dois repertórios terminológicos (Nanociência/Nanotecnologia e Biocombustíveis), representativos da área de Engenharias, no nível da estrutura interna dos seus termos constitutivos, verificando-se os principais processos de construção dos termos. A partir dessa descrição morfológica, organizar-se-á uma base de dados de maneira que seja possível a implementação computacional e a disponibilização on-line dos dados obtidos. Posteriormente, para os processos de maior produtividade, estabelecer-se-ão as Regras de Formação de Palavras, conforme o Modelo SILEX. Este modelo de análise morfológica foi concebido por Danielle Corbin (1987, 1991, 1997, 2004) e posteriormente desenvolvido por sua equipe de trabalho. Esta pesquisa busca preencher uma lacuna, na medida em que analisa e descreve os processos morfológicos possíveis no âmbito de terminologias em português (variante brasileira) e examina as mais produtivas regras pelas quais os termos são formados.

Palavras-Chave: Análise Morfológica, Terminologia, Morfologia, Modelo SILEX.

Abstract

The researches in Linguistics and Language fields are extremely rich, consistent and efficient in Brazil, however, still aren't expressive in the Brazilian researches the issues involving morphology and terminology. In this context, this research aims to describe the structure of two terminological repertoire (Nanoscience / Nanotechnology and Biofuels), representing the area of Engineering, at the level of the internal structure of its constituent terms, checking out the main processes of terms construction. Using this morphological description was organized a database for future computational implementation and on-line data availability. Later, for the higher productivity processes the Rules of Words Formation were established, as established by SILEX Model. This model of morphological analysis was designed by Danielle Corbin (1987, 1991, 1997, 2004) and developed by his team. This research aims to fill a gap, analyzing and describing the possible morphological processes in Portuguese (Brazilian variant) terminology and examining the most productive rules by which words are formed.

Keywords: Morphological Analysis, Terminology, Morphology, SILEX Model.

Lista de Siglas

Adj: Adjetivo

Adv: Advérbio

N: Substantivo

Prep: Preposição

SP: Sintagma preposicionado

V: Verbo

Lista de Tabelas

Tabela 1: Exemplos de candidatos a termos obtidos a partir de padrões (ALMEIDA & VALE, 2008)	40
Tabela 2: Caracterização da base e do produto (ROCHA, 2008, p. 130).....	55
Tabela 3: Itens lexicais não-construídos da terminologia da Nanociência/Nanotecnologia	95
Tabela 4: Itens lexicais não-construídos da terminologia do Biodiesel	95
Tabela 5: Derivação prefixal - Nanociência/Nanotecnologia.....	101
Tabela 6: Derivação sufixal – Nanociência/Nanotecnologia	112
Tabela 7: Composição culta – Nanociência/Nanotecnologia	118
Tabela 8: Casos de recomposição com nano- na terminologia de N&N.....	119
Tabela 9: Derivação prefixal – Biocombustíveis	120
Tabela 10: Derivação sufixal – Biocombustíveis.....	122
Tabela 11: Composição culta – Biocombustíveis.....	123
Tabela 12: Frequência total de derivação sufixal – Nanociência/Nanotecnologia cf. Tabela 6	124
Tabela 13: Frequência total de composição culta – Nanociência/Nanotecnologia cf. Tabela 7	124
Tabela 14: Frequência total de derivação prefixal – Nanociência/Nanotecnologia cf. Tabela 6	125
Tabela 15: Frequência total de derivação sufixal – Biocombustíveis cf. Tabela 9	126
Tabela 16: Frequência total de composição culta – Biocombustíveis cf. Tabela 10.....	126
Tabela 17: Frequência total de derivação prefixal – Biocombustíveis cf. Tabela 8.....	127
Tabela 18: Valores semânticos do elemento “nano-”, segundo Antunes & Correia (no prelo)	137

Sumário

1. Introdução	12
2. Perspectiva teórica	18
2.1 Terminologia	18
2.2 Teoria Comunicativa da Terminologia	23
2.3 Terminologia e Morfologia	27
2.4 Estudos morfológicos de terminologias	32
2.4.1 Kasama <i>et al.</i> (2008): a prefixação por nano-	34
2.4.2 Almeida & Vale (2008): o uso de conhecimento morfológico na extração semiautomática de termos	37
2.4.3 Coleti & Almeida (2010): descrição morfológica da terminologia de N&N	41
2.5 Morfologia: abordagem gerativa	44
2.5.1 Conceitos básicos de morfologia gerativa, segundo Rocha (2008)	46
2.5.2 Alguns estudos de morfologia gerativa	48
2.5.2.1 Basílio (1999)	49
2.5.2.2 Rocha (2008)	50
2.5.2.3 Rio-Torto (1998)	56
2.5.2.4 Correia (2004)	58
2.5.2.5 Correia (2006)	64
2.5.3. Modelo SILEX	71
2.5.4 Processos de inovação lexical em português	78
2.5.4.1 Derivação	80
2.5.4.1.1 Derivação afixal	80
2.5.4.1.1.1 Derivação prefixal	81
2.5.4.1.1.2 Derivação sufixal	82
2.5.4.1.1.3 Derivação parassintética	82
2.5.4.1.2 Conversão	83
2.5.4.1.3 Derivação regressiva	83
2.5.4.2 Composição	83
2.5.4.2.1 Composição morfológica	83
2.5.4.2.2 Recomposição	84
2.5.4.2.2 Composição por lexicalização de sintagmas	84
2.5.4.3 Processos de formação	85
2.5.4.3.1 Amálgama	85
2.5.4.3.2 Truncação	85
2.5.4.3.3 Sigla e acronímia	86
2.5.4.4 Lexicalização de formas flexionadas	86
2.5.4.5 Lexicalização de unidades discursivas	86
3. Metodologia	87
3.1 Seleção e delimitação dos repertórios terminológicos	89
3.2 Análises: inicial e profunda	95
4. Base de dados morfológicos: os resultados da análise inicial dos processos morfológicos	97
4.1 Descrição morfológica da terminologia de Nanociência/Nanotecnologia	98
4.2 Descrição morfológica da terminologia de Biocombustíveis	119

4.3 Resultados da descrição morfológica da terminologia de Nanociência/Nanotecnologia.....	123
4.4 Resultados da descrição morfológica da terminologia de Biocombustíveis	125
5. Regras de Construção de Palavras: os resultados da análise profunda dos processos morfológicos mais recorrentes.....	127
5.1 Derivação sufixal: “-ão” (Nanociência/Nanotecnologia)	128
5.2 Derivação sufixal: "-or" (Biocombustíveis)	130
5.3 Derivação prefixal "super/sobre-" (Nanociência/Nanotecnologia)	132
5.4 Derivação prefixal: "des-" (Biocombustíveis).....	133
5.5 Composição morfológica: "nan(o)-" (Nanociência/Nanotecnologia)	134
5.6 Composição morfológica: “hidro-” (Biocombustíveis).....	138
6. Considerações finais	139
7. Referencias bibliográficas	142

1. Introdução

A inovação científica e tecnológica impõe uma constante inovação lexical a todas as línguas vivas, a “(...) capacidade natural de renovação do léxico de uma língua pela criação, importação ou reutilização de palavras já existentes, atribuindo-lhes novos significados, é comum a todas as línguas vivas e ocorre privilegiadamente nas linguagens especializadas como resultado de necessidades denominativas.” (ANTUNES & CORREIA, no prelo¹).

Por isso, considerando que diversas são as pesquisas em Morfologia e em Terminologia, mas em menor quantidade são as pesquisas que tratam de ambas as áreas, busca-se contribuir com o preenchimento da lacuna existente na disponibilização de dados morfolexicais no cenário do português do Brasil, na medida em que são analisados e descritos os processos morfológicos possíveis no âmbito de terminologias na variante brasileira já que “mais do que a forma aparente, é a estrutura da unidade lexical aquela que importa ao lexicólogo e ao morfologista e, portanto, aquela que determina o significado da unidade construída” (CORREIA, 2004, p. 336)

Os resultados obtidos da descrição e análise morfológica dessas terminologias, originalmente multidisciplinares, representarão significativamente o domínio científico das Engenharias e, se implementadas em sistemas híbridos de extração automática de terminologias, contribuirão com o aperfeiçoamento desses sistemas.

O estudo sistemático dos repertórios terminológicos que compõem as terminologias da Nanociência/Nanotecnologia (*Terminologia em Língua Portuguesa da Nanociência e Nanotecnologia: Sistematização do Repertório Vocabular e Elaboração de Dicionário-*

¹ Nas transcrições literais de obras escritas em português variante europeia, manteve-se a grafia original dos textos.

Piloto – NanoTerm; apoio CNPq/Processo n°. 400506/2006-8²) e Biocombustíveis (*Terminologia de Biocombustíveis: descrição semântica e morfológica com vistas à sistematização*; apoio CNPq/Processo n°. 473414/2007-4³) vem sendo desenvolvido, respectivamente desde 2006 e 2008, pelo Grupo de Estudos e Pesquisas em Terminologia – GETerm⁴ da Universidade Federal de São Carlos – UFSCar, sob coordenação da Profa. Dra. Gladis Maria de Barcellos Almeida, em parceria com o Núcleo Interinstitucional de Linguística Computacional – NILC⁵, sediado na Universidade de São Paulo – USP, *campus* de São Carlos.

A escolha do GETerm por esses domínios científicos visa à produção de produtos terminológicos que satisfaçam demandas reais, visto que são áreas em crescente expansão no mundo todo e que não encontram, no cenário brasileiro, repertórios vocabulares sistematizados.

A seguir, apresentam-se resumidamente ambos os domínios: Nanociência e Nanotecnologia (doravante N&N) e Biocombustíveis (doravante BioC).

Fazem parte dos estudos em Nanociência e Nanotecnologia as capacidades de medir, manusear e organizar a matéria em nanoescala, já que nessa escala a matéria apresenta propriedades, fenômenos e processos únicos (SENAI, 2004).

Nanociência e Nanotecnologia (N&N) são atualmente áreas centrais das atividades de pesquisa, desenvolvimento e inovação (nos setores de Saúde, Meio Ambiente, Agropecuária, Transportes, Informática, Comunicações, Educação e outros) em crescente

² Maiores informações em http://www.geterm.ufscar.br/geterm2/?page_id=16

³ Maiores informações em http://www.geterm.ufscar.br/geterm2/?page_id=111

⁴ Maiores informações em <http://www.geterm.ufscar.br>

⁵ Maiores informações em <http://www.nilc.icmc.usp.br>

expansão no mundo todo. Investimentos aplicados nessa área de conhecimento aumentam a cada ano.

Nano-, prefixo grego, adotado na 11ª Conferência Geral de Pesos e Medidas⁶ (Resolução nº 12 de 1960) equivale a um multiplicador 10^{-9} , ou seja, um bilionésimo da unidade indicada; assim, um nanômetro corresponde a 10^{-9} metros ($1 \text{ nm} = 10^{-9} \text{ m}$). Apenas para que se tenha uma ideia dessa pequenez: o diâmetro de um fio de cabelo humano mede cerca de 30.000 nanômetros, já “um minúsculo vírus, invisível a olho nu, se apresenta como uma incrível entidade com cerca de 200 nm” (TOMA & ARAKI, 2005). É desse mundo do *muito pequeno* que tratam a Nanociência e a Nanotecnologia.

A origem do estudo das nanotecnologias é atribuída ao físico Richard Feynman. Feynman, em discurso para a Sociedade de Física Americana em 1959, alertou para a possibilidade de se fabricar, algum dia, materiais e dispositivos de acordo com as especificações de seus átomos. Entretanto, o termo *nanotecnologia* só veio a ser cunhado em 1974 por Norio Taniguchi, da Universidade de Tóquio, para diferenciar os trabalhos de engenharia no domínio da microescala dos trabalhos em submicroescala, que denominou *nanotecnologia* (SENAI, 2004).

Com relação aos biocombustíveis, importa registrar que o início do século XXI foi marcado por uma nova revolução energética: trata-se da mudança do uso e produção das energias fósseis para as energias de biomassa (biocombustíveis) (PINO, 2010).

Os biocombustíveis são considerados uma fonte de energia superior, com custos menores, além de serem renováveis, enquanto a escassez de energia fóssil é, atualmente,

⁶ A CGPM é uma das três organizações criadas para avaliar e gerir o Sistema Internacional de Unidades (SI), nos termos da Convenção do Metro (1875).

anunciada por diversos geólogos já que “a partir de 1981 o mundo começou a usar mais petróleo do que descobria” (PINO, p. 11, 2010).

A apreensão em torno das mudanças climáticas, causadas pela emissão excessiva de gases de efeito estufa, oriundos de energias fósseis, é outra preocupação atual concernente à sobreposição das energias fósseis por energias de biomassa, visto que os resíduos oriundos de sua queima não interferem no efeito estufa (PINO, 2010).

Neste contexto, o Brasil figura como o país mais competitivo do mundo para a geração de biocombustíveis, tendo aumentado os investimentos em energias de biomassa em todo o país (PINO, 2010).

Para acompanhar esse desenvolvimento científico e tecnológico, além de investimentos financeiros expressivos e formação de recursos humanos especializados, é preponderante a sistematização de repertórios vocabulares em língua portuguesa (LP). Com este fim, esta pesquisa se insere nos domínios da Linguística, nomeadamente na Descrição do Português, nível Morfológico. Aspectos terminológicos (com ênfase para a Teoria Comunicativa da Terminologia) também são tratados, dado que o objeto desta investigação são termos⁷ que integram um vocabulário especializado.

A metodologia do trabalho terminológico adotada no âmbito do GETerm, norteadora dos trabalhos citados e outros, está em consonância com os postulados de uma Terminologia de orientação descritiva, fundamentada em princípios da Linguística, cujo exemplo mais consolidado é a Teoria Comunicativa da Terminologia (TCT) (CABRÉ, 1999, 2003). Filiar-se teoricamente à TCT significa, fundamentalmente, abandonar “o tratamento prescritivo das terminologias em favor de enfoques descritivos capazes de

⁷ Nessa pesquisa adotam-se como sinônimos as seguintes expressões: termo, unidade terminológica (UT) unidade léxica (especializada).

entender o léxico especializado como um elemento natural das línguas naturais” (KRIEGER & BEVILACQUA, 2005).

Por isso, concebem-se como questões terminológicas fundamentais i) que os léxicos de especialidade não compõem línguas artificiais, apartadas da língua geral, mas integram a língua natural e geral, de forma que ii) os termos são signos linguísticos que podem realizar-se como termo ou como palavra, de acordo com a situação comunicativa.

A Terminologia, que desempenha papel fundamental no processo de divulgação dos novos conhecimentos científicos gerados, conheceu um desenvolvimento intenso nas últimas décadas, acompanhando o avanço científico e tecnológico. Suas bases teóricas têm sido revistas e novos modelos propostos; diferentes campos de atuação estão surgindo e novos desafios estão se apresentando.

Compreendendo “as unidades da terminologia como unidades do léxico associadas a um valor especializado ativado pelas condições discursivas e pragmáticas” (tradução nossa, CABRÉ, 2006), esta pesquisa tem como objetivo geral descrever e analisar os aspectos morfológicos de uma série de termos extraídos das terminologias em análise, evidenciando, assim, a integração entre Terminologia e Morfologia, integração essa cada vez mais urgente e promissora; e como objetivos específicos os seguintes:

1. descrever a estrutura dos vocabulários da N&N e dos BioC (especificamente o vocabulário do biodiesel) no nível da estrutura interna dos seus termos constitutivos, verificando-se os principais processos de construção dos termos nesses dois vocabulários;
2. propor padrões morfológicos dos processos mais produtivos de ambas as terminologias, os quais serão organizados numa base de dados, de maneira que seja possível a sua posterior implementação computacional e disponibilização on-line.

Como esta pesquisa gira em torno de dois grandes eixos, a Terminologia e a Morfologia, no que se refere à fundamentação teórica, elegeu-se a Teoria Comunicativa da Terminologia (CABRÉ, 1999, 2003) para tratar dos aspectos terminológicos; e o Modelo SILEX (CORBIN, 1987, 1991, 1997, 2004; CORREIA, 2004 e 2006; RIO-TORTO, 1998) que orientará todas as análises morfológicas, já que, para esse modelo, “as palavras construídas são portadoras de um significado previsível, linguístico, que lhes é conferido pela sua estrutura; desse modo, torna-se mais evidente a relação que existe entre a unidade lexical e sua referência⁸” (CORREIA, 2004, p. 21). Assim, o estudo das palavras construídas proporciona uma melhor análise da construção do significado descritivo ou referencial das unidades lexicais, no âmbito do modelo adotado (CORREIA, 2004), enfoque extremamente útil para o estudo de terminologias.

O texto que ora se apresenta está assim organizado: são apresentadas na seção 2, a seguir, as perspectivas teóricas norteadoras da pesquisa, com ênfase para Terminologia, Morfologia e especial atenção dada ao Modelo SILEX de análise morfológica. Descrevem-se na seção 3 os procedimentos metodológicos. Os resultados obtidos com essa pesquisa são apresentados inicialmente na seção 4, na qual se propõe a base de dados morfológicos e, na seção 5, as Regras de Construção de Palavras. Nas seções 6 e 7, apresentam-se, respectivamente, as considerações finais e as referências bibliográficas.

⁸ Correia (2004) concebe a referência “como objeto mental, sendo distinta, portanto, da denotação (relação entre a referência e a realidade extra-linguística)”. (CORREIA, 2004, p. 21)

2. Perspectiva teórica

Nesta seção, estabelece-se a perspectiva teórica que embasa esta pesquisa, a qual está dividida em dois grandes eixos: a Terminologia e a Morfologia.

Assim, é apresentada inicialmente uma breve incursão referente à Teoria Comunicativa da Terminologia, dado que o objeto desta investigação são termos que integram vocabulários especializados, obtidos a partir de um método fundamentado nessa teoria. Na sequência, são apresentados trabalhos que contribuíram com a aproximação dos estudos de Terminologia e Morfologia, bem como aqueles que relatam sua aplicação. A partir disso, serão apresentados alguns estudos do léxico no Brasil e em Portugal realizados a partir da perspectiva gerativa e o Modelo SILEX, que sustentará as descrições e análises feitas ao longo do trabalho.

2.1 Terminologia

Sem terminologia não se faz ciência, nem se descreve uma técnica, nem se exerce uma profissão especializada.

(CABRÉ, 1999)

Terminologia é um termo polissêmico, podendo significar, segundo a ISO 1087 (2000), a “ciência que estuda a estrutura, a formação, o desenvolvimento, o uso e a gestão das terminologias em diferentes domínios” ou um “conjunto de designações pertencente a uma língua de especialidade”. Nesta seção, abordar-se-á a Terminologia como ciência (conforme a primeira acepção da ISO 1087).

A Terminologia alinha-se à Lexicologia, à Lexicografia, à Morfologia e à Semântica (KRIEGER & FINATTO, 2004), mas mantém sua autonomia, sobretudo no que se refere ao objeto de análise e ao método de trabalho. Seu objetivo principal é “dar conta do funcionamento das unidades lexicais especializadas em situações comunicativas profissionais, acadêmicas ou científicas” (LORENTE, 2004).

Criar novas palavras para nomear o mundo, o conhecimento humano, a ciência, a prática profissional e as novas tecnologias é uma antiga necessidade humana, no entanto, muito recente é o surgimento de um campo de estudos dedicado à Terminologia, o qual começa a ser estabelecido a partir da segunda metade do século XX, com o tardio interesse dos linguistas sobre a constituição e o funcionamento dos termos técnico-científicos.

Na verdade, não existe ciência encerrada em si mesma, sem formas próprias de expressão. É necessário, então, comunicar ciência. E, mais uma vez, a língua, sob um figurino especializado, é a protagonista que desempenha o papel de ajudar a escrever a ciência. Explica-se, assim, também, o papel das terminologias na expressão dos saberes humanos (BORGES, 1998⁹ *apud* KRIEGER & FINATTO, 2004).

Desde o início do fazer terminológico, é notável o interesse de especialistas de diferentes áreas do conhecimento, esse interesse se deve à necessidade de se relacionar as denominações aos seus respectivos conceitos científicos, sobretudo no século XIX em que o desenvolvimento das ciências faz com que os cientistas busquem cada vez mais entender e descrever as regras de formação dos termos de cada domínio de especialidade (CABRÉ, 1993).

⁹ BORGES, M. F. *Identificação de sintagmas terminológicos em Geociências*. Porto Alegre: UFRGS, 1998. Dissertação de Mestrado (inclui volume de anexos com 254 pp.)

O estabelecimento da Terminologia como disciplina científica¹⁰ que estuda os termos de uma área de especialidade ocorre a partir dos trabalhos do engenheiro eletrônico Eugen Wüster (Wieselburg 1899 – Viena 1977) (BARROS, 2004 e CABRÉ, 2006), que publica na Universidade Técnica de Stuttgart (Alemanha), em 1931, sua tese de doutorado intitulada *A normalização internacional da terminologia técnica*¹¹ (CABRÉ, 1996). Com esse trabalho, Wüster inaugura a chamada Teoria Geral da Terminologia (TGT)¹², estabelecendo suas bases, que foram posteriormente desenvolvidas pela Escola de Viena, da qual o principal expoente foi o próprio Wüster (CABRÉ, 1993).

A necessidade de diálogo evidencia que, para Wüster, a terminologia de uma área é, em sua natureza, a expressão de um conhecimento científico, logicamente estruturado. Nesse sentido, os termos refletem fundamentos conceituais, bem como representam a apreensão da essência dos fenômenos estudados pelas especializações. Por isso, a colaboração dos especialistas complementa o círculo de vozes necessárias para explicar os termos (KRIEGER & FINATTO, 2004).

A partir da concepção positivista de que os termos são designações de conhecimentos científicos que expressam conceitos,¹³ Wüster propôs normalizar os termos para uniformizar a comunicação entre diferentes áreas de especialidades, já que, segundo ele, o conceito seria universal, atemporal e paradigmático, alterando-se entre diferentes línguas apenas a etiqueta léxica (CABRÉ, 1993; KRIEGER & FINATTO, 2004).

¹⁰ Ou “campo de conhecimento autônomo” segundo Cabré (2006).

¹¹ Tradução nossa.

¹² “Essa teoria, que se tornou um marco na história da área, está registrada em obra póstuma, intitulada *Introdução à Teoria Geral da Terminologia e à Lexicografia Terminológica*. A versão original foi publicada na Alemanha, em 1979, logo após a morte de Wüster, por seu discípulo Helmut Felber, que reuniu manuscritos e notas das aulas de Terminologia que o mestre havia ministrado na Universidade de Viena entre 1972 e 1974” (KRIEGER & FINATTO, 2004).

¹³ Em detrimento da concepção de que os termos seriam elementos das línguas naturais que expressam significados.

O objetivo de Wüster era conseguir uma comunicação eficiente e sem ambiguidade nas interações profissionais, para isso, propôs uma terminologia normalizada (CABRÉ, 2006). Assim surge o conceito de normalização em terminologia (FELBER, 1996 *apud* ALMEIDA, 2003).

Wüster desvinculou a teoria da Terminologia de uma teoria de língua geral, uma vez que, para ele, a Terminologia deveria ser capaz de controlar a evolução da língua, podendo, assim, prescrever os melhores usos e as melhores denominações para cada área de especialidade.

Seu trabalho está baseado em uma perspectiva onomasiológica (do conceito ao termo), com conceitos estáticos e bem definidos em um sistema estrutural. Para Wüster, os termos e conceitos mantêm uma relação biunívoca (um termo deve ser relacionado a apenas um conceito, assim como um conceito deve ser relacionado a apenas um termo) e devem ser estudados sincronicamente.

A difusão da TGT suscita um maior interesse pela Terminologia nos domínios especializados e influencia a criação do comitê técnico “Terminologia: princípios e coordenação” (TC37) da Federação Internacional das Associações Nacionais de Normalizadores, a precursora da atual ISO (*International Standardization Organization*) (GALINSKY, 1996).

Segundo Barros (2004), paralelamente às atividades de Wüster, desenvolvem-se também na ex-URSS estudos terminológicos. D. S. Lotte (1898-1950) é o fundador desta escola, chamada de Escola Russa de Terminologia. Ainda na mesma época, na Checoslováquia, foi formado outro pólo de excelência em Terminologia. Posteriormente, a Terminologia chegou à França e ao Canadá. Na fase seguinte, alcançou à América Latina e Portugal (BARROS, 2004).

É importante ressaltar, no entanto, que somente a partir da década de 50, com a Terminologia deixando de ser vista apenas como um instrumento de normalização de termos para tornar-se também um instrumento de comunicação é que ela se torna objeto de interesse dos linguistas (CABRÉ, 1993). “Com esse novo enfoque, a Terminologia se consolida como disciplina autônoma, com teoria e metodologia próprias, gerando, assim, produtos terminológicos de atestada qualidade científica”, conforme Almeida (2003).

As escolas de Terminologia podem ser agrupadas em dois diferentes grupos: (i) as escolas pioneiras que se caracterizaram pelo enfoque cognitivo do fenômeno terminológico e (ii) as que se desenvolvem a partir de uma visão do funcionamento linguístico dos termos. No primeiro grupo, que congrega as clássicas escolas de Viena, de Praga e a russa, prevalece uma perspectiva normativa sobre as terminologias. Com o avanço da Linguística, as demais escolas (do segundo grupo) adotam linhas de enfoque descritivo sobre o léxico especializado (KRIEGER & FINATTO, 2004).

Temmerman (2000) destaca como principais reações críticas à Escola de Viena os trabalhos de Juan C. Sager, Peter Weinsenhofer, Britta Zawada & Piet Swanepoel, Ingrid Meyer, Jean-Claude Boulanger, François Gaudin, Kyo Kageura e Maria Teresa Cabré. Esta última dará início à construção da Teoria Comunicativa da Terminologia, perspectiva teórica adotada neste trabalho e apresentada a seguir.

Pode-se destacar, também, a Socioterminologia, postulada por François Gaudin, que critica a inoperância de dicionários e glossários entre outros produtos terminológicos ao não expressar a real empregabilidade dos termos, por não registrar as variações denominativas e conceituais, deixando assim de atender às necessidades informacionais dos usuários. Gaudin propõe, então, a suplantação do artificialismo normalizador pelo exame dos termos em seu real contexto de ocorrência, tendo como consequência o reconhecimento da

variação e da sinonímia na comunicação especializada, assim a variação passa a ser tida como o eixo central para o desenvolvimento da Socioterminologia (KRIEGER & FINTATTO, 2004).

Como contestação da Escola de Viena, surge também a Teoria Sociocognitiva da Terminologia (TST), postulada por Rita Temmerman (1998 e 2000). A Teoria proposta por Temmerman propõe um procedimento para a descrição terminológica com base na análise de informação textual. A TST “parte de unidades de interpretação que geralmente apresentam estrutura prototípica” (TEMMERMAN, 2004). Esta teoria propõe que “dependendo do tipo de unidade e do nível e tipo de especialização do emissor e receptor na comunicação, o que é informação mais ou menos essencial para uma definição irá variar” (TEMMERMAN, 2004). A necessidade de se descrever a sinônimia e a polissemia é defendida a partir do posicionamento de que estas “são funcionais no desenvolvimento da compreensão” (TEMMERMAN, 2004). A TST defende ainda um posicionamento diacrônico já que as “unidades de interpretação evoluem constantemente” e a adoção de modelos cognitivos ao postular que estes possuem papel garantido no “desenvolvimento de novas ideias, o que significa que os termos são motivados” (TEMMERMAN, 2004) em oposição a uma relação arbitrária entre conceito e termo.

Apresenta-se, a seguir, a Teoria Comunicativa da Terminologia, que embasa esta pesquisa no que se refere aos aspectos terminológicos.

2.2 Teoria Comunicativa da Terminologia

Questionamentos sobre a insuficiência dos postulados da terminologia clássica, a chamada Teoria Geral da Terminologia (TGT), fez com que novos paradigmas fossem

propostos. Desses, merece destaque a Teoria Comunicativa da Terminologia (TCT) que, em pouco tempo, passou a ser referência teórica em grande parte das pesquisas terminológicas realizadas no Brasil, isso devido ao fato de que esta teoria descritiva de base linguística melhor se adequou a realidade brasileira, a saber, a de um país monolíngue, no entanto, com grande variedade dialetal (ALMEIDA, 2006).

O que motiva o surgimento da TCT remonta a um cenário político que viveu a Espanha durante a ditadura de Franco, regime totalitário de base ideológica fascista, que vigorou na Espanha entre 1939 e 1976. Durante esses quase 40 anos, a língua catalã foi proibida de ser ensinada na escola e de ser utilizada em todas as comunicações oficiais. Após a queda da ditadura, os catalães conseguem retomar o plano de fortalecimento da língua, incluindo sua modernização, o ensino da competência escrita, principalmente, e a propagação de seu uso nos meios de comunicação. Nesse contexto, são retomadas as atividades terminológicas e, em 1985, é criado o TERMCAT, que coordenaria a terminologia catalã (CABRÉ, 1993). A partir de então, começa a ser delineada a Teoria Comunicativa da Terminologia (TCT) pelo grupo IULATERM da Universidade Pompeu Fabra em Barcelona (Espanha), coordenado por Maria Teresa Cabré.

Os paradigmas da TCT são estabelecidos em Cabré (1998)¹⁴, (1999)¹⁵ e (2003)¹⁶, entretanto, a autora já acenava com uma diferente perspectiva de abordagem da Terminologia em sua obra anterior Cabré (1993)¹⁷ (ALMEIDA, 2006).

