



Programa de
Pós-Graduação em
Linguística

**ANÁLISE MORFOLÓGICA DOS TERMOS DAS PRAGAS DA
CANA-DE-AÇÚCAR: SUBSÍDIOS PARA A ORGANIZAÇÃO DE UMA
BASE DE DADOS MORFOLÓGICOS PARA O PORTUGUÊS**

Carolina Cardoso de Oliveira

SÃO CARLOS
2016



Universidade Federal de São Carlos

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

**ANÁLISE MORFOLÓGICA DOS TERMOS DAS PRAGAS DA CANA-DE-AÇÚCAR:
SUBSÍDIOS PARA ORGANIZAÇÃO DE UMA BASE DE DADOS MORFOLÓGICOS
PARA O PORTUGUÊS**

CAROLINA CARDOSO DE OLIVEIRA

Bolsista Capes

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do Título de Mestra em Linguística.

Orientadora: Profa. Dra. Gladis Maria de Barcellos Almeida.

São Carlos – São Paulo – Brasil

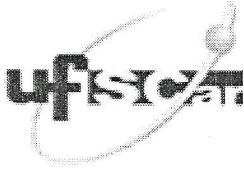
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

O48a Oliveira, Carolina Cardoso de
Análise morfológica dos termos das pragas da cana-
de-açúcar : subsídios para organização de uma base de
dados morfológicos para o português / Carolina
Cardoso de Oliveira. -- São Carlos : UFSCar, 2016.
87 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2016.

1. Morfologia derivacional. 2. Terminologia. 3.
Pragas da cana-de-açúcar. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Carolina Cardoso de Oliveira, realizada em 05/05/2016:

Profa. Dra. Gladis Maria de Barcellos Almeida
UFSCar

Profa. Dra. Flavia Bezerra de Menezes Hirata Vale
UFSCar

Prof. Dr. Thiago Alexandre Salgueiro Pardo
USP

À minha avó Adília que durante essa jornada passou a me olhar através das estrelas e me ensinou a cada casaco, blusa e meia de lã que me presenteava que para alcançar um objetivo eu apenas preciso de paciência, dedicação, amor e humildade.

Minha eterna saudade.

Agradecimentos

Mal posso acreditar que chegou o momento de relembrar a trajetória que resultou nessa dissertação. Por causa de tantas pessoas esse caminho foi único e inesquecível e agora quero agradecer àquelas que me acompanharam tão de perto.

A Deus, meu melhor Amigo, pelo amor incondicional e por se fazer tão presente em todos os momentos. Também a Nossa Senhora, minha Mãe e Protetora desde sempre.

Ao meu pai Hélio, minha mãe Luzia e minha irmã Débora: vocês são minha origem, meu lar, meu refúgio. Sem vocês eu jamais conseguiria chegar até aqui.

À minha família, composta por pessoas tão especiais que vibram e comemoram cada pequena conquista junto comigo.

À minha orientadora Gladis, que desde a graduação me acolheu, agradeço pela compreensão, paciência, pelos ensinamentos que levarei para vida toda e pelo exemplo de profissionalismo.

À professora Flávia e ao professor Thiago pela fina leitura do relatório de qualificação, pelos questionamentos, correções e sugestões que permitiram melhor desenvolvimento do trabalho. Agradeço também por aceitarem novamente o convite para participar desse momento especial de aprendizado.

Aos amigos que fazem a diferença: Maria Beatriz, Letícia, Marco, Jaqueline, Dayse, Eld, Stela, Gleice e Daniela.

À Capes pelo auxílio financeiro durante o mestrado.

Ao PPGL, ao DL e principalmente à UFSCar por ser praticamente a minha casa durante tanto tempo e me propiciar a melhor fase da minha vida.

Ao meu pequenino e tão amado Bob, pelo companheirismo e amor gratuito que alegam meu dia a dia.

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar.

Mas o mar seria menor se lhe faltasse uma gota”.

(Madre Teresa de Calcutá)

Resumo

As pesquisas em Terminologia são de extrema importância para a disseminação do conhecimento. Nesse sentido, é necessário que se façam cada vez mais estudos com o objetivo de aprimorar o trabalho terminológico e, para isso, a informática se tornou uma aliada, auxiliando o terminólogo na realização de suas tarefas. Uma ferramenta computacional que aperfeiçoa o trabalho em Terminologia são os extratores automáticos de candidatos a termos. Quanto mais finas e detalhadas forem as informações linguísticas presentes nesses sistemas, maior a precisão da lista de candidatos que eles irão fornecer ao terminólogo. Com o objetivo de oferecer informações para enriquecer as bases de dados utilizadas nessas ferramentas, propomos nesta dissertação descrever morfológicamente os termos do domínio das *Pragas e doenças da cana-de-açúcar* e verificar os processos de formação de palavras mais produtivos. Os termos aqui considerados foram obtidos no Trabalho de Conclusão de Curso (TCC), intitulado *A terminologia das pragas da cana-de-açúcar*, desenvolvido em 2013. São 135 termos que foram analisados segundo a classificação proposta por Correia & Almeida (2012). Desse total: 135 são importação de palavras, 16 são lexicalizações sintagmáticas, 10 são formados por derivação sufixal, apenas dois casos de extensão semântica, um de derivação não afixal e um de composição morfossintática. Com a descrição morfológica aqui realizada, espera-se enriquecer a base de dados morfológicas em elaboração pelo GETerm, oferecendo, assim, subsídios para a futura implementação em extratores semiautomáticos de termos que se utilizam de conhecimento linguístico, de maneira a melhorar o seu desempenho. Também esperamos contribuir para o avanço de estudos morfológicos de terminologias do português.

Palavras-chave: morfologia derivacional, terminologia, pragas da cana-de-açúcar.

Abstract

Research in Terminology is extremely important for the dissemination of knowledge. This way, more studies with the objective of improving the terminological work are necessary and for this purpose, computer science has become an ally, helping the terminologist in his tasks. A computational tool that improves research in Terminology is the automatic extractor of candidate terms. The more refined and detailed the linguistic information present in these systems is, the more precise the list of candidates provided to the terminologist will be. Focusing on offering information to enrich the database used in these tools, this research carries out a morphological description of the terms of the domain of *Plagues and sugarcane diseases* and a verification of the most productive processes of word formation. The terms considered in this research were obtained in an end of course project, entitled *The terminology of sugarcane plagues*, developed in 2013. There are 135 terms which were analyzed according to the classification proposed by Correia and Almeida (2012). From this total: 135 are importation of words, 16 are syntagmatic lexicalizations, 10 are formed by suffixal derivation, only 2 cases of semantic extension, 1 of non-affixal derivation and 1 of morphosyntactic composition. With the proposed morphological description, we expect to enhance the morphological database being elaborated by GETerm, offering, this way, resources for the future implementation of semiautomatic extractors of terms that make use of linguistic knowledge, improving its performance. We also hope to contribute with the advance in morphological studies in Portuguese terminologies.

Keywords: derivational morphology, terminology, sugarcane plagues.

Lista de siglas, abreviaturas e reduções

Acrôn. – acrônimo
Adj – adjetivo
Adp. – adaptado, adaptação
Al. – alemão
Amálg. – amálgama
Ár. – árabe
Consoa. De Lig. – consoante de ligação
El. Comp. – elemento de composição
Esp. – espanhol
Ext. Sem. – extensão semântica
Fam. – família
Freq. – frequência
Germ. – germânico
Gr. – grego
Id. – idem
Infl. – influência
It. – italiano
Lat. – latino
Lat. Cien. – latim científico
L.F.F – lexicalização de formas flexionadas
L.S – lexicalização de sintagmas
L.U.D. – lexicalização de unidades discursivas
Masc. – masculino
Mdv. – medieval
Morf. – morfológica
Morfos - morfossintática
N – nome
N.A. – não afixal
Or. – origem
Or. Controv. – origem controversa
Paras. – parassintética
Pre. – prefixal
Pref. – prefixo
Prep – preposição
Prov. – provavelmente
Rad. – radical
Recomp. – recomposição
Regr. – regressivo
Suf. – sufixo/ sufixal
SPrep – sintagma preposicionado
Trunc. – truncação

Índice de figuras, quadro e tabelas

Figura 1 Principais recursos teóricos e metodológicos de que o estudo em PLN dispõe.	p. 41
Figura 2 Organização das fases que compõem a metodologia.	p. 43
Figura 3 Esquema de possível arquitetura de um sistema de PLN.	p. 45
Figura 4 Fragmento da planilha de análise morfológica dos termos.	p. 62
Figura 5 Acepções do termo “amarelinho”.	p. 63
Quadro 1 Progresso dos estudos em PLN.	p. 42
Tabela 1 Ocorrência dos processos de formação de palavras do domínio da Nanociência/Nanotecnologia.	p. 49
Tabela 2 Ocorrência dos processos de formação de palavras do domínio da Educação a Distância.	p. 55
Tabela 3 Ocorrência dos processos de formação de palavras do domínio da Fisioterapia.	p. 57
Tabela 4 Comparação entre os trabalhos descritos.	p. 58
Tabela 5 Dados estatísticos do corpus.	p. 60
Tabela 6 Processos de formação de palavras encontrados.	p. 65

Índice

1 Introdução	p. 12
2 Pressupostos teóricos	p. 17
2.1 Contextualização e importância da Terminologia	p. 17
2.1.1 <i>A Teoria Geral da Terminologia (TGT)</i>	p. 19
2.1.2 <i>Questionamentos à TGT</i>	p. 22
2.1.3 <i>Contribuições de Cabré à Teoria Comunicativa da Terminologia (TCT)</i>	p. 23
2.2 Morfologia: conceito e história	p. 28
2.2.1 <i>Morfologia flexional e derivacional</i>	p. 31
2.2.2 <i>Morfologia derivacional e o enriquecimento do léxico</i>	p. 32
3 Metodologia	p. 40
3.1 A metodologia de Dias-da-Silva	p. 40
3.2 Descrição dos trabalhos já realizados	p. 48
3.3 Percurso até o conjunto terminológico	p. 59
3.3.1 <i>Compilação do corpus</i>	p. 59
3.3.2 <i>Extração dos termos</i>	p. 60
3.3.3 <i>Validação das listas</i>	p. 61
3.3.4 <i>Seleção dos termos e a descrição morfológica</i>	p. 61
4 Análise dos dados	p. 65
4.1 Importação de palavras	p. 65
4.2 Lexicalização sintagmática	p. 66
4.3 Derivação sufixal	p. 68
4.4 Comparação com os trabalhos já realizados	p. 69
5 Conclusão	p. 71
6 Referências	p. 74
7 Apêndice	p. 78

Introdução

A parceria entre profissionais e pesquisadores da Embrapa e Universidade Federal de São Carlos (UFSCar) foi a força motriz que possibilitou o desenvolvimento desta pesquisa. Essa parceria se deu especificamente entre a unidade Embrapa Informática Agropecuária, situada em Campinas, e o Grupo de Estudos e Pesquisas em Terminologia (GETerm),¹ do Departamento de Letras da UFSCar.

O primeiro passo desse percurso foi a realização de um estágio no GETerm, mas a serviço da Embrapa, cuja tarefa era redigir definições de termos referentes à Agropecuária. A partir daí, surgiu a proposta de um Trabalho de Conclusão de Curso (TCC), que compreendeu as fases iniciais de um trabalho terminológico, ou seja, compilação e limpeza de *corpus*, extração e validação de termos e organização de uma estrutura conceitual.² O domínio escolhido para o TCC foi o das pragas e doenças da cana-de-açúcar, devido à importância dessa cultura para o Brasil e, ainda, por ser o estado de São Paulo o maior produtor do país, com 426.572.099 toneladas produzidas entre 1993 e 2010, segundo o Instituto Brasileiro de Geografia e Estatística (2013), num total de 717.462.101 produzidas por todos os estados brasileiros considerados.

Cultivada desde o período colonial, a cultura da cana-de-açúcar se expandiu progressivamente e hoje assume um dos primeiros lugares na lista das culturas mais importantes para a economia brasileira. Segundo o Ministério da Agricultura, “o Brasil não é apenas o maior produtor de cana. É também o primeiro do mundo na produção de açúcar e etanol e conquista, cada vez mais, o mercado externo com o uso do biocombustível como alternativa energética.” (BRASIL, 2016, s/p). Ainda segundo o Ministério,

responsável por mais da metade do açúcar comercializado no mundo, o país deve alcançar taxa média de aumento da produção de 3,25% até 2018/19, e colher 47,34 milhões de toneladas do produto, o que corresponde a um acréscimo de 14,6 milhões de toneladas em relação ao período 2007/2008. Para as exportações, o volume previsto para 2019 é de 32,6 milhões de toneladas (BRASIL, 2016, s/p).

¹ O Grupo foi criado em 1999 no Departamento de Letras da UFSCar e conta com uma equipe multidisciplinar, envolvendo linguistas, informatas e especialistas de domínio. Tem como parceiros o Núcleo Interinstitucional de Linguística Computacional (NILC), sediado no Instituto de Ciências Matemáticas e da Computação (ICMC) da USP/São Carlos, e a Embrapa Informática Agropecuária, unidade Campinas (SP). O GETerm tem como objetivos estudar conteúdos pertinentes à Terminologia/Terminografia e desenvolver pesquisas que gerem produtos terminológicos em língua portuguesa, tais como glossários, dicionários, enciclopédias e assemelhados, que satisfaçam demandas reais.

² Cf. Oliveira (2013).

Com relação ao etanol, a previsão de produção “para 2019 é de 58,8 bilhões de litros, mais que o dobro da registrada em 2008. O consumo interno está projetado em 50 bilhões de litros e as exportações em 8,8 bilhões” (BRASIL, 2016, s/p).

No início da colonização do Brasil, o principal produto originado da cana-de-açúcar era o açúcar, entretanto, na década de 1970, iniciou-se a crise do petróleo, o que provocou a ascensão do etanol como alternativa. Esse processo se deu com a elevação do preço do petróleo e os incentivos realizados pelo Governo, a fim de estimular a produção e o consumo do etanol, visto que é uma fonte de energia renovável. A principal ação nesse sentido foi a criação do Programa Nacional do Álcool (Proálcool), que ofereceu benefícios aos produtores de cana-de-açúcar e às indústrias automobilísticas, tais como incentivos fiscais e empréstimos bancários com juros abaixo da taxa de mercado à época. Desde então, o álcool tem sido a escolha de grande parte dos brasileiros no momento de abastecer seu veículo (BIODIESEL BR, 2006).

Com o cultivo da cana-de-açúcar sempre em alta, compreende-se que pesquisas e trabalhos com o objetivo de aprimorar as plantações e todos os procedimentos que essa cultura demanda devem ser constantemente realizados.

Um problema que ainda prejudica o setor sucroenergético nacional são as pragas. Segundo o Glossário de Termos Fitossanitários presente na Norma Internacional para Medida Fitossanitária – NIMF 5 –, praga é “qualquer espécie, raça ou biótipo vegetal ou animal ou agente patogênico que cause danos para as plantas ou para produtos vegetais” (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA ALIMENTAÇÃO E AGRICULTURA, 2009, p. 15). Na cana-de-açúcar, pragas podem ocorrer nas folhas, no colmo e na raiz.

De acordo com Rossetto & Santiago (2016), “mundialmente, a cana-de-açúcar contabiliza perdas de aproximadamente 20% ao ano, considerando somente o ataque de pragas.” Ainda segundo a Agência, “a cana-de-açúcar pode ser atacada por mais de 80 espécies de pragas, sendo que algumas delas, como alguns besouros e cupins, muitas vezes são observadas nas lavouras somente após terem causado danos, uma vez que são pragas de solo e, por isso, de difícil observação.” Para se ter uma ideia, “para uma produtividade de 80 toneladas por hectare, as perdas ocasionadas pela broca para cada 1% de intensidade de infestação são de 616 quilos de cana, 28 quilos de açúcar e 16 litros de álcool, aproximadamente.” Já “os prejuízos causados pela cigarrinha-da-folha têm chegado a 17,5% de perda no processo industrial, quando a população de adultos chega a 0,7 indivíduos por colmo.” Há ainda a espécie *Heterotermes tenuis*, que chega a causar perdas “da ordem de dez

toneladas por hectare por ano, sobretudo em solos arenosos.” Outra praga conhecida é a cigarrinha-da-raiz, que “chega a causar perdas de 11% na produtividade agrícola e 1,5% na produção de açúcar.” (ROSSETTO & SANTIAGO, 2016, s/p).

Podemos constatar, a partir dos dados apresentados, a importância do conhecimento sobre as pragas e as doenças por elas provocadas. A fim de colaborar para que esse conhecimento esteja organizado e disponível para especialistas, pesquisadores, produtores e demais interessados é que foi realizado o TCC.

Por meio de estudos terminológicos, é possível que o conhecimento, antes restrito apenas a especialistas, possa ser acessível a leigos. Em outras palavras, a Terminologia possibilita o processo de comunicação de especialistas entre si, de especialistas com leigos e dos leigos entre si.

O objeto principal de uma pesquisa terminológica é o termo e é nele que nosso trabalho de mestrado se concentra, o qual tem como principal objetivo a análise morfológica dos termos das pragas da cana-de-açúcar que foram coletados e validados pelo especialista de domínio durante o TCC.

É relevante mencionarmos que a partir da década de 1980 a informática passou a atuar juntamente com os estudos terminológicos. Muitas ferramentas computacionais foram desenvolvidas para auxiliar e otimizar a pesquisa em Terminologia, entre elas os extratores automáticos de termos. Esses extratores funcionam a partir do fornecimento de um *corpus* e geram uma lista de termos ou candidatos a termos. Tais ferramentas podem se basear apenas em métodos estatísticos³, gerando a lista dos candidatos a termos mais frequentes no *corpus*. Nesse caso, é necessária uma revisão cuidadosa por parte do linguista, pois muitos candidatos elencados pelo extrator podem não ser considerados termos. No referido TCC, por exemplo, em que foi utilizado esse tipo de extrator, tivemos casos como *dependendo* e *de acordo com* na lista de candidatos.