¹⁴ CABRÉ, M. T. et al. *La terminologia hoy: replanteamiento o diversificación*. Organon, v. 12, n.26, p. 33-41, 1998

¹⁵ CABRÉ, M. T. *La terminología: representación y comunicación – elementos para una teoría de base comunicativa y otros artículos*. Barcelona, Institut Universitari de Lingüística Aplicada, 369 p., 1999

¹⁶ “Theories of Terminology - their Description, Prescription and Explanation”, publicado em 2003 na revista *Terminology* (v.9, n.2, p.63-199).

¹⁷ CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones* (trad. castellana de Carles Tebé). Barcelona, Editorial Antàrdida/Empúries, 529 p., 1993.

Abandonar “o tratamento prescritivo das terminologias em favor de enfoques descritivos capazes de entender o léxico especializado como um elemento natural das línguas naturais” (KRIEGER & BEVILACQUA, 2005 *apud* ALMEIDA *et al.*, 2007a) é objetivo do terminólogo filiado teoricamente à TCT.

A partir de fundamentos epistemológicos distintos, a TCT procura descrever as unidades terminológicas a partir de uma perspectiva linguística, com o principal objetivo de facilitar a comunicação especializada não só entre os especialistas de domínio, mas também entre estes e o público em geral. Nesse sentido, é dado maior relevo à comunicação e não à normalização terminológica, “bem como na compreensão de que as unidades terminológicas formam parte da linguagem natural e da gramática das línguas” (KRIEGER & FINATTO, 2004).

Os termos, para a TCT, não pertencem a um domínio, mas são empregados em cada domínio com um valor específico (CABRÉ, 1999). Ou seja, para a TCT, tanto os termos quanto os textos especializados contêm características próprias da língua geral: nesse sentido, pode-se afirmar que as terminologias não constituem línguas artificiais apartadas, como postulava a TGT. Ou ainda, “(...) não existe, pois, um conjunto de termos isolados constituindo uma língua marginal à língua geral; o que há são signos da língua natural que se realizam ora como palavras, ora como termos” (ALMEIDA, 2003).

Outro aspecto relevante é a não uniformidade do conhecimento especializado, assim, os termos não são unidades isoladas constituindo um sistema próprio, mas sim unidades que se incorporam ao léxico do falante enquanto este adquire o conhecimento especializado (CABRÉ, 1999). “Conseqüentemente, o conteúdo de um termo não é fixo, mas relativo, variando conforme o cenário comunicativo em que se inscreve” (KRIEGER & FINATTO, 2004).

A Terminologia deve ser capaz de explicar a interdisciplinaridade e a multidimensionalidade dos termos, levando em conta que um conceito pode formar parte da estrutura conceitual de distintas disciplinas, conservando, alterando ou matizando suas características e, conseqüentemente, a diversidade de visões que os especialistas envolvidos têm desses termos (CABRÉ, 1999).

Segundo Cabré (1999), a Terminologia é um campo interdisciplinar, construído a partir de três teorias:

1. **teoria do conhecimento:** que explica como se conceitualiza a realidade;
2. **teoria da comunicação:** que descreve, a partir de critérios explícitos, os tipos de situações em que se produzem as unidades de conhecimento especializado;
3. **teoria da linguagem:** que dá conta das unidades terminológicas (UTs) propriamente ditas dentro da linguagem natural, considerando que as UTs participam de todas as características dessa linguagem, porém, singularizando seu caráter terminológico e explicando como se ativa esse caráter na comunicação. A teoria da linguagem deve considerar três vertentes das UTs:
 - suas características gramaticais, semânticas, textuais e pragmáticas;
 - o uso que os especialistas fazem dessas UTs, explicando em que situações são empregadas e que valor adquirem em cada caso;
 - a aquisição da linguagem em geral e das terminologias em particular, ou seja, como o falante aprende uma especialidade e como adquire o conhecimento especializado e suas respectivas unidades de expressão e comunicação.

A finalidade aplicada da compilação e análise das terminologias é bastante diversa e permite muitas aplicações, mas em todas elas se ativa a dupla função dos termos: a

representação do conhecimento especializado e a sua **transferência**, ainda que em graus e maneiras distintos e em situações também diversas (CABRÉ, 1999).

A TCT requer que, tanto do ponto de vista teórico como metodológico, seja contemplada a variação linguística em toda a sua dimensão, já que é uma característica constitutiva da linguagem natural (CABRÉ, 1999).

Os conceitos, influenciados pelos fatores socioculturais e linguísticos de uma comunidade, bem como pelos canais comunicativos por onde circulam e se difundem, apresentam diferentes graus de especialização (CABRÉ *et al.*, 1998 *apud* ALMEIDA, 2003). Nesse sentido, Aubert (1996) afirma: “A comunidade de usuários das linguagens de especialidade, tanto quanto a comunidade e a língua em geral, não constituem um todo uniforme, mas se subdividem em grupos variados, com necessidades, pressupostos e motivações também variados (...)”. Assim, as UTs estão sujeitas a variações conceituais e denominativas.

A importância social da Terminologia está determinada pelas necessidades da sociedade, marcada pela expansão do conhecimento especializado e pelo plurilinguismo necessário.

2.3 Terminologia e Morfologia

A aproximação entre Morfologia e Terminologia se dá pelo léxico, ou mais especificamente, pela unidade léxica, que pode ser considerada um ponto forte de intersecção entre ambas as disciplinas. Isso fica evidente no surgimento de novas unidades, já que os neologismos constituem tema fundamental de estudos em Terminologia e Morfologia.

A criação lexical tem como uma de suas principais motivações o avanço científico e tecnológico, pois devido ao desenvolvimento de novas teorias, técnicas, produtos ou processos novas denominações são necessárias, assim, criam-se novas unidades léxicas que serão empregadas nos contextos comunicativos de especialidade para nomear as inovações surgidas. Estas novas unidades léxicas constituirão o repertório terminológico do domínio de especialidade.

Alves (1984) enfatiza esse aspecto, por meio de uma citação de Quemada (1971)¹⁸:

“uma língua de cultura moderna, necessariamente científica e técnica, não deve ver na neologia lexical apenas um mal inevitável. É a primeira condição a partir da qual o idioma pode permanecer um instrumento de comunicação nacional, mesmo internacional, e não ser apenas uma língua viva. Deve até considerar a criatividade lexical como parte responsável pela sua riqueza imediata, como sinal evidente de sua vitalidade” (QUEMADA *apud* ALVES, 1984, p. 119).

Também o morfema se estabelece como ponto de ligação entre Morfologia e Terminologia, já que é empregado na construção de denominações terminológicas e pode ser usado inclusive para identificar novos termos. Para ilustrar esse aspecto, , apresentam-se a seguir alguns dos diversos morfemas de origem greco-latina encontrados na terminologia da Fisioterapia¹⁹ e os seus respectivos exemplos (ALMEIDA, 2006):

a) artr(i/o)-: elemento de composição que significa 'juntura' ou 'articulação'

Exemplos: *artrite* e *artrogripose*;

b) -ite: sufixo muito produtivo nos nomes médicos, designando doença inflamatória de órgão, tecido, ou outros a que se refere o radical. Exemplos: *bronquite*, *bursite* e *dermatite*;

¹⁸ QUEMADA, B. A propos de la néologie. *La Banque des Mot*, Paris: 2, 137-50, 1971

¹⁹ O GETerm desenvolve desde 2001 o projeto DicFisio com o objetivo de construir um Dicionário de Fisioterapia, para maiores informações consulte http://www.geterm.ufscar.br/geterm2/?page_id=135

c) -óide: sufixo que indica 'aspecto ou forma de', 'semelhante a' ou ainda 'relativo a'

Exemplos: *deltóide, escafóide, quelóide*.

d) -ose: sufixo muito produtivo na fisioterapia, utilizado para formar os substantivos de diversos processos patológicos e doenças, agregando-se a uma base que designa quer o órgão, quer parte dele. Exemplos: *osteoporose* e *trombose*.

Nota-se, também, que as estruturas e combinações morfemáticas recorrentes na língua geral transbordam para as comunicações especializadas, como bem ilustra a citação de Correia a seguir:

Deste facto decorre que não apenas iremos encontrar inúmeros termos que têm por base unidades da língua corrente que se transferem para registos linguísticos especializados, como, além disso, iremos encontrar os mecanismos de construção de palavras que encontramos no léxico da língua corrente também nas unidades terminológicas. (CORREIA, 2004, p. 3)

Além desses pontos fortes de intersecção – a unidade léxica e o morfema –, pode-se ressaltar também a estreita relação que existe entre palavra e termo, haja vista que, muitas vezes, os termos são unidades da língua geral associadas a um valor especializado apenas em situações discursivas e pragmáticas específicas, ou seja, quando utilizadas em âmbitos de especialidades (CABRÉ, 2006).

A esses pontos de intersecção, pode-se acrescentar a seguinte proposição de Cabré (2005):

“(…) a observação dos dados empíricos própria do trabalho científico levou-me a constatar que entre os termos e as chamadas palavras do léxico das línguas se davam mais coincidências que divergências. E esta constatação me levou a explicar os termos e palavras a partir de uma mesma teoria das unidades léxicas, sem negar a especificidades dos termos, basicamente nos aspectos relativos a sua semântica e seu uso (CABRÉ, 2005, p. 13, tradução nossa)²⁰

²⁰ “(…) la observación de los datos empíricos propia del trabajo científico me condujeron a constatar que entre los términos y las llamadas palabras del léxico de las lenguas se daban más coincidencias que divergencias. Y

A integração dos estudos de Terminologia e Morfologia propiciou um mais amplo entendimento dos mecanismos que, relacionados à neologia terminológica, contribuem “para explicitar os processos envolvidos na constituição e uso das terminologias dos diferentes domínios de especialidade.” (CORREIA, 2006, p. 32), já que “a terminologia de um determinado domínio apenas se encontrará completamente descrita quando a estrutura dos termos se encontrar suficientemente caracterizada, dado que essa terminologia também se caracteriza pelas estruturas morfológicas dos termos que a compõem” (CORREIA, 2006, p. 32).

Correia (2006) defende que a estrutura interna dos termos é uma marca das condições sócio-históricas em que foram produzidos, servindo a objetivos comunicativos específicos. A escolha dos operadores morfológicos “resulta da necessidade de denominar os conceitos de determinada forma, determinada pelas condições sócio-históricas em que esta terminologia foi criada e é/foi usada, pela própria natureza dos conceitos em causa e pela visão que o falante tem ou se quer evidenciar relativamente às entidades denominadas” (CORREIA, 2004a, p. 41). Na terminologia da Fisioterapia, por exemplo, as formações eruditas são bastante produtivas, pois essa área recebe muitos termos emprestados da Medicina, cujos morfemas têm origem grega ou latina, como já mencionamos anteriormente.

Segundo Correia (2006), por mais que o significado de uma palavra derivada dependa significativamente da Regra de Construção de Palavras (RCP) que lhe deu origem, ainda assim é verdade que de uma mesma RCP fazem parte mecanismos e afixos

esta constatación me llevó a la propuesta de explicar términos y palabras desde una misma teoría de las unidades léxicas, sin negar la especificidad de los términos, básicamente en los aspectos relativos a su semántica y su uso.”

diferentes, resultando em construções com significados ligeiramente diferentes. Essas variações de significado são obtidas consciente ou inconscientemente, de acordo com a forma como se deseja caracterizar o referente, ou ainda, de acordo com as características do referente que se pretende destacar. Para ilustrar esse fenômeno, podemos recorrer à terminologia da N&N que apresenta adjetivos denominais construídos por meio de sufixos diferentes, os chamados adjetivos relacionais, que evidenciam relações diferentes (com o nome que modificam) consoante o sufixo escolhido para a sua construção: veja-se os casos de adjetivos em –al (coloidal, espectral, orbital), em –ic(o) (atômico, iônico, térmico), em –oso (argiloso, gasoso, poroso), todos pertencentes ao paradigma afixal da mesma Regra de Construção de Palavras.

Dessa forma, o conhecimento morfológico sobre terminologias permite um melhor e mais amplo entendimento de suas significações como também auxilia na identificação das terminologias, constituindo-se assim num método *bootstrapping*²¹, no qual a análise dos termos fornece material linguístico útil para a obtenção de termos (extração terminológica) de maneira cada vez mais eficiente, reforçando a pertinência da Morfologia para a Terminologia. Nessa direção, Almeida & Vale (2008) afirmam que:

“Dentre os três métodos de extração, os mais eficientes são justamente os que dependem de conhecimento linguístico, conhecimento este que constitui, sobretudo, uma exaustiva descrição morfológica da terminologia em estudo, o que significa que se deve proceder a essa descrição antes e durante a extração e não depois, já que essa descrição será imprescindível para a identificação dos candidatos a termos” (ALMEIDA & VALE, 2008, p. 485).

²¹ Um processo iterativo e incremental, no qual um resultado obtido é transformado em matéria para um novo experimento ou processo contribuindo assim para o refinamento/melhoramento do novo resultado a ser obtido, e assim sucessivamente.

No caso desta pesquisa, no entanto, as terminologias já se encontram completamente extraídas por meio estatístico, com uso de *stoplists*²² e validadas por especialistas humanos. Ressalte-se, também, que ambas as terminologias desenvolvidas no GETerm já foram objeto de outras duas pesquisas de mestrado, sempre numa perspectiva semântica²³, de maneira que a análise e a descrição agora realizadas, de enfoque morfológico, servirão como base para trabalhos futuros de extração terminológica e/ou outros experimentos voltados para Terminologia, Morfologia e Processamento Automático da Língua Natural (PLN).

Por isso tudo, antes de se iniciar a apresentação dos estudos morfológicos derivacionais no Brasil e em Portugal e a descrição do Modelo SILEX, serão apresentados estudos morfológicos de terminologias, especialmente aqueles desenvolvidos no âmbito do GETerm.

2.4 Estudos morfológicos de terminologias

Não se pode deixar de observar que cada terminologia possui estruturas morfológicas características. Exemplo disso pode ser encontrado no trabalho de Almeida & Vale (2006). Os autores abordam a variação dos padrões morfológicos específicos em três diferentes terminologias. Na terminologia de Revestimento Cerâmico, observam alta frequência de lexias complexas, tais como *argila refratária aluminosa, análise granulométrica por peneiramento, resíduo em malha*, etc.) e lexias simples com morfemas

²² A *stoplist* é formada por uma lista de palavras que devem ser evitadas pelo programa na geração dos candidatos a termos, pois não são relevantes para o intuito da pesquisa.

²³ A primeira pesquisa envolvendo a terminologia de N&N foi a dissertação de mestrado intitulada *Estruturação do Conhecimento e Relações Semânticas: uma ontologia para o domínio da Nanociência e Nanotecnologia*, desenvolvida por Deni Yuzo Kasama e defendida na UNESP- São José do Rio Preto em 2007. A segunda, tendo como objeto a terminologia dos BioC-biodiesel, foi a dissertação de Douglas Henrique Perez Pino, intitulada *Aspectos semânticos da terminologia do biodiesel*, defendida em 2008 na UFSCar.

que servem como identificadores de termos, é o caso dos sufixos derivacionais *-agem*, *-ção* e *-mento* que podem indicar que o conceito veiculado é um processo, como exemplo dessa proposição os autores citam: *secagem*, *moagem*, *britagem*, *ensilagem*, etc; *trituração*, *atomização*, *defloculação*, etc; *peneiramento*, *destorroamento*, *envelhecimento*, etc. Já na terminologia da Fisioterapia, os autores apontam como produtivas as formações eruditas (entenda-se compostos com recurso a radicais greco-latinos), ou seja, os termos cujos morfemas, tanto radicais quanto afixos, têm origem grega ou latina, como por exemplo: *artr(i/o)-* (do gr. *árthron*), *-óide* (do gr. *-(o)eîdos*, pelo lat. cient. *-îdes*), *-ose* (do lat. cient. *-osis* < gr. *-ōsis*), etc., como já citado anteriormente. A terminologia da N&N, por sua vez, apresenta lexias simples e complexas, tendo como característica principal a alta frequência do prefixo *nano-* (18.370 ocorrências num *corpus* de 2.565.490 palavras), como seria de esperar, tais como: *nanocompósitos*, *nanossistema biológico*, *nanotubo de carbono*, *nanopartícula de dióxido de titânio*, etc.

A variação dos padrões morfológicos, como observada nos exemplos anteriores, demanda uma ampla e detalhada descrição morfológica do maior número de terminologias possível, perfazendo todas as grandes áreas de conhecimento, para que se possa ter uma visão suficientemente abrangente das possibilidades de realização morfológica em diferentes terminologias e, dessa forma, circundar o fenômeno com vistas a melhorias nas aplicações em PLN, nomeadamente na extração automática de terminologias para o português. Nesse sentido, as áreas do saber (Nanociência/Nanotecnologia e Biocombustíveis) abordadas nesta pesquisa representam, de alguma forma, as Ciências Exatas e Engenharias, e a terminologia da Fisioterapia, pertencente à área da Saúde, é

objeto de outra pesquisa de mestrado²⁴, também no âmbito do GETerm, e sua descrição terminológica integrará a base de dados iniciada com esta pesquisa. Importa chamar atenção para o fato de que não se trabalhou ainda no GETerm com qualquer terminologia pertencente à grande área de Ciências Humanas; no entanto, certamente será objeto de projetos futuros, já que contribuir com a descrição morfológica de terminologias do português é meta do referido grupo de pesquisa.

Nas subseções seguintes, apresentar-se-ão alguns trabalhos envolvendo estudos morfológicos de terminologias.

2.4.1 Kasama *et al.* (2008): a prefixação por nano-

Kasama *et al.* retomam o trabalho de Alves (2000)²⁵ para tratar da denominação da neologia em contextos terminológicos. Esta recebe denominação própria: “*neonímia*, de acordo com Rondeau (1984²⁶), e *neotermo*, segundo Boulanger (1984²⁷)” (KASAMA *et al.*, 2008). Apesar da denominação própria, os processos envolvidos na formação de neologismos ou neónimos/neotermos são os mesmos, segundo Alves (2000) citada por Kasama *et al.* (2008).

²⁴ A pesquisa citada constitui o projeto de mestrado de Daniela Ferreira de Mattos, intitulada “Descrição e análise morfológica da terminologia da Fisioterapia: subsídios para organização de uma base de dados morfológicos de terminologias do português do Brasil” (processo FAPESP: 2011/05546-6, vigência: 01/08/2011 a 28/02/2013), também com orientação de Gladis Maria de Barcellos Almeida e coorientação de Margarita Correia.

²⁵ ALVES, I. M. Um estudo sobre a neologia lexical: os microssistemas prefixais do português contemporâneo. Tese de Livre-Docência. São Paulo: USP, 2000.

²⁶ RONDEAU, G. Introduction à la terminologie. 2e éd. Québec, Gaetan Morin, 1984.

²⁷ BOULANGER, J. C. L'évolution du concept de Néologie, de la linguistique aux industries de la langue. In Schaetzen, C. de (Org.). Terminologie diachronique. Paris/Bruxelles: Conseil International de la Langue Française/Ministère de la Communauté de Belgique. 1989

Os processos de criação lexical são frequentes e de diferentes tipologias, sua observação e análise permitem comprovar a afirmação de Humbley (2000),²⁸ citado em Kasama et al:

“a neologia nasce por definição no discurso, e se tentamos aplicá-la na língua, convém naturalmente estudar as condições que presidem seu nascimento. A palavra já "colocada no discurso" tem talvez mais chance de se implantar, em condições análogas, que uma criação „*in vitro*” (HUMBLEY, 2000, p.176 *apud* Kasama et al. 2008, p. 9, tradução nossa).²⁹

Os autores selecionam para análise o processo de prefixação. Discutem inicialmente a falta de clareza na definição do que seria um elemento de prefixação ou um elemento de composição. Na sequência, retomam Correia & Lemos (2005)³⁰ para estabelecer os três eixos semânticos nos quais se estabelece a prefixação: i) negação, oposição ou privação, ii) localização espaço-temporal e/ou iii) quantificação, intensificação ou avaliação. Kasama et al., observando o repertório terminológico de Nanociência/Nanotecnologia, propõem que o emprego do prefixo *nano-* ocorre geralmente no eixo semântico da quantificação; no entanto, destacam ter encontrado outras motivações para o uso deste prefixo, como é o caso do termo *nanocormia* que não possui escala nanométrica, como se espera, mas remete à pequenez anormal do tronco humano devido a uma anomalia de desenvolvimento. Este exemplo se assemelha ao que já foi comentado acerca do termo *nanopadronização*, no item 2.3.

²⁸ HUMBLEY, J. La néologie: avenir des langues. In: CABRÉ, M.T.; FREIXA, J.; SOLÉ, E. (orgs.) *Laneologia en el tombant de segle*. Barcelona: Institut Universitari de Lingüística Aplicada, 2000.

²⁹ *la néologie naît par définition dans le discours, et si on cherche à l'implanter dans la langue, il convient naturellement de bien étudier les conditions qui président à sa naissance. Un mot déjà « mis en discours » a peut-être plus de chance de s'implanter, dans des conditions analogues, q'une création « in vitro ».*

³⁰ CORREIA, M.; LEMOS, L. S. P. *Inovação lexical em português*. (Cadernos de Língua Portuguesa 4). Lisboa: Edições Colibri e Associação de Professores de Português, 2005.

Kasama *et al.* (2008) fazem um levantamento de dados no corpus³¹ do projeto NanoTerm, valendo-se do software Unitex³², em busca de itens lexicais iniciados pelo prefixo *nano-*, bem como o seu contexto no corpus. A busca pelo contexto justifica-se, pois é a partir do contexto (especificamente dos contextos que possuam informações definitórias ou explicativas sobre os termos) que os autores depreenderão o significado dos termos, de forma a validar ou não suas proposições acerca do prefixo em estudo.

A prefixação em *nano-* ocorre com maior frequência, segundo os autores, como forma de quantificação, remetendo à escala nanométrica. Outros empregos para o prefixo *nano-* também são identificados. Um desses é o uso do prefixo para indicar campos de aplicação da N&N como em *nanoeletrônica* e *nanoengenharia*. Exemplos de contextos obtidos no corpus são citados por Kasama *et al.* para ilustrar:

“A nanoeletrônica, um dos mais promissores campos de aplicação das nanotecnologias, enfrenta desafios particulares em relação ao seu desenvolvimento” e “Ciências dos materiais, bioengenharia e engenharia elétrica, estão todas, rapidamente, tornando-se componentes de uma superdisciplina: a nanoengenharia” (KASAMA *et al.*, 2008, p. 7).

No entanto, os sentidos para além da quantificação ocorrem, segundo Kasama *et al.* (2008), principalmente nos casos de truncação ou truncação seguida de composição e não de prefixação. Para os casos de truncação, os autores postulam que os termos *nanocoisas*, *nanoapetrecho*, *nanoassuntos* e *nanopreocupações* podem ser parafraseados por “X relativo/a(s) à Nanotecnologia”, nos quais X é a base da lexia prefixada. Dessa forma, por

³¹ O corpus em língua portuguesa de N&N foi elaborado no GETerm entre 2006 e 2007 por Joel Sossai Coleti e Daniela Ferreira de Mattos, ambos com bolsa PIBIC/CNPq. O corpus contém textos dos seguintes gêneros textuais: científico de divulgação, informativo, técnico-administrativo e outros. Compõem o corpus 1.057 textos de 57 fontes diferentes, totalizando 2.739.621 palavras.

³² Unitex é um sistema de processamento de corpus, baseado na tecnologia autômato-orientada. É um software criado no LADL (Laboratoire d'Automatique Documentaire et Linguistique), sob a direção de Maurice Gross. O Unitex permite análises nos níveis da morfologia, do léxico e da sintaxe. O programa pode ser obtido gratuitamente em: <http://www-igm.univmlv.fr/~unitex/>.

exemplo, *nanocoisas* denomina “coisas relativas à Nanotecnologia” de forma que o prefixo *nano-* representaria uma forma reduzida da palavra *nanotecnologia* que se junta a substantivos ou adjetivos. Esse fenômeno, segundo Alves (2007), recebe o nome de truncação. Já a truncação seguida de composição ocorre nos neologismos *nanopesquisadores* e *nanoengenheiros*, segundo Kasama *et al.* Esses termos denominam pesquisadores e engenheiros que atuam na área de Nanociência/Nanotecnologia e não pesquisadores e engenheiros de medidas nanométricas. De forma que o que se verifica é que “nesses casos há dois processos sequenciais envolvidos, primeiro a truncação, depois a junção dessa forma reduzida a uma base autônoma, constituindo, portanto, um processo de composição” (KASAMA *et al.*, 2008, p. 8).

São também objetos de investigação dos autores a alta produtividade do prefixo *nano-* e sua autonomia. Uma das justificativas levantadas pelos autores para a grande quantidade de ocorrências do prefixo *nano-* é “o índice de modernidade atribuído a uma lexia quando esta vem precedida desse prefixo, como se o pesquisador preferisse nanoengenharia a simplesmente engenharia.” (KASAMA *et al.*, 2008, p.7).

A observação do prefixo *nano-* ocorrendo autonomamente em contextos textuais obtidos no corpus, demonstram que esse prefixo é empregado nesta terminologia na posição de substantivos ou adjetivos. Exemplos apresentados pelos autores incluem: “a tecnologia nano”, “a escala nano”, “criar a nano no Brasil”, “empregos em nano nos próximos cinco anos” e “uma vez que a nano consiste”.

2.4.2 Almeida & Vale (2008): o uso de conhecimento morfológico na extração semiautomática de termos

A partir da análise das terminologias desenvolvidas pelo GETerm,³³ os autores observam padrões morfológicos específicos, os quais, devidamente descritos, podem funcionar como um recurso útil para a tarefa de extração semiautomática de termos em corpora. No artigo em pauta, os autores realizam um experimento utilizando o corpus de Nanociência/Nanotecnologia.

Buscando demonstrar a utilidade do conhecimento linguístico, particularmente o morfológico, na identificação de candidatos a termos, os autores recorrem à utilização do Unitex (cf. nota 29). Esse software permite a incorporação desse conhecimento para a realização de buscas extremamente complexas pela forma exata, ou pela forma canônica, ou por categorias gramaticais ou, ainda, por meio de formantes, realizadas por expressões regulares ou por meio de grafos. Os mesmos procedimentos de buscas foram replicados para todos os termos de maior frequência: *material* (7.198 ocorrências), *processo* (5.496 ocorrências), *sistema* (4.735 ocorrências), *amostra* (6.520 ocorrências), *estrutura* (3.926 ocorrências), *propriedade* (2.739 ocorrências) e *tecnologia* (2.847 ocorrências).

Para a explicitação dos procedimentos, os termos foram substituídos por X, sendo que nesta posição foram verificados todos os termos anteriormente citados, como se pode observar na Tabela 1. Os exemplos trazidos a seguir dizem respeito ao termo *material*.

A expressão regular <X><A> (ou <material><A>) traz todas as formas flexionadas de *material* (singular ou plural) na posição de base da lexia complexa, seguida de um adjetivo (A). A busca por *material* retornou 1.989 ocorrências, tais como *materiais amorfos* e *materiais moleculares*.

³³ Para maiores informações sobre as terminologias em estudo pelo GETerm pode-se consultar o site do grupo em www.geterm.ufscar.br

Um termo em todas as suas formas flexionadas seguido de uma palavra não existente no dicionário próprio do Unitex é o resultado que se obtém com a expressão regular <X><!DIC>. Essa expressão é promissora por incorporar palavras não registradas no próprio léxico do Unitex, sendo provavelmente neologismos ou pertencentes às linguagens de especialidade. Foram obtidas 722 ocorrências de expressões desse tipo, por exemplo, *material adsorvente* e *materiais biocerâmicos*.

A expressão <X>de<!DIC><A>, que busca itens léxicos seguidos de um sintagma preposicionado (que não consta do léxico do Unitex) mais adjetivo identificou alguns termos, mas também trouxe 2 ocorrências formadas por erros: *materiais de dimensões muito* e *materiais de dimensões normais* recuperadas apenas devido ao erro de grafia da palavra *dimensões*, que faz com que o dicionário do Unitex não a reconheça. A classificação de *muito* como adjetivo também corrobora para um resultado não satisfatório.

Todas as formas flexionadas de *material* seguidas da preposição *de* mais substantivo é o resultado esperado quando se utilizada na busca a seguinte expressão regular: <X>de<N>. Essa expressão, embora produtiva (353 ocorrências), retornou uma grande quantidade de sequência de palavras que não constituíam termos, tais como, *material de baixa* e *material de consumo*. Como exemplo de termos recuperados, podem-se citar: *material de eletrodo* e *material de enchimento*.

<X>de<N><A> é a sintaxe de busca que retorna todas as ocorrências de *material* seguidas de um sintagma preposicionado mais um adjetivo. A expressão recuperou 127 ocorrências, entre elas estão *material de banca direita* e *material de cor escura*. Assim como o caso anterior, muitos resultados não constituíam termos.

A expressão <X>de<N><!DIC> visou recuperar todas as ocorrências de *material* seguidas de um sintagma preposicionado mais qualquer item inexistente no dicionário do

Unitex. Essa expressão obteve três resultados, a saber, *materiais de compostos III-V*, *materiais de fase perovskita* e *materiais de símbolos Å*.

A Tabela 1 exemplifica os candidatos a termos obtidos a partir das sintaxes propostas no trabalho.