Atualmente, existem ferramentas que aceitam como entrada padrões morfossintáticos do tipo:

- N + Adj – *entomologia agrícola*
- N + Prep + N – *extrato de levedura*
- N + Adj + Prep + N – *fixação biológica de nitrogênio*
- N + Prep + N + Adj – *macacão de algodão hidrorrepelente*

³ Os demais métodos utilizados por esses extratores serão descritos na seção 3.1

Entretanto, a menor unidade considerada nos padrões acima é a palavra. Se aprofundarmos a descrição e chegarmos ao nível do morfema, poderemos enriquecer ainda mais os extratores com conhecimento linguístico, melhorando, então, seu desempenho. Isso é possível, pois em muitos casos os morfemas são utilizados na construção de novos termos e se os extratores foram capazes de reconhecê-los, a identificação dos termos será mais precisa. Uma área que possui informações morfológicas características é a Fisioterapia, que apresenta grande quantidade de termos formados por morfemas como: *-ite* (sufixo que designa doença inflamatória de órgão) e *-ose* (sufixo utilizado na nomeação de doenças) (COLETI, 2012). Se o extrator contar com uma base de dados rica e variada ele identificará quais morfemas são recorrentes nas terminologias e então poderá estabelecer uma espécie de padrão. O extrator será apto a reconhecer que palavras compostas por sufixos como *-mento*, *-ão*, *-agem* têm grande chance de serem termos enquanto que palavras compostas por sufixos como *-ndo*, por exemplo, certamente não serão.

Assim, os objetivos deste trabalho são:

1. analisar morfológicamente um conjunto de 135 termos relativos às pragas e doenças que acometem a cana-de-açúcar, de modo a identificar os processos de formação de palavras mais produtivos nesse domínio;
2. elaborar uma planilha que apresente todos os dados em colunas, de forma a facilitar a representação para eventual implementação computacional;
3. oferecer conhecimento sobre a morfologia derivacional do português, tendo como objeto de análise um léxico especializado.

Ressalte-se que no GETerm já foram realizadas as seguintes pesquisas de descrição morfológica de terminologias, as quais serão descritas na seção referente à metodologia:

- *Aspectos morfológicos dos termos da Nanociência e Nanotecnologia* (2009-2010). Trabalho de Conclusão de Curso (Graduação em Bacharelado em Linguística – UFSCar) desenvolvido por Joel Sossai Coleti, com orientação de Gladis Maria de Barcellos Almeida;
- *Base de dados morfológicos de terminologias do português do Brasil. Descrição e análise morfológica com vistas à disponibilização on-line* (2010-2012). Dissertação

(Mestrado em Linguística – UFSCar) desenvolvida por Joel Sossai Coleti,⁴ com orientação de Gladis Maria de Barcellos Almeida;

- *Descrição das unidades terminológicas da Educação a Distância em português do Brasil no nível morfológico* (2012-2013). Trabalho de Conclusão de Curso (Graduação em Bacharelado em Linguística – UFSCar) desenvolvido por Ana Catarina Gianoti, com orientação de Ariani Di Felippo;⁵
- *Descrição e análise morfológica da terminologia da Fisioterapia: subsídios para organização de uma base de dados morfológicos de terminologias do português*⁶ (2011-2013). Dissertação (Mestrado em Linguística – UFSCar) desenvolvida por Daniela Ferreira de Mattos, com orientação de Gladis Maria de Barcellos Almeida.

A análise morfológica do conjunto de termos sobre as pragas e doenças da cana-de-açúcar permitirá:

- o aprofundamento do conhecimento no nível morfológico sobre a terminologia das pragas da cana-de-açúcar, visto que poderemos identificar os processos de formação de palavras mais frequentes nesse domínio;
- o enriquecimento da descrição do próprio português brasileiro, já que o termo é um signo linguístico;
- a complementação de bases lexicais utilizadas em sistemas de Processamento Automático das Línguas Naturais (PLN), na medida em que poderemos disponibilizar conhecimento linguístico de nível morfológico.

Acreditamos que, quanto maior for o número de terminologias descritas em seu nível morfológico, mais conhecimento linguístico estará disponível para a melhoria do desempenho dos sistemas de extração automática de termos em língua portuguesa.

O trabalho está organizado da seguinte maneira: a seção dois é dedicada aos pressupostos teóricos, onde tratamos da Terminologia e da Morfologia. Na seção três, explicitamos a metodologia empregada. Na seção quatro, expomos a nossa análise. Na seção cinco, apresentamos nossas conclusões seguidas das referências bibliográficas e encerramos com o Apêndice, no qual estão os termos e nossa análise morfológica.

⁴ O aluno usufruiu de bolsa de mestrado da Fapesp (processo 2009/12824-2).

⁵ Docente da UFSCar e pesquisadora do GETerm. Mais informações disponíveis em <<http://lattes.cnpq.br/8648412103197455>>.

⁶ A aluna usufruiu de bolsa de mestrado da Fapesp (processo 2011/05546-6).

Pressupostos teóricos

Esta seção abordará dois temas diretamente relacionados com este trabalho: Terminologia e Morfologia. Na primeira subseção, contextualizaremos a Terminologia enquanto área de estudo, por meio de um breve percurso histórico até chegar à Teoria Comunicativa da Terminologia (TCT), que norteou o processo que envolveu a extração de termos aqui descritos. Por fim, dedicaremos uma subseção à Morfologia, abordando seu conceito, história, algumas classificações de processos de formação de palavras sugeridas por diferentes autores e a classificação que guiou este trabalho.

2.1 Contextualização e importância da Terminologia

A comunicação humana se dá, principalmente, por meio do uso da língua. Ao se apropriar das palavras, o homem consegue expressar seus sentimentos, suas necessidades, suas ideias. Através da linguagem é possível aprender novos conteúdos e conhecer diferentes ciências.

Atualmente, saber compreender diferentes tipos de texto, de vocabulário e de termos é fundamental. Podemos afirmar que essa habilidade é hoje mais importante que nos tempos antigos, pois os avanços são diários. As descobertas científicas, referentes às áreas mais variadas, acontecem diariamente. Todo esse conhecimento não se restringe mais a centros de pesquisa ou a poucos cientistas. Graças, principalmente, à internet, um enorme número de pessoas pode ter acesso a essas informações. Sendo assim, os estudos terminológicos assumem relevância diferenciada, pois através deles o acesso ao conhecimento é facilitado.

Para esclarecer a nomenclatura utilizada, o “T” maiúsculo (Terminologia) refere-se à disciplina enquanto teoria e prática; e o “t” minúsculo (terminologia), ao conjunto de termos de determinada área de especialidade (ALMEIDA, 2012).

Segundo Barros (2004):

O objeto de estudo da Terminologia é o conjunto de termos de um domínio e dos conceitos (ou noções) por eles designados. Os estudos terminológicos fornecem as bases teóricas e metodológicas para inúmeras pesquisas, tendo aplicações nos ensinamentos das línguas (materna e estrangeira), na tradução, na elaboração de obras terminográficas (dicionários especializados), no ensino de disciplinas técnicas e científicas, na documentação, no jornalismo científico, na produção industrial e nas políticas linguísticas [...] (BARROS, 2004, p. 34).

Se hoje a Terminologia pode ser considerada uma subárea da Linguística, antigamente não era assim. Ela surgiu a partir de pesquisas em áreas como a Botânica e a Química no século XIX. Nessa época, os próprios especialistas é que descreviam seus repertórios terminológicos (BARROS, 2004).

Sendo considerada por alguns especialistas como pioneira nos estudos terminológicos, a ex-União Soviética inicia o caminho científico da Terminologia a partir de 1930. O principal nome é o cientista D. S. Lotte, que foi responsável pelo Comitê de Normalização Terminológica do Instituto de Normalização do Conselho de Ministros da então URSS. Ele participou da criação da Comissão de Terminologia Técnica da Academia de Ciências e, em 1937, escreveu *Tarefas e Métodos do Trabalho de Sistematização da Terminologia Técnica*, que só foi publicado em 1961. Seu principal interesse nos estudos terminológicos era a elaboração de métodos de trabalho com fundamentos em teoria de Terminologia. A escola soviética caracteriza-se por uma concepção linguística da Terminologia e concilia teoria e prática, considerando o termo como sendo um elemento linguístico de uso em discursos técnicos e científicos não monossêmicos (BARROS, 2004).

Outro pesquisador que também é considerado o precursor da Terminologia e fundador da Escola de Viena é o austríaco Eugen Wüster (1898-1977), que foi engenheiro, industrial e professor. Na década de 1920, vigoravam na Alemanha estudos de termos técnicos e científicos, objetivando sua normalização. O princípio básico desses estudos era delimitar a univocidade entre termo e conceito, pois assim não haveria ruídos e desentendimentos entre especialistas (BARROS, 2004). De acordo com ele, para cada conceito existiria apenas um termo, ou seja, excluiu de seus estudos a polissemia, a sinonímia e a homonímia.

Dessa forma, é possível reconhecer o aspecto normativo que Wüster conferiu à Terminologia, pois a principal função dessa nova disciplina seria fornecer subsídios científicos para a eliminação da ambiguidade em textos técnicos. A busca pela normatização das terminologias fez com que Wüster desenvolvesse a Teoria Geral da Terminologia (TGT), descrita no próximo item, considerada até hoje a “pedra fundamental” da Terminologia enquanto disciplina (BARROS, 2004).

2.1.1 A Teoria Geral da Terminologia (TGT)

O interesse de estudo de Wüster recaía sobre a ciência da informação e a desambiguação da comunicação técnica e científica. Ele organizou um dicionário de termos padronizados em francês, inglês e alemão, denominado *The machine tool. An interlingual dictionary of basic concepts* (1968), com o intuito de que se tornasse um modelo de dicionário técnico.

Seus principais objetivos no que diz respeito aos estudos terminológicos foram:

- eliminar a ambiguidade da linguagem técnica através de padronizações;
- convencer os usuários da linguagem técnica sobre os benefícios da padronização;
- conferir à Terminologia caráter científico e autonomia enquanto disciplina.

A fim de alcançar esses objetivos, ele buscou elaborar princípios internacionais de padronização para a tarefa de descrição dos termos e também princípios gerais de Terminologia, já que ele a considerava uma ciência autônoma, não pertencente a outra área, como a Linguística, por exemplo. Wüster também quis que fosse criado um centro internacional responsável pela organização e disseminação de informações referentes à Terminologia. Esse centro acabou se tornando o que conhecemos hoje sob o nome de Infoterm, que atualmente está sob a responsabilidade da Unesco.

Segundo Cabré (2003), Wüster não idealizava criar uma teoria da Terminologia, mas sim uma espécie de manual/guia para as tarefas que fossem realizadas futuramente. Foram seus seguidores que atribuíram aos seus trabalhos e princípios o nome de Teoria Geral da Terminologia (TGT).

Os princípios propostos por Wüster diferenciavam a Terminologia dos estudos linguísticos, pois:

- priorizava o conceito em detrimento do termo;
- não admitia polissemia, ou seja, os conceitos deviam ser descritos de forma precisa, pois não eram considerados outros sentidos para o mesmo conceito;
- postulava que para cada termo há um conceito, o que significa que não existe em sua concepção, a possibilidade de sinônimos;
- concentrava-se apenas no nível lexical, não admitindo que os demais níveis linguísticos pudessem conter informações relevantes para a Terminologia;
- guiava seu trabalho sob uma abordagem sincrônica, desconsiderando informações etimológicas, por exemplo;

- priorizava registros escritos aos orais;
- priorizava padrões internacionais para designação;
- utilizava apenas a abordagem onomasiológica (partia do conceito para o termo). Os conceitos devem ser organizados em ordem sistemática dentro de uma hierarquia conceitual e depois cada língua atribui uma “etiqueta léxica” para designá-los, tendo em vista que, para ele, os conceitos são universais.

Sendo assim, é possível concluir que, na visão de Wüster, os termos não são considerados unidades da língua, mas sim unidades que integram campos específicos do saber e são regidos por padrões internacionais.

Wüster também não admitia que diferenças culturais pudessem interferir nas designações dos conceitos. Para ele, assim como o conceito é universal também as designações deveriam ser. Ele idealizava que pudesse existir uma língua internacional e exclusiva para uso técnico e científico. Como ele viu que isso não seria possível, ele passou a defender a padronização de termos técnicos.

Por muito tempo, a TGT não recebeu críticas e questionamentos e Cabré (2003) justifica esse fato, primeiramente, pela data em que a Terminologia se solidificou como disciplina autônoma, na década de 50, ou seja, pode ser considerada uma disciplina recente. A segunda razão é a falta de artigos e discussões com argumentos consistentes o suficiente para oferecer acréscimos significativos à teoria. Também não havia, segundo a autora, um verdadeiro debate sobre os princípios da TGT, que eram considerados invioláveis.

Ressalte-se, ainda segundo a autora, o desinteresse inicial de profissionais de áreas como a Linguística, Psicologia e Filosofia. Pode-se inferir que, nesse período, a Terminologia atuava apenas para sanar necessidades restritas de determinados campos da ciência.

Como consequência disso, a Terminologia não era pensada como uma área que necessitasse de reflexões teóricas a seu respeito e era vista efetivamente como uma prática que ajudaria a solucionar problemas específicos de outras atividades como a tradução, por exemplo. De acordo com Cabré (2003), mesmo nessas atividades práticas, ficava evidente que o princípio de padronização, idealizado por Wüster e defendido na TGT, simplesmente não acontecia. Ainda assim, havia muita resistência em se aceitar de fato que essa teoria carecia de contribuições.

Uma das razões possíveis para essa resistência, segundo Cabré (2003), era que a principal preocupação dos pesquisadores na época era consolidar a Terminologia como uma

disciplina autônoma e legítima. Por esse motivo, era necessário haver controle e centralização sobre a teoria vigente a fim de evitar que, caso essa teoria sofresse muitas contestações, o caráter científico da Terminologia viesse também a ser questionado.

A toda homogeneidade presente no desenvolvimento da Terminologia nas décadas de 60, 70 e 80, seguiu-se uma fase de diversas manifestações, apresentações de novas ideias e questionamentos.

As principais críticas à TGT vieram de três grandes áreas do conhecimento: Ciência Cognitiva, Linguística e Comunicação.

Os adeptos da Ciência Cognitiva questionavam a separação estabelecida entre língua geral e especializada e defendiam que a língua geral contribuía para a aquisição da linguagem especializada.

Os linguistas também questionavam essa separação e, após terem analisado as línguas especializadas, passaram a defender que a língua geral e de especialidade pudessem ser integradas uma à outra.

Os pesquisadores da área da Comunicação também buscavam integrar a linguagem de especialidade aos esquemas e *frames* utilizados na língua geral a fim de que a língua específica não necessitasse ser tratada como um tipo diferente de comunicação.

Essas novas ideias sobre os estudos terminológicos foram sendo levadas em consideração, pois os mais tradicionais reconheciam a possibilidade de contribuição advinda de novos grupos e ideias. Cabré (2003) menciona um estudo feito em 1993 por Laurén e Picht. Nesse trabalho, foram comparadas algumas escolas e teorias: a escola de Viena (representada por Wüster), a escola soviética, a escola de Praga e pesquisas realizadas no Canadá, Alemanha e Escandinávia. A conclusão dessa comparação é que não é possível separar, isolar uma escola/teoria das outras, pois elas estão interligadas e compartilham grande parte de suposições teóricas. Elas se diferenciam quanto às suas prioridades e interesses.

Sendo assim, o que se pode constatar é que, em meio a tantas teorias particulares, havia a necessidade de criação de uma teoria ampla que de alguma forma unificasse todas as demais. Uma teoria que não descartasse totalmente as bases iniciadas por Wüster, mas também abarcasse discussões e contribuições de pesquisas recentes.

Segundo Cabré (2003), a teoria não deveria ser apenas prescritiva, mas essencialmente descritiva. Ela afirma que a teoria desenvolvida por Wüster não era totalmente prescritiva, mas também não podia ser considerada descritiva.

De acordo com ela, toda teoria deve ter por objetivo a descrição de determinado objeto de estudo. Tal descrição deve ser tão detalhada e consistente a ponto de a teoria chegar a ser prescritiva em relação ao que ainda não foi descrito.

2.1.2 *Questionamentos à TGT*

Um dos aspectos da teoria de Wüster que sempre foi alvo de questionamentos e críticas diz respeito ao fato de que as informações que constituíram seu objeto de estudo e descrição eram conteúdos presentes em dicionários. Dessa forma, ela pode ser considerada tendenciosa, visto que nessas obras os termos e suas respectivas acepções são semanticamente precisos, sem ambiguidades e padronizados, desconsiderando o uso efetivo dos mesmos.

Os termos, quando estudados em uso, tanto na comunicação entre especialistas entre si ou com os leigos, na modalidade oral e mesmo na escrita, não se encaixam na teoria de Wüster. A realidade é que existe variação nas designações, e fatores como cultura e idioma influenciam os discursos especializados, não permitindo que existam equivalentes perfeitos entre diferentes línguas.

Outro aspecto que também é questionado é sobre o objetivo da Terminologia. Para Wüster, esse objetivo era evitar a ambiguidade na comunicação profissional/técnica em nível internacional. Por isso a tarefa principal de um terminólogo, nesse caso, é a padronização dos conceitos. Não é sem razão que na TGT há prioridade do conceito em relação à sua designação. Nessa perspectiva, o conceito é independente da sua designação e ambos devem ser padronizados em diferentes línguas.

Segundo Cabré (2003), esse fato é questionável, pois a Terminologia não existe somente para isso, mas sim para outras finalidades. Para exemplificar seu ponto de vista, ela cita a tradução. Para que ocorra uma tradução bem elaborada, principalmente de discursos técnicos, o tradutor deve dominar os termos, seus conceitos, designações e variantes, ou seja, os discursos especializados das línguas com as quais está lidando.