Itens lexicais pesquisados						
sintaxes de busca	processo	sistema	amostra	estrutura	Propriedade	tecnologia
X><A>	processo corrosivo	sistema capilar	amostra aglomerada	estrutura amorfa	Propriedade bactericida	tecnologia aeroespacial
X><!DIC>	processo catalítico	sistema agroalimentar	amostra pirolisada	estrutura dendrítica	Propriedade fotoativa	tecnologia MEMS
X>de<!DI <C><A>	processo de citrato amorfo	sistema de freagem regenerativo	amostras de preformas porosas	estruturas de nanotubos suspensas	propriedades de intervalo metálica	tecnologia de microcanais inteligentes
X>de<N>	processo de acoplamento	sistema de espectroscopia	amostra de acompanhamento	estruturas de multicamadas	propriedade de autoafinidade	tecnologias de recobrimento
X>de<N> <A>	processos de terapia celular	sistemas de feixes combinados	amostras de diamante nitrogenado	estrutura de rede casada	propriedades de chaveamento molecular	tecnologia de plasmas frios
X>de<N> <!DIC>	processo de realinhamento cristalográfico	sistemas de servomecanismos avançados	amostras de cobalto eletrodepositado	estruturas de carbono diamantífero	propriedades de anisotropia magneto-cristalina	tecnologia de tecido non-woven

Tabela 1: Exemplos de candidatos a termos obtidos a partir de padrões (ALMEIDA & VALE, 2008)

O texto de Almeida & Vale (2008) apresenta os procedimentos utilizados que se mostram úteis na busca de candidatos a termos, a partir de conhecimento morfológico. Importante destacar a conclusão dos autores: “Esse tipo de busca, como se observou, pode ser uma boa ferramenta para a listagem de candidatos a termos. Entretanto, o sucesso dessa busca pode ser determinado pela qualidade dos recursos linguísticos que servem de base para ela” (ALMEIDA & VALE, 2008, p. 497).

2.4.3 Coleti & Almeida (2010): descrição morfológica da terminologia de N&N

Neste trabalho, os autores propõem uma análise representativa, mas não exaustiva, da terminologia de Nanociência/Nanotecnologia. Foram descritos 295 termos e analisados 28, de um total de 3.069 termos, visando enriquecer a descrição morfológica do português a partir de um léxico especializado, de forma a contribuir com o refinamento de conhecimento linguístico útil para aplicação em sistemas híbridos de extração automática de terminologias.

Para o desenvolvimento do trabalho, Coleti & Almeida (2010) utilizaram a tipologia dos processos de formação de palavras proposta por Alves (2007).

Os processos mais produtivos observados na terminologia de N&N são detalhadamente discutidos pelos autores, no entanto, não serão retomados aqui por não estarem no foco desta pesquisa, são eles: os neologismos sintáticos, os neologismos formados por empréstimos e a conversão³⁴.

³⁴ Segundo Alves (2007) Neologismo sintático é a combinação de elementos já existentes no sistema linguístico, neste processo, estariam os neologismos formados por derivação e composição. O empréstimo é o neologismo formado a partir da utilização de bases de língua estrangeira. Já a conversão ou derivação imprópria é um tipo de formação lexical pelo qual uma unidade léxica sofre alterações em sua distribuição sem que haja manifestação de mudanças formais.

Analisando 10% dos termos mais frequentes de cada n-grama, de unigramas³⁵ até pentagramas³⁶, e excluindo-se os candidatos a termos equivocadamente validados como termos, foram descritos 295 termos. Com esta descrição, Coleti & Almeida (2010) observaram:

- i) O processo sintático de derivação prefixal ocorreu 53 vezes, já a derivação sufixal ocorreu 78 vezes. Foram observados os seguintes prefixos: *ab-* (*absorção*), *ad-* (*adsorção*), *bio-* (*biotecnologia*), *de-* (*decomposição*), *des-* (*dessorção*), *di-* (*dielétrico*), *infra-* (*infravermelho*), *micro-* (*microesfera*), *nano-* (*material nanoestruturado*), *pre-* (*preforma porosa*), *re-* (*moagem reativa*), *semi-* (*laser semiconductor*). Quanto aos sufixos, foram identificados: *-ado(a)* (*anodizado, amostra de espinélio dopada*), *-agem* (*armazenagem de hidrogênio*), *-al* (*barreira de potencial*), *-ar* (*controle escalar em malha*), *-ção* (*anodização*), *-dor* (*catalisador*), *-eto* (*iodeto*), *-ico* (*catódico*), *-(i)dade* (*condutividade*), *-io* (*germânio*), *-ito* (*crystalito*), *-ivo* (*dispositivo*), *-lito* (*eletrólito*), *-mento* (*acoplamento*), *-nte* (*retardante de chama*), *-on* (*fóton*), *-or* (*semiconductor*), *-oso* (*silício poroso*), *-ura* (*espessura do filme*).
- ii) O processo sintático de composição subordinativa foi encontrado em 11 termos, são eles: *eletroquímica, equiaxial, espectroscopia, espectroscopia Raman, fotodetector, impedância eletroquímica, litografia, morfologia, precipitador eletrostático, desvio padrão e sol-gel*; já a composição sintagmática registrou 129 ocorrências, como exemplos podem ser citados

³⁵ Termos formados por uma palavra.

³⁶ Termos formados por cinco palavras.

os seguintes termos: *amostra de espinélio dopada, área superficial específica, armazenagem de hidrogênio, atividade catalítica, banda de condução, barreira de potencial, cadeia polimérica, campo de Stokes e conformação por spray.*

- iii) O processo sintático de composição por sigla ou acronímica ocorreu em 27 itens, tais como *AFM, DRX e DSC.*
- iv) Foram observadas duas conversões: *dielétrico e precipitado.*
- v) Os estrangeirismos ocorreram 18 vezes, na maior parte das vezes em siglas (*conformação por spray e LASER escravo, por exemplo*).

Coleti & Almeida (2010) destacaram o fato de terem sido analisados 295 palavras da terminologia da Nanociência/Nanotecnologia e ter-se obtido um total de 318 repetições dos processos de formação de palavras identificados, já que em um termo podem ocorrer diferentes processos neológicos.

Das 28 análises apresentadas no artigo, destacam-se, a seguir, três, a título de ilustração:

- i) O sufixo *-dade* transforma uma base adjetiva em um substantivo, segundo Correia & Lemos (2005). Na terminologia analisada, ocorre apenas um caso de sufixação com esta partícula: *condutividade*, já que este substantivo é formado a partir do adjetivo “condutivo” acrescido do sufixo *-dade*.
- ii) Diferentemente das partículas *nano-* e *micro-*, o prefixo *semi-* não altera a dimensão do referente associado à base, mas sim reduz sua intensidade ou capacidade, assim, *semicondutor* é um condutor de capacidade reduzida.

- iii) Padrões morfossintáticos de grande produtividade observados: *N + A* e *N + prep. (+ det) + N*, fato que confirma diversos estudos terminológicos que apontaram a produtividade dessas formações em vocabulários especializados. São exemplos do sintagma *N + A*: *atividade catalítica*, *cadeia polimérica*, *campo claro* e *campo elétrico*. Já para o sintagma *N + prep (+det) + N* pode-se exemplificar com os seguintes termos: *armazenagem de hidrogênio*, *banda de condução*, *barreira de potencial*, *campo de bombeio*, *comprimento de onda*, *condição de anodização*, *conformação por spray*, *corpo de prova*, *diâmetro da nanoesfera*, *espessura do filme*, *resistência à corrosão* e outros.

Como o foco da nossa pesquisa é a análise morfológica de base gerativa, a seguir tratar-se-á da Morfologia sob essa perspectiva.

2.5 Morfologia: abordagem gerativa

Morfologia é entendida como o nível de descrição linguística que se ocupa do estudo da estrutura interna das palavras (=morfemas), a organização das palavras em classes e sua distribuição, bem como os processos envolvidos na sua derivação e formação. Considera-se, pois, que a unidade mínima de análise seja o morfema; e unidade máxima, a palavra, como bem define Cabral (1979):

...parte da gramática que descreve as unidades mínimas de significado, sua distribuição, variantes e classificação, conforme as estruturas onde ocorrem, a ordem que ocupam, os processos na formação de palavras e suas classes. (CABRAL, 1979, p. 134)

Segundo Rocha (2008), “É possível distinguir quatro grandes correntes ou escolas que procuraram descrever e analisar o componente morfológico das línguas: o descritivismo, o historicismo, o estruturalismo e o gerativismo” (ROCHA, 2008, p. 24). Por ter sido a base da Morfologia Construcional e do Modelo SILEX, apenas o gerativismo será abordado neste texto, ainda que brevemente, pois não constitui o foco principal deste trabalho.

A escolha da abordagem gerativa numa pesquisa envolvendo a morfologia introduz uma série de mudanças na forma de se analisar o objeto, como é de praxe em toda a mudança de perspectiva teórica:

“Na gramática tradicional, assim como no estruturalismo, a morfologia derivacional é definida como a parte da gramática de uma língua que descreve a formação e estrutura das palavras. Numa abordagem gerativa, podemos dizer que a morfologia derivacional é a parte da gramática que dá conta da competência do falante nativo no léxico de sua língua” (BASÍLIO, 1980, p. 7).

A Morfologia, na perspectiva gerativa, concentra-se na capacidade do usuário de uma língua em formar novas palavras. Na perspectiva gerativa, portanto, estuda-se a capacidade do falante nativo i) em criar novas palavras, por meio de regras de construção de palavras, reconhecendo sua estrutura, ii) tanto quanto, rejeitar determinadas formações ou palavras.

A morfologia gerativa, através do modelo Elemento e Processo, propõe uma reformulação dos princípios que sustentaram a noção estruturalista de morfema. Nesta nova perspectiva, as palavras são criadas por um conjunto de relações entre formas lexicais ou por processos de construções de novas palavras a partir das palavras existentes. Razão pela

qual Anderson (1992³⁷ *apud* ROCHA, 2008) vai postular que a Morfologia deve ocupar-se com a descrição do conjunto de regras que descrevam a modificação das formas existentes e não com um léxico de afixos.

As bases da Gramática Gerativo-Transformacional, origem da morfologia gerativa, foram lançadas por Noam Chomsky com o livro *Syntactic Structures* em 1957, posteriormente, diversos conceitos foram sendo discutidos, analisados, remodelados e/ou fixados, de forma que, a seguir, são expostos apenas os conceitos básicos de morfologia gerativa, visto que traçar uma ampla e completa perspectiva teórica desta abordagem não é o foco desse trabalho.

2.5.1 Conceitos básicos de morfologia gerativa, segundo Rocha (2008)

A Morfologia Gerativa não se ocupa de uma gramática prescritiva, normativa, histórica ou descritiva, mas sim de uma gramática subjacente, ou seja, uma gramática internalizada, implícita, que todo falante nativo de uma língua sabe como utilizar, embora nem sempre saiba explicitá-la. Por exemplo, ao encontrar em um texto o termo *anodização*, o falante nativo de língua portuguesa certamente reconhecerá que este termo tem origem no verbo *anodizar*, embora, muito provavelmente não saiba explicitar os processos gramaticais que levaram de uma palavra a outra, tampouco identifique seu significado. “Essa gramática [a subjacente], neste caso específico, permite ao falante analisar uma palavra e, a partir dessa análise, criar novas palavras, e, a partir de regras que formam novas palavras, analisar palavras preexistentes” (GRODT, 2009, p. 11).

Competência lexical é o conhecimento do léxico que um falante nativo possui sobre sua língua simplesmente por tê-la como língua materna. Este conhecimento compreende o

³⁷ ANDERSON, S. R. A. *Morphous Morphology*. Cambridge: Cambridge University Press, 1992.

domínio do repertório lexical e de formas linguísticas diversas (afixos, desinências e outras) além da capacidade de estabelecer relações entre eles. Não se trata, pois, de um dicionário da língua.

A caracterização da base³⁸ e do produto³⁹ (com a especificação de que tipo de base poderá receber determinado afixo) estabelece uma regra morfológica. Uma das tarefas pertinentes à morfologia gerativa é a explicitação dessas regras.

Grande quantidade de autores desta corrente teórica procura evidenciar a separação das regras morfológicas das regras sintáticas. Segundo Aronoff (1976), citado em Rocha (2008):

“[...] Regras de Formação de Palavras são regras do léxico e como tais operam totalmente dentro do léxico. São totalmente separadas das outras regras da gramática, embora não o sejam de outros componentes da gramática. Uma Regra de Formação de Palavras pode fazer referências a propriedades sintáticas, semânticas e fonológicas das palavras, mas não a regras sintáticas, semânticas ou fonológicas” (ARONOFF, 1976⁴⁰, p. 46 *apud* ROCHA, 2008, p. 38).

Outra distinção retomada por Rocha (2008) é a diferenciação entre Regras de Análise Estrutural (RAEs) e Regras de Formação de Palavras (RFPs). A primeira (RAEs) é empregada pelo falante na análise de palavras já existentes (institucionalizadas ou dicionarizadas) ou derivadas, enquanto a segunda (RFPs) é utilizada na produção de novos itens lexicais e pode ser acionada pelo falante a qualquer momento, com base em uma relação paradigmática (ex: *eletrocromismo*, *fotocromismo*, *sinergismo*, *magnetismo*, etc).

A formalização da análise da estrutura das palavras que o falante tem a capacidade de fazer se dá, nesta abordagem, da seguinte maneira:

³⁸ Base é a “seqüência fônica recorrente, a partir da qual se forma uma nova palavra, ou através da qual se constata que uma palavra é morfologicamente complexa” (ROCHA, 2008, p. 98).

³⁹ Entende-se como produto a palavra formada pelas regras morfológicas.

⁴⁰ ARONOFF, M. Word Formation in Generative Grammar. Cambridge: The MIT Press, 1976

[[X] a] Y] b

A RAE de *anodização*⁴¹, portanto, será:

[[anodizar] v -ção] s

Já a formalização de uma RFP pode ser exemplificada como:

[X] a → [[X] a Y] b

Assim, a RFP de *condensador*⁴² será:

[condensar] v → [[condensar] v -dor] s

É importante observar que uma grande quantidade de palavras não existe na língua, entretanto, poderia ser produzida por meio de RFPs. Essas palavras, inexistentes, mas previstas por RFPs, são chamadas de palavras possíveis. “(...) há muitas palavras que uma gramática pode gerar em uma língua que, acidentalmente ou assistematicamente, nunca aparecem” (ARONOFF, 1976, p. 18 *apud* ROCHA, 2008, p. 43).

A produtividade lexical é a possibilidade de surgimento de novos itens lexicais na linguagem coloquial, formal, técnica, científica, jornalística, literária ou qualquer outra modalidade de linguagem. A regularidade das formações as torna transparentes morfológica e semanticamente, no entanto, também há espaços para irregularidades, principalmente em formações já cristalizadas.

2.5.2 Alguns estudos de morfologia gerativa

Nesta seção, apresentam-se inicialmente alguns dos estudos de morfologia gerativa desenvolvidos no Brasil, de autoria de Margarida Basilio e Luiz Carlos Rocha. Na

⁴¹ Exemplo obtido da terminologia de Nanociência/Nanotecnologia, portanto, não apresentado por Rocha (2008).

⁴² Idem nota 38.

sequência, enfocam-se os estudos realizados em Portugal por Margarita Correia⁴³ e Graça Maria Rio-Torto.

2.5.2.1 Basílio (1999)

A morfologia é definida por Basílio (1999) como o estudo da estrutura interna das palavras de uma língua, bem como de suas variações de forma. A morfologia, de acordo com a autora, tem dificuldades de definir seu objeto de estudo, devido à múltipla pertinência da palavra como unidade lexical, gramatical, fonológica, entre outras.

O quadro contemporâneo apresentado por Basílio (1999) demonstra que, no Brasil, abordagens sociolinguísticas, funcionalistas e discursivas de fenômenos morfológicos, bem como abordagens morfológicas de temas flexionais, são minoritárias, predominando problemas de representação lexical, dentro de uma abordagem gerativa (BASÍLIO, 1999).

“O predomínio da abordagem gerativa decorre da suplantação do estruturalismo no desenvolvimento dos estudos lingüísticos no Brasil e da pouca relevância dada à estrutura vocabular em abordagens funcionalistas. Quanto à concentração no léxico, é possível que se deva em grande parte a interesses de lingüistas que lideraram a formação de novas gerações, embora deva também ser considerado o fato de que a flexão já havia sido abordada no período de influência dos estruturalistas” (BASÍLIO, 1999, p. 56).

Basílio postula que a maior parte dos trabalhos em morfologia no país se concentra na descrição do português, no entanto, também podem ser observados trabalhos em áreas menos centrais, tais como Aquisição da Linguagem, Fonologia Lexical, Descrição de

⁴³ Margarita Correia, coorientadora desta pesquisa de mestrado, desenvolveu sua tese de doutoramento intitulada *A denominação das qualidades: contributos para a compreensão da estrutura do léxico português* (2000), sob orientação de Danielle Corbin (Universidade de Lille III) e João Malaca Casteleiro (Universidade de Lisboa). A obra de Correia (2004), citada amiúde, constitui a versão editada e reformulada da referida tese.

Línguas Indígenas e Processamento de Linguagem Natural, entre outras, o que, segundo a autora, reforça e amplia a morfologia como campo de pesquisa no Brasil na última década.

A pesquisa em desenvolvimento neste mestrado põe em foco, assim como a maior parte dos trabalhos em morfologia, a descrição do português, especificamente a descrição dos processos de formação de termos, no entanto, o faz de forma integrada com o Processamento de Linguagem Natural, possibilitando sua aplicação em sistemas automatizados. Por isso, mantém-se em conformidade com os trabalhos desenvolvidos na área no cenário brasileiro, de acordo com Basílio (1999).

De acordo com a autora, “(...) a Morfologia lexical brasileira se preocupa significativamente com a investigação do fator semântico nos processos de formação de palavras, o que certamente é uma abordagem minoritária e marginalizada na Morfologia gerativa” (BASÍLIO, 1999, p. 57). Isso acaba por reforçar a pertinência do Modelo SILEX que, apesar de figurar no quadro teórico do gerativismo considera os diferentes níveis linguísticos, inclusive o semântico.

2.5.2.2 Rocha (2008)

Em sua obra, intitulada *Estruturas morfológicas do português*, o autor tem como objetivo explicitar a gramática subjacente (segundo a Morfologia Gerativa) no âmbito da morfologia lexical do português. Para isso, apresenta inicialmente um breve percurso histórico dos estudos linguísticos referentes à morfologia, abordando o Descritivismo, o Historicismo, o Estruturalismo e, por fim, o Gerativismo. Na sequência, o autor detalha os conceitos fundamentais para compreensão da Morfologia segundo uma abordagem gerativa.

No capítulo seguinte, o autor aborda o estudo da Morfologia nas gramáticas brasileiras, discutindo a constante confusão entre os planos sincrônico e diacrônico, além da impressão de que o estudo da sufixação está restrito à apresentação de uma lista de sufixos.

Já no terceiro capítulo, é feita uma ampla discussão sobre o léxico, tratando de morfemas, lexemas, idiossincrasias, polissemia, homofonia, empréstimos e hibridismos. Também são tratadas questões fonológicas, sintáticas, semânticas e discursivas referentes à lista de entradas lexicais de um falante, ou seja, o seu léxico mental.

A partir dessas proposições, o autor inicia, no capítulo quarto, a discussão sobre as razões pelas quais se formam novas palavras na língua.

A partir do capítulo quinto, Rocha (2008) começa a expor os processos de formação de palavras, iniciando pela derivação sufixal que, segundo ele, é o processo “mais rico e diversificado da língua portuguesa” (p. 95). No capítulo sexto, o autor analisa exaustivamente uma Regra de Formação de Palavra (RFP) para construção de nomes agentivos em *-eiro*. Sua intenção, nesse capítulo, é oferecer um modelo de análise que possa servir de guia para trabalhos acadêmicos em morfologia gerativa, confirmando a proposta didática do livro.

No capítulo sétimo, o autor trata da derivação prefixal, sempre apresentando uma profusão de exemplos que enriquecem sobremaneira sua exposição. Na sequência, apresenta outros processos de formação de palavras, assim divididos: derivação, composição e onomatopeia. Na derivação, o autor inclui: derivação parassintética, conversiva, siglada, truncada e regressiva. Na composição, inclui: composição erudita e reavaliação morfológica.

Nos capítulos 9, 10 e 11, o autor trata, respectivamente, da flexão e derivação, do gênero do substantivo e do grau e aspecto em português, perfazendo, pois, um completo manual de Morfologia.

A seguir, apresentam-se algumas ideias desenvolvidas por Rocha (2008), as quais estão diretamente ligadas a esta pesquisa.

O autor assinala, por exemplo, que a criação de uma nova palavra pode ser uma formação esporádica ou institucionalizada, ambas atendendo às RFPs da língua. Formação esporádica é aquela criada por uma necessidade comunicativa imediata, de acordo com as RFPs da língua em que é criada. A formação institucionalizada é aquela familiar a um grupo de indivíduos, mesmo o âmbito dessa familiaridade podendo variar significativamente.

A formação, inicialmente esporádica, pode institucionalizar-se em razão de alguns fatores, tais como: 1) o prestígio do criador da palavra; 2) o poder da mídia; 3) o fato de que muitos processos são, em dado momento, mais chamativos que outros; 4) ou simplesmente por necessidade histórica, cultural ou pragmática. No caso desta pesquisa, todos os termos analisados são formações institucionalizadas, amplamente conhecidas em suas áreas de aplicação e institucionalizadas para atender à necessidade de denominação das descobertas e avanços científicos.

O autor também aborda o fenômeno da lexicalização, que para ele é algum tipo de irregularidade ou desvio de uma regra em relação aos aspectos fonológico, morfológico ou semântico:

“(...) a lexicalização se caracteriza pelo fato de um lexema apresentar uma determinada estrutura diferente daquela prevista pela aplicação de sua respectiva RAE [Regra de Análise Estrutural]⁴⁴” (ROCHA, 2008, p. 84).

O autor ainda classifica a lexicalização em cinco tipos:

- categorial: aquela em que se verifica irregularidade na categoria da palavra base no processo de formação de uma nova palavra;
- prosódica: aquela em que se verifica irregularidade na pronúncia (tonicidade) da palavra formada devido à herança de padrões lexicais do latim, de fases tardias do latim e de outras línguas;
- estrutural: aquela em que se verifica irregularidade na estrutura do novo vocábulo, com relação à sua respectiva RAE;
- rizomórfica: aquela em que se verifica irregularidade na raiz da nova palavra;
- semântica: aquela em que se verifica irregularidade no significado do novo vocábulo.

Para o autor, “o processo de derivação caracteriza-se pela formação de um produto (P) a partir de uma base (B): $B \rightarrow P$ ” (ROCHA, 2008, p. 98). Nesta abordagem, o falante nativo de uma língua é capaz de formar novas palavras a partir de palavras já existentes na língua através de seu conhecimento lexical e não somente pela soma de radicais e afixos, como preconizava a linguística estrutural.

⁴⁴ Conforme apresentado anteriormente, segundo o autor, a RAE diz respeito à análise da estrutura de uma palavra que um falante é capaz de fazer, permitindo, portanto, que ele reconheça uma nova palavra formada a partir da identificação do significado de suas partes.

Um dos aspectos absolutamente relevantes para esta pesquisa é a análise que Rocha (2008) propõe para a regra $S \rightarrow S_{-EIRO}$, que possibilita a construção de nomes agentivos em *-eiro*.

Nessa análise, empregou-se a regra em que o produto designa um nome agentivo, tais como: *leiteiro, sanfoneiro, confeitoiro, banqueiro, doleiro, grafiteiro, roqueiro, etc.* Não são considerados, portanto, os sufixos homófonos de *-eiro*, tais como i) árvore ou arbusto (*limoeiro, tomateiro, etc.*), ii) lugar ou recipiente (*banheiro, saleiro, moedeiro, etc.*), iii) coletivo ou conjunto (*braseiro, berreiro, etc.*), iv) gentílico (*brasileiro, mineiro, etc.*), v) formador de adjetivos (*caseiro, grosseiro, etc.*) e vi) objeto (*chuveiro, chaveiro, etc.*).

Rocha (2008) toma aleatoriamente 150 substantivos para observar se em todos eles poderia ser aplicada a regra em análise. Ao final deste experimento, o autor constata que nem todos os substantivos permitem a aplicação dessa regra, portanto, definir a categoria da base apenas como “substantivo” se mostra insuficiente, de tal forma que se torna necessário identificar os tipos de substantivo que aceitam a aplicação desta regra por meio de subcategorias que podem ser fonética, morfológica, sintática e/ou outras, tais como: próprio, comum, concreto, abstrato ou outros.

Nesse caso, a RFP não se aplica a substantivos abstratos, a substantivos que nomeiam agente ou indivíduos e a substantivos compostos. O autor então sintetiza a regra por meio da Tabela 2.

Caracterização da base	Caracterização do produto
+ substantivo - abstrato - agente - palavra composta	+ substantivo + agente

Ainda assim, observa o autor, restam bases que satisfazem as exigências de caracterização da base, mas não há produtos correspondentes. Essa inexistência dos produtos correspondentes se deve a três fatores, segundo Rocha (2008): restrições, bloqueio e inércia morfológica. As restrições podem ser:

- a) fonológicas: sequência desagradável de fonemas, tais como *bandeireiro*, *dinheireiro*, etc;
- b) paradigmáticas: quando se opta por outra RFP para designar a mesma coisa, por exemplo, *sambeiro* não ocorre devido à existência de *sambista*;
- c) pragmáticas: culturalmente dispensáveis. Rocha (2008) exemplifica afirmando que conhecemos o profissional responsável pelos serviços de copa, o *copeiro*, no entanto, desconhecemos profissionais responsáveis pela sala ou corredor, de modo que não conhecemos, portanto, *saleiro* e *corredorzeiro*.
- d) discursivas: não são adequadas a determinados discursos, por apresentarem, por exemplo, sentido pejorativo. Por exemplo: *vetorzinho* e *dinamicazinha* certamente não serão encontradas em um contexto de produção científica em física, já que a sufixação em *-inho* geralmente não faz parte de discursos técnicos.

Já o bloqueio acontece quando uma formação possível não se concretiza devido à existência de outra forma lexical que carregue o mesmo significado. Acontece bloqueio,

por exemplo, quando constatamos a inexistência da palavra *denteiro* única e exclusivamente devido à presença da palavra *dentista*⁴⁵.

A inercia morfológica refere-se ao fato de não haver nenhum motivo contrário à existência de uma dada palavra e, mesmo assim, esta formação não é criada pelos falantes de uma língua, mesmo com a existência e adequabilidade à RFP.

2.5.2.3 Rio-Torto (1998)

Em Rio-Torto (1998), versão revista e atualizada de Rio-Torto (1994)⁴⁶, apresenta-se um quadro geral das regras de formação de palavras no português, “evitando a pulverização dos produtos e das RCPs e concentrando o seu número, de modo a obter um modelo de componente lexical mais coeso” (CORREIA, 2004, p. 121).

Rio-Torto considera que o tratamento computadorizado do léxico carece de um amplo conhecimento do sistema derivacional do português, ainda a ser descrito. Rio-Torto propõe-se então “dar a conhecer alguns dos paradigmas derivacionais do português europeu contemporâneo e equacionar alguns aspectos da problemática que a identificação destes suscita (...)” (RIO-TORTO, 1998, p. 109).

O quadro teórico e metodológico do Modelo SILEX, é brevemente delineado com algumas breves considerações, entre elas:

⁴⁵ O exemplo trazido do texto de Rocha (2008) refere-se ao bloqueio paradigmático. O autor aborda ainda o bloqueio heterônimo (ex: *ensinador* é bloqueado por *professor*, trata-se da mesma denominação a partir de bases diferentes), bloqueio homofônico (quando não ocorre o surgimento de uma nova palavra por essa apresentar o mesmo aspecto fonético de outra palavra já existente, por exemplo, a partir de *terra* não se cria *terreiro* (a pessoa que trabalha na terra ou com terra) devido à existência de *terreiro* (“quintal”) e o bloqueio parônimo (a existência de parônimos bloqueia o surgimento de certas palavras, como é o caso de *cabeleiro* – com origem em *cabelo* – bloqueado por *cabeleiro* – que tem como base *cabeleira*).

⁴⁶ RIO-TORTO, G. M. Regras de formação de palavras em português: achegas para um quadro geral. In: *Diacrítica*, n° 9, pp. 319 – 342, 1994.

- “São essencialmente de natureza aditiva as operações que sustentam os principais tipos de processos genolexicais que o português convoca: a derivação e a composição.” (RIO-TORTO, 1998, p. 109-110).
- “A composição define-se pela concatenação de pelo menos duas bases, cada uma das quais pode ser [+/- autônoma]; a derivação consiste na adjunção de uma base a de um operador afixal” (RIO-TORTO, 1998, p. 110).
- “No âmbito de cada paradigma podem actuar diversas operações morfológicas (...). Ao serviço de cada paradigma podem estar diferentes operadores afixais” (RIO-TORTO, 1998, p. 110).
- “(...) a operação semântica que preside a cada RFP não substitui nem esgota a complexidade semântica do produto construído. Daí também a necessidade de delimitar os diversos níveis de significação presentes em cada produto derivacional” (RIO-TORTO, 1998, p. 111).
- “Por certo que a estrutura interna de um derivado não representa apenas o resultado da simples adição entre base e afixo. Todavia, a estrutura compósita dum produto lexical não pode deixar de se ancorar na co-articulação das dimensões morfológica e semântica das partes nele envolvidas” (RIO-TORTO, 1998, p. 111).