A partir da década de 80, emergiram novas teorias a respeito da Terminologia, teorias essas que eram embasadas em estudos linguísticos. São elas: Socioterminologia, Terminologia de Base Textual, Teoria Sociocognitiva da Terminologia e a Teoria Comunicativa da Terminologia.

A Socioterminologia tem por objetivo a descrição dos termos em seus contextos de uso, abarcando as modalidades escrita e oral, como também os registros formal e informal. A

Socioterminologia aceita a existência da variação terminológica e também considera estudos diacrônicos e sincrônicos (KAMIKAWACHI, 2009).

A Terminologia de Base Textual concentra-se na análise detalhada e profunda de textos especializados, considerando vários aspectos do texto como fazendo parte de um “jeito de dizer” em determinada área ou discurso especializado. Nesse sentido, não é o termo o objeto principal de análise, mas o texto terminológico (KAMIKAWACHI, 2009).

Já a Teoria Sociocognitiva da Terminologia fundamenta-se na semântica cognitiva e defende que, para além de os termos serem descritos levando-se em consideração os contextos em que são empregados, considera que, para uma descrição terminológica plausível, seja necessário conciliar três perspectivas: a nominalista, a mentalista e a realista. A primeira (nominalista) considera que a unidade constitui o sentido do mundo; a segunda (mentalista) postula que a unidade seja uma ideia que existe na mente; e a última (realista) entende que a unidade seja uma entidade externa à língua que existe no universo (TEMMERMAN 2000, p. 224, apud MARTÍNEZ, 2003).

Cabré (2003) traz uma avaliação de Budin (2001) sobre as novas teorias que foram surgindo como alternativas à TGT, como a Socioterminologia e a Terminologia de Base Textual. Nessa avaliação, o autor conclui que essas teorias possuem pouca consistência teórica. Para ele, a emergência dessas teorias é apenas o começo para construção de uma teoria terminológica coerente. O que podemos concluir é que as contribuições dessas teorias devem ser consideradas, mas também aprofundadas e enriquecidas.

A Teoria Comunicativa da Terminologia será apresentada na próxima seção com maior detalhamento, pois foi a teoria que guiou o processo que resultou no conjunto terminológico descrito nesse trabalho.

2.1.3 *Contribuições de Cabré à Teoria Comunicativa da Terminologia (TCT)*

Cabré, juntamente com seu grupo de pesquisa IULATERM,⁷ trabalhou desde 1996 na elaboração de uma teoria terminológica que pudesse abranger diferentes aspectos sobre os termos. Para iniciar a explicação de sua teoria, Cabré (2003) apresenta dois principais pressupostos: o primeiro é que a Terminologia é simultaneamente um conjunto de técnicas para sanar determinadas necessidades e um campo do conhecimento. O segundo diz respeito

⁷ IULATERM (Léxico, terminologia, discurso especializado e engenharia linguística) é um grupo de pesquisa do Instituto Universitário de Linguística Aplicada da Universidade Pompeu Fabra, Barcelona (Espanha), criado em 1994, paralelamente à criação do próprio Instituto.

ao fato de as unidades terminológicas (=termos) serem o objeto de estudo/análise/descrição preferencial da Terminologia.

A Terminologia auxilia todas as tarefas relacionadas à representação e à transferência do conhecimento especializado como: tradução técnica, ensino de línguas para determinados fins, padronização técnica etc. Todas as profissões lidam com discurso especializado. Para conseguir dar conta de tantas possíveis aplicações é que a Terminologia enquanto campo de conhecimento deve possuir uma teoria que tenha bem definido qual seu objeto central de estudo.

Cabré afirma que os termos são unidades do conhecimento (perspectiva cognitiva), unidades linguísticas (perspectiva linguística) e unidades comunicativas (perspectiva comunicativa), razão pela qual a descrição do termo deve abranger sempre esses três aspectos, pois somente assim podem-se diferenciar as unidades terminológicas (termos) das unidades lexicais (palavras). Embora seja possível sempre essa diferenciação, isso não faz com que os termos sejam considerados unidades separadas da língua, ao contrário, para a autora, os termos são unidades que integram a língua geral, como qualquer palavra.

Do ponto de vista cognitivo, para uma unidade lexical ser reconhecida como termo, ela deve preencher os seguintes requisitos:

- depender de um contexto temático;
- ocupar um lugar determinado em uma estrutura conceitual;
- seu significado específico ser definido de acordo com o lugar que ocupa nessa estrutura;
- seu significado ser considerado uma propriedade do termo;
- esse significado específico deve ser explicitamente fixado, reconhecido e disseminado na comunidade temática da qual faz parte.

Considerando a perspectiva linguística, o termo se diferenciará das demais unidades, pois:

- ele possui as mesmas estruturas sintáticas e lexicais das palavras da língua geral, obedecendo aos mesmos processos de formação de palavras;
- ele pertence às categorias semânticas como: evento, entidade, propriedade;
- seu significado é determinado a partir de sua inserção em uma área específica.

Da perspectiva comunicativa, o termo obedece às seguintes condições:

- ocorre em discursos especializados;
- é normalmente denotativo, não costuma dar margens para interpretações variadas;

- é adquirido por meio de um processo de aprendizado e aplicado pelos especialistas em seus campos de trabalho.

Podemos constatar que a unidade terminológica pode ser considerada um poliedro com três diferentes, porém inseparáveis, pontos de vista. Cabré (2003) então defende que uma teoria da Terminologia deva abranger todos esses aspectos das unidades terminológicas. Isso não acontece na TGT, visto que seu objeto central de estudo são os conceitos e suas relações e não as unidades terminológicas. A TGT adota uma abordagem onomasiológica, na qual o conceito não só é independente de sua designação, mas a precede. Do ponto de vista linguístico, o objeto de estudo central deve ser o termo que é concebido como signo linguístico. Isso significa que forma e conteúdo são inseparáveis, e para Wüster eles são independentes.

Cabré (2003) questiona se é possível conciliar essas três concepções a respeito do objeto da Terminologia. Ela propõe um modelo denominado “teoria das portas,” que deve permitir que se alcance o objeto central, o termo, por meio de diferentes formas. Assim, considerando novamente o termo como o poliedro com três lados (cognitivo/linguístico/comunicativo), é possível transformar cada lado em uma porta de acesso ao termo.

Vale lembrar que a concepção de termo da TGT, que prioriza o aspecto conceitual, coincide com a perspectiva cognitiva e, portanto, pode se tornar uma porta de acesso ao termo, juntamente com os demais aspectos do poliedro, que não devem ser deixados de lado, já que, segundo a autora, uma teoria pode apresentar apenas uma porta de entrada, mas não pode negar a multidimensionalidade do objeto. Esse reconhecimento era ausente na TGT, por isso Cabré passou a buscar o desenvolvimento de uma teoria ampla que pudesse abranger toda essa complexidade a respeito do objeto central de estudo.

A fim de explicar a fundamentação de uma teoria terminológica baseada em conhecimentos linguísticos, a autora elabora algumas questões que ajudarão nesse processo.

A questão preliminar é: qual o ponto de partida dessa nova teoria? Ela parte de uma base teórica ampla que lida com os termos a partir de diferentes concepções e posições, a fim de respeitar a multidimensionalidade dos mesmos. Essa teoria lida com o termo pela porta linguística, o que significa que os termos serão descritos por intermédio de uma teoria da língua natural.

A próxima questão se refere ao contexto em que as unidades terminológicas devem ser estudadas. Segundo a autora, deve ser no contexto do discurso especializado. O significado do termo deve ser preciso e restrito ao domínio do saber no qual é inserido a fim de que a comunicação e a transferência de conhecimento especializado seja eficiente e clara.

Outra questão relevante é como deve ser uma teoria linguística para que seja possível o estudo terminológico a partir dela. Cabré defende que uma teoria da língua natural considera as unidades terminológicas como unidades lexicais com significados especiais. Isso significa que qualquer unidade lexical possui potencial para ser um termo. O fator determinante no processo de reconhecimento de um termo é seu contexto de uso, ou seja, como e em qual discurso está sendo utilizado. Segundo a autora, a teoria também não deve considerar apenas os aspectos gramaticais da língua, mas também os aspectos sociais, políticos e, principalmente, pragmáticos. Em outras palavras, deve ser uma teoria linguística de caráter funcional e cognitivo.

É com base nesses pressupostos que a Teoria Comunicativa da Terminologia (TCT), elaborada por Maria Teresa Cabré e o IULATERM, desponta como uma das respostas teórico-metodológicas viáveis para os questionamentos já explicitados sobre a TGT no que diz respeito ao reconhecimento da Terminologia como disciplina linguística (ALMEIDA, 2012).

Segundo Almeida (2006), há um conjunto de pressupostos que particularizam a TCT em relação às demais teorias terminológicas:

- a) o objeto central da Terminologia são as unidades terminológicas e não os conceitos. Eleger as unidades como objeto central significa reforçar uma perspectiva linguística e uma abordagem semasiológica;
- b) não há uma diferença *a priori* entre termo e palavra, o que há são signos linguísticos que podem realizar-se no discurso como termo ou palavra dependendo da situação comunicativa;
- c) os níveis lexical, morfológico, sintático e textual podem veicular conhecimento especializado;
- d) os termos devem ser observados no seu ambiente natural de ocorrência, ou seja, nos discursos especializados;
- e) a variação conceitual e denominativa deve ser considerada;
- f) do ponto de vista cognitivo, as unidades terminológicas: i) estão subordinadas a um contexto temático, ii) ocupam um lugar preciso num mapa conceitual, iii) o seu significado específico é determinado pelo lugar que ocupam nesse mapa (CABRÉ, 2003, apud ALMEIDA, 2006, p. 86).

O objeto central da Terminologia são as unidades terminológicas e não os conceitos: ao acessar a Terminologia pela porta linguística deve-se considerar que o mais importante não é apenas aquilo que o termo representa, mas sim que o termo é a união daquilo

que significa com a sua forma. Em outras palavras, é um signo linguístico composto por um significante e um significado.

Não há uma diferença *a priori* entre termo e palavra: para a TCT, o termo é um signo linguístico que pode assumir um valor especializado quando utilizado em determinado domínio. Isso significa que o conjunto de termos não representa um conjunto linguístico separado do sistema linguístico geral. Segundo Almeida (2003, p. 17), “[...] 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”.

Os níveis lexical, morfológico, sintático e textual podem veicular conhecimento especializado: a partir de pesquisas já realizadas, é possível constatar que é possível identificar padrões linguísticos que representam conhecimento técnico. Esses padrões não aparecem apenas no nível lexical, mas também podem ocorrer no nível morfológico, sintático e textual. Podemos encontrar, por exemplo, a presença excessiva de determinados elementos mórficos (nível morfológico), estruturas oracionais próprias (nível sintático), ou mesmo fórmulas retóricas particulares de cada ciência ou domínio (nível textual).

Os termos devem ser observados no seu ambiente natural de ocorrência: para que seja possível a distinção entre termo e palavra, o contexto de uso é fundamental. Em nosso trabalho, temos o exemplo de “amarelinho” que, em contextos de língua geral, remete à cor amarela enquanto que no contexto da cana-de-açúcar se refere a uma doença. Por isso, em trabalhos orientados pela TCT, a compilação de *corpus* é de grande importância, pois é ele que permitirá o reconhecimento do termo. É no discurso especializado que o termo deve ser analisado.

A variação conceitual e denominativa deve ser considerada: a partir do momento em que se utiliza um *corpus* como fonte para buscar os termos, o terminólogo deve estar atento para identificar casos em que termos diferentes designam o mesmo conceito e termos iguais que designam conceitos diferentes, pois assim como acontece na língua geral, a língua de especialidade também admite sinonímia e polissemia/homonímia, ou seja, a variação linguística. Tal variação pode acontecer também devido ao meio em que circulam os termos, exemplo: possivelmente os agricultores que lidam com a cultura da cana-de-açúcar se referem

à “broca da cana” enquanto pesquisadores designam a mesma praga pelo nome científico *diatraea saccharalis*.

Do ponto de vista cognitivo, as unidades terminológicas: i) estão subordinadas a um contexto temático, ii) ocupam um lugar preciso num mapa conceitual, iii) o seu significado específico é determinado pelo lugar que ocupam nesse mapa: como já foi mencionado, um signo linguístico assume valor de termo quando é utilizado em uma determinada área para designar algo próprio daquela área, de maneira que é justamente o contexto temático que vai dar ao signo o estatuto de termo. Ressalte-se que os termos que funcionam em dado contexto temático estão dispostos numa espécie de estrutura conceitual, de forma que a posição do conceito nessa estrutura revela seu significado. Como exemplo, podemos citar a estrutura conceitual das pragas e doenças elaborado no TCC que antecedeu esse trabalho. Por meio dele, pudemos inferir que **cigarrinha-da-raiz** é um tipo de praga de raiz que, por sua vez, é um agente causador de **doença insetológica** que é um tipo de doença da cana-de-açúcar.

Após a exposição desses pressupostos, assumimos que todo o percurso terminológico que seguimos e que culmina aqui nesta pesquisa teve como orientação de base a TCT.

2.2 Morfologia: conceito e história

Adotaremos neste trabalho o conceito de Cabral (1979, p. 134), segundo o qual a Morfologia é entendida como “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”.

O termo *morfologia* tem origem grega e significa: *morph-* (=forma) e *-logia* < [*log(o)* + *ia*] (= ciência, arte, estudo, tratamento sistemático de um tema) (FERREIRA, 2009; HOUAISS, 2001). Seu emprego não se restringe apenas à ciência linguística, mas também a outras áreas do conhecimento, notadamente a Biologia. A propósito disso, vale ressaltar que, segundo Aronoff & Fudeman (2005), a criação do termo *morfologia* é geralmente atribuída ao poeta, romancista, ensaísta e dramaturgo Johann Wolfgang Von Goethe (1749-1832), no contexto da Biologia, por volta de 1830, tendo em vista que o eminente escritor alemão também fez incursões pelas ciências naturais.

Até então, o termo *morfologia* não fazia parte dos estudos da linguagem, embora o conhecimento envolvendo a flexão e a derivação das palavras fosse considerado pela gramática greco-latina, que era dividida em três partes: flexão, derivação e sintaxe.

Por volta de 1860, *morfologia* passou a ser um termo empregado também nos estudos linguísticos. O seu uso nesse campo deveu-se muito à influência do darwinismo, posto que, assim como postula a teoria evolucionista: “os gramáticos e filósofos alimentavam o sonho de descobrir a origem da linguagem através do estudo da evolução das palavras em indo-europeu.” (LAROCA, 2004, p. 12). Ganha relevo, assim, o estudo dos processos de formação de palavras a partir de uma abordagem diacrônica, já que os estudiosos da linguagem consideravam seus elementos mínimos como suas respectivas formas originárias (LAROCA, 2004).

Ainda na busca pelas formas originárias, há outro fator que também influenciou o desenvolvimento da Morfologia: a descoberta do sânscrito no fim do século XVIII, descoberta essa atribuída a William Jones (1746-1794). Observando a descrição da antiga língua falada na Índia por meio da notável gramática de Pāṇini (século VI a.C.), foi possível perceber a sofisticação do sânscrito com respeito, principalmente, à Fonética/Fonologia e à Morfologia. No que tange a esta última, observou-se que as palavras poderiam ser segmentadas em unidades menores, como raízes e afixos. Além disso, ao contrastar paradigmas morfológicos do sânscrito com o do grego, do latim e o do alemão, chegou-se à conclusão de que todas poderiam pertencer a uma única família, como bem descreve Franz Bopp na sua obra intitulada *Sistema da conjugação do Sânscrito* (SAUSSURE, 1972, p. 8).

De acordo com Rocha (2008), podemos identificar quatro correntes principais que se ocuparam do nível morfológico: Descritivismo, Historicismo, Estruturalismo e Gerativismo.

O Descritivismo abrange os estudos gramaticais dos filósofos gregos, que pretendiam entender a língua por meio da lógica. Seus principais questionamentos envolviam as regularidades e irregularidades entre os conceitos e as línguas. Por esse motivo, eles buscaram estabelecer paradigmas e, por isso, a esse modelo de estudo se deu o nome de “Elemento e Paradigma” (ROCHA, 2008).

O Historicismo compreende os estudos de caráter filológico, ou seja, desejava-se encontrar a origem das línguas, saber qual idioma originou qual. Tais estudos tinham caráter diacrônico e os produtos desses estudos eram as chamadas *gramáticas históricas* ou *gramáticas comparativas*. Nessa época, os traços morfológicos começaram a surgir no

cenário linguístico, entretanto, o foco era a evolução das línguas, e não os processos de formação das palavras em si (ROCHA, 2008).

A partir de 1916, ano de publicação do *Cours de Linguistique Générale* de Ferdinand de Saussure, os estudos morfológicos passaram a conquistar mais espaço na Linguística. O cerne do Estruturalismo europeu foi comprovar que na língua o valor se dá na relação dos elementos entre si e, por isso, a língua é considerada um sistema de valores. Uma palavra, por exemplo, adquire sentido a partir da relação entre seus constituintes mínimos (ROCHA, 2008).

É importante enfatizar que, nos Estados Unidos, o Estruturalismo teve um amplo desenvolvimento. Os principais nomes do Estruturalismo norte-americano são Edward Sapir e Leonard Bloomfield, que empreenderam muitos esforços na descrição de línguas indígenas daquele país. Nesse processo, eles não só estudaram as estruturas dessas línguas, mas passaram a descrevê-las até o nível do morfema. Ressalte-se, ainda, a contribuição de Bloomfield (1933) ao conceituar o morfema como sendo a menor unidade significativa da palavra. A afirmação de que a combinatória de morfemas se dá no nível da palavra e a combinatória de palavras se dá no nível da sentença deve-se a Bloomfield. Nesse sentido, o autor já anunciava a autonomia da Morfologia em relação à Sintaxe.

De acordo com Rocha (2008), do ponto de vista da Morfologia,

[...] podemos dizer que a visão estruturalista desenvolveu com bastante rigor as técnicas de apreensão dos morfemas e essa foi a sua preocupação básica como movimento linguístico. Em síntese, o Estruturalismo preocupou-se em: a) fazer a segmentação dos morfemas; b) proceder à classificação dos morfemas (ROCHA, 2008, p. 28).

No Brasil, a figura de destaque no Estruturalismo é Mattoso Camara Jr., que publicou suas obras nas décadas de 1940, 1950 e 1960.

As bases dessa vertente do Estruturalismo (sobretudo aquele praticado por Bloomfield) eram influenciadas pela psicologia behaviorista, ou comportamentalismo. Trata-se de uma linha da Psicologia que defende a teoria de estímulos e respostas. Alguns estruturalistas americanos assumiam que a língua fora criada para responder as necessidades (estímulos) do ambiente dos falantes. Para contradizer essa ideia, surge o Gerativismo de Chomsky, que lança em 1957 seu livro *Syntactic Structures*, marco inicial dessa corrente. Chomsky refutava a ideia de que a língua fosse uma resposta a qualquer estímulo: “[...] a linguagem humana é livre de controle de estímulos e não serve a uma função meramente comunicativa, mas é antes um instrumento para a livre expressão do pensamento e para a

resposta apropriada às novas situações” (CHOMSKY, 1972, p. 23 apud ROCHA, 2008, p. 29).

Podemos constatar que, para o Gerativismo, a linguagem humana é, em sua essência, “órgão do pensamento [...] e só secundariamente serve à finalidade de comunicação social” (ROCHA, 2008, p. 29). As ideias gerativistas inovam os estudos linguísticos desde a concepção da própria língua. A língua deixa de ser um instrumento para “se confundir com a essência do próprio homem” (ROCHA, 2008, p. 30).

A Morfologia não foi o ponto central da teoria gerativista, mas, sim, a Sintaxe. Apenas a partir de 1970 o Gerativismo reconheceu a autonomia da Morfologia e sua importância para a teoria, que afirmava existir um componente lexical no qual operam regras morfológicas. Ao contrário do Estruturalismo, que buscava descrever línguas já “prontas”, o Gerativismo se deteve em descobrir como funciona o mecanismo do falante de reconhecer palavras, rejeitá-las, criar novos termos, estabelecer relações entre eles. O objetivo era descobrir quais são as regras que orientam os falantes. A tal modelo de abordagem linguística denominou-se *Elemento e Processo* (ROCHA, 2008, p. 31).

Ressalte-se que a morfologia derivacional (ou lexical) teve um amplo desenvolvimento na abordagem gerativa a partir da década de 1970. Destacam-se especialmente os trabalhos de Jackendoff e Aronoff. No Brasil, assume-se que a primeira pesquisadora a tratar do tema pela ótica gerativa foi Margarida Basílio, que publica sua tese de doutorado em 1980, a qual se intitula *Estruturas lexicais do português: uma abordagem gerativa* (LAROCCA, 2004, p. 12).

2.2.1 *Morfologia flexional e derivacional*

Considera-se que os estudos morfológicos podem ser divididos em duas grandes partes: a morfologia flexional (ou gramatical) e a morfologia derivacional (ou lexical).

A morfologia flexional tem como objeto o estudo das relações entre as diferentes formas de um mesmo paradigma (LAROCCA, 2004), e como tarefa a de “determinar os morfemas que servem às diversas categorias flexionais que a língua utiliza” (BORBA, 1991, p. 166). Nesse sentido, assume-se que a flexão tem motivações morfossintáticas e é obrigatória nas sentenças, como em “*O menino é arteiro*”, pois o artigo, o substantivo e o adjetivo devem seguir um determinado padrão, que no caso aqui é o masculino singular.

A morfologia derivacional trata da estrutura das palavras e dos seus processos de formação, observando a relação entre as formas de paradigmas diferentes, como, por exemplo, *belo* (adjetivo) e *beleza* (substantivo) (LAROCA, 2004). Nesse sentido, a morfologia derivacional cumpre um importante papel no enriquecimento do léxico. Ressaltamos que o foco nesta pesquisa é a morfologia derivacional, pois ela se ocupa dos processos de formação das palavras, que é o interesse do trabalho.

2.2.2 Morfologia derivacional e o enriquecimento do léxico

O léxico de uma língua é composto não somente por palavras, mas também pelas unidades mínimas de significado que compõem uma palavra, tais como raízes e afixos. Em geral, assume-se que as raízes carregam a significação lexical (ou referencial); e os afixos, a significação gramatical ou instrucional (CORREIA & ALMEIDA, 2012).

As unidades de significado lexical, ou raízes, não são autônomas, pois não podem ocorrer sozinhas. Elas só atuam como componentes de outras palavras e não admitem flexão. São exemplos de raízes *agro-* e *clor(o)-*. Já os afixos (prefixos e sufixos) “são unidades que têm apenas significado gramatical ou instrucional e que, associadas a unidades de significado lexical ou referencial, permitem a construção de novas palavras” (CORREIA & ALMEIDA, 2012, p. 12).

Assim como as partes constituintes assumem significados de ordem lexical e gramatical, também as palavras em si podem ser classificadas dessa maneira. As palavras de significado lexical são as chamadas *palavras de classe aberta*, tais como substantivos, verbos, adjetivos e advérbios, visto que admitem novas formações (neologismos). As palavras de significado gramatical constituem a chamada *classe fechada* e incluem preposições, conjunções, artigos, pronomes e numerais (CORREIA & ALMEIDA, 2012).

Os mecanismos utilizados pelo léxico de uma língua para a criação de palavras novas são muito semelhantes. O que varia, em geral, é a classificação sugerida pelos autores. A seguir, apresentamos algumas classificações já arroladas em Almeida (1995).

Para Matoré (1953), as palavras podem ser criadas de três modos: 1) por meio de uma palavra nova – nesse item estariam inseridas as criações *ex nihilo*, as criações onomatopéicas, as de nomes próprios, as formações por sufixação e prefixação e os empréstimos; 2) por meio de uma palavra já existente no léxico, mas com uma aceção nova; 3) por meio de uma mudança de categoria gramatical.

Biderman (2001, p. 203) distingue dois tipos de processos: 1) o conceptual, “acepção nova que se incorpora ao campo semasiológico de um significante qualquer”; 2) o formal, que se constitui numa “palavra nova introduzida no idioma”, podendo ser tanto um termo vernáculo como um empréstimo estrangeiro. Dentre os estrangeirismos, a autora reconhece ainda três tipos: o decalque, a adaptação da forma estrangeira à fonética e à ortografia portuguesas, variante brasileira, e a incorporação do vocábulo com a sua grafia e fonética originais.

Guilbert (1975) propõe quatro tipos de processos: 1) o fonológico, que diz respeito à formação da substância do significado e à sua transcrição; 2) o sintagmático, que engloba todos os modos de formação que implicam a combinação de elementos diferentes, ela é morfossintática e reúne também todas as formas de derivação; 3) o semântico, que se refere a uma mudança semântica sem a criação de uma forma significante nova, sendo do domínio do significado; 4) o empréstimo, com suas diferentes formas de manifestação.

Segundo Rondeau (1983), os processos podem ser agrupados em três categorias: 1) os formais, que envolvem a criação de uma nova palavra; 2) os semânticos, que preveem a ligação de um novo significado a um significante já existente; 3) os empréstimos, envolvendo a entrada numa língua de um “signo linguístico completo”, ou seja, com seu significante e significado.

Já Alves (2004, p. 11) divide em seis os processos de formação de palavras, a saber: 1) fonológico, que trata da criação de um item léxico cujo significante seja totalmente inédito, ou seja, que tenha sido criado com base em nenhuma palavra já existente; 2) sintático, o qual diz respeito à combinatória de elementos já existentes no sistema linguístico – neste processo neológico, estariam os neologismos formados por derivação prefixal, derivação sufixal, composição, composição sintagmática e composição por siglas ou acronímica; 3) conversão ou derivação imprópria, que é um tipo de formação lexical pelo qual uma unidade léxica sofre alterações em sua distribuição sem que haja manifestação de mudanças formais; 4) semântico, que trata da criação de um novo elemento a partir de uma transformação semântica manifestada num item lexical; 5) os formados por empréstimos, incluídos aqui aqueles que se utilizam de bases de língua estrangeira; e 6) os processos menos produtivos, mas que também contribuem para o enriquecimento lexical da língua – truncação, palavra-valise, reduplicação e derivação regressiva. A truncação é um tipo de abreviação em que uma parte da sequência lexical, geralmente a final, é eliminada (ex.: *depressão* > *deprê*). Palavra-valise é um tipo de redução em que duas bases são privadas de parte de seus elementos para constituírem um

novo item léxico (ex.: das bases *português* e *espanhol* resultaria o neologismo *portunhol*). Reduplicação é um recurso morfológico em que uma mesma base é repetida duas ou mais vezes a fim de constituir um novo item léxico (ex.: “O tranca-tranca pelo bloco asiático está a mil no início da era.” F, 21/02/89 apud Alves (2004)). E, finalmente, derivação regressiva, que é um tipo de criação lexical em que ocorre a supressão de um elemento, considerado de caráter sufixal (ex.: *amassar* > *amasso*).

Nesta pesquisa, adotaremos a classificação de Correia & Almeida (2012), cuja explanação faremos a seguir. A escolha dessa classificação se justifica por apresentar mais subespecificações de tipos de processos e, quanto mais especificações, maior é a probabilidade de se encontrar características da terminologia em questão.

De acordo com as autoras, os processos disponíveis para a inovação lexical são divididos em três grandes tipos:

1. criação de palavras *ex nihilo*;
2. onomatopeias;
3. construção de palavras dentro do sistema do português.

Este último processo se subdivide em:

- 3.1 derivação (não afixal ou conversão, prefixal, sufixal e parassintética);
- 3.2 composição (morfológica, morfossintática e recomposição);
- 3.3 processos deformacionais (amalgamas, truncação, siglas e acronímia);
- 3.4 lexicalização de estruturas (lexicalização de formas flexionadas, de sintagmas, de unidades discursivas);
- 3.5 extensão semântica;
- 3.6 importação de palavras.

A seguir, faremos a explanação detalhada de cada um deles.

1. Criação de palavras *ex nihilo*: *ex nihilo* significa “do nada”, ou seja, afirmar que uma palavra é uma criação *ex nihilo* é o mesmo que dizer que ela foi criada sem nenhum tipo de motivação já existente. São raras as formações *ex nihilo*, pois, segundo Correia & Almeida (2012),

[...] existe evidência de que, no léxico mental, as palavras se encontram armazenadas em função das relações de vários tipos que estabelecem entre si, relações formais, morfológicas, semânticas e referenciais, combinatórias, por consequência, parece óbvio o papel da motivação na construção de novas unidades. A inexistência dessa motivação dificulta o armazenamento na memória, a recuperação da informação e o processamento das palavras, por isso o processo contraria os princípios básicos do funcionamento psicológico e neurológico da linguagem humana (CORREIA & ALMEIDA, 2012, p. 34).

2. Onomatopeias: trata-se de “unidades léxicas cuja forma pretende reproduzir um som da realidade.” (CORREIA & ALMEIDA, 2012, p. 34). Bechara (1999, p. 74, grifos do autor) enriquece um pouco mais a definição quando afirma: “Exemplos bem conhecidos podem ser: *bem-te-vi*, *zum-zum-zum*, *tic-tac* etc.”.

3. Construção de palavras dentro do sistema do português:

3.1. Derivação: segundo Correia & Almeida (2012), a derivação é o processo mais produtivo e regular nas línguas românicas. Segundo as autoras, é importante ressaltar que a derivação “é um processo binário”, o que significa a união de apenas uma base ou radical derivacional a um afixo de cada vez. Isso não impede que o processo não possa se repetir na formação de uma palavra, como, por exemplo, *desvalorizar*. Primeiramente, temos a base *valor* na qual se junta o afixo *-izar*, formando *valorizar*. Em seguida, temos o segundo processo que adiciona no início da palavra o afixo *des-*, gerando, assim, a palavra *desvalorizar*. Os elementos que fazem parte da derivação são a base ou radical (unidade que contém o significado lexical) e os afixos.

Os afixos derivacionais no português são uma classe fechada, podendo ser prefixos e sufixos, e

[...] são portadores de uma instrução categorial e semântica que lhes permite, em conjunto com a informação da base e com a informação da regra a que pertencem, determinar a categoria do derivado, construir o seu significado e em muitos casos delimitar a sua capacidade referencial, isto é, as classe de entidades extralinguísticas que é capaz de denominar (CORREIA & ALMEIDA, 2012, p. 41).

De acordo com as autoras, a derivação pode ser de dois tipos:

- **derivação não afixal ou conversão** – existe a mudança de categoria, mas não há qualquer tipo de adição de afixos. Exemplos: *genérico*, no sentido de “medicamento comercializado com o nome do princípio ativo que o integra e que deve prevalecer sobre a marca comercial” (FERREIRA, 2009), pois, com este sentido, a palavra deixa de ser um adjetivo para ser um substantivo (também denominado *derivação imprópria*); *entregar* > *entrega*, *combater* > *combate*, *fugir* > *fuga* (também denominado *derivação regressiva*).

- **derivação afixal** – neste tipo de derivação, há a junção de um afixo a uma base. São três os processos desse tipo de derivação:
 - derivação sufixal ou sufixação – esse processo possui as seguintes regularidades: o sufixo sempre será adicionado à direita da base e vai determinar a categoria do derivado, assim como a sílaba tônica da palavra. Exemplo: *amarelo* > *amarelinho*;
 - derivação prefixal ou prefixação – assim como a sufixação, esse processo também possui algumas regularidades: ser adicionado sempre à esquerda da base e ter conteúdo menos gramatical que o sufixo, além de nem sempre determinar a categoria do derivado como a sufixação. Exemplo: *biótico* > *antibiótico*;
 - derivação parassintética ou parassíntese – consiste na adição simultânea de dois afixos à base. Para diferenciar dos casos em que há mais de um processo para gerar a palavra, basta verificar se não existe uma fase intermediária; por exemplo: *alaranjado* → não existe *alaranja*, portanto, a junção é simultânea do prefixo *a-* e do sufixo *-ado*, constituindo-se, portanto, num caso de parassíntese, *laranja* > *alaranjado*;

3.2. Composição: a composição pode ser dividida em morfológica, morfossintática e em recomposição:

- **composição morfológica** – também denominada *composição neoclássica* ou *erudita*, “consiste na construção de palavras compostas a partir de unidades de significado lexical e não autônomas. “Na construção de compostos morfológicos intervém, frequentemente, uma vogal de ligação, *-o-* ou *-i-*.” (CORREIA & ALMEIDA, 2012, p. 52). Exemplos: *ec(o-) + -s- + sistema* > *ecossistema*, *agr- + -i- + cultura* > *agricultura* etc.
- **composição morfossintática** – são as palavras compostas geralmente por *nome + nome*, nas quais o composto é hipônimo de um dos nomes que o constitui. Por exemplo: *seguro-desemprego* é um tipo de *seguro*. Exemplos: *couve-flor*, *operação-tartaruga*, *cirurgião-dentista* etc.
- **recomposição** – consiste na junção de falsos prefixos ou pseudoprefixos a uma base. Esses falsos prefixos aparentam ser prefixos, mas são a forma reduzida

de outras palavras. Por exemplo: *eco-* em *ecoturismo* (redução de *ecologia*). Correia & Almeida (2012, p. 56) citam os pseudoprefixos mais frequentes: *agro-* em *agropecuária* (redução de *agricultura*), *euro-* em *eurodeputados* (redução de europeu), *foto-* em *fotobiografia* (redução de *fotografia*), *narco-* em *narcotráfico* (redução de *narcótico*). Um exemplo que diferencia a recomposição da composição morfológica e derivação prefixal é o uso do prefixo “nano”, que significa um bilionésimo, ou seja, algo muito pequeno. No caso do termo “nanolaboratório de engenharia” se trata de uma recomposição, pois nano- não é um prefixo, já que seria irreal um laboratório com essa dimensão. Nesse caso é a redução da palavra “nanotecnologia”.

3.3. Processos deformacionais: segundo Correia & Almeida (2012, p. 56), trata-se dos “processos envolvidos na construção de amálgamas, trunções, siglas e acrônimos, já que constituem uma deturpação da integridade morfológica dos radicais envolvidos na construção de seus produtos.” Observe-se o detalhamento a seguir:

- **amálgama** – junção de pedaços de outras palavras. Exemplos: *aborrecimento* + *adolescente* > *aborrescente* etc.
- **trunção ou abreviação vocabular** – redução de uma palavra. Exemplos: *metropolitano* > *metrô*, *otorrinolaringologia* > *otorrino*, *churrasco* > *chúrras*, *psicologia* > *psico* etc.
- **siglas e acrônimos** – as siglas são as junções das iniciais das palavras que integram um sintagma denominativo, como, por exemplo: *CPF* (*Cadastro de Pessoa Física*) e *USP* (*Universidade de São Paulo*). As siglas são frequentes no discurso especializado e por isso passam a ser termos reais dos repertórios terminológicos. O acrônimo, por sua vez, é a junção de letras que não necessariamente são as primeiras de cada palavra que integra o sintagma. Essa junção torna-se uma palavra, pois possui a mesma estrutura da língua; por exemplo: *UFSCar* (*Universidade Federal de São Carlos*) e *GETerm* (*Grupo de estudos e pesquisas em Terminologia*).