Após essas considerações, Rio-Torto (1998) discorre sobre as regularidades derivacionais observáveis entre produtos e bases para a análise das operações semântico-categoriais do sistema de formação de palavras do português. Os produtos derivacionais alvos de sistematização da autora são: produtos deverbais, produtos verbais, nomes não

deverbais, adjetivos não deverbais, nomes isocategoriais, adjetivos isocategoriais e verbos isocategoriais. Dessas proposições a autora conclui: “impõe-se, portanto, o estabelecimento de diferentes graus de sistematicidade e/ou de regularidade e o reconhecimento dos diferentes tipos de padrões composicionais em jogo” (RIO-TORTO, 1998, p. 115) acrescentando que “(...) só o conhecimento do sistema de formação de palavras, na sua globalidade, permite apurar quais as relações derivacionais sistêmicas e quais as que não representam mais do que regularidades superficiais, não inscritas na arquitectura da língua” (RIO-TORTO, 1998, p. 115).

Rio-Torto (1998), a respeito das sistematizações, destaca que a complexidade linguística das línguas vernáculas não podem ser reduzidas a esquemas rígidos de sistematizações, já que a própria assistematicidade é uma característica dessas línguas. Dessa forma, a autora propõe um quadro apenas provisório e não exaustivo de paradigmas de formação de palavras em português e das regras de formação de palavras em português.

“Mais do que o número de regras identificadas e de variáveis que cada uma consigna, importa salientar as interferências que frequentemente se verificam entre produtos e paradigmas derivacionais. A análise efectuada aponta para a necessidade de se considerar que o quadro de RFP assenta numa estrutura de parecença de família, em que as regras mantêm relações não apenas bilaterais, mas também plurilaterais” (RIO-TORTO, 1998, p. 131)

2.5.2.4 Correia (2004)

Na obra *Denominação e construção de palavras*, Correia (2004) tem como objetivo descrever, por meio de uma abordagem lexicológica, o modo como se constrói a denominação das qualidades em português.

“(...) o estudo das palavras construídas é um espaço privilegiado para a observação da construção do significado, em particular do significado descritivo ou referencial das unidades lexicais. Tal acontece porque, no quadro do modelo de análise adoptado, se concebe que as palavras construídas são portadoras de um

significado previsível, linguístico, que lhes é conferido pela sua estrutura; desse modo, torna-se mais evidente a relação que existe entre a unidade lexical e sua referencia” (CORREIA, 2004, p. 21).

O modelo de análise referido pela autora no excerto anterior e adotado por ela em sua pesquisa é o SILEX, no âmbito da Morfologia Construcional.

Com relação ainda ao fenômeno da denominação, a autora postula também que:

“a denominação caracteriza-se também pelo facto de ser uma relação em grande parte não explicável linguisticamente. Dito de outro modo, entende-se que a língua, o sistema linguístico, pode fornecer diversas denominações possíveis para uma mesma classe de entidades (...). O facto de numa língua (ou num determinado registo) se escolher uma denominação possível e não outra é um fenómeno não-linguístico” (CORREIA, 2004, p. 24).

Inicialmente, a autora tece breves considerações sobre as diferentes fases do Modelo SILEX⁴⁷ e sobre os dados utilizados para análise.

A autora aproxima a compreensão de unidades construídas transparentes da compreensão das estruturas sintáticas, de forma que, para compreendê-las, bastaria interpretar sua estrutura e seus constituintes. “Deste modo, entende-se que, da competência lexical, faz parte a competência derivacional (...)” (CORREIA, 2004, p. 42).

Correia (2004) apresenta uma visão global dos tipos de estruturas morfológicas passíveis de constituírem denominações de qualidades em português para, em seguida, iniciar sua análise (nos capítulos 3, 4 e 5 de sua livro), desconsiderando as palavras não-construídas, sejam simples ou complexas.

Palavras complexas não-construídas são “aquelas que apresentam estrutura compósita, mas em relação às quais apenas parcialmente existe concordância entre a estrutura e o significado” (CORREIA, 2004, p. 100). Por isso, pode-se dizer que são

⁴⁷ O Modelo SILEX será apresentado na seção 2.5.3, inclusive abordando as considerações desta autora, motivo pelo qual o modelo não será agora discutido.

palavras não- construídas em português aquelas que são resultado de importações⁴⁸, podendo apresentar estruturas simples ou complexas nas suas línguas de origem.

As palavras oriundas de importações podem não corresponder fonológica, morfológica e sintaticamente ao sistema linguístico de acolhimento. Exemplos dessas importações podem ser, entre outros: *ketchup* e *shopping center*. A maior ou menor correspondência de cada palavra faz com que Correia (2004) proponha a criação dos subconjuntos seguintes:

- a) subconjunto das importações alógenas de fato: reúne unidades completamente opacas em português por serem provenientes, em geral, de línguas não românicas, portadoras, portanto, de estrutura fonológica, morfológica, sintática e semântica diferentes do português. Este conjunto será tratado pelo Modelo SILEX como palavras de estruturas simples, unidades que não se valem de nenhuma RCP sendo, portanto, impossível inferir quaisquer informações de significação a partir de sua forma;
- b) subconjunto das importações parcialmente alógenas: agrupa as palavras apenas parcialmente opacas em língua portuguesa por serem parcialmente semelhantes ao sistema linguístico do português. Este conjunto será tratado pelo Modelo SILEX como o conjunto das “palavras complexas não-construídas”. A estrutura das palavras deste subconjunto faz com que não possam ser tratadas como palavras construídas, uma vez que não preenchem os requisitos definidos para a delimitação das bases e

⁴⁸ “Importar palavras é sempre trazer de fora (de uma língua A) para dentro (para uma língua B)” (A.M.R. Andrade, 1995, p. 89-90 *apud* Correia, 2004, p. 101).

exibem terminação sufixóide não correspondente com nenhum sufixo da língua portuguesa, além disso seu significado não corresponde ao esperado.

Correia (2004) apresenta ainda o conceito de *internacionalismos*, palavras que apresentam “estrutura semelhante em diferentes línguas e que por isso são mais facilmente compreensíveis e traduzíveis” (CORREIA, 2004, p. 107).

Os internacionalismos assumem esse estatuto quando tanto o sufixo quanto a base têm formas cognatas em diferentes línguas, sendo totalmente transparentes nessas línguas. Os internacionalismos carregariam unidades de formas semelhantes em diferentes línguas românicas e em inglês, devido a sua origem latina ou grega, e comportamento semelhante em todas elas. Segundo Correia (2004), “(...) existe, em cada uma das línguas mencionadas (as românicas e o inglês), um conjunto de mecanismos de formação de palavras que é muito semelhante em todas elas. No entanto, não é forçoso que estes processos sejam exatamente iguais em todas as línguas.⁴⁹” (CORREIA, 2004, p. 108).

As palavras com estrutura analisável por possuírem base e forma sufixal reconhecível em língua portuguesa, embora não tenham sido construídas nesta língua, carregariam, segundo Corbin (em comunicação pessoal, segundo CORREIA, 2004, p. 106) sufixos fósseis, ou seja, “não são sufixos produtivos no estado da língua que é alvo de descrição, embora, do ponto de vista da analisabilidade, eles se comportem como os sufixos efetivamente produtivos nessa sincronia” (CORREIA, 2004, p. 107). São citados como exemplos: *-ícia, -ície, -or e -tude*.

Após essas considerações, Correia (2004) ressalta que:

⁴⁹ Esta última frase consta em nota de rodapé na fonte consultada.

“(…) para além de descrever a competência derivacional dos falantes, se está a descrever uma faixa do léxico actual da língua e que este léxico actual apresenta palavras que entraram nele em épocas diferentes e de modos e proveniências diferentes, então, a postura a adoptar não pode ser estritamente sincronista e os dados da história da língua, e da história das palavras em particular, terão que ser tidos em consideração. Esta é a postura assumida neste trabalho, em coerência, de resto, com a postural actual do modelo SILEX” (CORREIA, 2004, p. 109).

Correia (2004) aprofunda a discussão em torno dos elementos periféricos da categoria dos nomes de qualidade construídos, tratando dos substantivos denominais⁵⁰ e dos substantivos deverbais⁵¹, nesses casos o significado predicativo, referente aos nomes de qualidade, são derivados, ou seja, não são o significado principal. Correia (2004) assinala ainda a presença de substantivos deadjectivais resultantes de conversão⁵² como palavras construídas possíveis para denominar qualidades em português.

Como elemento central, no entanto, Correia (2004) afirma que os nomes de qualidade em português contemporâneo são prototipicamente “substantivos deadjectivais derivados por sufixação com recurso a sufixos como *-ia*, *-idade*, *-ismo* [considerados sufixos internacionais], *-ez* e *-eza*, *-idão*, *-ura*, *-ice* e *-eira* [considerados sufixos autóctones]” (CORREIA, 2004, p. 169).

⁵⁰ A autora verificou que esses substantivos podem ser construídos “(…) pela RCP que permite a construção de nomes de estatuto / condição / atividade, ou pela regra que permite a construção de nomes colectivos. No caso dos nomes de estatuto / condição / atividade, os derivados afectados são os construídos por meio dos sufixos *-ado/-ato*, *-ia* e *-tura*, todos eles operando sobre substantivos que denominam entidades humanas por meio do seu estatuto social ou profissional. No caso dos nomes colectivos, os derivados considerados são os construídos por meio dos sufixos *-ada*, *-agem* e *-aria* sobre as bases que denominam tipos humanos e aos quais associamos estereótipos” (Correia, 2004, p. 217).

⁵¹ Correia (2004) afirma que “(…) esses substantivos assumem o papel de nomes de qualidade sempre que a qualidade denominada é vista como resultando de um processo e são, portanto, denominadas, enquanto adjectivos, por adjectivos deverbais. Como a generalidade dos adjectivos deverbais de significado “passivo” (...) resistem a construir-se em bases para a construção de nomes deadjectivais, o nome deverbal equivalente acaba por ocupar o lugar do nome de qualidade deadjectival que é impossível construir” (Correia, 2004, p. 217).

⁵² “Concluiu-se que esse processo de formação se encontra disponível em português, mas apenas em casos muito particulares, a saber, quando a construção de um nome de qualidade por sufixação se revela impossível (ou, pelo menos, pouco viável), devido a condicionantes que podem ou não ser dependentes da gramática da língua” (Correia, 2004, p. 241).

A autora se propõe a abordar com mais especificidade e detalhamento os sufixos -*ia*, -*idade*, -*ismo* e -*ice* por serem aqueles que apresentaram maior produtividade no corpus analisado e por serem os que apresentam maior disponibilidade para a construção de novas palavras.

Os sufixos -*ia* e -*idade*⁵³ selecionam de suas bases o mesmo tipo de informação, segundo Correia (2004), o que faz com que cubram o mesmo tipo de informação do ponto de vista semântico. Esses sufixos selecionam das bases (adjetivos) a que são acrescidas suas propriedades mais objetivas e é isto que explica sua presença na construção da grande maioria dos nomes de qualidade que ocorrem nos domínios de especialidade, segundo a autora. “Existem, portanto, palavras que, pela sua estrutura, se encontram „predispostas“ a tornarem-se termos (...), ao passo que outras resistem a assumir o estatuto de termos (...)” (CORREIA, 2004, p. 246).

Considerado o sufixo internacional por excelência, -*ismo*⁵⁴ é, no entanto, um sufixo atípico, já que opera sobre bases adjetivais fundamentalmente, mas também em bases nominais ou verbais. Segundo a autora, “os produtos da sufixação por meio de -*ismo* são basicamente nomes de qualidade, parafraseáveis por „o fato de ser X_{Adj}“, “(...) o significado descritivo construído da maioria destes nomes de qualidade permite-lhes assumir um significado referencial tal que podem denominar doutrinas, ideologias, comportamentos sistematicamente assumidos” (CORREIA, 2004, p. 278).

Também os sufixos autóctones são tratados pela autora⁵⁵. O sufixo -*ice* possui vários pontos em comum com o sufixo internacional -*ismo*, no nível do funcionamento. Pontos em comum também são encontrados entre os sufixos -*ice* e -*eira*, na medida em que

⁵³ A análise e a descrição completa desses sufixos encontram-se em Correia (2004), páginas 246 a 277.

⁵⁴ A análise e a descrição completa desse sufixo encontram-se em Correia (2004), páginas 278 a 298.

⁵⁵ A análise e a descrição completa desses sufixos encontram-se em Correia (2004), páginas 299 a 331.

tratam das qualidades de forma menos objetiva. Os demais sufixos (-ez e -eza, -idão, -ura) permitem formar palavras a partir de adjetivos de estrutura simples, tratam das qualidades de forma mais objetiva e, ainda, são sufixos pouco disponíveis em português para a construção de novos nomes de qualidade, segundo Correia (2004).

Por fim, a autora afirma:

“um dos aspectos que considero mais atraentes na mais recente produção teórica de D. Corbin é o facto de esta autora ter conseguido ultrapassar as limitações iniciais do seu modelo, estando agora muito mais próxima de uma teoria geral do léxico do que propriamente de um modelo de morfologia derivacional (1987) ou construcional (sobretudo a partir de 1991)” (CORREIA, 2004, p. 335).

2.5.2.5 Correia (2006)

Neste trabalho, também com base no modelo de morfologia construcional SILEX, a autora busca verificar quais os principais mecanismos derivacionais envolvidos na construção de termos de Marinha, bem como fazer o levantamento dos principais afixos envolvidos nessa construção.

Correia (2004a) obteve as unidades derivadas do *Dicionário de Linguagem de Marinha Antiga e Actual* (3ª edição, de 1990). A autora justifica a fonte das terminologias (um dicionário) em detrimento de um corpus “na medida em que aquilo que se encontra em observação é a estrutura interna dos termos e não tanto o seu uso” (CORREIA, 2006, p. 33).

A autora inicialmente exclui do repertório terminológico em análise os termos antigos, gírias e aqueles que apresentam no dicionário o mesmo significado que é empregado na língua geral. Na sequência, os dados foram organizados de acordo com a sua

categoria, de modo a obter uma imagem mais clara dos mecanismos derivacionais mais produtivos na terminologia da náutica.

Na primeira análise dos dados recolhidos (item 2 do artigo de CORREIA, 2006), a autora detalha as motivações que a levaram a excluir um grande número de termos de estrutura simples da língua corrente, itens constantes do dicionário com sentidos metafóricos ou metonímicos, e também termos compostos resultantes da lexicalização de sintagmas, além das siglas. A autora aponta, ainda, que a prefixação não se revelou um processo muito produtivo no repertório terminológico analisado.

No item seguinte, a autora apresenta os processos mais produtivos, segundo a primeira análise dos dados, e conclui já no item subsequente que:

Comparando estes dados com os de outras terminologias, verifica-se que os operadores morfológicos aqui presentes diferem significativamente dos que se encontram na maioria dos vocabulários especializados. Os afixos envolvidos na construção de termos da terminologia náutica não são muito correntes em terminologias científicas e técnicas, dado, aparentemente, não darem a ver os conceitos com a objectividade que seria esperável de palavras que expressam conceitos de natureza científica e/ou técnica (CORREIA, 2006, p. 36).

Já no item 3, Correia (2006) explica as razões para a ocorrência dos fenômenos em análise, partindo dos contextos histórico e social da construção das terminologias, mas sem se limitar a eles.

A autora retoma trabalho anterior, a saber, Correia (2004), no qual analisou a construção dos nomes de qualidade em português, para ilustrar que as escolhas feitas dentro de uma RCP podem possuir motivação linguística ou extralinguística. Como fatos linguísticos são citados, por exemplo, a origem e a estrutura interna e fonológica da base e a incompatibilidade entre afixos. Como fatos extralinguísticos são citados fatores de natureza

semântico-referencial, como o fato de os sufixos selecionarem das bases traços distintos, tais como: propriedades comportamentais humanas, denominação de sistemas de ideias, de comportamentos ou de atuações e, ainda, propriedades relacionadas com o saber científico sobre as entidades denominadas (CORREIA, 2006).

“Finalmente, também factores comunicativos podem estar implicados na selecção de um derivado construído com um afixo e não com outro. Por exemplo, quando em vez de anglicidade ou anglicismo utilizo inglesice ou inglesada, embora todas estas palavras tenham como significado estrutural “o facto de ser anglo- / inglês”, é certo que quero dar a ver essa qualidade de forma diferente, focalizando formas diferentes de ver a qualidade de quem ou do que é inglês – ex.: inglesice e inglesada focalizam as propriedades estereotípicas que habitualmente associamos aos ingleses” (CORREIA, 2006, p. 41)

Conclui-se, portanto, que a seleção de uma determinada estrutura morfológica, em detrimento de outra, igualmente possível de se escolhida de acordo com a RCP em questão, é motivada, consciente ou inconsciente, pelo anseio de se mostrar o referente de uma determinada maneira e de focalizar certos traços característicos (CORREIA, 2006).

A partir dessa conclusão, a autora prossegue seu estudo, buscando responder agora as seguintes questões:

“i) Quando da construção dos termos em análise neste trabalho, houve motivação para a escolha dos operadores recenseados e não de outros? ii) Por que razão são escolhidos determinados processos ou operadores morfológicos e não outros? iii) Poderá a escolha destes mecanismos e operadores morfológicos ter um objetivo denominativo e/ou comunicativo específico(s)? iv) A escolha destes mecanismos e operadores será mais apta a denominar determinado tipo de conceitos?” (CORREIA, 2006, p. 41).

Dessa forma, Correia (2006) inicia uma reanálise e explicação de três casos significativos e representativos (mencionando apenas as estruturas mais produtivas): nomes em *-eir-*, verbos denominais e nomes construídos com sufixos avaliativos.

Para tratar dos nomes em *-eir-*, a autora apresenta, inicialmente, os quatro mecanismos disponíveis em português para denominar profissionais e especialistas de diferentes domínios. Podem-se destacar: o recurso a nomes derivados de verbais em *-or* ou em *-eir-*, construídos sobre o tema do particípio passado dos verbos, e o recurso a nomes derivados denominais em *-eir-*, em *-ista* ou em *-ári-*. A escolha por um mecanismo ou outro, como já relatado, não é, obviamente, ingênua, mas procura evidenciar as profissões de formas distintas. Assim, “o sufixo *-eir-* opera sobre bases nominais, prioritariamente nomes concretos, seleccionando das suas bases as propriedades pragmáticas, isto é, aquelas que têm a ver directamente com a relação que o ser humano tem com o objecto em causa” (CORREIA, 2006, p. 43). É por isso que “os nomes em *-eir-* que ocorrem no corpus denominam, como vimos, por um lado, marinheiros com uma função específica na manobra do barco e, por outro, barcos especialmente destinados à concretização de determinado objetivo” (CORREIA, 2006, p. 43).

Quanto ao segundo caso, o dos verbos denominais, a autora afirma que este reforça a perspectiva pragmática e económica da náutica, representada por esta terminologia, “os marinheiros envolvidos na empresa dos Descobrimentos e nas viagens marítimas eram fundamentalmente gente do povo, pouco instruída, não lhes sendo acessível, portanto, o domínio de uma terminologia muito elaborada” (CORREIA, 2006, p. 38). Os verbos denominais são, então, utilizados para nomear as ações praticadas durante a preparação da viagem e manobra do navio. “Nestes verbos, as bases denominam, na maioria dos casos, partes ou instrumentos do barco, focalizando, portanto, a relação concreta que se estabelece com essa entidade, objecto ou instrumento da acção praticada – exs.: *arpoar* < *arpão* («fisgar com o arpão»); *xaretar* < *xareta* («defender com a xareta ou xaretas») (...)” (CORREIA, 2006, p. 44). Já que é grande a quantidade de partes e de instrumentos

diferentes de um navio, todos com nomes suficientemente bem diferenciados, a forma mais econômica de denominar as diferentes ações é, portanto, fazê-lo através da focalização direta do objeto da ação ou do seu instrumento, evitando ambiguidades na comunicação (CORREIA, 2006). A autora afirma que pretende-se, assim, utilizar a linguagem de forma figurada buscando a sua economia, visto que na maior parte dos casos as ordens dentro de um navio são gritadas, sendo imprescindíveis enunciados curtos.

O terceiro e último caso, os substantivos dotados de sufixos avaliativos, serve à denominação de conceitos tidos como semelhantes a outros, justificando a seleção da base, “podendo essa semelhança advir do facto de os conceitos terem entre si uma relação de hiperonímia / hiponímia, ou apenas uma relação de semelhança que resulta de um processo metafórico inerente” (CORREIA, 2006, p. 46). “Neste sentido, pode afirmar-se que a existência destes avaliativos serve a figuratividade da linguagem e, ao mesmo tempo, a sua economia deste tipo de linguagem” (CORREIA, 2006, p. 46). Os nomes com avaliativos são, ainda, usados para expressar níveis de semelhança entre a entidade que corresponde ao conceito normal e a que corresponde ao conceito denominado pelo avaliativo, mas apenas conceitualizados de forma intuitiva, ou seja, aproximada (CORREIA, 2006).

Os nomes avaliativos detectados no corpus e apresentados pela autora denominam, principalmente, partes do barco ou instrumentos e equipamentos diversos utilizados a bordo. São utilizados como i) hipônimo de uma base, portanto, parafraseável como “tipo de X”. “Este é [por exemplo] o caso de *velacho* (< *vela*), que denomina uma vela específica, a saber, a vela da gávea do mastro do traquete” (CORREIA, 2006, p. 45). E ii) derivados não hipônimos, mas resultantes de um processo metafórico inerente a sua construção. Este “é o caso de *veleta* (< *vela*), que não é um tipo de vela, mas antes uma «pequena bandeira, ou

uma pequena figura plana de metal, que se coloca ao tope de um mastro para indicar a direcção do vento»” (CORREIA, 2006, p. 45).

Aproximando-se da conclusão final do trabalho, a autora julga oportuno abordar os sufixos que ocorrem mais frequentemente nos restantes subdomínios da Náutica e que não são os que mais ocorreram nos subdomínios até agora em estudo (Navegação, Construção Naval e parcialmente Meteorologia), identificando que tipos de sufixos se podem encontrar em cada caso.

Referindo-se a Correia (2004), a autora delimita dois tipos de sufixos disponíveis em português contemporâneo: os sufixos internacionais e os sufixos autóctones.

Os sufixos internacionais aparecem em estruturas analisáveis em português, portando significado relacional típico de sufixos do português, estabelecem com suas bases uma relação semântica que é reproduzível em outras estruturas semelhantes, são integráveis em paradigmas sufixais próprios de RCPs do português, sendo, portanto, interpretáveis como sufixos do português, no entanto, podem ou não ter sido construídos em português (CORREIA, 2006).

“Tendo tido origem latina ou grega, os sufixos internacionais apresentam sufixos cognatos em diversas línguas e comportamentos semelhantes em todas aquelas em que surgem, (...) constituem frequentemente internacionalismos, isto é, unidades que têm uma estrutura semelhante em diferentes línguas e que, por isso, são mais facilmente compreensíveis e traduzíveis” (CORREIA, 2006, p. 47).

Já os sufixos autóctones são de origem latina e, por isso, apresentam sufixos cognatos nas demais línguas românicas (sendo que alguns sufixos autóctones não ocorrem em todas as línguas românicas, inclusive alguns apenas ocorrem em português, não apresentando formas próximas em nenhuma das restantes línguas românicas.). Os sufixos autóctones não apresentam formas cognatas em inglês e, tampouco, apresentam

comportamento idêntico nas diversas línguas românicas, tais como os sufixos internacionais (CORREIA, 2006).

“A análise dos nomes de qualidade construídos em português permitiu verificar que os nomes construídos com sufixos internacionais dão, geralmente, a ver a qualidade que denotam de forma mais objectiva, ao passo que os nomes construídos com sufixos autóctones dão a ver as qualidades de forma menos objectiva, intervindo na denominação da qualidade aspectos como conhecimento estereotípico, atitude do enunciador, etc” (CORREIA, 2006, p. 48).

Assim, Correia (2006) levanta a hipótese, por ela mesma descartada na sequência, de que os sufixos utilizados nas terminologias científicas e técnicas seriam sobretudo os sufixos internacionais, por serem mais objetivos, enquanto que os sufixos utilizados na língua geral seriam sobretudo os sufixos autóctones, por serem menos objetivos. Essa hipótese é descartada no próprio estudo da terminologia da Náutica, com a presença nesta terminologia, por exemplo, dos sufixos *-eir-* e de sufixos avaliativos. A autora sugere como trabalhos futuros a busca de outras explicações para o fenômeno (CORREIA, 2006).

Correia (2006) observa que a maior parte dos sufixos presentes nos vocabulários “usados a bordo, dos subdomínios da Navegação, da Construção Naval e da Meteorologia são sufixos autóctones. Em contrapartida, os vocabulários de outras especialidades da Náutica (Política de Pesca e Águas Territoriais, Políticas Ambientais e Poluição, Comércio Internacional Marítimo e a vertente mais científica da Meteorologia) apresentam, sobretudo, derivados com sufixos internacionais” (CORREIA, 2006, p. 48), justificando estas ocorrências pelo fato de os profissionais deste último grupo possuírem elevada formação acadêmica e domínio de vários idiomas (especialmente o inglês), como já se mencionou anteriormente, já que “todos estes subdomínios são intrinsecamente internacionais, desenvolvidos e discutidos no âmbito de organizações internacionais, o que,

por si só, favorece a difusão de internacionalismos” (CORREIA, 2006, pp. 48-49) . “Os termos destes subdomínios em português correspondem, portanto, na maioria dos casos, a réplicas, decalques morfológicos ou morfossintáticos, de termos construídos primeiramente em língua inglesa” (CORREIA, 2006, p. 49).

“A adoção de estruturas semelhantes em línguas distintas, dando lugar à existência de internacionalismos, tem a vantagem de facilitar a transmissão de conhecimentos entre línguas, como é sabido, e, por isso, tem sido um procedimento adoptado no seio de instituições internacionais, como é o caso da União Europeia. Porém, esta prática pode ter também consequências, na minha opinião, dignas de reflexão e pode ser tanto mais preocupante quanto se assiste na comunicação profissional, actualmente, à substituição, consciente e/ou inconsciente, de termos autóctones por termos que apresentam estruturas mais próximas das dos equivalentes termos ingleses, para denominar os mesmos conceitos ou os que evoluíram a partir de conceitos antigos” (CORREIA, 2006, p. 49).

A autora preocupa-se com o predomínio do inglês como língua de comunicação científica internacional, principalmente, por acreditar na substituição das inúmeras opções de regras e operadores de construção de palavras autóctones das diferentes línguas por um pequeno número de operadores dotados de formas cognatas em inglês. No entanto, a autora não apresenta uma visão fundamentalista das línguas, admitindo que este fenómeno possa fazer parte da evolução natural das línguas (CORREIA, 2006).

Por fim, em sua conclusão, a autora ressalta a recente aproximação da Linguística com a Terminologia, buscando “ter mostrado o tipo de contributos que a nova linguística pode trazer ao desenvolvimento, sobretudo teórico, mas também aplicado, da Terminologia” (CORREIA, 2006, p. 50).

2.5.3. Modelo SILEX

O modelo teórico adotado nesta pesquisa foi concebido e desenvolvido por Danielle Corbin (Universidade de Lille, França) na década de 1980, e aperfeiçoado pela autora e pelos membros de sua equipe anos subsequentes.

Como sua unidade de pesquisa na ocasião denominava-se *Syntaxe, Interprétation et LEXique – SILEX*⁵⁶, por um processo metonímico de associação, os próprios integrantes do grupo passaram a denominar o próprio modelo morfológico de SILEX. Razão pela qual se adotou esse nome no decorrer de todo o trabalho.

O SILEX é “um modelo de „morfologia construcional” (e não apenas derivacional), que permite o tratamento da semântica e da referência das palavras construídas morfológica ou não-morfológica.” (CORREIA, 2004, p. 29), conferindo-lhe flexibilidade e eficiência, na medida em que é capaz de explicar tanto as palavras construídas como as não-construídas.

O termo „morfologia construcional” proposto por Danielle Corbin nos seus trabalhos, sobretudo a partir de 1992, em substituição de „morfologia derivacional”, pretende ser mais descritivo daquilo que a autora entende ser o objecto de estudo desta disciplina, isto é, a construção de palavras, não apenas por derivação, mas com recurso a outros processos de construção, tais como a composição, os processos deformacionais ou a lexicalização de sintagmas (CORREIA, 2004, p. 1)

Segundo Correia (2004), esse modelo tem sido aplicado com sucesso na análise e descrição do léxico construído de várias línguas românicas, particularmente do francês, língua primeira de aplicação, mas também do português e do grego moderno. No português, cabem citar os trabalhos norteadores de Graça Maria Rio-Torto (Universidade de Coimbra) e da própria Margarita Correia (Universidade de Lisboa), ambas de Portugal.