3.4. Lexicalização de estruturas: “processo pelo qual determinadas unidades construídas em outros componentes da gramática (sintático, morfológico, discursivo) se transformam em unidades lexicais” (CORREIA & ALMEIDA, 2012, p. 59). São três os tipos de lexicalização de estruturas:

- **lexicalização de formas flexionadas** – segundo Correia & Almeida (2012, p. 59) é “[...] a entrada no componente lexical de unidades cujo lema corresponde a uma forma flexionada de outra palavra”. Exemplos: nominalização deverbal (*olhar* – verbo, (*o*) *olhar* – substantivo) e adjetivação de formas no particípio passado (*queimado*, *aguado*). Ainda segundo as autoras, “alguns nomes de ação são também formas de particípio passado,” em construções com o verbo suporte *dar*. Exemplos: (dar) uma *estudada*, (dar) uma *limpada*, (dar) uma *lida*, (dar) uma *caminhada* (CORREIA & ALMEIDA, 2012, p. 59).
- **lexicalização de sintagmas** (ou *compostos sintagmáticos*) – é quando sintagmas linguísticos passam a funcionar como itens léxicos da língua. Exemplos: *identificação taxonômica*, *linhagem endofítica*, *gorgulho da cana-de-açúcar* etc. Esses compostos sintagmáticos geralmente seguem os seguintes padrões:
 - N + Adj (*sistema antioxidante*);
 - Adj + N (*azul bebê*);
 - N + SPrep (*sala de espera*);
 - Prep + N (*sem-terra*).
- **lexicalização de unidades discursivas** – ocorre quando frases ou segmentos das frases passam a funcionar como itens léxicos da língua, “assumindo a categoria morfossintática dos nomes.” Exemplos: *maria vai com as outras*, *bom de bico*, *sem noção* (CORREIA & ALMEIDA, 2012, p. 62).

3.5. Extensão semântica: consiste na atribuição de um novo significado a um termo já existente, o que provoca a polissemia linguística, que, segundo Correia & Almeida (2012, p. 63), “é um dos mecanismos que mais contribuem para a economia dos sistemas linguísticos”. Como exemplo, podemos citar a palavra *vírus*, proveniente da área de saúde e, desde o advento do computador pessoal, passou a ser associada a um “programa estranho ao sistema de computador capaz de copiar e instalar a si mesmo, geralmente concebido para provocar

efeitos nocivos ou estranhos à funcionalidade do sistema ou aos dados nele armazenados” (FERREIRA, 2009).

3.6. Importação de palavras: o intercâmbio de palavras estrangeiras é natural entre as línguas, especialmente com a globalização e a facilidade do acesso à informação. As palavras importadas são denominadas *empréstimos*. Estabelece-se, em geral, uma diferenciação entre *estrangeirismo* e *empréstimo*. Estrangeirismo são palavras de outro idioma que não sofreram qualquer tipo de alteração na língua de chegada, por exemplo, *shopping center*. Empréstimo é uma palavra importada, adaptada ao idioma receptor, por exemplo, *estresse*. Entretanto, por meio do estudo dessas palavras importadas, constatou-se que todas sofrem algum tipo de modificação/adaptação quando são inseridas em outra língua. Essas modificações/adaptações podem ser de ordem fonológica, morfológica, sintática ou semântica (CORREIA & ALMEIDA, 2012). Nesse sentido, as autoras questionam se essa diferenciação entre estrangeirismo e empréstimos é de fato produtiva. Segundo Correia & Almeida (2012), o termo *empréstimo* inclui tanto as unidades lexicais que foram importadas de outro idioma (empréstimo externo) como também o “processo de transferência de um registro linguístico para outro dentro da mesma língua (‘empréstimo interno’)” (CORREIA & ALMEIDA, 2012, p. 70).

A seguir, detalharemos a metodologia utilizada na pesquisa, explicitaremos trabalhos de descrição morfológica de terminologias já realizados no âmbito do GETerm e também traçaremos o percurso terminológico que originou o conjunto de termos descritos nesse trabalho.

Metodologia

Nesta pesquisa, utilizamos como orientação metodológica as fases da pesquisa em Processamento da Língua Natural (PLN) de Dias-da-Silva, apresentada em 3.1, e orientamos por outras pesquisas de morfologia derivacional já desenvolvidas no GETerm, detalhadas em 3.2.

3.1 A metodologia de Dias-da-Silva

Para levar a cabo nossa análise morfológica, iremos nos orientar segundo a metodologia proposta por Dias-da-Silva (1996). O autor apresenta uma organização clara de como deve se dar o tratamento lexical e sua relevância para estudos em PLN. Essa escolha também se deve ao fato de que os principais objetivos deste trabalho são o enriquecimento de bases de dados e a disponibilização desses dados para o eventual aprimoramento de programas híbridos de extração automática de termos. Sendo assim, encontramos na metodologia apresentada pelo autor um caminho que permite o desenvolvimento, ao mesmo tempo, específico e contextualizado da pesquisa. Específico, pois essa metodologia reconhece o lugar e a função da natureza do nosso trabalho; contextualizado, pois ela permite a visualização do todo em um trabalho de PLN, possibilitando antever os passos que serão tomados adiante e que não cabem mais ao linguista, mas, sim, aos profissionais da Informática.

Dias-da-Silva (1996) demonstra como o PLN se compõe de conhecimentos originários de áreas diversificadas e por isso é necessária a criação de uma metodologia que alcance todos esses campos do saber. Isso porque, desde a criação dos computadores, a grande ocupação dos cientistas tem sido desenvolver meios para que a máquina consiga decifrar instruções e executá-las. Nesse sentido, foram criadas as linguagens de programação. No início, a estrutura dessas linguagens era extremamente lógica e pouco parecida com a linguagem humana. Com aprofundamento dos estudos na área e a disseminação dos computadores por leigos, o desafio de fazer com que a máquina fosse capaz de entender a linguagem humana foi tomando lugar de destaque e inquietando cada vez mais pesquisadores. Segundo Dias-da-Silva (1996):

A grande meta prevista para as pesquisas dessa natureza é indiscutivelmente ousada: projetar e implementar sistemas computacionais avançados em que a comunicação entre o homem e o computador possa realizar-se por meio de línguas naturais, e não por meio de instruções e comandos codificados numa linguagem de programação artificialmente construída por programadores (DIAS-DA-SILVA, 1996, p. 13).

Vejamos a Figura 1, em que o autor sintetiza os “principais recursos teóricos e metodológicos de que o estudo em PLN dispõe” (DIAS-DA-SILVA, 1996, p. 83).

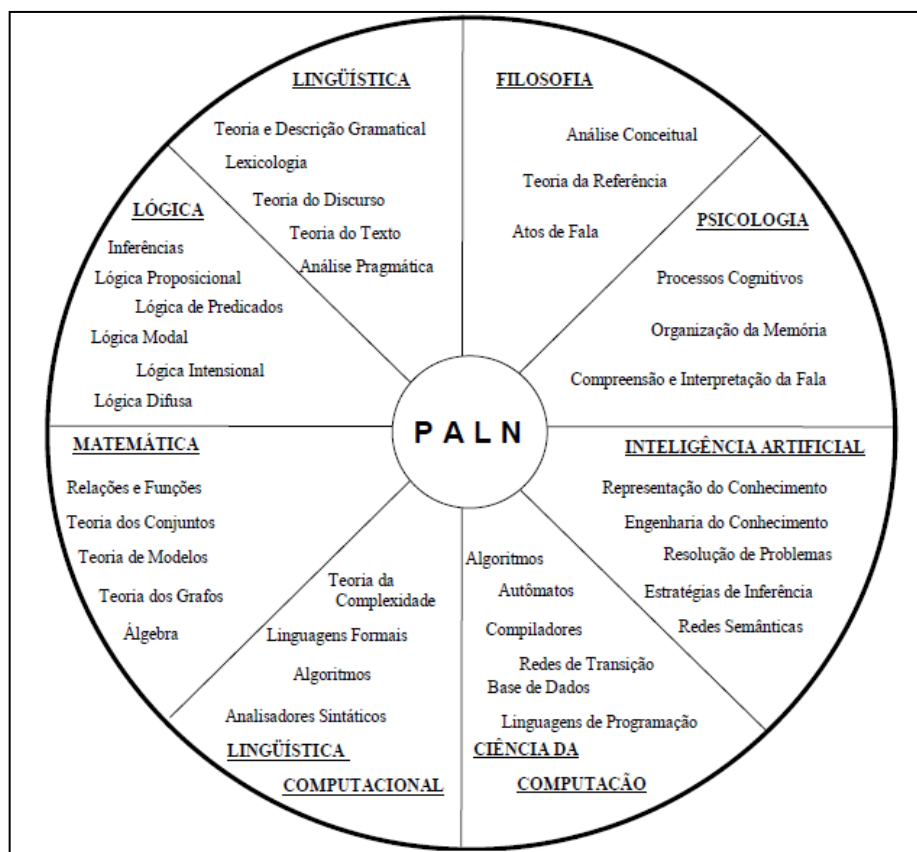


Figura 1 Principais recursos teóricos e metodológicos de que o estudo em PLN dispõe.
Fonte: Dias-da-Silva (1996, p. 83).

Dias-da-Silva (1996) apresenta uma citação de Sanders e Sanders (1989) que evidencia a necessidade do trabalho colaborativo entre especialistas da linguagem e da computação:

Os cientistas da computação sabem propor e gerenciar projetos de software. Eles dispõem de equipamentos e ferramentas de programação de vanguarda, e as linguagens simbólicas são seu material de trabalho. Por outro lado, entretanto, eles frequentemente não dominam os conhecimentos linguísticos [...]. É evidente que o trabalho é condição essencial. Entretanto a comunicação entre especialistas diversos não é tarefa fácil [...] (SANDERS e SANDERS, 1989, p. 30 apud DIAS-DA-SILVA, 1996, p. 84).

Pode-se constatar que a realização do desafio ousado explicitado por Dias-da-Silva (1996) depende da cooperação entre cientistas da computação e cientistas da linguagem, já que a descrição da língua, em todos os seus níveis (fonético/fonológico, morfológico, sintático, semântico, pragmático/discursivo) é de grande importância para o desenvolvimento de ferramentas e recursos computacionais. O grau de complexidade na tarefa de descrição desses níveis cresce do primeiro até o último, ou seja, do nível mais superficial (fonético/fonológico) até o mais profundo (pragmático/discursivo). Essas descrições compõem uma base de dados a qual um sistema de PLN pode recorrer quando necessário para execução de determinadas tarefas. Essas tarefas serão requisitadas dependendo do sistema de PLN em questão. Alguns exemplos desses sistemas são sistemas de manipulação de bases de dados, sistemas tutores, sistemas de automação de tarefas administrativas e gerenciais, sistemas de programação automática de computadores, sistema de processamento automático de textos e informações e sistemas de tradução automática (DIAS-DA-SILVA, 1996).

Observe-se, a seguir, o Quadro 1, baseado em Dias-da-Silva (1996), que resume o progressos estudos em PLN até a década de 90.

Quadro 1 Progresso dos estudos em PLN.

Anos 50	<ul style="list-style-type: none"> ✓ Sistematização computacional de classes de palavras presentes na gramática tradicional ✓ Início do processo de identificação de alguns tipos de constituintes da oração ✓ Destaque para os trabalhos em tradução automática
Anos 60	<ul style="list-style-type: none"> ✓ Criação de analisadores sintáticos ✓ Início da elaboração de formalizações de significado para compor redes semânticas
Anos 70	<ul style="list-style-type: none"> ✓ Primeiras implementações de gramáticas e analisadores sintáticos com base na gramática gerativo-transformacional ✓ Investigação de meios para formalizar conhecimento dos níveis pragmático/discursivo
Anos 80	<ul style="list-style-type: none"> ✓ Elaboração de teorias linguísticas a partir de estudos de PLN, exemplos: gramática sintagmática generalizada e gramática léxico-funcional
Anos 90	<ul style="list-style-type: none"> ✓ Busca pelo desenvolvimento de sistemas capazes de processar conhecimentos linguísticos e extralinguísticos

Fonte: adaptado de Dias-da-Silva (1996).

Podemos constatar, por meio do quadro, o percurso que vem sendo traçado para se alcançar o objetivo mais complexo que é a comunicação entre homem e máquina apenas por meio da linguagem natural. Nesse sentido, juntamente com o autor, reforçamos a necessidade da descrição de todos os níveis linguísticos, a fim de tornar os sistemas de PLN robustos e completos.

Para abranger tanto aspectos teóricos quanto metodológicos de áreas diferentes, Dias-da-Silva propõe uma metodologia trifásica composta pelas fases: linguística, representacional e implementacional. Na Figura 2, o autor esquematiza o funcionamento dessas fases.

DOMÍNIOS	PROBLEMAS	RECURSOS
Linguístico	explicitar o conhecimento e o uso linguístico	teorias da competência e do desempenho
↓↑	↓↑	↓↑
Representacional	representá-los	linguagens formais de representação
↓↑	↓↑	↓↑
Implementacional	codificar as representações	linguagens de programação e sistemas computacionais

Figura 2 Organização das fases que compõem a metodologia.
Fonte: Dias-da-Silva (1996, p. 89).

No domínio linguístico, o objetivo central é produzir descrições de língua, incluindo seus fenômenos, suas regras, seus conceitos e como ela é usada. No domínio representacional, o objetivo central é representar o conhecimento descrito no domínio linguístico. Essa representação se dá a partir da criação de sistemas formais que possam ser processados pela máquina. Por fim, no domínio implementacional, o objetivo central é codificar as representações do domínio representacional para se chegar a linguagens de programação e implementação em sistemas. O autor propõe que as tarefas de cada domínio sejam cumpridas de maneira sucessiva, progressiva e cíclica (DIAS-DA-SILVA, 1996).

Sendo assim, para que um sistema de PLN seja capaz de interpretar um texto, ele deve segmentar o texto em sentenças, as sentenças em palavras. De forma semelhante, para que um sistema de PLN seja capaz de gerar um texto, ele deve ser capaz de formar, identificar e escolher palavras, formar sentenças e, por fim, o texto. Para que tais tarefas sejam cumpridas satisfatoriamente, existem recursos linguísticos na arquitetura desses sistemas, a saber: léxico, gramática, modelo de domínio e modelo de usuário.

O léxico é o conjunto de informações e atributos de cada palavra. Por exemplo, se o sistema detectar a palavra “glicose”, ele poderá recorrer ao léxico e colher informações morfológicas e sintáticas que irão capacitá-lo a afirmar que “glicose” é um substantivo feminino singular composto pelo radical *glic(i/o)-* e pelo sufixo *-ose*. O léxico pode também oferecer informações não apenas morfológicas e sintáticas, mas também semânticas e pragmáticas/discursivas (DIAS-DA-SILVA et al., 2007).

A gramática é o conjunto das regras gramaticais e é o recurso que possibilita ao sistema validar uma sentença como aceitável em determinado idioma. No caso da sentença “*Ela fui comprar pão*”, o sistema não a reconheceria devido à conjugação errônea do verbo “ir”.

O modelo de domínio fornece informações para a interpretação e a geração de textos, como, por exemplo, que “pedra” é “inanimado”, logo, o sistema saberá que a palavra “pedra” não pode ocorrer com verbos como “correr”, “comer” e “chorar”. Esse recurso também é importante para eliminar a ambiguidade, por exemplo, da palavra “manga”, que pode ser parte de uma camisa ou uma fruta (DIAS-DA-SILVA et al., 2007).

O modelo de usuário compreende informações a respeito do usuário do sistema em questão. Nas palavras de Dias-da-Silva et al. (2007):

[...] o grau de informatividade na geração textual depende do que é relevante ao leitor e, portanto, irá implicar escolhas diversas de vocabulário, estruturas linguísticas etc.; o nível de conhecimento do assunto (superficial ou profundo) que o usuário apresenta pode levar a estruturas semânticas particulares, que, resultantes de um processo de *parsing*, podem auxiliar um sistema de consulta a, p.ex., fornecer respostas em grau adequado de clareza. Em geral, o conhecimento representado nesse módulo inclui as seguintes informações a respeito do usuário do sistema: seus objetivos, planos, preferências, intenções etc (DIAS-DA-SILVA, 2007, p. 36).

Na estrutura de um sistema de PLN, além desses recursos descritos, devem estar presentes as bases de conhecimento e os módulos de conhecimento que atuam sobre elas. As bases de conhecimento são subdivididas em base gramatical, base lexical e base conceitual. Já os módulos de conhecimento se subdividem em módulo de análise, de síntese e módulo especializado (DI FELIPPO & DIAS-DA-SILVA, 2006). A seguir, reproduziremos a Figura 3, extraída de Dias-da-Silva (1996), que representa como essas bases e módulos se organizam e interagem.

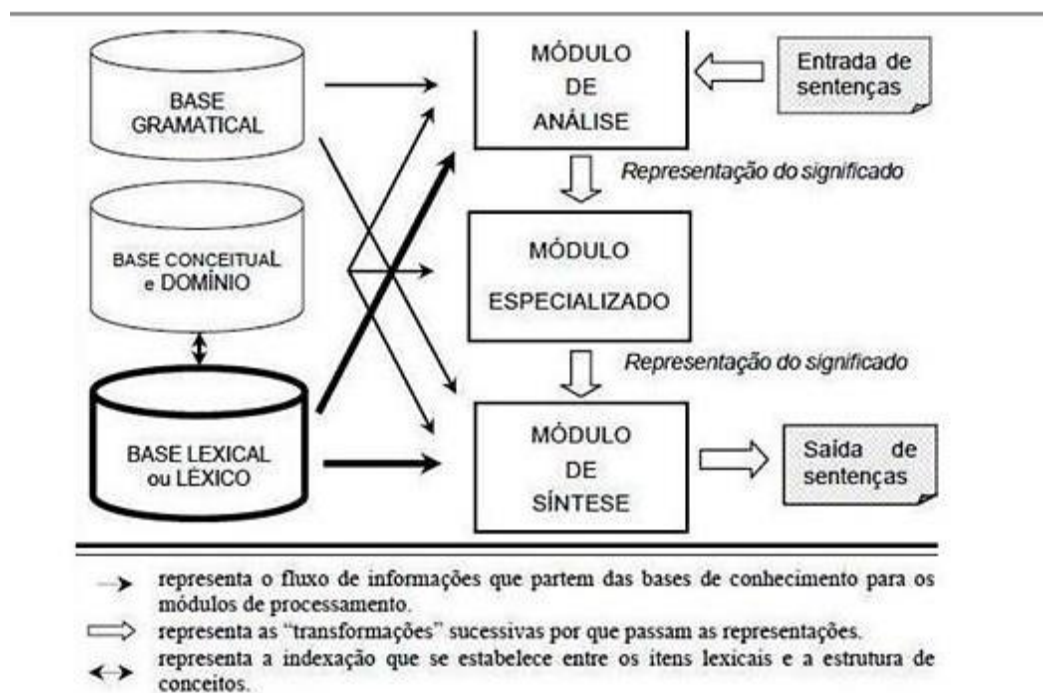


Figura 3 Esquema de possível arquitetura de um sistema de PLN.
 Fonte: Dias-da-Silva (1996, p. 89).

O módulo de análise é composto por analisadores morfológicos e sintáticos e por interpretadores semânticos e pragmáticos/discursivos, e tem por objetivo elaborar a representação interna do significado das sentenças inseridas no sistema. O módulo de síntese é responsável por transformar as representações elaboradas pelo módulo de análise em frases (DI FELIPPO & DIAS-DA-SILVA, 2006).

Como podemos constatar, a partir do esquema de Dias-da-Silva (1996), o módulo de síntese requer a representação dos módulos de análise e especializado. O módulo de análise, por sua vez, requer informações fornecidas pela base gramatical. Essas informações consistem nas regras sintáticas da língua. A base conceitual fornece informações aos três módulos por meio de uma rede de conceitos inter-relacionados por hiponímia e hiperonímia, por exemplo. São informações que descrevem objetos, eventos, atributos e propriedades em ordem hierárquica. Por fim, a base lexical é aquela que fornece aos módulos de análise e de síntese as informações morfológicas, sintáticas, semânticas e pragmáticas/discursivas dos itens lexicais (DI FELIPPO & DIAS-DA-SILVA, 2006).

Essa base é o recurso denominado *léxico*, já citado anteriormente, e nosso trabalho também poderá contribuir com o seu aprimoramento. Vejamos, a seguir, alguns tipos de produtos lexicográficos resultantes de trabalhos feitos por profissionais do PLN, os quais também poderão se enriquecer com as descrições desta pesquisa:

- Dicionários Legíveis com Máquina (DLM): dicionários cujas estruturas internas são semelhantes aos dicionários impressos, contando com o verbete, seguido da categoria e da definição. São produzidos para o uso humano, o que implica pouca linguagem formal, já que um ser humano compreende a linguagem natural. Dessa maneira, seu uso para sistemas de PLN é restrito, pois a máquina não possui a mesma capacidade de compreensão da língua natural como o ser humano, portanto, ela não compreenderia o conteúdo dos DLMs. Esses dicionários contam com versões impressas e digitais (CORREIA, 1994);
- Dicionários Tratáveis por Máquina (DTM): dicionários com o conteúdo de um DLM traduzido em linguagem formal, ou seja, a máquina será capaz de processar as informações, pois elas estarão escritas de outra maneira (CORREIA, 1994);
- Base de Dados Lexicais (BDL): segundo Correia (1994), a BDL

[...] é uma estrutura computacional concebida de modo a ser capaz de suportar os mais variados tipos de conhecimento sobre cada unidade lexical, permitindo estabelecer conexões, quer entre unidades lexicais distintas, quer entre características pertencentes a unidades lexicais distintas. Por outras palavras, a estruturação de uma BDL permitir-nos-á observar as unidades lexicais sob os mais variados prismas e aceder a elas das mais variadas formas possíveis (CORREIA, 1994, p. 4).

- bases digitais lexicais: é um conceito definido por Correia (1994) como genérico, pois engloba todos os tipos de bases lexicais que contenham informações dos itens lexicais que possam ser convertidas em diferentes formatos e servir para variados objetivos (CORREIA, 1994).

É possível reconhecer, a partir desses dicionários e bases, a importância da descrição da língua em todos os seus níveis para que as ferramentas de PLN tenham melhor desempenho. No caso deste trabalho, em que faremos uma descrição do nível morfológico de uma terminologia, poderemos além de aprimorar os extratores automáticos de candidatos a termos, oferecer subsídios para o enriquecimento dessas bases, a fim de que um sistema de PLN que trate de textos (tanto de língua geral quanto de terminologias), esteja apto a identificar a estrutura interna das palavras de um texto de entrada.

Contextualizando nosso trabalho na metodologia trifásica de que tratamos, podemos situá-lo no domínio linguístico e, a partir da implementação dos nossos resultados, poderemos fornecer informações de ordem morfológica às Bases de Dados Lexicais (BDL) e ao léxico

dos sistemas de PLN, especificamente os extratores automáticos de termos, sistemas de PLN que objetivamos enriquecer com os resultados do trabalho.

A tarefa de extração dos candidatos a termos é uma das mais importantes de um trabalho terminológico e, se feito manualmente, demanda um longo período da parte do terminólogo, o que certamente atrasa o trabalho e, muitas vezes, o inviabiliza.

Devido a essa dificuldade, a partir dos anos 80, surgiram os primeiros esboços de extratores semiautomáticos, já que na mesma época já estavam sendo elaborados programas que lidavam com *corpora* terminológicos (OLIVEIRA, 2006). Segundo o autor: “Um extrator automático de termos é um conjunto de programas dedicados ao reconhecimento e extração de termos que aparecem em um corpus de especialidade.” (OLIVEIRA, 2006, p. 59). A partir dessa concepção, podemos considerar um extrator como um tipo de sistema de PLN.

São três os métodos de extração que o programa pode utilizar: o estatístico, o linguístico e o híbrido.

O método estatístico é aquele cujo único critério empregado para selecionar o conjunto de termos é a frequência com que cada termo ocorre no *corpus*. De acordo com Oliveira (2006), é o método mais simples de aplicar e também o que mais pode trazer palavras da língua geral. Por isso, muitos dos extratores estatísticos utilizam uma *stoplist*, uma lista de palavras que são próprias da língua geral e que não veiculam conhecimento especializado. O ambiente computacional que utilizamos em nosso TCC, o e-Termos,⁸ possui uma *stoplist* padrão com palavras como: *nessa, nesse, introdução, conclusão, metodologia, conforme, muito, muitos* etc. O programa também aceita a inserção de *stoplists*, ou seja, o usuário pode criar uma que se adeque melhor às características do *corpus* com que está trabalhando e inseri-la no ambiente. Dessa forma, o resultado da extração conterà menos ruído. Um exemplo desse tipo de extrator é o Pacote NSP⁹ (*N-gram Statistics Package*) (OLIVEIRA, 2006).

O método linguístico considera as informações dos níveis linguísticos (morfológico e/ou sintático e/ou semântico). Um ponto negativo desse critério é que sua aplicação se restringe a cada idioma em particular. Assim como o método anterior, o linguístico também necessita de uma *stoplist*, pois muitas vezes os padrões das terminologias podem coincidir

⁸ Ambiente computacional colaborativo *Web* de acesso livre e gratuito, dedicado à gestão terminológica. Resultado de um projeto de pesquisa envolvendo três instituições públicas brasileiras (USP, UFSCar e Embrapa), seu principal objetivo é viabilizar a criação de produtos terminológicos, sejam eles para fins de pesquisa acadêmica, de divulgação ou de ensino, por meio da (semi)automatização das etapas do trabalho terminológico. Disponível em: <<https://www.etermos.cnptia.embrapa.br>>.

⁹ Mais informações podem ser obtidas em <<http://www.d.umn.edu/~tpederse/nsp.html>>.

com os da língua geral. Um exemplo é o padrão morfossintático “NOME + ADJETIVO”, recorrente em muitas terminologias, mas também na língua geral: “amarelecimento foliar” é um termo, mas “referência bibliográfica” não (pelo menos não no domínio com o qual trabalhamos) e ambos possuem o mesmo padrão. Alguns exemplos de extratores desse tipo são: NODALIDA e FASTR (BAGOT, 1999 apud OLIVEIRA, 2006).

O método híbrido é aquele que aplica primeiro o critério linguístico e depois o estatístico para a extração dos candidatos a termos. São considerados os mais eficientes, segundo Oliveira (2006). Nessa metodologia, o *corpus* deve ser etiquetado morfossintaticamente. Quanto mais detalhadas as descrições sobre diferentes terminologias, mais ricos poderão ser esses extratores automáticos de candidatos que utilizam conhecimento linguístico. Se o extrator conseguir distinguir morfemas greco-latinos nas palavras, por exemplo, isso já pode ser considerado um indicativo de que essa palavra pode ser um termo, já que nas pesquisas que descreveram morfologicamente diferentes terminologias verificou-se a recorrência desses morfemas.

Apresentaremos, na próxima seção, os trabalhos já realizados no GETerm (citados na Introdução) que também tiveram como objeto o aspecto morfológico de terminologias a fim de podermos compará-los entre si e com a terminologia aqui descrita e verificarmos as possíveis semelhanças, recorrências e diferenças.

3.2 Descrição dos trabalhos já realizados

O nosso trabalho busca oferecer informações morfológicas sobre a terminologia das pragas e doenças da cana-de-açúcar e faz parte de um conjunto de outros trabalhos que também já foram feitos com o mesmo intuito, mas em áreas diferentes.

Os trabalhos que apresentaremos a seguir são:

1. *Aspectos morfológicos dos termos da Nanociência e Nanotecnologia;*
2. *Base de dados morfológicos de terminologias do português do Brasil. Descrição e análise morfológica com vistas à disponibilização on-line;*
3. *Descrição das unidades terminológicas da Educação a Distância em português do Brasil no nível morfológico;*
4. *Descrição e análise morfológica da terminologia da Fisioterapia: subsídios para organização de uma base de dados morfológicos de terminologias do português.*

Aspectos morfológicos dos termos da Nanociência e Nanotecnologia foi um Trabalho de Conclusão de Curso (TCC) desenvolvido por Joel Sossai Coleti, entre os anos de 2009-2010. Nesta pesquisa, publicada na revista *Filologia e Linguística Portuguesa* em 2010, descrevem-se e analisam-se os aspectos morfológicos de um conjunto de 295 termos mais frequentes da terminologia da Nanociência e Nanotecnologia. O objetivo do trabalho era enriquecer a descrição morfológica do português a partir de um léxico especializado, de maneira a contribuir com o refinamento de conhecimento linguístico útil para aplicação em sistemas híbridos de extração automática de terminologias.

Neste trabalho, os processos de formação de palavras identificados são apresentados na Tabela 1.

Tabela 1 Ocorrência dos processos de formação de palavras do domínio da Nanociência/Nanotecnologia.

Processo de formação	Número de ocorrências	Exemplos
composição sintagmática	129	<i>amostra de espinélio dopada, armazenagem de hidrogênio, atividade catalítica</i> etc.
derivação sufixal	79	<i>acoplamento, anodização, armazenagem de hidrogênio</i> etc. [morfemas de que fato ocorrem: <i>-ada, -ado, -agem, -al, -ar, -ção, -dor, -eira, -eto, -ico, -(i)dade, -io, -ito, -ivo, -lito, -mento, -nte, -on, -or, -oso, -ura</i>]
derivação prefixal	53	<i>absorção, decomposição</i> etc. [morfemas de que fato ocorrem: <i>ab-, ad-, bio-, de-, des-, di-, infra-, micro-, nano-, pre-, re-, semi-</i>]
composição acronímica	26	<i>filme de PET, imagem de MET, pastilha de SNO, PST</i> etc.
estrangueirismo	18	<i>conformação por spray, DNA, LASER escravo</i> etc.
composição subordinativa	9	<i>espectroscopia Raman, fotodetector, impedância eletroquímica</i> etc.
composição coordenativa	2	<i>desvio padrão, sol-gel</i>
conversão	2	<i>dielétrico, precipitado</i> (que ocorrem como substantivos)
TOTAL	318	

Fonte: adaptada de Coleti & Almeida (2010, p. 284).

Os autores chamam atenção para o fato de que foram analisados 295 termos, mas o total de processos observados é 318. Isso ocorre porque muitas vezes em um único item léxico observam-se mais de um processo, exemplos:

- *amostra de espinélio dopada*: incidem aqui os processos de derivação sufixal (no item *dopada*) e de composição sintagmática.

- *LASER escravo*: neste item, ocorrem os processos de sigla (*LASER*), estrangeirismo (*Light Amplification by Stimulated Emission of Radiation*) e composição sintagmática.

Ao final do artigo, os autores apresentam os padrões morfolexicais observados nessa terminologia:

- com relação à sufixação, destacam-se os seguintes morfemas: *-ado*; *-agem*, *-ção* e *-mento*; *-ico*; *-or*; *-dade*; *-al*; os quais poderiam gerar os seguintes padrões de busca, nos quais X representa a base léxica:
 - X + *-ado*
 - X + *-agem*
 - X + *-ção*
 - X + *-mento*
 - X + *-ico*
 - X + *-or*
 - X + *-dade*
 - X + *-al*
- com relação à prefixação, destacam-se os seguintes morfemas: *nano-*, *micro-*, *semi-*, *des-*, *infra-*, os quais poderiam gerar os seguintes padrões de busca, nos quais X representa a base léxica:
 - *nano-* + X
 - *micro-* + X
 - *semi-* + X
 - *des-* + X
 - *infra-* + X
- com relação aos compostos subordinativos, embora sendo menos frequentes no grupo analisado, constatou-se grande incidência de morfemas greco-latinos, como se pôde observar nas formações:
 - eletr(i/o)- + química > *eletroquímica*
 - equi- + ax(i)- + -al > *equiaxial*
 - espectr(o)- + -scop- + -ia > *espectroscopia*

- fot(o)- + detector > *fotodetector*
 - lit(o)- + -grafia > *litografia*
 - morf(o)- + -logia > *morfologia*
 - ele(c)tr(o)- + -stat(o)- + -ico > *eletrostático*
- no que tange aos composto coordenativos, observaram-se apenas as ocorrências *desvio padrão* e *sol-gel* no grupo selecionado para análise, demonstrando que não é um processo produtivo;
- no que se refere aos compostos sintagmáticos, constatou-se a grande produtividade das estruturas N + A e N + Prep (+ Det) + N. Em menor ocorrência, mas não menos importante, têm-se as estruturas: N + N, N + Prep (+ Det) + N + A, N + A + Prep + N, N + A + A, N + Prep + N + N e, finalmente, N + Prep + A + N;
- no que diz respeito às siglas ou formações acronímicas (26 ocorrências), observou-se que elas constituem itens autônomos ou integram composições sintagmáticas. Constatou-se, também, que correspondem a formas expandidas em português e em inglês, como em *DRX* (= *Difração de Raios X*) e *DSC* (= *Dye-sensitized Solar Cell*);
- com relação aos empréstimos, observou-se baixa frequência, entretanto, 100% dos casos (18) são provenientes da língua inglesa.

O segundo trabalho, intitulado *Base de dados morfológicos de terminologias do português do Brasil. Descrição e análise morfológica com vistas à disponibilização on-line*, corresponde à dissertação de mestrado de Joel Sossai Coleti e foi desenvolvida entre os anos de 2010 e 2012.

Coleti (2012) descreveu em sua dissertação os aspectos morfológicos do domínio de Nanociência e Nanotecnologia e Biocombustíveis (especialmente o biodiesel), com a diferença de que, nesta pesquisa, utilizou como base teórica o modelo SILEX, desenvolvido em 1987 por Daniele Corbin. Esse modelo originou-se nos estudos gerativos, por isso uma de suas características é considerar a previsibilidade das regras de formação de palavras e não os produtos criados.

Assim como em seu TCC, este trabalho teve como principal objetivo verificar os processos mais produtivos de formação de termos de modo a aprimorar os sistemas híbridos de extração automática de terminologias.

Para esta pesquisa, foi considerado um conjunto de 1.398 termos da Nanociência e Nanotecnologia e 578 de Biodiesel, totalizando 1.976 termos que foram analisados. De acordo com os critérios adotados na pesquisa, foi feita uma delimitação, de maneira que nesse total não estivessem contidos empréstimos, nomes próprios, siglas e acrônimos como também aqueles termos complexos formados por duas, três, quatro ou mais palavras. A prioridade foi dada aos processos que ocorrem no interior de cada palavra.

Para sua análise, o autor distinguiu duas etapas: análise inicial e profunda. Na análise inicial, foi feita a classificação morfológica de todos os termos de acordo com a tipologia proposta por Correia & Lemos (2005). Em seguida, organizou uma listagem dos termos por processo morfológico apenas para os processos de maior pertinência para a pesquisa que são: sufixação, prefixação e composição morfológica. Depois disso, fez a identificação do morfema de maior frequência em cada processo morfológico.

A análise profunda compreendeu: análise criteriosa dos itens de maior frequência em cada processo morfológico para responder as questões:

- Que significado tem os morfemas?
- Que significados passam a ter os produtos formados quando estes se associam a determinada base?
- Há alguma interferência da classe gramatical da base a qual o morfema se associa?
- Que tipo de generalizações se podem fazer para gerar regras?

Finalmente, o autor propôs as Regras de Construção de Palavras (RCPs) para os casos selecionados.

Na análise inicial dos termos de Nanociência e Nanotecnologia, foi constatada a predominância de derivação sufixal, seguida pela composição morfológica e depois pela derivação prefixal. Os casos de sufixação somam 514 ocorrências e, como exemplo, temos os seguintes sufixos mais frequentes: *-(ç/s/ss)ão* (aparecem em 149 termos: *aceleração*, *aglomeração*), *-mento* (presente em 56 termos: *confinamento*, *empacotamento*) e *-ico* (ocorre em 47 termos: *carboxílico*, *elástico*). A prefixação ocorre em 100 termos e os três prefixos

mais frequentes são: *super-/sobre-* (17 termos: *superabsorvente, sobreposição*), *inter-* (12 termos: *interação, intermolecular*) e *des-* (8 termos: *descarga, descontaminação*).

Os termos resultantes de composição morfológica totalizam 290, sendo os três morfemas mais frequentes: *nan(o)-* (75 termos: *nanoadesivo, nanoagradado*), *micro-* (57 termos: *microbomba, micróbio*) e *foto-* (25 termos: *fotoalinhador, fotocatalisador*). Na análise inicial dos termos do Biodiesel, ocorreu uma distribuição de processos semelhante à terminologia anterior: predominância de sufixação, seguida pela composição e prefixação.