⁵⁶ Unidade de pesquisa do CNRS (UMR), que funcionou e foi liderada por Danielle Corbin até 2000, ano da sua morte. Esta unidade funcionava na Universidade de Lille III (França). Foi integrada ao laboratório associado do CNRS STL – Savoirs, Textes, Langage (UMR 8163), que funciona também na Universidade de Lille – <http://stl.recherche.univ-lille3.fr/>

A escolha desse modelo teórico para esta pesquisa não é ingênua: o objetivo do Modelo SILEX é construir uma teoria sincrônica do léxico capaz de atribuir uma estrutura e uma interpretação adequadas às palavras construídas, atestadas ou não nos dicionários, de modo a caracterizar a natureza da “gramaticalidade lexical” e de determinar as restrições das regras de formação de palavras (cf. CORBIN, 1987, p. 1).

Este modelo parte da hipótese fundamental de que, além das irregularidades de variados tipos observáveis na parte atestada do léxico das palavras construídas, este subconjunto lexical obedece a um conjunto hierarquizado de regras e princípios, com natureza, conteúdo e campo de aplicação que devem ser determinados pelo linguista (cf. CORBIN, 1987, p. 1). Nesse sentido, o modelo assume-se como um modelo associativo e estratificado.

Por „modelo associativo” entende-se aquele cujas Regras de Construção de Palavras (RCPs) permitem construir conjuntamente a estrutura morfológica e a interpretação semântica das palavras construídas; ou seja, uma palavra construída por uma RCP apresenta uma motivação recíproca entre a sua forma (ou estrutura) e o seu significado. Essa motivação é necessária no quadro associativo adotado: entende-se que a adequação de uma gramática se mede designadamente pela sua capacidade de dar conta dessa motivação (cf. CORBIN, 1997).

O SILEX é um „modelo estratificado” porque apresenta um componente lexical da gramática composto por vários níveis, ao longo dos quais se vai construindo o significado das palavras construídas. A estratificação manifesta-se no fato de a estrutura interna do componente lexical e a correspondente hierarquia de operações nas palavras construídas refletirem a complexidade das relações entre forma e significado características das

palavras construídas, relações essas que não podem ser reduzidas à mera dicotomia regular vs. irregular (cf. CORBIN, 1989, p. 31).

Segundo a autora, a originalidade do modelo de morfologia construcional que propõe reside no refinamento da tipologia das associações entre forma e significado características das palavras construídas e na mudança da habitual ordem de prioridades presente nos trabalhos de morfologia derivacional: em vez de propor análises baseadas na evidência do léxico observável, propõe uma análise baseada na estratificação e na reconstrução do léxico descritível; em vez de dar prioridade à análise morfológica sobre a análise semântica, propõe uma análise que associa forma e significado (cf. CORBIN, 1989, p. 49).

Uma das questões a que os trabalhos enquadrados neste modelo pretende responder é a seguinte: para denominar um conceito, entre as várias possibilidades que a língua oferece, a escolha de uma ou outra estrutura morfológica não é indiferente – por que, então, se escolhe uma determinada estrutura em detrimento de outra?

De acordo com este modelo, as unidades lexicais⁵⁷ são portadoras, entre outros, de significado descritivo, de significado referencial e de significado enunciativo.

O significado descritivo e o significado referencial caracterizam as unidades que dão nome, que permitem denominar entidades que existem fora da própria linguagem. Esses significados permitem a unidades de categorias maiores (substantivos, adjetivos, verbos e alguns advérbios) referirem por elas mesmas, independentemente dos usos que delas se fazem. O significado descritivo é intensional por natureza, mas tem, ao mesmo tempo, poder referencial, na medida em que as propriedades semânticas se encontram

⁵⁷ As unidades lexicais, a par das unidades infralexicais (*grosso modo*, afixos e elementos não-autônomos de composição), constituem os indivíduos lexicais de uma língua. Apenas algumas das unidades lexicais têm significado referencial, dado que apenas estas são capazes de nomear categorias referenciais.

reunidas em categorias (semânticas) cujos objetos mentais que lhes correspondem – os referentes – constituem a sua extensão, as categorias referenciais correspondentes.

As categorias semânticas definem-se por propriedades exclusivamente semânticas e intralinguísticas e são apenas caracterizáveis através da sua intensão; à noção de categoria semântica corresponde o significado descritivo. As categorias referenciais representam a adaptação das categorias semânticas à referência, em função da apreensão perceptual do mundo e da apreensão cultural das coisas. As categorias referenciais são caracterizáveis em intensão e em extensão; à noção de categoria referencial corresponde a noção de significado referencial.

Por conseguinte, o significado enunciativo tem a ver com a posição que o enunciador assume em relação ao(s) referente(s) das unidades lexicais que utiliza no seu discurso e pode exprimir os seguintes tipos de relações entre o enunciador e o(s) referente(s) em causa: *a*) avaliação das dimensões (tamanho, intensidade, etc.) do(s) referente(s) em relação a uma norma social pré-estabelecida e partilhada pelo enunciador; *b*) avaliação de tipo axiológico, favorável ou desfavorável, positiva ou negativa, em relação ao(s) referente(s), incluem-se neste item os chamados sentidos 'pejorativo' e 'valorativo' atribuídos às unidades lexicais; *c*) proximidade em relação ao(s) referente(s), essa proximidade pode ser de tipo factual (locativo) ou apenas de tipo afetivo. Os recursos morfológicos disponíveis para a construção dos tipos de significado enunciativo mencionados em *a*) e *b*) são, normalmente, abrangidos pela chamada construção de avaliativos (cf. CORREIA, 1999, p. 33-38).

Acredita-se que o aprofundamento na análise do significado das unidades lexicais torna este modelo particularmente eficiente para a descrição das terminologias selecionadas para esta pesquisa, assim como da sua capacidade de nomear os conceitos

próprios desses domínios. Já que este modelo também prevê a distinção entre significado de base e significado derivado, ele é também apto para a descrição dos processos que levam determinadas unidades a tornarem-se aptas para denominar conceitos próprios de domínios especializados. Em resumo, este modelo oferece o instrumental teórico necessário para dar conta não apenas da estrutura morfológica dos termos, mas também da polissemia e da polirreferência⁵⁸ das unidades que integram as terminologias em estudo.

O estudo da polissemia tem estado no centro das atenções de linguistas de diferentes filiações, em particular daqueles que participam no desenvolvimento de bases de dados lexicais concebidas para funcionarem como suporte de programas de processamento de linguagem natural (PLN) (cf. PUSTEJOVSKI, 1995). A idéia de base destes trabalhos é explorar as potencialidades do seguinte fato: é muito mais fácil e econômico descrever uma regra que implica que, sempre que uma unidade lexical apresenta o significado *X*, apresenta concomitantemente o significado *Y*, do que levar a cabo essa descrição de forma atomística, para cada unidade lexical tratada. Em outras palavras, desde que se detecte a regra que padroniza uma transferência de significado, é mais econômico afetar a aplicação dessa regra a todas as palavras que a sofrem (gerando automaticamente os significados derivados) do que descrever, para cada forma, tantas entradas homônimas quantos os significados que essa forma apresente.

⁵⁸ Polirreferência compreende a situação na qual a uma mesma forma, portadora de um determinado significado descritivo, estão associados diferentes significados referenciais. Distingue-se, então, no quadro deste modelo, polirreferência de polissemia. A polissemia ocorre sempre que entre os significados apresentados por uma mesma unidade lexical é possível estabelecer elos semânticos, resultantes do fato de que, a partir de um significado de base, se construíram diversos significados derivados, por meio de operações de derivação semântica.

Ressalte-se, finalmente, que esse modelo se enquadra no quadro teórico gerativo, na medida em que pretende descrever a competência derivacional dos falantes de uma língua, como se pode observar na citação a seguir:

Qualquer pessoa que tenha adquirido o conhecimento de uma língua interiorizou um sistema de regras que determina as conexões som-significado para uma infinidade de palavras construídas. Naturalmente, aquele que conhece perfeitamente uma língua tem pouco ou nenhum conhecimento consciente das regras que usa constantemente ao falar, ouvir, escrever ou ler, ou no seu monólogo interior. É esse sistema de regras que o torna capaz de produzir e de interpretar palavras construídas que nunca encontrou antes. Normalmente se negligencia este fato notável de que, no discurso normal, cotidiano, cada um compreenda e produza palavras construídas novas sem nenhuma consciência da sua novidade, embora essas palavras construídas normais apenas sejam semelhantes às que foram produzidas ou encontradas anteriormente pelo fato de serem construídas e interpretadas pela mesma gramática, o mesmo sistema interiorizado de regras. (CORBIN, 1987, p. 47-48. Trad. de Margarita Correia)⁵⁹

Embora o Modelo SILEX tenha todas as características para conformar-se com a perspectiva gerativa, segundo Correia (2004), Corbin, a partir da década de 1990, distancia-se dessa perspectiva devido aos desenvolvimentos que a escola gerativa vinha apresentando naqueles anos. Ainda de acordo com Correia (2004), esse distanciamento se deve fundamentalmente pelas seguintes razões:

- por um lado, no quadro do modelo SILEX, o significado das palavras construídas encontra-se no cerne da discussão morfológica, ao passo que as teorias morfológicas que actualmente se reclamam gerativas se interessam mais pela forma das palavras construídas do que pelo seu significado;
- por outro lado, no quadro do modelo SILEX não se parte do princípio de que as palavras construídas sejam micro-frases, ou frases em modelo

⁵⁹ “Toute personne qui a acquis la connaissance d’une langue a intériorisé un système de règles qui détermine des connexions son-sens pour une infinité de mots construits. Naturellement, celui qui connaît parfaitement une langue a peu ou n’a pas de connaissance consciente des règles qu’il utilise constamment lorsqu’il parle, écoute, écrit ou lit, ou dans le monologue intérieur. C’est ce système des règles qui le rend capable de produire et d’interpréter des mots construits qu’il n’a jamais rencontrés auparavant. On néglige trop souvent ce fait remarquable que, dans le discours normal, quotidien, chacun comprend et produit des mots construits nouveaux sans aucune conscience de leur nouveauté, bien que ces mots construits normaux ne soient semblables à ceux qui ont été produits ou rencontrés antérieurement qu’en ce qui ils sont formés et interprétés par la même grammaire, le même système intériorisé de règles. ”

reduzido, analisáveis com recurso a dispositivos próprios da sintaxe: nomeadamente, não se raciocina em termos de cabeças (*heads*) nem em termos de percolação (*percolation*) para explicar a estrutura das palavras construídas (...) (CORREIA, 2004, p. 33).

Ainda assim, considera-se nesta pesquisa o Modelo SILEX como integrante de uma abordagem gerativa *lato sensu*, por duas razões fundamentais: 1) o modelo está mais preocupado em descrever as regras que permitem a construção de palavras do que as palavras construídas, ou seja, o foco está mais no processo do que no produto; 2) como consequência, o modelo passa a considerar fortemente a competência lexical do falante, que permite que ele construa palavras novas como também compreenda as palavras construídas, como ilustra claramente a citação de Corbin (1987) acima.

A análise da estrutura morfológica dos termos, enquadrada pelo Modelo SILEX, pretende, portanto, responder, entre outras questões, quais os processos de construção de palavras mais produtivos na construção de termos dos domínios aqui analisados.

2.5.4 Processos de inovação lexical em português

Correia & Lemos (2005) sintetizam a ideia consensualmente aceita de que “uma das características universais da linguagem é a mudança. Qualquer língua é caracterizada pela mudança e pela inovação. Todas as línguas evoluem necessariamente ao longo do tempo e a ausência de evolução significa para elas a sua morte” (CORREIA & LEMOS, 2005, p. 10). Essa mudança afeta a língua de forma geral, no entanto, é observável majoritariamente no nível do léxico, de forma específica.

A mudança lexical decorre do desaparecimento e do surgimento de palavras, o primeiro refere-se às palavras que, por razões extralinguísticas, deixam de ser utilizadas em uma língua, os chamados arcaísmos, enquanto que o segundo surge da necessidade de

denominação de novos conceitos e realiza-se por meio de diferentes fenômenos de criação lexical. Desses fenômenos, o desuso (desaparecimento) e a inovação (surgimento), apenas o segundo está no âmbito desta pesquisa, por isso apenas os processos por meio dos quais ocorre a inovação lexical serão abordados na sequência.

Para além das inovações *ex nihilo* (a partir do nada), as línguas vivas possuem três alternativas de inovação lexical:

- 1) a construção de palavras a partir de regras da própria língua;
- 2) a atribuição de novos significados a palavras já existentes, os chamados neologismos semânticos;
- 3) a importação de palavras de outras línguas, o que se convencionou chamar de estrangeirismos⁶⁰.

Nesta pesquisa, dado o seu foco e os objetivos traçados, importa estudar apenas a primeira alternativa elencada.

A construção de palavras a partir de regras da própria língua constitui-se como mecanismo de inovação lexical e realiza-se da forma que se descreve abaixo:

“(…) parte-se de elementos pré-existentes (unidades lexicais e afixos) e, com base num conjunto de regras interiorizadas e partilhadas pelos falantes, juntam-se esses elementos, de modo a obter novos itens, de estrutura normalmente transparente, ou seja, itens ou palavras cuja estrutura morfológica é perceptível e cujo significado é coerente com essa estrutura.” (CORREIA & LEMOS, 2005, p. 24).

Nesse sentido, são os mecanismos encontrados nas regras interiorizadas e compartilhadas pelos falantes que serão detalhados nos itens seguintes⁶¹ e que,

⁶⁰ Em comunicação pessoal, Margarita Correia acrescentou a esta lista a formação de onomatopeias.

⁶¹ O estabelecimento, a denominação e a caracterização dos diferentes processos morfológicos não são consensuais, mas de acordo com o modelo ou abordagem teóricos adotados são, por vezes, destoantes. Por

posteriormente, constituirão a base das análises morfológicas propostas. Os exemplos citados são resultados da análise morfológica empreendida na terminologia em análise de Nanociência e Nanotecnologia e Biocombustíveis.

2.5.4.1 Derivação

A derivação é o mais disponível, produtivo e típico processo de inovação lexical do português e de todas as línguas românicas, segundo Correia & Lemos (2005). Semelhante a outras classificações, as autoras apresentam os seguintes tipos de derivação: afixal, subdividida em sufixal e prefixal; parassintética; imprópria (ou conversão) e regressiva. Todos esses tipos serão abordados na sequência.

2.5.4.1.1 Derivação afixal

Correia & Lemos (2005), postulam que “ao estudar a derivação afixal importa ter em atenção que este é tipicamente um processo binário, isto é, em cada processo derivacional apenas intervêm, de cada vez, uma base ou radical derivacional e um afixo.” (CORREIA & LEMOS, 2005, p. 24). Assim, com exceção dos derivados parassintéticos, os produtos com mais de um processo derivacional devem ser entendidos como frutos não de um, mas de vários processos derivacionais encadeados.

Nos processos derivacionais afixais, à base se junta um afixo. A base deve ser o único elemento de significado lexical⁶² envolvido no processo, já que quando dois ou mais

isso, adotam-se nesta pesquisa as denominações e conceitos propostos por Correia & Lemos (2005), por estarem de acordo o modelo morfológico adotado (Modelo SILEX).

⁶² “Entende-se que as “unidades de significado lexical” são aquelas que permitem (...) denominar entidades que existem fora da própria linguagem. Estas unidades são, principalmente, substantivos, adjetivos e verbos, constituindo, portanto, classes abertas de palavras.” (Correia & Lemos, 2005, p. 26).

elementos de significado lexical estiverem envolvidos no processo de criação lexical, tratar-se-á de composição e não derivação.

Os afixos derivacionais, conjunto limitado em cada língua, caracterizam-se por terem apenas significado gramatical, relacional ou instrucional⁶³. São afixos derivacionais, segundo Correia & Lemos (2005) os prefixos, sufixos, circunfixos (elementos da derivação parassintética) e os segmentos epentéticos (elementos de adequação, como o “l” em *chaleira*, ou o “t” em *cafeteira*), o que tradicionalmente se denomina *consoante de ligação*.

A derivação afixal é um conjunto de três processos diferentes, explicitados a seguir.

2.5.4.1.1.1 Derivação prefixal

Os prefixos envolvidos neste processo possuem as seguintes características, de acordo com o proposto por Correia & Lemos (2005):

- ocorrem sempre à esquerda da base;
- correspondem, frequentemente, a antigas proposições e advérbios latinos e gregos;
- possuem conteúdo semântico menos gramatical do que os sufixos.

A prefixação, tipicamente, motiva diferentes acréscimos de significação ao elemento base, tais como: negação (*insolúvel*), oposição ou privação (*anticongelante*); localização espaço-temporal (*pré-incubação*); quantificação ou intensificação (*polipropileno*).

⁶³ “Por seu turno, as unidades de significado gramatical (ou relacional, ou instrucional) são aquelas cujo significado não permite estabelecer uma relação directa com a realidade extralinguística, consistindo, sobretudo, na capacidade de estabelecer relações de variado tipo entre as unidades de significado lexical. São unidades de significado gramatical as preposições, as conjunções, os artigos, os pronomes, e, também, os afixos. Estas unidades constituem classes fechadas.” (Correia & Lemos, 2005, p. 26).

2.5.4.1.1.2 Derivação sufixal

Os processos de sufixação se valem de afixos com as seguintes características, conforme proposto por Correia & Lemos (2005):

- ocorrem sempre à direita da base;
- na maior parte dos casos, determinam a categoria do derivado;
- determina a sílaba tônica da palavra;
- impulsiona os seguintes processos em língua portuguesa: verbalização, nominalização, adjetivação, adverbialização.

A sufixação de uma base pode provocar sua verbalização, nominalização (*pureza*, *dureza*) adjetivação (*molecular*, *nuclear*) ou adverbialização.

2.5.4.1.1.3 Derivação parassintética

Para grande parte dos manuais de Morfologia Derivacional, a parassíntese é o processo de junção simultânea de um prefixo e um sufixo a uma base, desde que a junção de ambos seja necessária para a construção da palavra, isto é, a palavra não se constitui como signo linguístico independente na ausência do prefixo ou do sufixo. A palavra *infelizmente*, por exemplo, não seria um caso de parassíntese, já que são possíveis as formas *infeliz* (prescindindo do sufixo *-mente*) e *felizmente* (prescindindo do prefixo *in-*). Para outros autores, a derivação parassintética pode ser compreendida como a afixação de um afixo descontínuo, por exemplo, “(...) num verbo como envelhecer, interviria apenas um afixo, representável como *en ... ec(er)*, que se juntaria a uma base, o radical *velh-_{Adj.}*” (CORREIA & LEMOS, 2005, p. 33).

As palavras obtidas como produtos da parassíntese são verbos ou substantivos derivados de nomes ou adjetivos.

2.5.4.1.2 Conversão

Este tipo de derivação consiste basicamente em uma mudança de categoria sem a introdução de qualquer afixo ou qualquer outra modificação.

Por não ocorrer mudanças no significante, não constitui objeto de interesse desta pesquisa.

2.5.4.1.3 Derivação regressiva

Neste tipo de derivação, ocorre a supressão de um segmento da base. Tem-se como base do processo os verbos e, como produtos resultantes, nomes de ação com a significação “ato ou efeito de (verbo base)”.

2.5.4.2 Composição

A partir da estrutura interna das palavras compostas, Correia & Lemos (2005) dividem a composição em dois processos: a composição morfológica e a lexicalização de sintagmas, ambas serão descritas a seguir.

2.5.4.2.1 Composição morfológica

Este processo, típico em terminologias, consiste na junção de radicais gregos ou latinos (*auto*, *geo*, *ribo*, *macro*, ...) a uma palavra autônoma (*sensor*, *processador*, *litografia*, ...) ou a uma unidade lexical não autônoma (*gênese*, *lise*, *xido*, ...). Entre elas podem ocorrer a vogal de ligação “o” ou “i”.

Diversas são as diferentes denominações recebidas por este processo, tais como: composição morfológica, composição morfológica culta, composição neoclássica, composição erudita, composição por temas. E ainda, sob a luz do Modelo SILEX: composição por arqueoconstituintes, composição por constituintes patrimoniais.

2.5.4.2.2 Recomposição

Este processo consiste na junção de elementos de recomposição (prefixóides ou pseudoprefixos ou, ainda, fractoconstituintes) a uma palavra autônoma ou a uma unidade lexical não autônoma. Entre elas podem ocorrer a vogal de ligação “o” ou “i”.

Elementos de recomposição ou fractoconstituintes são o resultado de truncação de unidades maiores. Esses elementos revelam a existência em português de elementos de composição homônimos. São exemplos desses elementos: “*agro-* (de *agricultura*, em *agropecuária*), *eco-* (de *ecologia*, em *ecoponto*), *euro-* (de *uropeu*, em *eurodeputados*), *foto-* (de *fotografia*, em *fotobiografia*), *narco-* (de *narcótico*, em *narcotráfico*), ou *petro-* (de *petróleo*, em *petrodólares*)” (CORREIA & LEMOS, 2005, p. 43).

2.5.4.2.2 Composição por lexicalização de sintagmas

Ao se transformarem em unidades lexicais, os sintagmas passam por uma especialização semântica, ou seja, o seu significado deixa de ser composicional, formado a partir da soma dos significados dos elementos que compõem a unidade lexical.

Os compostos sintagmáticos são, segundo Correia & Lemos (2005), sobretudo, nomes cuja sintaxe é frequentemente correspondente a uma das seguintes estruturas sintáticas:

- N + Adj
- Adj + N
- N + SP
- N + N
- V + N
- Prep + N
- Adj + Adj
- Adv + Adj

2.5.4.3 Processos deformacionais

São processos deformacionais as amálgamas, as trunicações, as siglas e os acrónimos. Todos esses diferentes processos integram este grupo já que em todos ocorre “um atentado à integridade fonológica dos radicais **envolvidos** na construção dos seus produtos.” (CORREIA & LEMOS, 2005, p. 43). Cada processo é descrito detalhadamente na sequência.

2.5.4.3.1 Amálgama

Amálgamas são unidades construídas com partes de outras palavras, tal como *diodiesel* formada a partir de *diesel* + *biodiesel*.

2.5.4.3.2 Truncação

A truncação é o processo no qual ocorre a redução de uma palavra para apenas parte dessa palavra, por exemplo, *otorrino* de *otorrinolaringologista*, tornando-a mais fácil de se memorizar e de se utilizar.

2.5.4.3.3 Sigla e acronímia

As siglas são formadas pela junção das iniciais de um sintagma, como em *MEV* (*Microscópio Eletrônico de Varredura*), por exemplo. Já os acrônimos são formados pelas letras ou grupos de letras iniciais de um sintagma, além disso, são pronunciados como uma palavra, enquanto as siglas são pronunciadas letra por letra, como exemplo de acrônimo pode-se citar a palavra *CANOLA* formada a partir de "*CAnadian Oil, Low Acid*".

2.5.4.4 Lexicalização de formas flexionadas

Segundo Correia & Lemos (2005), entende-se como lexicalização de formas flexionadas “a entrada na componente lexical de unidades cujo lema corresponde a uma forma flexionada de outra palavra.” (CORREIA & LEMOS, 2005, p. 46).

Constituem esse fenômeno a nominalização de verbal e o particípio passado, como se pode observar em:

- aguado_{p.p.v} (aguar = transformar, dissolver em água) → aguado_A (estragado, imperfeito, malgrado) – particípio passado.

2.5.4.5 Lexicalização de unidades discursivas

Algumas palavras podem, ainda, ser formadas pela “lexicalização de unidades discursivas (frases e partes de frases) que passam para a componente lexical, assumindo a

categoria morfossintáctica de nomes, trazendo as suas marcas flexionais próprias.” (CORREIA & LEMOS, 2005, p. 47). São exemplos desse processo: *dois-em-um* e *maria-vai-com-as-outras*.

É a partir dessa classificação dos processos de inovação lexical em língua portuguesa que as terminologias serão analisadas, conforme procedimento metodológico exposto a seguir, procurando-se i) identificar quais os processos envolvidos e entre eles quais os mais produtivos, ii) proceder a uma análise profunda dos processos mais produtivos a fim de iii) estabelecer-se as respectivas RCPs.

3. Metodologia

No campo do processamento automático das línguas naturais (PLN), os sistemas (*softwares*) de PLN são formados por ferramentas computacionais e recursos linguísticos. As ferramentas computacionais (aqui entendidas como os códigos de programação e demais componentes computacionais do *software*) pertencem ao domínio da Ciência da Computação e, por isso, não serão tratados aqui. Já por recursos linguísticos compreendem-se: Gramática, Léxico ou Corpora.

Nessa pesquisa, o conhecimento linguístico obtido a partir da descrição dos repertórios terminológicos em análise encontra-se no domínio do Léxico.

O léxico, enquanto recurso de PLN, pode assumir diferentes concepções. A concepção adotada nessa pesquisa é a de Base de Dados Lexicais (BDL). A BDL pode ser compreendida como um arquivo, repositório de conhecimento linguístico. O conhecimento linguístico armazenado em um léxico não é fixo, podendo compreender apenas um ou

vários níveis linguísticos, tais como: morfológico, sintático, semântico, pragmático e discursivo, de acordo com as especificidades e objetivos de cada sistema de PLN (DI FELIPPO & DIAS DA SILVA, 2006).

Nesta pesquisa, busca-se a construção de um repositório de conhecimento linguístico de nível morfológico envolvendo duas terminologias que integram a grande área das Engenharias, conhecimento este que poderá suprir sistemas de processamento automático de linguagem natural (PLN) que demandem conhecimentos morfológicos para suas ações. Para tanto, iniciar-se-á com a análise e descrição morfológica dos domínios da Nanociência/Nanotecnologia e Biocombustíveis-Biodiesel⁶⁴, conforme será apresentado nas subseções a seguir.

A criação dessa base se faz necessária devido à insuficiente quantidade de recursos linguísticos em nível morfológico para o português, em um formato adequado para processamento computacional em sistemas de PLN, o que tem representado baixíssimo rendimento nas tarefas de recuperação de informação, nomeadamente, na extração automática de terminologias.

A metodologia genérica de pesquisas em PLN, proposta por Dias-da-Silva (2006), estabelece três domínios de atuação, a saber:

- domínio linguístico: investigação e descrição dos fatos da língua em diferentes dimensões (níveis);
- domínio representacional: formalização do conhecimento descrito no domínio linguístico;

⁶⁴ Conforme discutido ao longo da seção 2, cada terminologia possui estruturas morfológicas características, com grande variação nos padrões morfológicos de formação de palavras, assim, com esta pesquisa que ora se apresenta, objetiva-se fazer a descrição morfológica de dois vocabulários, de maneira a colaborar com a construção da BDL.

- domínio implementacional: codificação do conhecimento formalizado em uma linguagem de programação.

A instanciação do domínio representacional revela que a formalização do conhecimento linguístico descrito no domínio anterior implica na representação do conhecimento em um formato compatível com o domínio implementacional.

A metodologia pertinente ao domínio linguístico é detalhada a seguir, desde a seleção dos vocabulários, objetos de análise, até a análise propriamente dita.

3.1 Seleção e delimitação dos repertórios terminológicos

Conforme já mencionado na seção 1, as terminologias analisadas por essa pesquisa são:

- **Nanociência/Nanotecnologia:** essa terminologia teve como origem o projeto *Terminologia em Língua Portuguesa da Nanociência e Nanotecnologia: Sistematização do Repertório Vocabular e Elaboração de Dicionário-Piloto – NanoTerm*⁶⁵,
- **Biocombustíveis:** terminologia oriunda do projeto *Terminologia de Biocombustíveis: descrição semântica e morfológica com vistas à sistematização*⁶⁶.

⁶⁵ Apoio CNPq, processo n°. 400506/2006-8, coordenado por Gladis Maria de Barcellos Almeida, desenvolvido por Joel Sossai Coleti (UFSCar) e Daniela Ferreira de Mattos (UFSCar), com a colaboração de Sandra Maria Aluísio (USP), Ariani Di Felippo (UFSCar), Luiz Carlos Genoves Junior (GOOGLE), Arnaldo Candido Júnior (USP), Claudia Zavaglia (UNESP), Deni Kasama (UNESP), Osvaldo Novais de Oliveira Junior (USP) e Leandro Henrique Mendonça de Oliveira (EMBRAPA).

⁶⁶ Apoio CNPq, processo n°. 473414/2007-4, coordenado por Gladis Maria de Barcellos Almeida, desenvolvido por Douglas Henrique Perez Pino (UFSCar) e Fernando Alves (UFSCar), com a colaboração de Sandra Maria Aluísio (USP), Margarita Correia (Universidade de Lisboa, Portugal), Ariani Di Felippo (UFSCar), Arnaldo Candido Júnior (USP) e Felipe Gomes (USP).

Para a escolha desses domínios, valeu-se da disponibilidade dos mesmos aliada a sua adequabilidade aos objetivos da pesquisa. Como extração de repertórios terminológicos não é um dos objetivos desta pesquisa, necessitava-se de repertórios terminológicos prontos, coerentes com as teorias norteadoras da pesquisa, especificamente a Teoria Comunicativa da Terminologia, e aptos para reuso. A opção pela análise de repertórios terminológicos obtidos de corpora tem como motivação a possibilidade de se observar no uso real a maneira como essas terminologias têm sido empregadas. Além disso, examinando o termo em seu contexto de ocorrência no corpus, é possível recuperar diversas informações de ordem semântica.

Dentre outros repertórios terminológicos produzidos pelo GETerm que poderiam ser objeto de reuso, optou-se por esses dois tendo em vista que são terminologias multidisciplinares da área de Engenharia, por isso, com a utilização destes repertórios, acredita-se iniciar a composição de uma larga base de dados com uma cobertura inicial de, ao menos, uma grande área do saber científico, a das Engenharias.

As terminologias totalizam 3.996 termos, sendo 3.069 oriundos da Nanociência/Nanotecnologia e 927 dos Biocombustíveis. Essa diferença de tamanho das duas terminologias deve-se ao fato de que somente foram utilizados nesta pesquisa os termos do biodiesel, já que a lista dos termos do etanol ainda não havia sido validada pelos especialistas de domínio quando esta pesquisa se iniciou. Mas essa diferença quantitativa entre as terminologias não é problemática, tendo em vista que o interesse aqui está na precisão e não na revocação (*recall*).

Dada a grande quantidade de termos, para a viabilização da realização dessa pesquisa, fez-se necessário a delimitação de um recorte para análise e descrição.

Inicialmente, propôs-se um recorte com base em critérios quantitativos. Para efeito de ilustração, cogitou-se a seleção de 10% dos termos mais frequentes em cada uma das listas. No entanto, à medida que se estudava o modelo de análise adotado, verificou-se que essa primeira forma de delimitação das unidades não se mostrava adequada.

A formação de palavras, vista pelo Modelo SILEX, ocorre através de Regras de Construção de Palavras, compostas cada uma por vários e diferentes processos morfolexicais, assim, é mais pertinente e coerente descrever as Regras de Construção de Palavras praticadas nessas terminologias e não os itens lexicais mais frequentes e recorrentes, uma vez que se estaria focalizando o produto e não o processo⁶⁷, o que absolutamente contraria o modelo teórico adotado. Dessa forma, optou-se por não se fazer uma delimitação com base em critérios quantitativos, mas sim qualitativos, os quais dizem respeito à delimitação dos processos a serem analisados.

Apesar de o Modelo SILEX ser capaz de estabelecer RCPs por meio de vários e diferentes processos morfolexicais, optou-se por analisar apenas os processos morfológicos que ocorrem no interior de um mesmo indivíduo lexical. Assim, o primeiro desbaste realizado foi retirar de ambas as terminologias os termos compostos por duas, três, quatro ou mais palavras⁶⁸ (os chamados bigramas, trigramas, tetragramas e assim sucessivamente) Dessa forma, nesse primeiro recorte, os repertórios terminológicos foram reduzidos a 1.558 termos de Nanociência/Nanotecnologia e 580 termos de Biodiesel, totalizando 2.138 termos (unigramas).

⁶⁷ A descrição das RFPs presentes em um determinado conjunto de termos, de acordo com o recorte terminológico originalmente proposto, poderia ser repetitiva (caso os termos mais frequentes fossem formados majoritariamente pelas mesmas regras) e limitada (caso não cobrisse as RFPs dos termos de frequências menores).

⁶⁸ Palavra aqui entendida como sendo uma sequência de caracteres separada por espaços em branco, tal como é de consenso no campo dos estudos em PLN.

A seguir, em consonância com o modelo teórico adotado, foram excluídas do repertório a ser analisado as palavras não construídas morfológicamente, ou seja, os empréstimos, as palavras simples e as palavras complexas não construídas, os nomes próprios, as siglas e os acrônimos.

Apresentam-se, na tabela 3 os itens lexicais excluídos do repertório terminológico da Nanociência/Nanotecnologia.

Termos	Justificativa
AURUM BANDGAP BOTTOM BOTTOM-UP COATING CLUSTER DIP-COATING DISPLAY DUPLEX FILLERS FLASH FREEZE-DRYING FUNDING GAP GIGABYTE HEAT-TREATMENT INGAP INK-JET INPUT IN-SITU ION-PLATING LAB-ON-A-CHIP LASER LEAP-FROG LIFT-OFF NANOBOT PIPE SCANNING SCATTERING SEED SOFTWARE SPIN SPIN COATING SPIN-COATING SPIN-FLOP SPIN-OFF STOKES START-UP SPRAY SPRAY-COATING	Importações parcialmente ou completamente opacas.

SPUTTERING SYNCHROTRON TERABIT TOP-DOWN VITRO	
ADP AFM AIF ASTM ATP BET BIAS BIT CCD CCT CDS CDSE CHEMFET CHIP CMOS CVD DLC DNA DRX DSC DTA EDA EDS EDX EELS FDA FEG-SEM FET GAP GBITS GEIM GIBBS GPS HDG HF ISFET ITO JST LED LIMN MBE MC MD MEMS MET MEV MFA MHZ MI MICA MIT	Siglas e acrônimos

NEMS NIST NOR NPS NRI NSF NTT OCAS OCDE OD OLED PAMAM PCA PDMS PEG PEM PEO PHMB PIPE PLED PMMA POEA P-OLED PPM PPY PSA PST PTP PVD QFD REES RGB RMA RMN RN RNA SAW SAXS SEM-FEG SIMS SOLVIS SPM SQW SRG STM STM-UHV TFT-LCD TG TIN UV-VIS VBE VIM WET XAFS XPS XRD	
---	--

ANGSTRÖNS BRILLOUIN DREXLER ESCHERICHIA RAMAN RIETVELD RIKEN VAN DER WAALS	Substantivos próprios
---	-----------------------

Tabela 3: Itens lexicais não-construídos da terminologia da Nanociência/Nanotecnologia

E, na Tabela 4, os itens lexicais excluídos do repertório terminológico do Biodiesel.

Termos	Justificativa
CANOLA	Sigla
DIESEL	Substantivo próprio

Tabela 4: Itens lexicais não-construídos da terminologia do Biodiesel

Podem-se observar, pelas tabelas 3 e 4, 45 importações parcialmente ou completamente opacas, grande quantidade de siglas e acrônimos (107), além de 8 substantivos próprios, totalizando 160 exclusões no repertório terminológico da Nanociência/Nanotecnologia.

A análise da terminologia do Biodiesel identificou apenas uma sigla e um substantivo próprio, somando 2 exclusões.

Totalizam-se, assim, 162 exclusões o que reduz o repertório terminológico para 1.976 termos.

Delimitado o repertório terminológico, todos os termos foram submetidos à análise inicial, que subsidia a análise profunda, ambas objeto de descrição da subseção seguinte.

3.2 Análises: inicial e profunda

O termo *análise inicial* é utilizado para identificar o seguinte grupo de ações e procedimentos metodológicos que subsidiarão a *análise profunda*:

- análise morfológica de todos os termos⁶⁹, seguindo a classificação de processos de inovação lexical proposta por Correia & Lemos (2005), em conformidade com os pressupostos teóricos estabelecidos pelo Modelo SILEX;
- listagem dos termos por processo morfológico, apenas para os processos de maior pertinência para essa pesquisa, ou seja, aqueles que envolvem operações predominantemente morfológica, excluindo-se aqueles que tangem a sintaxe, a semântica e/ou outros níveis linguísticos. Serão considerados pertinentes, portanto, os seguintes processos:
 - derivação prefixal
 - derivação sufixal
 - composição morfológica
- identificação do morfema de maior frequência em cada processo morfológico.

Como etapa seguinte, tem-se o desenvolvimento das *análises profundas*, termo que denomina:

- análise criteriosa dos itens de maior frequência selecionados em cada processo morfológico pertinente, buscando responder, entre outras, as seguintes questões: i) Que significados têm os morfemas? ii) Que significados passam a ter os produtos formados quando estes se associam a determinada base? iii) Há alguma interferência da classe gramatical da base

⁶⁹ Exceto aqueles excluídos conforme o item 3.1

a qual o morfema de associa? iv) Que tipo de generalizações se podem fazer para gerar as regras?

- proposição das Regras de Construção de Palavras, as RCPs, para os casos selecionados.

4. Base de dados morfológicos: os resultados da análise inicial dos processos morfológicos

Apresentam-se, nessa seção, os resultados da análise inicial (explicitada no item 3.2). Correia & Lemos (2005), autoras do modelo de classificação morfológica dos processos de inovação lexical adotado, destacam que “as diferentes classes de palavras não constituem categorias estanques, constituindo antes um *continuum* (...)” (CORREIA & LEMOS, 2005, p. 27), faz-se prioritário destacar, portanto, que ainda que sob a luz deste modelo teórico a classificação proposta não encerra o tema, antes, dá início a discussões, possivelmente, controversas.

Apresenta-se, primeiramente, a descrição morfológica de Nanociência/Nanotecnologia para, em seguida, apresentar-se a descrição morfológica do Biodiesel.

Os termos apresentados a seguir como ocorrências dos processos analisados estão reproduzidos da forma como foram validados pelos especialistas de domínio dos projetos dos quais foram extraídos para esta análise..

4.1 Descrição morfológica da terminologia de Nanociência/Nanotecnologia

A classificação proposta é apresentada abaixo através de um conjunto de tabelas. Cada tabela é adaptada ao processo de formação de palavras nela representado, de forma que as informações pertinentes em cada caso pudessem ser satisfatoriamente mostradas. Como o objetivo último deste trabalho é identificar os operadores mais produtivos nos domínios da Nanociência e Nanotecnologia e dos Biocombustíveis, nas tabelas a seguir incluem-se os termos que apresentam o operador em questão, independentemente de a operação que o envolve ter sido a última a ocorrer. No entanto, nos casos em que o operador não foi o último a ser aplicado, o termo aparece em itálico.

Conforme detalhado no item (3.2), os processos de inovação lexical, anteriormente apresentados, que não estejam de acordo com os critérios selecionados para delimitação do objeto desta pesquisa não serão tratados.

É preciso ressaltar que todos os termos⁷⁰ bem como os afixos foram consultados nos dicionários Aurélio (FERREIRA, 2004) e Houaiss (HOUAISS & VILLAR, 2001), ambos da última edição, bem como em dois manuais de gramática, a saber: Rocha Lima (2000) e Cunha & Cintra (2001).

Derivação prefixal - Nanociência/Nanotecnologia		
Prefixos	Significação	Ocorrências
co-/com-/con-	1 - concomitância ou reunião	COAXIAL (1) COPOLÍMERO (1) COPRECIPITAÇÃO (1) COVALÊNCIA (1)
de-	1- ato ou efeito de decair 2- consumir, desfazer	DECAIMENTO (1) DECOMPOR (2)

⁷⁰ Apenas aqueles já dicionarizados.

	3- mudar	DECOMPOSIÇÃO (2) DEFLOCULANTE (2) DEFORMAÇÃO (3) DEGRADAÇÃO (2)
des-	1- separação 2- ação contrária, negação	DESAGLOMERADA (1) DESCARGA (2) DESCONTAMINAÇÃO (2) DESIDROGENAÇÃO (2) DESPOLUIÇÃO (2) DESSALINIZAÇÃO (2) DESSORÇÃO (2) DESSULFATAÇÃO (2)
di(s)-	1- separação; dispersão; 2- negação	DIFUSÃO (1) DISSOLUÇÃO (1) DISSIPACÃO (1) DISSIPADOR (1) DISSOCIAÇÃO (1) DISSOLVIDO (1) DISTRORÇÃO (2)
im-/in-	1- movimento para dentro	INCISÃO (1) INCRUSTAÇÃO (1) INCUBAÇÃO (1) INFUSÃO (1)
in-	1- negação	INORGÂNICO (1) INOXIDÁVEL (1) INSOLÚVEL (1) INVISIBILIDADE (1) INVISÍVEL (1)
inter-	1- entre dois ou mais	INTERAÇÃO (1) INTERAGENTE (1) INTERAGIR (1) INTERCONEXÃO (1) INTERDIFUSÃO (1) INTERESPAÇO (1) INTERFACE (1) INTERGRANULAR (1) INTERMETÁLICA (1) INTERMOLECULAR (1) INTERNUCLEAR (1) INTERPLANAR (1)
micro-	1- anteposto ao nome duma unidade de medida, forma o nome de uma unidade derivada um milhão de vezes menor que a primeira	MICRÔMETRO (1) MICRONANO (1)
nano-	1- anteposto ao nome duma unidade de medida, forma o	NANOGRAMA (1) NANOLITRO (1)

	nome de uma unidade derivada um milésimo milionésimo de vezes menor que a primeira	NANÔMETRO (1) NANOPEAGÂMETRO (1)
per-	1- em Química, exprime 'o mais alto grau de oxidação' ou 'excesso'	<i>PEROXIDAÇÃO</i> (1) PERÓXIDO (1)
pre-	1- anterioridade 2- designativos de períodos, fases, estágios.	PRÉ-CANCEROSO (2) PREFORMA (1) PRÉ-INCUBAÇÃO (1) PREPOLÍMERO (2) PRÉ-TRATAMENTO (1)
re-	1- repetição; 2- reciprocidade, decorrência	REAÇÃO (1) <i>RECOBRIMENTO</i> (1) <i>RECOMBINAÇÃO</i> (1) <i>RECOZIMENTO</i> (1) <i>RECRISTALIZAÇÃO</i> (1) <i>REPLICAÇÃO</i> (2) <i>REPRODUTIBILIDADE</i> (2) <i>RESFRIAMENTO</i> (2)
semi-	1- Redução	SEMICONDUTOR (1) SEMIMETÁLICO (1)
super-/sobre-	1- posição acima 2- abundância, excesso, grande quantidade,	<i>SOBRENADANTE</i> (1) <i>SOBREPOSIÇÃO</i> (1) <i>SUPERABSORVENTE</i> (2) <i>SUPERAQUECIMENTO</i> (2) <i>SUPERCONDUTIVIDADE</i> (2) SUPERCONDUTOR (2) SUPERCRÍTICO (2) SUPERGEL (2) SUPERLENTE (2) SUPERMOLÉCULA (2) <i>SUPERMOLECULAR</i> (2) SUPERPARAMAGNÉTICO (2) SUPERPLÁSTICO (2) SUPERPOSIÇÃO (1) <i>SUPERSATURAÇÃO</i> (2) <i>SUPERSATURADO</i> (2) SUPERTUBO (2)
sub-/sob-	1- menor que, tamanho inferior a	SUBATÔMICO (1) SUBMICROMÉTRICO (1) SUBNANOMÉTRICO (1)
trans-	1- referente a mudança ou	<i>TRANSFORMAÇÃO</i> (1)

	item modificado	<i>TRANSFORMADOR</i> (1) <i>TRANSGÊNICO</i> (1) <i>TRANSIMPEDÂNCIA</i> (1) <i>TRANSPASSIVA</i> (1)
ultra-	1- superior a referência normal	<i>ULTRACAPACITOR</i> (1) <i>ULTRAFINO</i> (1) <i>ULTRANANOCRISTALINO</i> (1) <i>ULTRASSOM</i> (1) <i>ULTRASSÔNICO</i> (1) <i>ULTRAVIOLETA</i> (1)

Tabela 5: Derivação prefixal - Nanociência/Nanotecnologia

Derivação sufixal – Nanociência/Nanotecnologia		
Sufixos	Significação	Ocorrências
-(d/t/s/ss)or (a)	1- agente, instrumento de ação	<i>ABSORVEDOR</i> (1) <i>ACELERADOR</i> (1) <i>ANTICONTAMINADOR</i> (1) <i>ASPIRADOR</i> (1) <i>ATUADOR</i> (1) <i>BIOSENSOR</i> (1) <i>CARREADOR</i> (1) <i>CATALISADOR</i> (1) <i>CHAVEADOR</i> (1) <i>COMUTADOR</i> (1) <i>CONDENSADOR</i> (1) <i>CONDICIONADOR</i> (1) <i>DISSIPADOR</i> (1) <i>ELETROCATALISADOR</i> (1) <i>ESPALHADOR</i> (1) <i>FOTOALINHADOR</i> (1) <i>FOTOCATALISADOR</i> (1) <i>FOTODETECTOR</i> (1) <i>FOTOGERADOR</i> (1) <i>INIBIDOR</i> (1) <i>ISOLADOR</i> (1) <i>MANIPULADOR</i> (1) <i>MICROAQUECEDOR</i> (1) <i>MICROPROCESSADOR</i> (1) <i>MONOCROMADOR</i> (1) <i>NANOATUADOR</i> (1) <i>NANOCATALISADOR</i> (1) <i>NANOELEVADOR</i> (1) <i>NANOINDENTADOR</i> (1) <i>NANOMANIPULADOR</i> (1)

		<i>NANOMARCADOR</i> (1) <i>NANOOSCILADOR</i> (1) OSCILADOR (1) POSICIONADOR (1) PRECIPITADOR (1) PROCESSADOR (1) RESONADOR (1) RESTAURADOR (1) SECADOR (1) SIMULADOR (1) SINALIZADOR (1) TRANSFORMADOR (1)
-dade	1- qualidade, propriedade	BIOCOMPATIBILIDADE (1) BIODISPONIBILIDADE (1) BIODIVERSIDADE (1) CITOTOXICIDADE (1) CONDUTIBILIDADE (1) CONDUTIVIDADE (1) CRISTALINIDADE (1) CROMATICIDADE (1) DUCTILIDADE (1) ELASTICIDADE (1) FLEXIBILIDADE (1) HETEROGENEIDADE (1) HIDROFOBICIDADE (1) HOMOGENEIDADE (1) INCOMBUSTIBILIDADE (1) INVISIBILIDADE (1) LONGEVIDADE (1) LUMINOSIDADE (1) MOBILIDADE (1) PERMEABILIDADE (1) POLARIDADE (1) POROSIDADE (1) PRODUTIVIDADE (1) QUIRALIDADE (1) RADIOATIVIDADE (1) REATIVIDADE (1) RELATIVIDADE (1) REPRODUTIBILIDADE (1) RESISTIVIDADE (1) RUGOSIDADE (1) SALINIDADE (1) SENSIBILIDADE (1) SINTERABILIDADE (1) SOLUBILIDADE (1) <i>SUPERCONDUTIVIDADE</i> (1) SUSCETIBILIDADE (1)

		TENACIDADE (1) TETRAGONALIDADE (1) TOXICIDADE (1) VISCOSIDADE (1)
-al	1- relação, pertinência; 2- coleção, quantidade	AXIAL (1) COAXIAL (1) COLOIDAL (1) EQUIAXIAL (1) ESPECTRAL (1) ESPIRAL (1) HELICOIDAL (1) HIDROCOLOIDAL (1) HIDROTERMAL (1) ICOSAEDRAL (1) INTERSTICIAL (1) MICROESTRUTURAL (1) MITOCONDRIAL (1) MONOCLONAL (2) ORBITAL (1) PERIODONTAL (1) SENOIDAL (1) SPECTRAL (1) SPINODAL (1) TETRAGONAL (1) TRANSVERSAL (2) TRIDIMENSIONAL (2) UNIAXIAL (2) UNIDIRECIONAL (2)
-ico	1- pertinência, relação, referência;	AGROQUÍMICO (1) ANÓDICO (1) ANTINEOPLÁSICO (1) ATÔMICO (1) AUTOTRÓFICO (1) BÁSICO (1) BIOTECNOLÓGICO (1) CARBOXÍLICO (1) CATIÔNICO (1) CATÓDICO (1) DIELÉTRICO (1) ELÁSTICO (1) ELASTOMÉRICO (1) ELÉTRICO (1) ELETROCALÓRICO (1) ELETROCRÔMICO (1) ENDOTÉRMICO (1) EXOTÉRMICO (1) FARMOQUÍMICO (1)

		<p>FENÓLICO (1) FERROELÁSTICO (1) FERROELÉTRICO (1) FOTOCRÔMICO (1) HETEROTRÓFICO (1) HIDROFÍLICO (1) IÔNICO (1) ISOPROPÍLICO (1) ISOTÉRMICO (1) MACROSCÓPICO (1) MARTENSÍTICO (1) MICROFLUÍDICO (1) POLIMÉRICO (1) POLIOLEFÍNICO (1) POLIVINÍLICO (1) PROTEICO (1) QUÂNTICO (1) REOLÓGICO (1) ROBÓTICO (1) SIMÉTRICO (1) SINÉRGICO (1) SUBATÔMICO (1) TAXONÔMICO (1) TECNOLÓGICO (1) TÉRMICO (1) TETRAÉDRICO (1) TRANSGÊNICO (1) ULTRASSÔNICO (1)</p>
-ito	1- (Química) indica ânion derivado de oxiácido de elemento que está em um de seus estados de oxidação mais baixos (como <i>sulfito</i> , p. ex.); 2) designa termos da química inorgânica (como <i>nitrito</i> , p. ex.)	<p>NITRITO (1) SULFITO (1)</p>
-ica	1- técnica, ciência	<p>ATOMÍSTICA (1) AUSTENÍTICA (1) BIOMIMÉTICA (1) BIOSSINTÉTICA (1) CINÉTICA (1) CRIOGÊNICA (1) DENTÍSTICA (1) GENÉTICA (1) GENÔMICA (1) ROBÓTICA (1) SPINTRÔNICA (1)</p>
-oso(a)	1- pertinência, relação,	ARGILOSO (1)

	referência;	<p>ENDOVENOSA (1) GASOSA (1) MESOPOROSO (1) NANOPOROSO (1) POROSO (1) PRÉ-CANCEROSO (1)</p>
-ano(a)	1- relativo a	<p>AMINOPROPILTRIETOXISILANO (1) <i>ANTIMICROBIANO</i> (1) BROWNIANO (1) CATENANO (1) MICROBIANO (1) ORGANOSILANO (1) POLISSILANO (1)</p>
-(ç/s/ss)ão	ação ou resultado dela	<p>ACELERAÇÃO (1) ACIDIFICAÇÃO (1) ADSORÇÃO-DESSORÇÃO (1) AGLOMERAÇÃO (1) AGLUTINAÇÃO (1) AMORFIZAÇÃO (1) ANODIZAÇÃO (1) APASSIVAÇÃO (1) ATOMIZAÇÃO (1) ATRAÇÃO (1) <i>AUTOFORMAÇÃO</i> (1) AUTORREPLICAÇÃO (1) BIOAPLICAÇÃO (1) BIODEGRADAÇÃO (1) BIOINTEGRAÇÃO (1) BIORREMEDIAÇÃO (1) CALCINAÇÃO (1) CALIBRAÇÃO (1) CAPACITAÇÃO (1) CARACTERIZAÇÃO (1) CENTRIFUGAÇÃO (1) CITOLocalização (1) COMPACTAÇÃO (1) CONDENSAÇÃO (1) CONDUÇÃO (1) COPRECIPITAÇÃO (1) CRISTALIZAÇÃO (1) DEFORMAÇÃO (1) DEGRADAÇÃO (1) DENSIFICAÇÃO (1) DESCONTAMINAÇÃO (1) DESIDROGENAÇÃO (1) DESPOLUIÇÃO (1) DESSALINIZAÇÃO (1) DESSULFATAÇÃO (1)</p>

		DESTILAÇÃO (1) DISCRETIZAÇÃO (1) DISSIPACÃO (1) DISSOCIAÇÃO (1) DISTORÇÃO (1) EBULIÇÃO (1) ELETRODEPOSIÇÃO (1) <i>ELETROFORMAÇÃO</i> (1) EMBOLIZAÇÃO (1) ESTERIFICAÇÃO (1) EVAPORAÇÃO (1) EXTRUSÃO (1) FABRICAÇÃO (1) FERMENTAÇÃO (1) FILTRAÇÃO (1) FIXAÇÃO (1) <i>FOTODEGRADAÇÃO</i> (1) <i>FOTOGRAVAÇÃO</i> (1) FRAGMENTAÇÃO (1) GALVANIZAÇÃO (1) GRANULAÇÃO (1) HIBRIDIZAÇÃO (1) HIDRODESSULFURIZAÇÃO (1) HIDROGENAÇÃO (1) HIDROISOMERIZAÇÃO (1) HOMOGENEIZAÇÃO (1) INCRUSTAÇÃO (1) INCUBAÇÃO (1) INSTRUMENTAÇÃO (1) INTERCALAÇÃO (1) IRRADIAÇÃO (1) ISOMERIZAÇÃO (1) LAMINAÇÃO (1) LIGAÇÃO (1) LINEARIZAÇÃO (1) LIOFILIZAÇÃO (1) LIXIVIAÇÃO (1) LUBRIFICAÇÃO (1) MAGNETIZAÇÃO (1) MANIPULAÇÃO (1) MATERIALIZAÇÃO (1) MEDIÇÃO (1) <i>MICROFABRICAÇÃO</i> (1) <i>MICROINCISÃO</i> (1) MINERAÇÃO (1) MINIATURIZAÇÃO (1) MODIFICAÇÃO (1) MODULAÇÃO (1)
--	--	---

		MUTAÇÃO (1) NANOCARACTERIZAÇÃO (1) NANOCARACTERIZAÇÃO (1) NANODISPERSÃO (1) NANOEMULSÃO (1) NANOENCAPSULAÇÃO (1) NANOESTRUTURAÇÃO (1) NANOEXPOSIÇÃO (1) NANOFABRICAÇÃO (1) NANOFILTRAÇÃO (1) NANOIMPRESSÃO (1) NANOMANIPULAÇÃO (1) NANOSSEGREGAÇÃO (1) NEBULIZAÇÃO (1) NITRETAÇÃO (1) NUCLEAÇÃO (1) OSCILAÇÃO (1) OSSEOCONDUÇÃO (1) OSSEOINTEGRAÇÃO (1) OXICLORAÇÃO (1) OXIDAÇÃO (1) PARAMETRIZAÇÃO (1) PASSIVAÇÃO (1) PENETRAÇÃO (1) PERFURAÇÃO (1) PERMEAÇÃO (1) PEROXIDAÇÃO (1) POLARIZAÇÃO (1) POLIESTERIFICAÇÃO (1) POLIMERIZAÇÃO (1) PRECIPITAÇÃO (1) PRÉ-INCUBAÇÃO (1) PROPAGAÇÃO (1) PROTONAÇÃO (1) PULVERIZAÇÃO (1) PURIFICAÇÃO (1) QUANTIFICAÇÃO (1) QUANTIZAÇÃO (1) RADIAÇÃO (1) RAMIFICAÇÃO (1) RECOMBINAÇÃO (1) RECRISTALIZAÇÃO (1) REFRIGERAÇÃO (1) REGENERAÇÃO (1) REGRESSÃO (1) RELAXAÇÃO (1) REMOÇÃO (1) REPLICAÇÃO (1)
--	--	---

		REPULSÃO (1) RETENÇÃO (1) RETICULAÇÃO (1) ROTAÇÃO (1) SATURAÇÃO (1) SEDIMENTAÇÃO (1) SEGREGAÇÃO (1) SIMULAÇÃO (1) SINTERIZAÇÃO (1) SOBREPOSIÇÃO (1) SOLIDIFICAÇÃO (1) SOLUBILIZAÇÃO (1) SUPERSATURAÇÃO (1) SUSTENTAÇÃO (1) TERMORREDUÇÃO (1) TRANSFORMAÇÃO (1) VAPORIZAÇÃO (1) VIBRAÇÃO (1)
-nte	1- ação característica ou própria ou agente de ação	ABSORVENTE (1) ANTILIGANTE (1) ANTIOXIDANTE (1) AUTOLIMPANTE (1) AUTORREPLICANTE (1) CORANTE (1) CORRENTE (1) DEFLOCULANTE (1) DOPANTE (1) EFLUENTE (1) ENCAPSULANTE (1) FERTILIZANTE (1) FLUORESCENTE (1) FOTOCORRENTE (1) FOTOLUMINESCENTE (1) INTERAGENTE (1) IONIZANTE (1) IONOCOVALENTE (1) ISOLANTE (1) LIGANTE (1) LUBRIFICANTE (1) LUMINESCENTE (1) NANOCOMPONENTE (1) OXIDANTE (1) PLASTIFICANTE (1) REAGENTE (1) RESISTENTE (1) SOBRENADANTE (1) SOLVENTE (1) SUPERABSORVENTE (1)

		SURFACTANTE (1)
-mento	1- ação ou resultado dela	ACOPLAMENTO (1) AFUNDAMENTO (1) ALARGAMENTO (1) APAGAMENTO (1) APRISIONAMENTO (1) AQUECIMENTO (1) ARMADILHAMENTO (1) BIOPROCESSAMENTO (1) BOMBEAMENTO (1) CHAVEAMENTO (1) COALESCIMENTO (1) CONFINAMENTO (1) CONGELAMENTO (1) DECAIMENTO (1) DERRAMAMENTO (1) DESLIZAMENTO (1) DESLOCAMENTO (1) DIMENSIONAMENTO (1) EMBAÇAMENTO (1) EMPACOTAMENTO (1) ENCADEAMENTO (1) ENCAPSULAMENTO (1) ESCANEAMENTO (1) ESCOAMENTO (1) ESFRIAMENTO (1) ESPAÇAMENTO (1) ESPALHAMENTO (1) ESTIRAMENTO (1) EXCITAMENTO (1) EXPERIMENTO (1) FECHAMENTO (1) GRAMPEAMENTO (1) HIDROTRATAMENTO (1) IMAGEAMENTO (1) INTUMESCIMENTO (1) ISOLAMENTO (1) JATEAMENTO (1) MAPEAMENTO (1) <i>MICRODESLOCAMENTO</i> (1) MODELAMENTO (1) MONITORAMENTO (1) POLIMENTO (1) <i>PRÉ-TRATAMENTO</i> (1) PROCEDIMENTO (1) PROCESSAMENTO (1) RASTREAMENTO (1) RECOBRIMENTO (1)