Os processos de sufixação somam 90 casos e os três sufixos mais frequentes são: *-(d/t/s)or(a)* (17 termos: *precipitador, esmagador*), *-mento* (11 termos: *amarelecimento, borbulhamento*) e *-(ç/s/ss)ão* (10 termos: *destoxificação, econeutralização*). Os casos de prefixação ocorrem em 14 termos, sendo os prefixos mais frequentes: *des-* (seis termos: *desproteínizar, destoxificação*), *re-* (três termos: *reação, replicabilidade*) e *trans-* (três termos: *transiluminador, transesterificação*). Por fim, os casos de composição acontecem em 41 termos, e os morfemas mais frequentes são: *hidro-* (nove casos: *hidrocarboneto, hidrociclone*), *bio-* (sete casos: *biodiesel, bioenergia*) e *poli-* (seis casos: *poliestireno, polissiloxano*).

Com esses dados, o autor selecionou os elementos mais frequentes e estabeleceu seis RCPs. Os elementos utilizados foram:

- *-(ç/s/ss)ão*
- *-(d/t/s)or(a)*
- *super-/sobre-*
- *des-*
- *nan(o)-*
- *hidro-*

A partir desses morfemas, foram organizadas as seguintes RCPs:

1. $V_{(t/i)} \rightarrow S_{-ção}$

V: verbo

T: transitivo

I: intransitivo

S: substantivo

-(ç/s/ss)ão: sufixo

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

V: verbo

T: transitivo

S: substantivo

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

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

S: substantivo

A: adjetivo

Super-/sobre-: prefixo

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

S: substantivo

A: adjetivo

V: verbo

Des-: prefixo

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

S: substantivo

Nano-: prefixo ou elemento de composição ou elemento de recomposição

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

S: substantivo

Hidro-: elemento de composição morfológica

Após cada proposição de regra, o autor faz uma vasta análise justificando cada uma delas e conclui afirmando que os resultados da pesquisa poderão enriquecer bases de dados morfológicos e melhorar ferramentas de extração de candidatos a termos.

A pesquisa intitulada *Descrição das unidades terminológicas da Educação a Distância em português do Brasil no nível morfológico*, foi realizada por Ana Catarina Gianoti em seu TCC, entre os anos de 2012 e 2013.

Neste trabalho, a autora analisa um conjunto de 465 termos da área Educação a Distância e, como tipologia, foi utilizada a classificação de processos de formação de palavras

proposta por Alves (2004). Como já mencionamos na seção 2.2.2, segundo essa proposta de classificação, são seis os processos de formação de palavras: fonológico, sintático, conversão (ou derivação imprópria), semântico, empréstimo e os processos menos produtivos (truncação, palavra-valise, reduplicação e derivação regressiva).

Os 465 termos descritos foram selecionados a partir de uma base de dados terminológicos do domínio da Educação a Distância, resultante do projeto denominado TermiNet.¹⁰

A seguir, apresentamos na Tabela 2 os processos identificados pela autora em ordem decrescente de frequência.

Tabela 2 Ocorrência dos processos de formação de palavras do domínio da Educação a Distância.

Processos de formação	Número de ocorrências
Composição sintagmática	321
Derivação sufixal	51
Composição subordinativa	33
Derivação prefixal	25
Empréstimo	22
Composição acronímica	15
Palavra-valise	6
Composição coordenativa	2
Truncação	0
Reduplicação	0
Derivação regressiva	0
Derivação parassintética	0
Semântico	0
Conversão	0
Total	467

Fonte: elaborada pela autora.

Em sua análise, a autora verificou que o sufixo *-ção* é o mais produtivo (*massificação, formatação*), seguido por *-dade* (*afetividade, conectividade*) e *-mento* (*compartilhamento, agrupamento*). Já os prefixos mais produtivos são: *tele-* (*telecurso, teleconferência*) e *inter-* (*interconexão, interatividade*).

As composições sintagmáticas, que somam a maioria, apresentaram os seguintes padrões: N+Adj (*abordagem colaborativa*), N+Prep+N (*acesso a internet*), N+Adj+Adj

¹⁰ Disponível em <<http://www.nilc.icmc.usp.br/~arianidf/terminet/index.html>>.

(*material didático impresso*), N+Adj+Prep+N (*abordagem orientada por objetivo*) e N+Prep+N+Adj (*produção de material didático*).

Em relação aos empréstimos, foi constatado que todos são originados do inglês (*benchmarking, designer, e-mail, PDF, blended*).

Com esta pesquisa, a autora trouxe uma importante contribuição, pois não havia ainda sido descrita no GETerm nenhuma terminologia de um domínio das Ciências Humanas. A conclusão é de que esta terminologia segue, em geral, os padrões morfológicos das demais, sobretudo no que se refere aos processos de formação de palavras mais frequentes (derivação e composição).

A última pesquisa diz respeito à dissertação 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* e realizada entre os anos de 2011 e 2013.

Com o objetivo de descrever a estrutura interna dos termos pertencentes à área da Fisioterapia, Mattos (2013) também visou a organização de uma base de dados que pudesse ser implementada em sistemas computacionais.

Assim como a dissertação de Coleti (2012), o modelo teórico escolhido foi o SILEX.

Para empreender a análise, a autora organizou uma tabela contendo os processos de formação dos termos e outra contendo todos os arqueoconstituintes¹¹ utilizados na formação dos termos. Nessa última, ela apresenta o significado descritivo dessas unidades e os termos relacionados.

Foi descrito um conjunto de 921 termos e constatou-se que: a sufixação é o processo mais preponderante, principalmente a presença dos sufixos: *-ia*, *-ite* e *-ose* (*artrite, bronquite, equimose, trombose, ultrassonografia, rizotomia*) e em segundo lugar está a composição, com destaque para os arqueoconstituintes gregos (*eletroencefalograma, hemograma*). A prefixação (*hipotermia, infravermelho*) e a conversão (*fratura, sutura, enxerto*) foram os menos encontrados.

A autora também classificou alguns termos como unidades simples e unidades complexas não construídas. As unidades simples são aquelas que não possuem estrutura interna construcional, como: *tosse, parafina, tipoia*. As unidades complexas não construídas “[...] são aquelas que, apesar de possuírem uma aparente estrutura morfológica, não

¹¹ Segundo Mattos (2013), são radicais de origem grega ou latina e não podem ocupar posições sintáticas, pois são formas presas como *agr(o)-* (agronomia).

apresentam um significado previsível a partir dela” (p. 59). São exemplos: *postura*, *prevenção*, *eritema*, *equilíbrio*. As siglas, acrônimos e importações também foram considerados como unidades simples. Alguns exemplos: *splint*, *geno recurvatum*, *leg-press*, *ADM*, *LER*, *RTCA*.

Os compostos sintagmáticos, assim como as unidades supracitadas, não foram detalhados em seu trabalho, pois a autora afirma que seu foco eram apenas as unidades construídas morfologicamente. Alguns dos padrões encontrados foram: N + Adj (*ação muscular*, *exercício ativo*) e N + Adj + Adj (*calor superficial condutivo*, *doença pulmonar restritiva*).

Considerando essas unidades que não foram detalhadas e as demais, temos os seguintes números para cada processo, organizados na Tabela 3.

Tabela 3 Ocorrência dos processos de formação de palavras do domínio da Fisioterapia.

Processos de formação	Número de ocorrências
Compostos sintagmáticos	436
Sufixação	233
Composição morfológica	74
Unidades complexas não construídas	54
Unidades simples	44
Prefixação	34
Importações	15
Siglas e acrônimos	10
Conversão	9
Total	909

Fonte: elaborada pela autora.

A autora conclui que a alta frequência de arquetipos gregos comprova a relação da Fisioterapia com a área médica. Foi constatado também um elevado número de compostos sintagmáticos, o que é comum em terminologias. Outra característica que pode ser considerada mais particular dessa área é a pouca importação de termos.

Após a explanação desses quatro trabalhos, organizamos uma tabela comparativa (Tabela 4), em que evidenciamos os processos de formação mais frequentes em cada um deles.

Tabela 4 Comparação entre os trabalhos descritos.

	NANOCIÊNCIA/ NANOTECNOLOGIA	BIODIESEL	EDUCAÇÃO A DISTÂNCIA	FISIOTERAPIA
COMPOSIÇÃO MORFOLÓGICA	2º processo mais produtivo, em 290 casos. Ex: nano-, micro-, foto-.	2º processo mais produtivo, em 41 casos. Ex: hidro-, bio-, poli-.	Ocorre em 35 casos. Ex: sócio-, ped(o)-, auto-.	3º processo mais produtivo, em 74 casos. Ex: termo-, eletr-, hidro-.
COMPOSIÇÃO SINTAGMÁTICA	Ocorre em 1511 casos, mas não é tratado pelo autor.	Ocorre em 347 casos, mas não é tratado pelo autor.	O mais produtivo, em 321 casos. Ex: educação superior, abordagem colaborativa.	É o processo mais produtivo, ocorre em 436 casos. Ex: músculo antagonista, articulação plana.
SUFIXAÇÃO	É o processo mais produtivo, ocorre em 514 casos. Ex: -(ç/s/ss)ão, -mento, -ico.	É o processo mais produtivo, ocorre em 90 casos. Ex: -(d)/(t)/(s)or(a), -mento, -(ç/s/ss)ão.	2º processo mais produtivo, ocorre em 51 casos. Os mais frequentes são: -são, -dade e -mento.	2º processo mais produtivo, ocorre em 233 casos. Ex: -ia, -ite e -ose.
PREFIXAÇÃO	3º processo mais frequente, ocorre em 100 casos. Ex: super/sobre-, inter-, des-.	3º processo mais frequente, ocorre em 14 casos. Ex: des-, re-, trans-.	4º processo mais frequente. Os mais frequentes são: tele- e inter-.	6º processo mais produtivo, ocorre em 34 casos. Ex: hipo-, infra-.

Fonte: elaborada pela autora.

Na Subseção 4.4 iremos comparar a descrição aqui realizada com os dados obtidos da análise do nosso trabalho e verificar as principais diferenças e semelhanças entre todos.

3.3 Percurso até o conjunto terminológico

Como já mencionamos na introdução, nosso trabalho vem em continuidade ao TCC realizado em 2013, no qual foi organizado uma estrutura conceitual do domínio das pragas e doenças da cana-de-açúcar. Para que essa estrutura pudesse ser construída, passamos pelas fases iniciais de um trabalho terminológico baseado em *corpus* seguido pela extração semiautomática dos candidatos a termos e, em seguida, validação desses candidatos por um especialista de domínio. Esse trabalho resultou em 941 termos validados pelo especialista. Desse total, identificamos aqueles referentes às pragas e doenças da cana-de-açúcar e chegamos a um conjunto de 135 termos que foram efetivamente analisados.

3.3.1 *Compilação do corpus*

Ao adotarmos uma teoria descritiva de base linguística, assumimos que os termos são signos linguísticos que podem ocorrer como termos ou como palavras. A língua de especialidade não é composta por unidades léxicas à parte da língua geral, mas sim por unidades que são signos linguísticos integrantes da língua natural (ALMEIDA, 2010). Para verificar o contexto de uso, é necessária a organização de um *corpus*, pois a partir dele é possível extrair os termos e suas variantes, observar as colocações e fraseologias, identificar as relações semânticas entre os termos e os traços semânticos do termo em diferentes contextos de uso. O uso de *corpus* garante à teoria o caráter descritivo (ALMEIDA, 2010).

Para garantir uma boa representatividade, o *corpus*, a partir do qual se obtiveram os termos que são aqui analisados, foi composto de textos dos seguintes gêneros: científico (teses, dissertações, artigos científicos, documento técnico), científico de divulgação (artigos de divulgação, notícias) e instrucional (bulas de pesticidas). A fim de facilitar a organização e consulta, os textos foram divididos em dois grandes grupos: científico e divulgação. No grupo científico de caráter acadêmico, estão as teses, dissertações e artigos científicos. O grupo de divulgação contém textos que circulam entre os indivíduos quer sejam especialistas na área ou leigos. Estão inclusos nesse grupo os artigos de divulgação, as notícias, as bulas de pesticidas e os documentos técnicos.

As fontes do *corpus* são: banco de teses e dissertações da ESALQ, base Scielo, Jornal da Cana, Revista Superinteressante,¹² *site* da revista Globo Rural,¹³ *site* do jornal O Globo,¹⁴

¹² Disponível em <<http://super.abril.com.br/>>.

Revista Ciência Hoje,¹⁵ *site* do Instituto Biológico de São Paulo,¹⁶ Revista Biotecnologia,¹⁷ Revista FAPESP, *site* do Ministério da Agricultura, *site* da Defesa Agropecuária de São Paulo,¹⁸ *sites* da Dupont, Ihara, Extraprática¹⁹ e Nortox. Apresentamos na Tabela 5 os dados estatísticos do *corpus*.

Tabela 5 Dados estatísticos do corpus.

Gênero do texto	Número de palavras	Número de arquivos
Dissertações	248.487	15
Teses	302.531	9
Artigos científicos	113.749	29
Artigos de divulgação	18.508	10
Notícias	22.154	35
Bulas	60.522	11
Documento técnico	1.752	1
Total	767.703	110

Fonte: elaborada pela autora.

3.3.2 *Extração dos termos*

Primeiramente foi criado um projeto chamado *Terminologia das Pragas da cana-de-açúcar* para abrir uma área no e-Termos. O *corpus* foi inserido no referido ambiente e, a partir da sua inserção, foi realizada a extração estatística de frequência simples das listas de uni, bi, tri, tetra e pentagramas.

Como o critério de extração considerava apenas critérios estatísticos, foi necessário fazer uma revisão a fim de retirar itens que certamente não eram termos. Alguns exemplos são: *molecular, contendo, estudo, possível, observados, dias após, não diferem, não há, pode ocorrer, o número de, em relação à, o período de, por meio da, horas após a exposição, para o controle de, as embalagens vazias não podem, os equipamentos de proteção individual, entre outros.*

¹³ Disponível em <<http://revistagloborural.globo.com/>>.

¹⁴ Disponível em <<http://oglobo.globo.com/>>.

¹⁵ Disponível em <<http://cienciahoje.uol.com.br/>>.

¹⁶ Disponível em <<http://www.biológico.sp.gov.br/>>.

¹⁷ Disponível em <<http://www.biotecnologia.com.br/>>.

¹⁸ Disponível em <<http://www.cda.sp.gov.br/>>.

¹⁹ Disponível em <https://www.extrapratca.com.br/BR_Docs/English/Instructions/71.pdf>.

Feita essa revisão, as listas passaram a conter os seguintes números de candidatos a termos:

- unigramas: 1.840 candidatos;
- bigramas: 477 candidatos;
- trigramas: 402 candidatos;
- tetragramas: 103 candidatos;
- pentagramas: 57 candidatos.

3.3.3 *Validação das listas*

Com as listas de candidatos a termos já limpas, elas foram direcionadas para um especialista do domínio,²⁰ que selecionou os termos que seriam relevantes para compor a estrutura conceitual. Sendo assim, dos 2.879 candidatos a termos, o especialista validou 941. Essa tarefa se deu com o envio, em formato .xls de uma tabela contendo esses candidatos a fim de que o especialista assinalasse na coluna ao lado do termo validado.

3.3.4 *Seleção dos termos e a descrição morfológica*

Os motivos que levaram a escolher esse conjunto de termos foram, em primeiro lugar, o contato anterior com a área, no já referido TCC, o que garantiu mais segurança e domínio no tratamento dos termos. Em segundo lugar, a necessidade de enriquecer a base de dados morfológicos com um domínio vinculado à agricultura, um dos segmentos motrizes da economia brasileira. Em terceiro lugar, o fato de já termos um conjunto terminológico proveniente de *corpus*, validado, apto para uso em diferentes perspectivas e tarefas.

Ao analisarmos os termos validados pelo especialista, pudemos constatar que muitos diziam respeito aos métodos de controle de pragas, elementos químicos, pesticidas, fungicidas, tipos de culturas. Sendo assim, para o trabalho do mestrado, decidimos por descrever apenas os termos que designam pragas e doenças da cana-de-açúcar, a fim de restringir o número do conjunto terminológico. Dessa forma, podíamos ter uma caracterização mais própria da área, observar quais eram os processos morfológicos mais frequentes neste universo e se havia alguma característica que particularizasse essa área, do ponto de vista da

²⁰ Ivo Pierozzi, biólogo e pesquisador na Embrapa.

sua morfologia, de forma a enriquecer a base de dados morfológicos que vem sendo construída no GETerm.

Sendo assim, o primeiro passo para a análise foi a inserção dos 135 termos selecionados em uma planilha no Microsoft Excel, da mesma forma como foi feito nos trabalhos de Coleti (2009; 2012), Gianoti (2013) e Mattos (2013). Nas primeiras linhas, estão os tipos de formação de palavras, segundo a classificação de Correia & Almeida (2012). Na primeira coluna, encontra-se a lista de termos; na segunda, as observações de caráter semântico, etimológico e/ou a segmentação morfemática do termo; na terceira, a frequência de cada termo no *corpus*, a fim de indicar a sua relevância, ao menos do ponto de vista quantitativo. Nas colunas seguintes estão os processos de formação de palavras, de maneira que cada processo observado no termo possa ser registrado ao longo da planilha. Cada processo identificado é marcado com o algarismo “1”, pois assim o programa realiza a soma na última linha de cada coluna e, ao final, poder-se-á ter uma ideia de qual(is) processo(s) predomina(m) na área.

A título de ilustração, apresentamos na Figura 4 um trecho dessa planilha, que está completa no Apêndice do presente trabalho em formato diferente por motivos de apresentação exigida.