		RECOZIMENTO (1) RESFRIAMENTO (1) REVESTIMENTO (1) SENSORIAMENTO (1) SEQUENCIAMENTO (1) <i>SUPERAQUECIMENTO</i> (1) TRAVAMENTO (1) TUNELAMENTO (1) VAZAMENTO (1)
-agem	1- ação ou resultado de ação;	AUTOMONTAGEM (1) CICLAGEM (1) CLONAGEM (1) DECAPAGEM (1) DOPAGEM (1) DOSAGEM (1) ESTOCAGEM (1) FILTRAGEM (1) <i>MICROUSINAGEM</i> (1) MOAGEM (1) MODELAGEM (1) MOLDAGEM (1) MONTAGEM (1) PROTOTIPAGEM (1) SECAGEM (1) <i>TELECLONAGEM</i> (1) <i>TERMOFORMAGEM</i> (1) USINAGEM (1)
-ita	1- elemento químico	AUSTENITA (1) BAINITA (1) BENTONITA (1) CASSITERITA (1) FERRIDRITA (1) FERRITA (1) HIDROXIAPATITA (1) MAGNETITA (1) MANGANITA (1) MARTENSITA (1) MONTMORILONITA (1) PEROVISQUITA (1) VERMICULITA (1) VOLASTONITA (1) ZEÓLITA (1)

-ato	1- Utilizado no domínio da química para identificar um elemento como sal	ALGINATO (1) ALUMINATO (1) CARBONATO (1) CITRATO (1) ESTEARATO (1) FOSFATO (1) HIDRATO (1) NITRATO (1) OXALATO (1) POLIACRILATO (1) SILICATO (1) TITANATO (1)
-ina	1 - substancia química	ACTINA (1) ADENINA (1) ADENOSINA (1) AMINA (1) BIPIRIDINA (1) CICLODEXTRINA (1) CITOSINA (1) DENTINA (1) FERROPORFIRINA (1) FIBRINA (1) HEMOGLOBINA (1) MIOSINA (1) MUCINA (1) POLIANILINA (1) PORFIRINA (1) RUTÊNIO-BIPIRIDINA (1)
-ivo	1- que apresenta (certa característica)	ABRASIVO (1) DISRUPTIVO (1) OXIDATIVO (1) REFRATIVO (1)
-ez(a)	1- denominador de propriedade/caraterística	ASPEREZA (1) DUREZA (1) MICRODUREZA (1) NANODUREZA (1) PUREZA (1)
-ose	1- denominador de elemento químico	APOPTOSE (1) CARBOXIMETILCELULOSE (1) CELULOSE (1) HIDROXIPROPILMETILCELULOSE (1) QUERATOSE (1)
-vel	1- passível de sofrer ou praticar determinada ação	FLEXÍVEL (1) INOXIDÁVEL (1) INVISÍVEL (1) METAESTÁVEL (1)

		SOLÚVEL (1) BIOCOMPATÍVEL (1) BIODEGRADÁVEL (1) BIORREABSORVÍVEL (1) BIOCOMBUSTÍVEL (1)
-ar	1- característica/propriedade	BIPOLAR (1) DIPOLAR (1) INTERGRANULAR (1) INTERMOLECULAR (1) INTERNUCLEAR (1) INTERPLANAR (1) MOLAR (1) MOLECULAR (1) NUCLEAR (1) SOLAR (1) SUPERMOLECULAR (1) SUPRAMOLECULAR (1) TEGUMENTAR (1) TURBOMOLECULAR (1)

Tabela 6: Derivação sufixal – Nanociência/Nanotecnologia

Composição culta – Nanociência/Nanotecnologia		
Elementos de composição	Significação	Ocorrências
fito-	1- relativo a vegetal ou planta	FITOPLÂNCTON (1) FITOQUÍMICA (1)
foto-	1 – relativo a luz	FOTOALINHADOR (1) FOTOCATALISADOR (1) FOTOCATÁLISE (1) FOTOCORRENTE (1) FOTOCRÔMICO (1) FOTOCROMISMO (1) FOTODEGRADAÇÃO (1) FOTODETECTOR (1) FOTODINÂMICA (1) FOTODIODO (1) FOTOELÉTRONS (1) FOTOELETROQUÍMICA (1) FOTOESTABILIDADE (1) FOTOFÍSICA (1) FOTOGERADOR (1) FOTOGRAFAR (1) FOTOGRAFIA (1) FOTOGRAVAÇÃO (1) FOTOGRAVADO (1) FOTOINDUZIDA (1)

		FOTOINTERRUPTOR (1) FOTOLITOGRAFIA (1) FOTOLUMINESCÊNCIA (1) FOTOLUMINESCENTE (1) FOTOMICROGRAFIA (1)
hidro-	1- relativo a água	HIDROCARBONETO (1) HIDROCLORETO (1) HIDRODESSULFURIZAÇÃO (1) HIDROFÍLICO (1) HIDRÓFILO (1) HIDROFOBICIDADE (1) HIDROFÓBICO (1) HIDROGEL (1) HIDROGENAÇÃO (1) HIDROGÊNIO (1) HIDROISOMERIZAÇÃO (1) HIDRÓLISE (1) HIDROQUINONA (1) HIDROTERMAL (1) HIDROTRATAMENTO (1) HIDRÓXIDO (1)
geo-	1- Relativo a Terra	GEOBACTÉRIA (1) GEOMAGNÉTICA (1)
micro-	1- indica ordem ou grandeza reduzida	MICROAERONAVE (1) MICROAGULHA (1) MICROANÁLISE (1) MICROAQUECEDOR (1) MICROBIANO (1) MICRÓBIO (1) MICROBOMBA (1) MICROCANAL (1) MICROCÁPSULA (1) MICROCAVIDADE (1) MICROCÉLULA (1) MICROCHIP (1) MICROCIRCUITO (1) MICROCONTATO (1) MICRODESLOCAMENTO (1) MICRODISPOSITIVO (1) MICRODUREZA (1) MICROELETRODOS (1) MICROELETRÔNICA (1) MICROENERGIA (1) MICROESCALA (1) MICROESFERA (1) MICROESTRUTURA (1) MICROESTRUTURAL (1) MICROFABRICAÇÃO (1)

		<p>MICROFABRICADO (1) MICROFITOPLÂNCTON (1) MICROFLUÍDICO (1) MICROFOTOLUMINESCÊNCIA (1) MICROGRAFIA (1) MICROGRAVIDADE (1) MICROINCISÃO (1) MICROLABORATÓRIO (1) MICROMÁQUINA (1) MICROMATERIAL (1) MICROMÉTRICO (1) MICROONDA (1) MICROÓPTICA (1) MICROORGANISMO (1) MICROPARTÍCULA (1) MICROPEÇA (1) MICROPENEIRA (1) MICROPLÂNCTON (1) MICROPORO (1) MICROPROCESSADOR (1) MICROQUANTIDADE (1) MICROQUÍMICO (1) MICROROBÔ (1) MICROSCOPIA (1) MICROSCÓPICO (1) MICROSCÓPIO (1) MICROSSISTEMA (1) MICROSSONDA (1) MICROTECNOLOGIA (1) MICROTÚBULO (1) MICROUSINAGEM (1) MICROVIGA (1)</p>
poli-	1- indica muitos ou diversos	<p>POLIACRILATO (1) POLIAMIDA (1) POLIANILINA (1) POLICETONA (1) POLICRISTALINO (1) POLIDISPERSO (1) POLIELETRÓLITO (1) POLIÉSTER (1) POLIESTERIFICAÇÃO (1) POLIESTIRENO (1) POLIETILENO (1) POLIETILENO-AMIDO (1) POLIETILENOGLICOL (1) POLIÓIS (1) POLIPROPILENO (1) POLISSACARÍDEO (1)</p>

		POLISSILANO (1) POLISULFONA (1) POLIURETÂNICA (1) POLIVINÍLICO (1)
macro-	1- indica ordem ou grandeza maior	MACROCÍCLICO (1) MACROCROPORO (1) MACROESTRUTURA (1) MACROMOLÉCULA (1) MACROMOLECULAR (1) MACROSCÓPICO (1)
agro-	1- relativo a prática agrícola	AGROENERGIA (1) AGROQUÍMICO (1)
auto-	1- por si próprio, de si mesmo;	AUTOFORMAÇÃO (1) AUTOLIMPANTE (1) AUTOMONTAGEM (1) AUTOREPLICAÇÃO (1) AUTOREPLICANTE (1) AUTOTRÓFICO (1)
cit(o)-	1- relativo a célula	CITOLocalização (1) CITOTOXICIDADE (1) CITOTÓXICO (1)
ele(c)tr(o)-	1- relativo a eletricidade	ELETROCALÓRICO (1) ELETROCATALISADOR (1) ELETROCATÁLISE (1) ELETROCRÔMICO (1) ELETROCROMISMO (1) ELETRODEPOSIÇÃO (1) ELETRODÍCO (1) ELETROFORMAÇÃO (1) ELETROLÍSE (1) <i>ELETROLÍTICA</i> (1) <i>ELETROLITO</i> (1) ELETROLUMINESCÊNCIA (1) ELETROMAGNÉTICA (1) ELETROMECAÂNICO (1) ELETRÔNICA (1) ELETROÓPTICO (1) ELETROQUÍMICA (1) ELETROSTÁTICA (1)
espectr(o)-	1- relativo a imagem	ESPECTRAL (1) <i>ESPECTROFOTOMETRIA</i> (1) <i>ESPECTROFOTOMÉTRICA</i> (1) ESPECTROFOTÔMETRO (1) ESPECTROMETRIA (1) ESPECTRÔMETRO (1) ESPECTROSCOPIA (1) <i>ESPECTROSCÓPICA</i> (1)

ferr(i)(o)-	1- indica presença ou reação a ferro	FERRIDRITA (1) FERRITA (1) FERRÍTICA (1) FERRITINA (1) FERROCENO (1) FERROELÁSTICO (1) FERROELÉTRICO (1) FERROFLUIDO (1)
fluor(o)-	1- indica presença ou reação a flúor	FLUORCARBONO (1) FLUORESCÊNCIA (1) FLUORESCENTE (1) FLUORÓFORO (1) FLUOROPOLÍMERO (1)
hemo-	1- relativo a sangue	HEMOCOMPATIBILIDADE (1) HEMOGLOBINA (1)
heter(o)-	1- diferente	HETEROJUNÇÃO (1) HETEROTRÓFICO (1)
hom(o)-	1- semelhante	HOMÓDINO (1) HOMOPOLÍMERO (1)
imun(o)-	1- livre de, isento	IMUNONANOCONCHA (1) IMUNOSSUPRIMIDOS (1)
mon(o)-	1- único	MONOCAMADA (1) MONOCLONAL (1) MONOCRISTAL (1) MONOCRISTALINO (1) MONODISPERSO (1) MONOFÁSICO (1) MONÔMERO (1) MONOMODO (1) MONÓXIDO (1)
mult(i)-	1- numeroso	MULTICAMADA (1) MULTIELETRÔNICO (1) MULTIESCALA (1) MULTIFERRÓICO (1) MULTIFOTÔNICA (1)
nan(o)-	1-"antepositivo, do gr. nánnos, ou 'anão', ocorre em cultismos do século XIX em diante, como nanico, nanismo, nanocefalia (...)" (HOUAISS & VILLAR, 2001)	NANOADESIVO (1) NANOAGREGADO (1) NANOANEL (1) NANOANTENA (1) NANOATUADOR (1) NANOBASTÃO (1) NANOBIOSSENSOR (1) NANOCÁPSULA (1) NANOCARRO (1) NANOCATALISADOR (1)

		<p> NANOCÉLULA (1) NANOCHAVE (1) NANOCINTO (1) NANOCIRCUITO (1) NANOCOMPLEXO (1) NANOCOMPÓSITO (1) NANOCOMPOSTO (1) NANOCONCHA (1) NANOCRISTAL (1) NANODIAMANTE (1) NANODISPOSITIVO (1) NANODUREZA (1) NANOELETRÓDO (1) NANO ELEVADOR (1) NANOEMULSÃO (1) NANOENCAPSULAÇÃO (1) NANOESCALA (1) NANOESFERA (1) NANOESPIRAL (1) NANOESPUMA (1) NANOESTRUTURA (1) NANOESTRUTURAÇÃO (1) NANOFABRICAÇÃO (1) NANOFIBRA (1) NANOFILME (1) NANOFILTRAÇÃO (1) NANOFIO (1) NANOFITA (1) NANOFÓSSIL (1) NANOFOTODIODO (1) NANOFÓTON (1) NANO GEL (1) NANO GUITARRA (1) NANO HÉLICE (1) NANOIMÃ (1) NANOIMPLANTE (1) NANOINDENTADOR (1) NANOJUNÇÃO (1) NANOLAMELA (1) NANOLÁTEX (1) NANOMÁQUINA (1) NANOMARCADOR (1) NANOMATERIAL (1) NANOMECANISMO (1) NANOMETAL (1) NANOMOLA (1) NANOMOTOR (1) NANO OBJETO (1) </p>
--	--	---

		NANOOSCILADOR (1) NANOPARTÍCULA (1) NANOPLÂNCTON (1) NANOPÓ (1) NANOPOLÍMERO (1) NANOPORO (1) NANOPRODUTO (1) NANORROBÔ (1) NANORROLO (1) NANOSCOPIA (1) NANOSCÓPIO (1) NANOSSENSOR (1) NANOSSILÍCIO (1) NANOSSONDA (1) NANOTECIDO (1) NANOTUBO (1) NANOVÁLVULA (1)
organo-	1- indica origem orgânica	ORGÂNICO-INORGÂNICO (1) ORGANOMETÁLICO (1)
pir(o)-	1- relacionado a fogo; ou calor	PIROCLORO (1) PIROELÉTRICO (1) PIRÓLISE (1)
tele-	1- longe, ao longe	TELECLONAGEM (1) TELESCÓPIO (1) TELETRANSPORTE (1)
term(o)-	1- relativo a calor	TERMODINÂMICA (1) TERMOFORMAGEM (1) TERMOGRAVIMETRIA (1) TERMOPLÁSTICO (1) TERMORREDUÇÃO (1)
top(o)-	1- relativo a lugar	TOPOGRAFIA (1) TOPOLOGIA (1)

Tabela 7: Composição culta – Nanociência/Nanotecnologia

Abre-se exclusivamente para o elemento *nano-* o precedente de demonstrar quais são os termos formados pelo processo de recomposição, justifica-se esta exceção pela grande quantidade de termos formados desta maneira e pelo alto índice de ocorrência no corpus desses termos.

Recomposição com “nano-”		
Elemento de	Significação	Ocorrências

Recomposição		
nano-	1- “relativo à nanociência / à nanotecnologia”, sendo essa relação de vários tipos (“produzido por”, “com técnicas próprias de”, etc.)	NANOBIOCIÊNCIA (1) NANOBIOLOGIA (1) NANOBIOFOTÔNICA (1) NANOBIOTECNOLOGIA (1) NANOCOSMÉTICO (1) NANOELETRÔNICA (1) NANOENGENHARIA (1) NANOÉTICA (1) NANOFOTÔNICA (1) NANOGLICBIOTECNOLOGIA (1) NANOIMPRESSÃO (1) NANOLITOGRAFIA (1) NANOMAGNETISMO (1) NANOMANIPULAÇÃO (1) NANOMEDICINA (1) NANOMETROLOGIA (1) NANOPATENTE (1) NANOQUÍMICA (1) NANOCHARACTERIZAÇÃO (1) NANOBIOESTRUTURA (1)

Tabela 8: Casos de recomposição com nano- na terminologia de N&N

4.2 Descrição morfológica da terminologia de Biocombustíveis

Assim como feito para a terminologia de Nanociência/Nanotecnologia, também a classificação proposta para os Biocombustíveis é apresentada por meio de tabelas, de acordo com cada processo de inovação morfológica. Novamente apresentam-se apenas os casos dos processos englobados pela pesquisa.

Derivação prefixal - Biocombustíveis		
Prefixos	Significação	Ocorrências
anti-	1- ação contrária	ANTICONGELANTE (1) ANTIDETONANTE (1)
des-	1- ação contrária	DESACIDIFICAR (1) DESCAROÇADEIRA (1) DESFIBRADOR (1) DESIDRATANTE (1) DESPROTEINIZAR (1) DESTOXIFICAÇÃO (1)

re-	1- repetição	REAÇÃO (1) REGASEIFICAÇÃO (1) REPLICABILIDADE (1)
trans-	1- mudança, transformação	TRANSESTERIFICAÇÃO (1) TRANSESTERIFICANTE (1) TRANSILUMINADOR (1)

Tabela 9: Derivação prefixal – Biocombustíveis

Derivação sufixal - Biocombustíveis		
Sufixos	Significação	Ocorrências
-(d)/(t)/(s)or(a)	1- agente ou instrumento de ação	AMOSTRADOR(A) (1) AQUECEDOR (1) BRANQUEADOR (1) CATALISADOR (1) CRAQUEADOR (1) DECANTADOR (1) DEGRADADOR(A) (1) DEGRADOR (1) DEFIBRADOR (1) ESCARIFICADOR (1) ESMAGADOR(A) (1) FOTOSENSIBILIZADOR (1) GASEIFICADOR (1) INCINERADOR (1) INOCULADOR (1) LAMINADOR(A) (1) PRECIPITADOR (1)
-dade	1- qualidade	BASICIDADE (1) CORROSIBILIDADE (1) CORROSIVIDADE (1) FITOTOXIDADE (1) HETEROCEASTICIDADE (1) HOMOCEASTICIDADE (1) OLEOSIDADE (1) PEDREGOSIDADE (1) PERECIBILIDADE (1) REPLICABILIDADE (1)
-al	1- cultura de vegetais	BABAÇUAL (1) DENDEZAL (1)
-ico	1- elemento químico	ETÍLICO (1) FENÓLICO (1) FENOLÓGICO (1) LINOLÊNICO (1) OLEICO (1)
-(ç/s/ss)ão	1- ação ou resultado dela	DESTOXIFICAÇÃO (1) ECONEUTRALIZAÇÃO (1)

		EVAPOTRANSPIRAÇÃO (1) FOSFORILAÇÃO (1) FOTODEGRADAÇÃO (1) FOTODIMERIZAÇÃO (1) FOTOXIDAÇÃO (1) REGASEIFICAÇÃO (1) SAPONIFICAÇÃO (1) TRANSESTERIFICAÇÃO (1)
-ase	1- elemento químico	CARBOXIPEPTIDASE (1) CATALASE (1) HIDROLASE (1) LIPASE (1) LIPOLASE (1) OXIGENASE (1) RIBONUCLEASE (1)
-nte	1- agente de ação	BIOCARBURANTE (1) DESIDRATANTE (1) EMULSIFICANTE (1) EMULSIONANTE (1) ESTABILIZANTE (1) GELIFICANTE (1) LIMIANTE (1) LUBRIFICANTE (1) TRANSILUMINANTE (1) VISCOSIFICANTE (1)
-mento	1- ação ou instrumento da ação	AMARELECIMENTO (1) BASCULAMENTO (1) BORBULHAMENTO (1) CARVOEJAMENTO (1) CISALHAMENTO (1) CRAQUEAMENTO (1) HIDROCRAQUEAMENTO (1) INCRUSTAMENTO (1) PROCESSAMENTO (1) RESFRIAMENTO (1) TURBILHONAMENTO (1)
-agem	1- ação	BRIQUETAGEM (1) CALAGEM (1) DEGOMAGEM (1)
-ato	1- sal químico	ACETATO (1) ESTEARATO (1) GLUCOSINOLATO (1) LINOLENATO (1) TIOSSULFATO (1)
-ina	1- substância química	ACROLEÍNA (1) DIOXINA (1) LECITINA (1) AMINA (1)

		METIONINA (1) BUTILAMINA (1) DINITROFENILHIDRAZINA (1) TRIFLURALINA (1) TRIGONELINA (1) FEOFITINA (1)
--	--	--

Tabela 10: Derivação sufixal – Biocombustíveis

Composição culta – Biocombustíveis		
elementos de composição	Significação	Ocorrências
Fit(o)-	1- relacionado a vegetal ou planta	FITOMASSA (1) FITORREMEDIO (1) FITOSSANITÁRIO (1) FITOTECNIA (1) FITOTOXIDADE (1)
Fot(o)-	1- relacionado a luz	<i>FOTODEGRADAÇÃO</i> (1) <i>FOTODIMERIZAÇÃO</i> (1) <i>FOTOSENSIBILIZADOR</i> (1) <i>FOTOXIDAÇÃO</i> (1)
Hex(a)-	1- indicador da quantidade “seis”	HEXADECANO (1) HEXAFLUORETO (1)
Hidr(o)-	1- indica a presença de átomo de hidrogênio ou água	HIDROCARBONETO (1) HIDROCICLONE (1) HIDROCLOROFLUORCARBONETO (1) HIDROCRAQUEAMENTO (1) HIDROFLUORCARBONO (1) HIDROGÊNIO (1) HIDROLASE (1) HIDRÓLISE (1) HIDROPEROXILA (1)
Ge(o)-	1- relacionado a terra	GEOMORFOL (1) GEOTERMAL (1)
Bi(o)-	1- relacionado a vida ou biológico	<i>BIOCARBURANTE</i> (1) <i>BIOCATALIZADOR</i> (1) BODIESEL (1) BIOENERGIA (1) BIOMASSA (1) BIOPLÁSTICO (1) BIOQUEROSENE (1)
Ricin(o)-	1- referente a ricino	RICINOCULTURA (1) RICINOQUÍMICA (1)
Micr(o)-	1- de tamanho reduzido	MICROEMULSÃO (1) MICROENCAPSULAR (1) MICRONUTRIENTE (1)

		MICROPROCESSAR (1) MICROTUBO (1)
poli-	1- indica a presença de muitos ou diversos	POLIACRILAMIDA (1) POLIAMINOESTIRENO (1) POLIESTIRENO (1) POLIETILENOGLICOL (1) POLIFENILPARAFENILENO (1) POLISSILOXANO (1)
Macr(o)-	1- de tamanho longo ou grande	MACROMOL (1) MACROPORO (1)

Tabela 11: Composição culta – Biocombustíveis

4.3 Resultados da descrição morfológica da terminologia de Nanociência/Nanotecnologia

A *análise inicial* da terminologia de Nanociência/Nanotecnologia revela, como pôde ser observado no item anterior, a predominância do processo de derivação sufixal, sobre as ocorrências de composição morfológica, esta em quantidade superior aos casos de derivação prefixal. As Tabelas 12, 13 e 14 apresentam esses dados detalhadamente.

-(ç/s/ss)ão	149
-mento	56
-ico	47
-(d)/(t)/(s)or(a)	42
-dade	40
-nte	31
-al	24
-agem	18
-ina	16
-ita	15
-ar	14
-ato	12
-ica	11
-vel	9
-ano(a)	7
-oso(a)	7
-ez(a)	5

-ose	5
-ivo	4
-ito	2
SUFIKAÇÃO	514

Tabela 12: Frequência total de derivação sufixal – Nanociência/Nanotecnologia cf. Tabela 6

nan(o)-	75
micro-	57
Foto	25
poli-	20
ele(c)tr(o)-	18
Hidro	16
mon(o)-	9
espectr(o)-	8
ferr(i)(o)-	8
auto-	6
macro-	6
term(o)-	5
fluor(o)-	5
mult(i)-	5
pir(o)-	3
cit(o)-	3
tele-	3
organo-	2
heter(o)-	2
hom(o)-	2
agro-	2
imun(o)-	2
fito	2
Geo	2
hemo-	2
top(o)-	2
COMPOSIÇÃO CULTA	290

Tabela 13: Frequência total de composição culta – Nanociência/Nanotecnologia cf. Tabela 7

super-/sobre-	17
---------------	----

inter-	12
des-	8
re-	8
di(s)-	7
de-	6
ultra-	6
trans-	5
in-	5
pre-	5
im-/in-	4
co-/com-/con-	4
nano-	4
sub-/sob-	3
per-	2
semi-	2
micro-	2
PREFIXAÇÃO	100

Tabela 14: Frequência total de derivação prefixal – Nanociência/Nanotecnologia cf. Tabela 6

Entre os casos de derivação sufixal, destaca-se a predominância dos sufixos em “-(ç/s/ss)ão”. No processo de composição morfológica, pode-se observar a predominância dos casos envolvendo o elemento de composição grego "nan(o)-". Já no tocante à derivação prefixal, o processo mais recorrente é aquele iniciado pelo prefixo "super/sobre-". Esses casos mais frequentes serão objeto de estudo da análise profunda, na seção 5.

4.4 Resultados da descrição morfológica da terminologia de Biocombustíveis

A *análise inicial* da terminologia de Biocombustíveis permitiu observar uma repetição quanto à frequência dos processos. Novamente se identificou a predominância dos processos de derivação sufixal, seguido da composição culta e, por fim, da derivação prefixal, como se pode observar nas tabelas 15, 16 e 17.

-(d)/(t)/(s)or(a)	17
-mento	11
-(ç/s/ss)ão	10
-dade	10
-ina	10
-nte	10
-ase	7
-ico	5
-ato	5
-agem	3
-al	2
SUFIXAÇÃO	90

Tabela 15: Frequência total de derivação sufixal – Biocombustíveis cf. Tabela 9

hidro	9
bio	7
poli	6
fito	5
foto	4
micro	2
hexa	2
geo	2
ricino	2
macro	2
COMPOSIÇÃO CULTA	41

Tabela 16: Frequência total de composição culta – Biocombustíveis cf. Tabela 10

des-	6
re	3
trans-	3
anti-	2
PREFIXAÇÃO	14

Entre os casos de sufixação, há ampla maioria de ocorrências em "-(d)/(t)/(s)or(a)". Já na composição morfológica, observa-se a predominância do elemento "hidro". O prefixo "des-" é o mais recorrente no campo da derivação prefixal.

5. Regras de Construção de Palavras: os resultados da análise profunda dos processos morfológicos mais recorrentes

Nesta seção, busca-se descrever alguns dos principais processos de construção dos termos das terminologias de Nanociência/Nanotecnologia e Biocombustíveis.

Partindo da análise classificatória e quantitativa anteriormente apresentada, serão tratados os casos dos sufixos, prefixos e elementos de composição morfológica de maior frequência nas terminologias analisadas.

Tendo em vista o modelo teórico adotado, nas análises seguintes, deter-se-á fundamentalmente nos seguintes aspectos (já elencados por CORREIA, 2004):

- breve historial do afixo ou elemento;
- tipos de base que seleciona, basicamente em função da sua origem, estrutura e denominação;
- tipos de significados apresentados pelos termos formados.

Antes de se iniciar as análises, retoma-se aqui a seguinte afirmação de Rocha (2008): “É preciso lembrar que ,gramática não é matemática”, ou seja, o que se deve levar em consideração são as tendências, as linhas gerais.” (ROCHA, 2008, p. 130). É a partir

dessa visão que se procede abaixo à análise dos dados majoritários, apontando direcionamentos para trabalhos futuros.

Além disso, busca-se tratar das condições de produtividade das RCPs e não das condições de produção. Em princípio, todas as palavras que satisfazem às exigências de uma RCP apresentam as condições de produtividade dessa regra e, por isso, deveriam servir de base para geração de novos produtos (palavras ou termos). No entanto, atender às condições de produtividade não garante a existência de produtos correspondentes para todas as bases possíveis, pois para algumas não existem condições de produção devido à incidência de restrições de ordem fonológica, morfológica, paradigmática, pragmática e/ou discursiva, além de bloqueios paradigmáticos, heterônimos, homofônicos e/ou parônimos e, ainda, da inércia morfológica – descritas no item 2.5.2.2 deste trabalho. Aqui importa descrever as condições de produtividade que viabilizam a inovação lexical, e não as condições de produção que muitas vezes limitam essa inovação, já que no vivo processo de enriquecimento lexical – especialmente terminológico – essas restrições podem oportunamente serem desconsideradas.

5.1 Derivação sufixal: “-ão” (Nanociência/Nanotecnologia)

Segundo Cunha & Cintra (2001), trata-se de um sufixo nominal, ou seja, forma substantivos a partir de verbos, atribuindo ao produto o sentido de “ação ou resultado dela” (CUNHA & CINTRA, 2001, p. 98).

Cunha & Cintra (2001) postulam que “os sufixos *-ção* e *-são* depreendem-se de substantivos deverbais, quase todos formados no próprio latim.” (CUNHA & CINTRA, 2001, p. 98).

O produto formado designa o nome da ação ou do resultado desta ação. São formações dicionarizadas, logo institucionalizadas, e recenseadas na terminologia em análise, tais como: *capacitação*, *evaporação*, *fabricação*, *modificação*, *mineração* e *saturação*, entre outros.

Para o estabelecimento da RCP, faz-se necessário a caracterização da base. Em primeiro lugar, deve-se verificar se todo e qualquer verbo pode vir a ser utilizado como base para aplicação da RCP em formulação.

As terminologias analisadas permitem constatar que é possível aplicar a regra tendo como base os verbos transitivos⁷¹ e intransitivos⁷², mas não os verbos impessoais⁷³ ou de ligação⁷⁴. No entanto, quanto à morfologia, não é possível estabelecer padrões de acordo com classes⁷⁵.

Dessa forma, retoma-se proposição de Rocha (2008):

“Conclui-se que a categorização da base, ou seja, a sua classificação em uma determinada categoria léxica, não é suficiente para definir o tipo de base específica da regra. Torna-se necessário, portanto, delimitar o tipo de substantivo [ou qualquer outra categoria], através de suas subcategorias. Essas subcategorias podem ser de qualquer natureza: fonética, morfológica, sintática, semântica, etc.” (ROCHA, 2008, p. 129).

Dessa forma, a base da RCP deverá ser um verbo que apresente os seguintes traços:

+ transitivo

+ intransitivo

- impessoal

⁷¹ Designam ações voluntárias, causadas por um ou mais indivíduos, e que afetam outro(s) indivíduo(s) ou alguma coisa, exigindo um ou mais objetos na ação.

⁷² Designam ações que não afetam outros indivíduos.

⁷³ São verbos que designam ações involuntárias. Geralmente (mas nem sempre) designam fenômenos da natureza e, portanto, não possuem sujeito nem objeto na oração.

⁷⁴ São os verbos que não designam ações; apenas servem para ligar o sujeito ao predicativo.

⁷⁵ Verbos regulares, irregulares, defectivos e/ou abundantes.

- ligação

Não se aplicam limitações quanto à sufixação de bases que tenham sido objeto de prefixação e/ou composição morfológica culminando na soma das diferentes significações do prefixo e sufixo ao elemento base, como se pode observar nos seguintes casos colhidos da terminologia analisada: *biodegradação*, *nanocaracterização*, *autoformação*, *descontaminação*, *eletrodeposição*, *microfabricação*, *recristalização*, *supersaturação*.

Ao atribuir ao produto o nome da ação ou do resultado desta ação, esta RCP seleciona como base os verbos que designam ações voluntárias ou involuntárias, ou seja, sejam elas i) praticadas por agentes, como em *caracterização*, *dessalinização*, *filtração*, *manipulação* e *nanoencapsulação* ou ii) simplesmente inevitáveis, ou seja, automáticas, como por exemplo a *biodegradação*, ou ainda a *autorreplicação*, cujo formante culto explicita uma ação involuntária.

Representação da regra:

$$V_{(T/I)} \rightarrow S_{-ção}$$

Em que:

V = verbo

T = transitivo

I = intransitivo

S = substantivo

-(ç/s/ss)ão = sufixo

5.2 Derivação sufixal: "-or" (Biocombustíveis)

Segundo Cunha & Cintra (2001), as consoantes “d”, “t” e “s” pertencem ao tema do participio latino. Assim, doravante, adotar-se-á apenas a forma “-or”.

Trata-se de um sufixo nominal que forma substantivos a partir de verbos, atribuindo ao produto o sentido de “agente, instrumento da ação” (CUNHA & CINTRA, 2001, p. 98)

do verbo de origem⁷⁶. Cunha & Cintra (2001) afirmam ainda que estas são formas eruditas, ou seja, “só ocorrem em palavras latinas ou formadas sobre o seu modelo.” (CUNHA E CINTRA, 2001, p. 98).

O produto formado designa o nome de um agente ou instrumento de ação. São formações institucionalizadas⁷⁷ (e recenseadas na terminologia em análise) do tipo: *catalisador, aquecedor, decantador, precipitador, aspensor, propulsor, esmagadora, laminadora e amostradora*.

A regra de formação de termos substantivos com o sufixo “-or” em que o produto denomina um agente ou instrumento de ação começa a ser estabelecida com a caracterização da base. Em primeiro lugar, deve-se verificar se todo e qualquer verbo pode servir de base à regra em questão.

A análise dos termos constantes do repertório terminológico em observação permite verificar que é possível aplicar a regra tendo como base os verbos transitivos, mas não os intransitivos, impessoais ou de ligação. No entanto, quanto à morfologia, não é possível estabelecer padrões de acordo com classes (cf. nota 72).

Dessa forma, a base da RCP deverá ser um verbo que apresente os seguintes traços:

+ transitivo

- intransitivo

- impessoal

- ligação

Não existem restrições à sufixação “-or” em bases verbais que tenham sido objeto de prefixação e/ou formação morfológica, como se observa nos casos seguintes (extraídos

⁷⁶ Não foram constadas quaisquer formas homófonas.

⁷⁷ Equivalente à dicionarizadas.

da terminologia em análise): *transiluminador*, *motogerador*, *fotosensibilizador*, *microaspersor* e *turbocompressor*.

Ao atribuir ao produto o significado de agente ou instrumento da ação, esta RCP seleciona como base os verbos que designam ações voluntárias e que afetam outro(s) indivíduo(s) ou coisa(s) – efeito da ação. São exemplos disso os seguintes termos, obtidos do repertório terminológico analisado: *craqueador*, *desfibrador*, *branqueador*, *degradador*, *modulador*, *escarificador*, *inoculador*, *gaseificador*, *incinerador*, *fumicultor*, *degradador(a)*, *extrusora*. A ação, no entanto, pode não ser única, como por exemplo, em *transiluminador*, formação na qual o produto denomina, para além da ação de iluminar, uma ação de “*transiluminar*”.

Representação da regra:

$$V_{(T)} \rightarrow S_{-(d/t/s)or(a)}$$

Em que:

V = verbo

T = transitivo

S = substantivo

-(d/t/s)or(a) = sufixo

5.3 Derivação prefixal "super/sobre-" (Nanociência/Nanotecnologia)

Segundo Rocha Lima (2000), os prefixos correspondem a preposições ou advérbios extintos ou vivos na língua, por isso seu significado é conhecido e preciso, modificando assim o sentido da palavra primitiva. Neste caso, trata-se de prefixo latino que adiciona à base o significado de “posição em cima, excesso”, segundo Cunha & Cintra (2001).

Na terminologia analisada, registram-se 19 ocorrências da prefixação por “super-“ e apenas 2 ocorrências do prefixo “sobre-”.

Dessa forma, a base da RCP deverá ser um substantivo ou adjetivo que apresente os seguintes traços:

+ substantivo

+ adjetivo

Apesar de serem encontrados registros do acréscimo de significado “posição em cima” (*sobrenadante* e *superposição*), observa-se a predominância do acréscimo de significado “excesso, abundância” (podem ser citados como exemplos, os casos de *superplástico*, *superlente*, *supersaturação* e *superabsorvente*).

Representação da regra:

$$S/A \rightarrow \text{super/sobre-}S/A$$

Em que:

S = substantivo

A = adjetivo

super/sobre- = prefixo

5.4 Derivação prefixal: "des-" (Biocombustíveis)

Prefixo de origem latina, tem o sentido de “separação, ação contrária”, segundo Cunha & Cintra (2001). Rocha Lima (2000) acrescenta como possibilidades de sentido, além das elencadas por Cunha & Cintra (2000), “privação” e “negação”.

A análise das bases revela a presença adjetivos, substantivos e verbos.

Dessa forma, a base da RCP deverá apresentar os seguintes traços:

+ substantivo

+ adjetivo

+ verbo

A análise revela a predominância de emprego com o valor de “ação contrária”, por isso mesmo, emprega-se este prefixo em bases que remetem a processos (*desacidificar, desproteínizar, destoxificação*), de forma que o principal sentido formado é por meio de um processo de desfazer/reverter o estado em que algo se encontra (por exemplo: *desidratante*, no qual o sentido formado é de reverter o estado *hidratado* em que algo se encontra).

Representação da regra:

$$S/A/V \rightarrow \text{des-}S/A/V$$

Em que:

S = substantivo

A = adjetivo

V = verbo

des- = prefixo

5.5 Composição morfológica: "nan(o)-" (Nanociência/Nanotecnologia)

Antes de se iniciar esta análise, é importante ponderar que as formações a partir de *nano-* são as mais numerosas e de maior frequência tanto no corpus quanto na terminologia de N&N, no entanto, esta formação pode ser dividida em três diferentes processos:

i) composição morfológica: *nano-*, nesses casos, ocorre como arqueoconstituente (na denominação do Modelo SILEX⁷⁸), presente em cultismos do século XIX em diante para explicitar que uma das dimensões do objeto ou ação estão em escala nanométrica.

ii) recomposição: casos em que se emprega *nano-* como fractoconstituente (segundo a denominação do Modelo SILEX⁷⁹), relativo à nanociência/nanotecnologia, sendo essa relação de vários tipos (“produzido por”, “com técnicas próprias de”, ...).

iii) derivação prefixal: nesses casos, tem-se *nano-* utilizado como prefixo, denominam unidades de medidas com dimensão 10^{-9} menor que a original

A composição morfológica, formada a partir do elemento de composição *nano-*, originou a maior quantidade de termos (75), indicando que a base (à qual o elemento se agrega) possui pelo menos uma de suas dimensões em escala nanométrica. As formas de recomposição têm a segunda maior ocorrência (19), e os termos formados indicam que a base (à qual o elemento se agrega) é relativa à nanociência/nanotecnologia, sendo essa relação de vários tipos (“produzido por”, “com técnicas próprias de”, etc.). As formas prefixais são pouco numerosas (4) e denominam unidades de medidas com dimensão 10^{-9} menor que a original.

Apesar dos diferentes processos morfológicos em que estão envolvidos, o elemento de composição, o elemento de recomposição e o prefixo possuem formas homógrafas e diferem apenas no aspecto semântico. Por isso, nesta pesquisa esses elementos serão analisados em grupo, sem estabelecer diferenciações, já que a principal motivação aqui é a descrição de uma RFP com o propósito de melhoria dos sistemas de extração automática de terminologias. Este fato pode ser exemplificado com o termo *nanobiologia*: uma análise

⁷⁸ Ou, simplesmente, elemento de composição.

⁷⁹ Também chamado de prefixóide ou, ainda, elemento de recomposição.

apenas formal não permite identificar qual o processo morfológico envolvido na formação do termo: i) no momento em que significa “biologia em escala nanométrica” trata-se de processo de composição morfológica, ii) já no momento em que significa um tipo de “biologia relacionada à nanociência/nanotecnologia,” trata-se de um processo de recomposição. Pelo exemplo, pode-se confirmar que não há qualquer tipo de diferença formal, apenas semântica. Corroboram esta afirmação as seguintes proposições de Antunes & Correia (2010):

“D. Corbin (1997) estabelece que a diferença entre um afixo e um constituinte de palavra composta é determinada pelo papel semântico que o morfema em causa desempenha na palavra. Assim, um mesmo morfema pode corresponder ora a uma unidade infralexical de significado descritivo (um arqueoconstituinte ou um fractoconstituinte), ora a uma unidade infralexical de significado instrucional (um afixo).” (ANTUNES & CORREIA, 2010⁸⁰).

“(…) mesmo um elemento que etimologicamente é um prefixo, *nano-*, pode vir a adquirir o estatuto de fractoconstituinte, intervindo, pois, na construção de compostos cultos (...)” (ANTUNES & CORREIA, 2010).

Diferentemente das demais RCPs apresentadas neste trabalho, o elemento *nan(o)-* em terminologias em língua portuguesa já foi objeto de investigação científica, destacadamente nos trabalhos de Antunes & Correia (2010 e NO PRELO).

Orientado pelo Modelo SILEX, o trabalho “O vocabulário da Nanotecnologia em português europeu: da linguagem de especialidade à língua corrente” de Antunes & Correia, que ainda permanece no prelo, mas cujo acesso foi disponibilizado pelas autoras, apresenta os diferentes valores semânticos que *nano-* pode assumir, a saber: i) 10^{-9} ; ii) muito pequeno, reduzido; iii) insignificante, despiciendo; iv) relativo à nanotecnologia. As autoras trazem, conforme tabela 18, exemplos de termos para cada acepção.

⁸⁰ Acesso direto aos originais, sem indicação de páginas.

Exemplos	Nano 1 10 ⁻⁹	Nano 2 muito pequeno, reduzido	Nano 3 insignificante, despiciendo	Nano 4 relativo à nanotecnologia
nanopessimismo	x	x	x	✓
nanopiadas	x	x	x	✓
nanocuriosos	x	x	x	✓
nanotretas	x	x	x	✓
nano projector	x	✓	x	x
nano-espião	x	✓	x	x
nanoministro	x	✓	✓	x
nanocérebro	x	✓	✓	x

Tabela 18: Valores semânticos do elemento “nano-”, segundo Antunes & Correia (no prelo)

Os valores semânticos apresentados por Antunes & Correia (no prelo), na tabela 18, referem-se tanto a termos quanto a palavras, já que as autoras consideraram também textos de língua geral para recolha e definição dos termos. Especificamente no cenário terminológico, deve-se excluir a significação apresentada pelo elemento “Nano 3”. Pode-se ainda estabelecer relações entre os valores semânticos retratados e os processos morfológicos que se valem do afixo *nano-*: a acepção número 1 é aquela obtida pelo processo de derivação prefixal (quanto anteposta a uma unidade de medida, como é o caso de *nanômetro*), já a acepção número 2 é aquela proveniente do processo de composição morfológica e, por fim, a acepção número 4 é obtida com a recomposição.

A análise da terminologia permite restringir a base da RCP a um substantivo com as seguintes características:

+ substantivo (+/- ação/processo)

Como exemplos de termos formados a partir de substantivos não deverbais, portanto, não indicam ação/processo, tem-se *nanococoncha*, *nanobastão*, *nanoanel*, *nanogel*, *nanotubo*. Por outro lado, os exemplos de formações a partir de substantivos deverbais,

que indicam ação/processo são, entre outros, *nanocaracterização*, *nanoimpressão* e *nanomanipulação*.

Representação da regra:

$$S \rightarrow \text{nano-}S$$

Em que:

S = substantivo

nano- = prefixo ou elemento de composição morfológica ou, ainda, elemento de recomposição (prefixóide)

5.6 Composição morfológica: “hidro-” (Biocombustíveis)

Segundo o Dicionário Aurélio (FERREIRA, 2004), o elemento *hidro-* pode se referir a: i) “água, líquido” (*hidrociclone*, *hidrolase* e *hidrólise*); ii) quando empregado na Química, indica a presença de átomo de hidrogênio (*hidroclorofluorcarboneto*, *hidrocarboneto*, *hidrofluorcarbono*, *hidrogênio*, *hidroperoxila*, *hidroxila* e *hidrocraqueamento*) e iii) refere-se a “suor”.

A referência a “suor” não é observada na terminologia. Observa-se que *hidro-* não revela a presença de água, mas sim que o processo é feito utilizando-se água ou outro tipo de líquido (*hidrociclone*, *hidrocraqueamento*, *hidrólise*). Por fim, pode ser empregado também para marcar a presença do átomo de hidrogênio em elementos ou compostos químicos, produzindo termos como *hidrocarboneto*, *hidrofluorcarbono*, *hidrogênio* e *hidroperoxila*.

A base da RCP deverá ser um substantivo que apresente os seguintes traços:

+ substantivo (+/- ação processo)

Representação da regra:

$$S \rightarrow \text{hidro-}S$$

Em que:

S = substantivo

hidro- = elemento de composição morfológica

6. Considerações finais

Neste trabalho, apresentou-se inicialmente uma breve incursão sobre a Teoria Comunicativa da Terminologia (TCT), dado que o objeto desta investigação foram termos que integram vocabulários especializados, obtidos a partir de um método fundamentado nessa teoria.

Ao filiar-se a TCT, adota-se uma visão na qual os termos não pertencem a um domínio, mas são empregados em cada domínio com um valor específico (CABRÉ, 1999), dessa forma, um modelo de análise morfológica inicialmente proposto para língua geral viabilizou-se como instrumento de análise de terminologias. A aproximação entre Morfologia e Terminologia se dá pelo léxico, isso fica evidente nos processos de formação de novas unidades, já que os neologismos constituem tema fundamental de estudos em Terminologia e Morfologia.

Posteriormente, foram apresentados alguns trabalhos voltados para o léxico geral e especializado tanto no Brasil como em Portugal, todos com um enfoque morfológico.

Também foi feita uma incursão à Morfologia de base gerativa, de maneira a introduzir o modelo SILEX, modelo construcional gerativo de análise morfológica,

escolhido para embasar as classificações e análises realizadas neste trabalho. A escolha deste modelo justifica-se por seu objetivo que é construir uma teoria sincrônica do léxico capaz de atribuir uma estrutura e uma interpretação adequadas às palavras construídas, atestadas ou não nos dicionários, de modo a caracterizar a natureza da “gramaticalidade lexical” e de determinar as restrições das regras de formação de palavras (cf. CORBIN, 1987, p. 1).

O estudo morfológico de terminologias não pode deixar de considerar os processos de inovação lexical, dessa forma os processos por meio dos quais ocorre a inovação lexical foram temas abordados neste trabalho.

Cada terminologia possui estruturas morfológicas características, por isso, optou-se por duas terminologias originalmente multidisciplinares, de forma que se obtivesse, ao menos, uma visão parcial, mas ampla, da grande área das Engenharias, são elas: Nanociência e Nanotecnologia e Biodiesel.

Nesse sentido, foi realizada uma análise quantitativa e qualitativa dos morfemas das terminologias em estudo, o que permitiu a construção de uma base de dados lexicais (BDL) de nível morfológico. Com a base de dados desenvolvido nesta pesquisa, inicia-se a criação de um repositório de informações morfológicas de terminologias com amplas possibilidades de aplicação em ferramentas computacionais de PLN, que demandem como recurso linguístico o conhecimento morfológico.

No domínio linguístico, foram estabelecidas e comentadas seis diferentes Regras de Construção de Palavras envolvendo os elementos mais frequentes, de maneira a demonstrar a viabilidade do modelo SILEX para a análise morfológica com vistas a aplicações computacionais. As seis RCPs propostas dizem respeito aos seguintes morfemas: - *(ç/s/ss)ão*, -*(d/t/s)or(a)*, *super/sobre-*, *des-*, *nan(o)-* e *hidro-*.

As RCPs propostas neste trabalho, envolvendo os morfemas acima são:

- $V_{(T/I)} \rightarrow S_{-(\zeta/s/ss)ão}$
- $V_{(T)} \rightarrow S_{-(d/t/s)or(a)}$
- $S/A \rightarrow_{super/sobre} S/A$
- $S/A/V \rightarrow_{des} S/A/V$
- $S \rightarrow_{nano} S$
- $S \rightarrow_{hidro} S$

Com o fim desta pesquisa, pode-se dar início ao trabalho de implementação computacional (fases representacional e computacional, conforme descritas na Metodologia) pelos membros do Núcleo Interinstitucional de Linguística Computacional, conforme previsto no projeto inicial.

Para a implementação do domínio representacional, a cargo do NILC, pretende-se desenvolver, como trabalhos futuros, um ambiente on-line a semelhança do “Vocabulário ortográfico português – VOP⁸¹” implementado pelo ILTEC⁸². Além da disponibilização da BDL desenvolvida neste trabalho, também outras BDL e conhecimentos morfológicos sistematizados poderão ser agrupados e disponibilizados neste ambiente on-line de forma a facilitar o seu compartilhamento e, conseqüentemente, sua implementação.

Os resultados obtidos satisfazem os resultados propostos inicialmente e, além disso, contribuem para o desenvolvimento de novas pesquisas na medida em que deixa em aberto outros temas também importantes. Um interessante trabalho futuro a partir desta pesquisa é avaliar se alguns dos empréstimos excluídos transformam-se em bases para construção de

⁸¹ <http://www.portaldalinguaportuguesa.org/>

⁸² O Instituto de Linguística Teórica e Computacional (ILTEC) é uma associação privada sem fins lucrativos cujos associados são a Fundação para a Ciência e a Tecnologia, a Universidade de Lisboa e a Universidade Nova da Lisboa (Unidade de I&D da FCT com o número 17/146).

RCPs vernáculas, semelhantes aos casos de **linkar** e **clusterização**, por exemplo, em que à base estrangeira *link* se une o sufixo vernáculo formador de verbo em *-ar*, e à base estrangeira *cluster* se junta o sufixo também vernáculo *-ização*, este último, resultado de um processo de duas fases: **clusterizar** > **clusterização**.

Além disso, buscar ampliar a descrição das RCPs dos casos não tratados nesta pesquisa é um trabalho futuro que certamente enriquecerá a descrição linguística e produzirá melhores resultados em PLN.

7. Referencias bibliográficas

ALMEIDA, G.M.B. O percurso da Terminologia: de atividade prática à consolidação de uma disciplina autônoma. Tradterm, São Paulo, v. v. 9, p. 211-222, 2003.

ALMEIDA, G.M.B. A Teoria Comunicativa da Terminologia e a sua prática. Alfa (Araraquara), v. 50, p. 81-97, 2006. Disponível em: <http://seer.fclar.unesp.br/alfa/article/view/1413/1114>

ALMEIDA, G.M.B. Os anglicismos e as linguagens de especialidade no português do Brasil. Debate Terminológico, RITerm - Paris (França), v. 2, p. 1-12, 2006.

ALMEIDA, G.M.B.; SOUZA, D.S.L.; PINO, D.H.P. A definição nos dicionários especializados: proposta metodológica. Debate Terminológico, RITerm, Paris (França), v. 3, p. 1-20, 2007.

ALMEIDA, G. M. B.; VALE, O. A. Do texto ao termo: interação entre Terminologia, Morfologia e Linguística de Corpus na extração semi-automática de termos. In: ISQUERDO, Aparecida Negri; FINATTO, Maria José Bocorny. (Orgs.). *As ciências do Léxico: Lexicologia, Lexicografia e Terminologia*. 1 ed. Campo Grande: Editora da UFMS, 2008, v. IV, p. 483-499.

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 linguística**. Calidoscópico (UNISINOS), v. 4, p. 156-178, 2006.

ALVES, I. M. A integração dos neologismos por empréstimo ao léxico português. *Alfa*, v. 28, São Paulo: EDUNESP, 1984.

ALVES, I. M. *Neologismo: Criação lexical*. 3a.ed. São Paulo: Ática, 2007.

ANDRADE, A. R. de. **Os Corpora Lingüísticos: uma nova forma de fazer Lexicografia?** In: ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGUISTICA, 18., 2003, Lisboa. Actas. Lisboa: Colibri, 2003, p. 103-110.

ANTUNES, M.; CORREIA, M. Novos formantes da língua portuguesa. Análise dos fractoconstituintes presentes no ONP. In: Alves, Ieda Maria (org.). Neologia e Neologismos em Diferentes Perspectivas. São Paulo: Paulistana, 2010.

ANTUNES, M.; CORREIA, M. O vocabulário da Nanotecnologia em português europeu: da linguagem de especialidade à língua corrente. Comunicação apresentada ao *XI Simpósio Ibero-Americano de Terminologia – RITerm* (Lima, Peru; Outubro de 2008), NO PRELO.

AUBERT, Francis Henrik. (1996) Introdução à Metodologia da Pesquisa Terminológica Bilíngüe - Cadernos de Terminologia, São Paulo: Humanitas Publicações - FFLCH/ USP.

BARROS, L.A. Curso básico de Terminologia. São Paulo: EDUSP, 2004.

BASÍLIO, M. Estruturas lexicais do português: uma abordagem gerativa. Petrópolis: Vozes, 1980

BASÍLIO, M. *Teoria Lexical*. São Paulo: Ática, 1989.

BASILIO, M. A morfologia no Brasil: indicadores e questões. *DELTA* [online]. 1999, vol.15, pp. 53-70 . ISSN 0102-4450.

BERBER SARDINHA, T. **Lingüística de Corpus: Histórico e Problemática**. D.E.L.T.A., vol.16, nº 2, 2000 (323-367)

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

CABRAL, L. S. Introdução à linguística. 4ª. edição revista e ampliada. Porto Alegre: Globo, 1979.

CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona, Editorial Antárdida/Empúries, 1993.

CABRÉ, M.T. **La terminología: representación y comunicación** – elementos para una teoría de base comunicativa y outros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, 1999.

CABRÉ, M. T. A modo de introducción. In: *La terminologia: Representación y comunicación*, Girona: Documenta Universitária, 2005.

CABRÉ, M. T. “Morfologia e Terminologia”. Em Feliu, Elena (ed.) *La morfologia a debate*. Jaén: Universidad e Jaén, 2006, 131-144.
Correa, 2004 <http://www.iltec.pt/pdf/wpapers/2004-mcorreia-barcelona.pdf>

COLETI, J. S. A terminologia da Nanociência e Nanotecnologia: descrição e análise morfológica. Trabalho de Conclusão de Curso. São Carlos: Curso de Letras, Universidade Federal de São Carlos, 2009

COLETI, Joel Sossai ; ALMEIDA, G. M. B. . Aspectos morfológicos da terminologia da Nanociência e Nanotecnologia. *Filologia e Lingüística Portuguesa*, v. 12/13, p. 1-12, 2010.

CORBIN, D. *Morphologie dérivationnelle et structuration du lexique*. 2 vols. Tubinga: Max Niemeyer Verlag, 1987.

CORBIN, D. Form, structure and meaning of constructed words in an associative and stratified lexical component. In: *Yearbook of Morphology 2*. Dordrecht: Foris Publications, 1989, p. 31-54.

CORBIN, D. La représentation d’une “famille” de mots dans le *Dictionnaire dérivationnel du français* et ses corrélats théoriques, méthodologiques et descriptifs. In: *Recherches linguistiques de Vincennes*, 1997 pp. 5-37 + errata.

CORREIA, M. Capítulo 1 – Introdução. In: *A denominação das qualidades em português – contributos para a compreensão da estrutura do léxico português*. Tese de doutoramento apresentada à Universidade de Lisboa, 1999.⁸³

CORREIA, M. *Denominação e construção de palavras*. Lisboa: Edições Colibri, 2004.

CORREIA, M. *Terminologia e morfologia: marcas morfológicas da génese do vocabulário da Náutica em português*. In: *La terminología en el siglo XXI – Actas del IX Simposio Iberoamericano de Terminología*. Barcelona (Espanha): IULA, 2004a. Disponível em www.iltec.pt/pdf/wpapers/2004-mcorreia-barcelona.pdf

CORREIA, M.; LEMOS, L. S. P. *Inovação lexical em português*. Lisboa: Colibri, 2005

CUNHA, C.; CINTRA, L. F. Nova gramática do português contemporâneo. Rio de Janeiro: Nova Fronteira, 2001.

DI FELIPPO, A.; DIAS DA SILVA, B. C. Dos “olhares” sobre o léxico: diferenças e semelhanças. In *A construção de dicionários e de bases de conhecimento lexical*, Araraquara: Laboratório Editorial e São Paulo: Cultura Acadêmica Editora, 2006.

DIAS-DA-SILVA, B.C. The WordNet.Br: an Exercise of human language technology research. In: *International WordNet Conference, 2007, South Jeju Island. Proceedings of the Third International WordNet Conference*. Brno: Masaryk University Press, 2007. v. 3. p. 301-303

⁸³ Este capítulo, que constitui uma apresentação global do Modelo SILEX, não foi publicado em CORREIA 2004, encontrando-se, portanto, inédito.

- FERREIRA, A. B. H. Novo dicionário Aurélio século XXI: dicionário eletrônico versão 5.0. Curitiba: Positivo, 2004.
- GALINSKI, C. La figura d'E. Wüster. In: CABRÉ, M. T. (org.) Terminologia - selecció de textos d'E. Wüster (trad. catalã de Anna C. i Galera e Teresa G. i Morell). Barcelona: Servei de Llengua Catalana - Universitat de Barcelona, 1996.
- GRODT, A. *Um estudo sobre produtividade derivacional no português falado no sul do Brasil*. Mestrado. Porto Alegre: Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, 2009
- HOUAISS, A.; VILLAR, M. S. Dicionário Houaiss da língua portuguesa. Rio de Janeiro: Objetiva, 2001.
- KASAMA, D. ; ALMEIDA, G. M. B. ; ZAVAGLIA, C. . A influência das novas tecnologias no léxico: processos de formação neológica no domínio da Nanociência e Nanotecnologia. *Debate Terminológico*, v. 4, p. 1-11, 2008.
- KRIEGER, M. G.; FINATTO, M.J.B. *Introdução à Terminologia: teoria e prática*. São Paulo: Editora Contexto, 2004.
- LORENTE, M. A lexicologia como ponto de encontro entre a gramática e a semântica. In: ISQUERDO, A. N. e KRIEGER, M.G. (orgs.) *As ciências do léxico*, vol. II. Campo Grande: Editora UFMS, 2004.
- PINO, D. H. P. *Aspectos semânticos da terminologia do biodiesel*. São Carlos: UFSCar, 2010.
- PUSTEJOVSKY, J. *The Generative Lexicon*. Cambridge: MIT Press, 1995.
- RIO-TORTO, G. M. *Regras de formação de palavras em português: achegas para um quadro geral*. In: RIO-TORTO, G. M. *Morfologia Derivacional – teoria e aplicação ao português*. Porto: Porto Editora, 1998.
- ROCHA, L. C. A. *Estruturas morfológicas do português*. São Paulo: WMF Martins Fontes, 2008
- ROCHA LIMA, C. H. *Gramática Normativa da Língua Portuguesa*, Rio de Janeiro: José Olympio Editora, 2000
- SENAI. *Nanotecnologias*. Brasília: SENAI, 2004
- TEMMERMAN. R. Terminology Theory and Terminography in a Natural Language Processing Environment. *Revue Française de Linguistique Appliquée*, v. III – 2, p. 29 – 46, 1998

TEMMERMAN, R. Towards new ways of terminology description. The sociocognitive approach. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2000.

TEMMERMAN, R. Teoria sociocognitiva da terminologia Presentació. In: KRIEGER M. G.; ARAÚJO L. (org,) *Cadernos de Tradução - A terminologia em foco* (trad. De Natacha Enzweiler e Luzia Araújo, Revisão: Tália Burgel). Porto Alegre: Instituto de Letras - UFRGS, 2004.

TOMA, Henrique E.; ARAKI Koiti. *O gigantesco e promissor mundo do muito pequeno*, 2005. Disponível em: <http://cienciahoje.uol.com.br/materia/view/3440>. Acessado em 18 de abril de 2007.