TERMOS	Observações	FREQ	ONOMA TÓPIA	DERIVAÇÃO			COMPOSIÇÃO			PROCESSOS DEFORMACIONAIS			LEXICALIZAÇÃO DE ESTRUTURAS			IMP PALAV SEM		
				AFIXAL			MORF	MORFOS	RECOMB	AMÁLG	TRUNC	SIG	ACRÔN	L.F.F	L.S		L.I.D	
				N.A.	PRE	SUF												PARAS
AMARELECIMENTO FOLIAR	amarelecer + -mento [amarelecer: amarelo + -ecer [amarelo: lat. hispânico <i>amarellus</i>] + [-ecer terminação de v. latinos] = amarelecer] + [-	105		0	0	0	1	0	0	0	0	0	0	0	0	0	1	
ACIDOVORAX AVENAE		11		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AGROBACTERIUM TUMEFACIENS		32		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AMARANTHUS HYBRIDUS	planta infestante	36		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
AMARELINHO	Também conhecido como "vírus do amarelecimento foliar da cana-de-açúcar" amarelo + -inho [amarelo: lat. hispânico <i>amarellus</i>] + [-inho: suf. lat.]	12		0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
ANTHONOMUS GRANDIS		18		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figura 4 Fragmento da planilha de análise morfológica dos termos.
Fonte: elaborada pela autora.

Para realizar a pesquisa pelas informações linguísticas presentes na coluna “observações” da planilha, foram consultadas as seguintes fontes: *Dicionário de Elementos*

Mórficos presente no *Dicionário Eletrônico Houaiss da Língua Portuguesa* em versão digital, *Dicionário Caudas Aulete*, disponibilizado online e o buscador *Google*. A preferência por essas obras se deu devido a maior facilidade de acesso e confiabilidade.

O primeiro passo, depois da inserção dos termos na tabela e a frequência com que aparecem no *corpus*, foi consultar ambos os dicionários e verificar como eles apresentavam a segmentação morfológica e se eram equivalentes. No conjunto terminológico descrito, não houve diferenças entre as informações das obras, então o passo seguinte foi a inserção dessas informações no campo “observações”. A partir daí, passamos a classificar cada termo no seu(s) processo(s) correspondente(s). Por exemplo, no caso do termo “amarelinho”; o primeiro passo foi consultar o dicionário geral, que forneceu as seguintes informações:

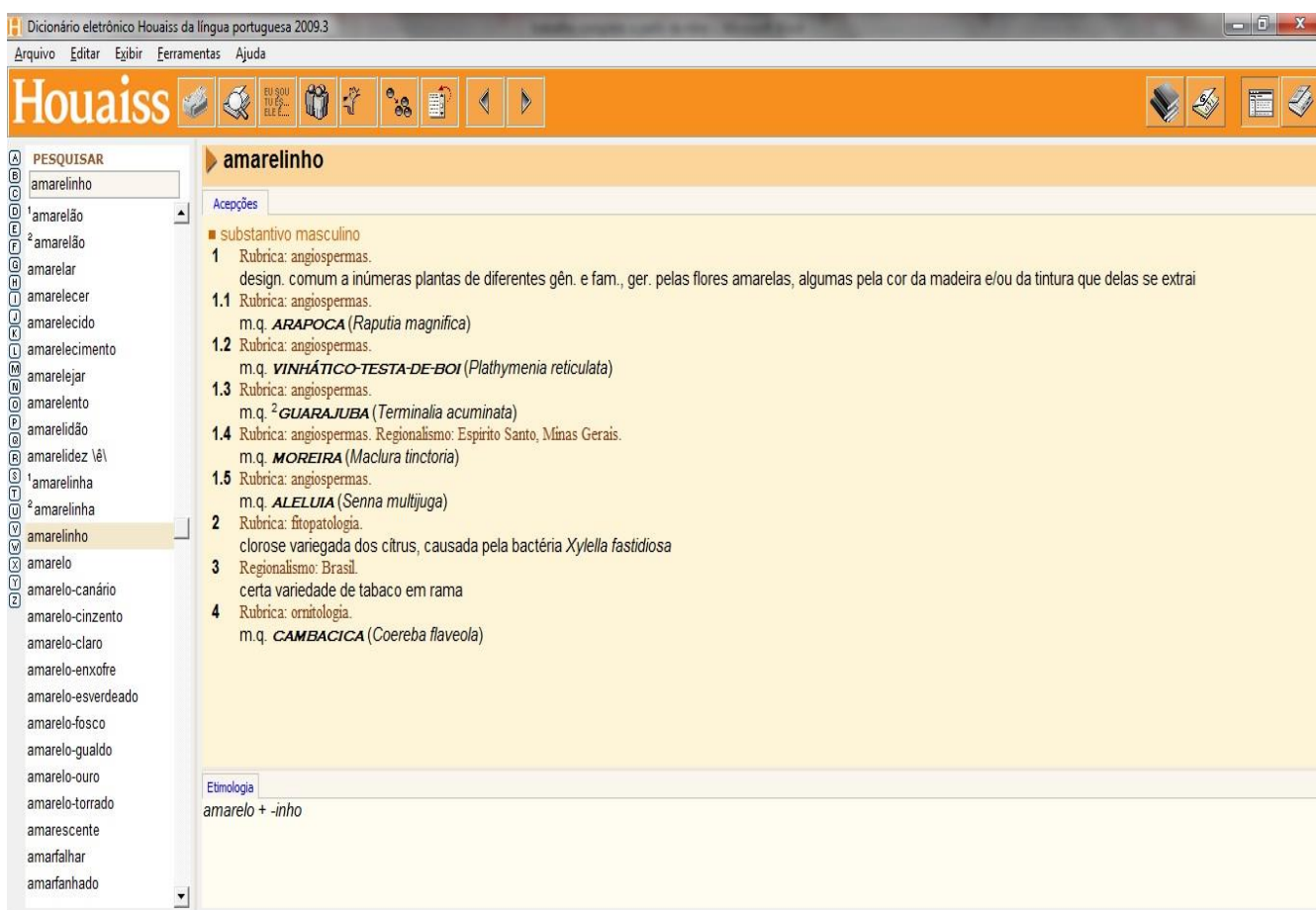


Figura 5 Acepções do termo “amarelinho”.

Fonte: Dicionário Eletrônico Houaiss

Como é possível ver na figura, o próprio dicionário fornece informações etimológicas sobre o termo, dessa forma sabemos que “amarelinho” é resultado da formação “amarelo” +

“-inho”. Em seguida, procuramos no mesmo dicionário o verbete “amarelo” para recuperar a sua etimologia (palavra proveniente do latim “*amarellus*”). Com essa informação, já podemos assinalar em nossa tabela o processo de formação: importação de palavras. O passo seguinte foi consultar o *Dicionário Eletrônico de Elementos Mórficos Houaiss* para sabermos mais informações sobre o morfema “-inho”. A partir dessa consulta, verificamos que se trata de um sufixo latino e então assinalamos também para esse termo o processo de formação: derivação sufixal. Como esse, há também outros casos em que um mesmo termo pode ser resultado de mais de um processo de formação de palavras.

A seguir apresentaremos os resultados das classificações e uma análise dos principais processos de formação presentes nessa terminologia.

Análise dos dados

Foi analisado um total de 135 termos e, na Tabela 6, apresentamos a quantidade de ocorrência dos processos de formação de palavras, em ordem decrescente. Destacamos que em um mesmo termo pode intervir mais de um processo.

Tabela 6 Processos de formação de palavras encontrados.

Processos de formação	Número de ocorrências
Importação de palavras	135
Lexicalização de sintagmas	15
Derivação sufixal	11
Extensão semântica	2
Derivação não afixal	1
Composição morfossintática	1

Fonte: elaborada pela autora.

Ressaltamos que os processos: onomatopeia, derivação prefixal, derivação parassintética, composição morfológica, recomposição, amálgama, truncação, sigla, acronímia, lexicalização de formas flexionadas, lexicalização de unidades discursivas não ocorreram. A seguir, detalharemos os três processos mais frequentes: importação de palavras, lexicalização de sintagmas e derivação sufixal.

4.1 Importação de palavras

As palavras importadas de outros idiomas somam a maioria significativa nesse conjunto terminológico e podemos afirmar que o predomínio são os latinismos. Como exemplo, temos: *diatraea saccharalis*, que aparece 299 vezes no *corpus* e é o nome científico da broca da cana-de-açúcar. Outro termo frequente é *Mahanarva fimbriolata*, que aparece 240 vezes no *corpus* e é o nome científico da cigarrinha-da-raiz.

Foram classificados também como importação de palavras todos os termos cujos radicais se originam diretamente de outro idioma como: mosca (latim: *musca*), carvão (latim: *carbonis*) lagarta (latim: *lacarta*). Esses termos somam 17 dos 135 classificados como

importação e desse total de 17, 14 designam pragas e 3 designam doenças. Além do latim há termos provenientes do tupi (cupim), umbundo (camundongo), espanhol (mariposa), italiano (mosaico) e catalão (broca).

4.2 Lexicalização sintagmática

Em segundo lugar estão os termos formados pela lexicalização de sintagmas, a seguir eles são listados seguidos de trechos retirados do corpus que exemplificam o contexto de uso de cada um:

- *Amarelecimento foliar* (N+Adj) – “A relevância econômica atual da cultura da cana-de-açúcar e a projeção da expansão da área plantada revela uma doença uma doença conhecida como síndrome do amarelecimento foliar da cana-de-açúcar ou amarelinho, que causou danos expressivos à cultura no estado de São Paulo.”
- *Broca gigante da cana-de-açúcar* (N+Adj+Prep+Det+N) – “A primeira ocorrência da broca gigante da cana-de-açúcar foi observada por Costa Lima em 1927 na região Nordeste”. *Cigarrinha da folha* (N+Prep+Det+N) – “A cultura da cana-de-açúcar é danificada por inúmeras pragas, destacando-se a cigarrinha da folha (...)”.
- *Cigarrinha da raiz* (N+Prep+Det+N) – “O controle da cigarrinha da raiz vinha sendo contido, particularmente pelo controle cultural.” *Coração morto* (N+Adj) – “São larvas que ao atacarem plantas novas, causam a morte da gema apical, sintoma conhecido como coração morto”.
- *Corda de viola* (N+Prep+N) – “Características da espécie merremia cissoides (lam.) hall. f. conhecidas popularmente por corda-de-viola, as espécies da família convolvulaceae são plantas nativas da América do Sul, têm ciclo biológico longo, terminando após a maturação das culturas; o que tende a criar problemas na colheita, principalmente em cana-de-açúcar, pois seus ramos se fixam aos colmos da cultura.”
- *Escaldadura das folhas* (N+Prep+Det+N) – “A escaldadura das folhas e o carvão tem nos toletes um eficiente meio de disseminação para novas áreas.”
- *Estria vermelha* (N+Adj) – “Entretanto as doenças causadas pelos fungos, como a “estria vermelha”, a “ferrugem da cana”, “mancha ocular” e “mancha amarela”, são relativamente fáceis de se identificar porque produzem sintomas evidentes nas folhas, uma vez que estão sempre associadas às condições favoráveis do clima e, como essas

doenças não colonizam toda a folha, é mais difícil de ocorrer a transmissão pela muda.”

- *Ferrugem alaranjada da cana-de-açúcar* (N+Adj+Prep+Det+N) – “Em dezembro de 2009, foi oficialmente detectado, na região de araraquara/sp, o primeiro foco da ferrugem alaranjada da cana-de-açúcar. a doença se manifestou de forma bastante agressiva(...)”
- *Mosca branca* (N+Adj) – “ Iniciar o tratamento logo após o aparecimento da praga. repetir se necessário em intervalo de 7 dias na mosca-branca : o nível de ação, a partir do qual o controle químico dos percevejos deve ser realizado, é de 1 indivíduo em média por planta.”
- *Mucuna-preta* (N+Adj) – “(...) áreas de cana com manejo de rotação de cultura com a mucuna-preta (s. aterrimum) durante as reformas dos canaviais.”
- *Picão-preto* (N+Adj) – “ Picão-preto tem pré-emergência em solo leve”.
- *Podridão vermelha* (N+Adj) – “Também é comum a entrada de diversos microrganismos, especialmente fungos, que causam a “ podridão vermelha ”, com redução do teor de açúcar nos colmos devido à inversão da sacarose armazenada na planta e sua transformação em glicose e levulose.”
- *Sorghum bicolor* (N+Adj) – “(...) dentre as 31 diferentes espécies de gramíneas, apenas sorghum bicolor (sorgo), zea mays (milho) e avena sativa l. (aveia) apresentaram sintomas quando inoculados com xanthomonas sp. agente causal da fev da cana-de-açúcar.”
- *Vírus da granulose* (N+ Prep+Det+N) – “O vírus da granulose de *diatraea saccharalis* (dsgv) foi citado pela primeira vez em larvas procedentes de Louisiana, EUA, em 1960”.
- *Vírus do mosaico* (N+Prep+Det+N) – “A cultura da cana-de-açúcar apresenta muitos problemas fitossanitários que prejudicam a produção, entre eles o causado pelo vírus do mosaico da cana-de-açúcar (sugarcane mosaic virus, scmV). o mosaico da cana-de-açúcar vem sendo controlado com o uso de variedades e híbridos resistentes.”

As formações que se destacam são N+Adj (8 ocorrências) e N+Prep+Det+N (5 ocorrências). Em seguida temos N+Adj+Prep+Det+N (2 ocorrências) e N+Prep+N (1 ocorrência) .

4.3 Derivação sufixal

Como pôde ser constatado pela Tabela 6, a derivação sufixal ocupa o terceiro lugar como processo mais frequente ocorrendo em 10 termos. Os casos de sufixação são listados e assim como na subseção anterior, seguem os trechos retirados do corpus para exemplificar seu uso.

- *Amarelecimento* – “O dano causado por esse inseto ocorre no rizoma da planta abaixo do nível do solo, local onde a larva se alimenta e se abriga, abrindo galerias circulares e longitudinais na base da brotação. esse ataque resulta no amarelecimento da folha e na morte do perfilho”.
- *Amarelinho* – “A síndrome do amarelecimento foliar, também conhecida por ” amarelinho”, foi relatada no Brasil em 1989, mas começou a se tornar problema sério a partir do início da década de 90. ”.
- *Cigarrinha da raiz* - “O controle da cigarrinha da raiz vinha sendo contido, particularmente pelo controle cultural. ”.
- *Cigarrinha da folha* - “A cultura da cana-de-açúcar é danificada por inúmeras pragas, destacando-se a cigarrinha da folha (...)”.
- *Escaldadura das folhas* - “A escaldadura das folhas e o carvão tem nos toletes um eficiente meio de disseminação para novas áreas. ”.
- *Ferrugem alaranjada da cana-de-açúcar* - “Em dezembro de 2009, foi oficialmente detectado, na região de Araraquara/sp, o primeiro foco da ferrugem alaranjada da cana-de-açúcar. a doença se manifestou de forma bastante agressiva(...)”
- *Levedura* – “(...) quanto à concentração de açúcares, o estresse induzido pelo aumento da osmolaridade externa leva à redução em crescimento e perda da viabilidade das células das leveduras, devido às perturbações no gradiente osmótico (...)”.
- *Podridão abacaxi* – “O controle da podridão abacaxi da cana-de-açúcar por meio da pulverização de fungicidas em mudas no sulco de plantio. nos últimos anos, o plantio mecanizado da cana-de-açúcar vem sendo cada vez mais utilizado no Brasil.”.
- *Podridão vermelha* - “Também é comum a entrada de diversos microrganismos, especialmente fungos, que causam a “podridão vermelha”, com redução do teor de açúcar nos colmos devido à inversão da sacarose armazenada na planta e sua transformação em glicose e levulose.”

- *Raquitismo* – “(...) as doenças que ocorrem na cana-de-açúcar têm se constituído em fatores que limitam o rendimento da cultura. mais de uma centena de doenças já foram identificadas (16), sendo o carvão, o mosaico, a escaldadura das folhas, o raquitismo e a ferrugem consideradas as mais importantes”.
- *Vírus da granulose* - “O vírus da granulose de *diatraea saccharalis* (dsgv) foi citado pela primeira vez em larvas procedentes de Louisiana, EUA, em 1960”.

Os sufixos presentes são: *-ar* (*foliar*), *-mento* (*amarelecimento*), *-inh(o/a)* (*amarelinho, cigarrinha*), *-ura* (*levedura*), *-dão* (*podridão*), *-ismo* (*raquitismo*) e *-ose* (*granulose*).

Podemos constatar, na formação dos termos por derivação sufixal, a seguinte recorrência: raiz latina + sufixo latino como em:

- Amarelinho: amarelo + -inho [amarelo: lat. hispânico *amarellus*] + [-inho: suf. lat.]
- Escaldadura: *escalda(r)* + *-ad(o)* + *-ura* [escaldar: lat. *excaldare*] + [-ado: suf. lat.] + [-ura: suf. lat.] = escaldadura + folhas: folha: lat. *folium*
- Levedura: levedo + *-ura* [levedo: lat. *Levitus*] + [-ura: suf. Lat.]

4.4 Comparação com os trabalhos já realizados

O principal diferencial da terminologia das pragas e doenças da cana-de-açúcar é o número elevado de importações de palavras do latim, que não ocorre na demais. A presença dos nomes científicos das pragas se destaca nesse conjunto terminológico e podemos listar duas razões. A primeira é que essa terminologia é a única dentre as que descrevemos que designa seres vivos e, ao se tratar de fungos, bactérias, insetos e planta, é comum que se remeta ao nome científico. A segunda razão é que esses termos foram extraídos de um *corpus* escrito e com grande número de textos do gênero científico e acadêmico. Sendo assim, é natural a presença dos nomes em latim, o que não aconteceria, por exemplo, se lidássemos com *corpus* na modalidade oral e com discursos dos agricultores em vez de pesquisadores.

Na terminologia da Educação a Distância, o processo de empréstimo é o quinto mais frequente, quase empatando com a prefixação, mas todos os termos estrangeiros são importados do inglês. Já na terminologia da Fisioterapia, o número de importações é pouco significativo e reduzido e no trabalho sobre Nanociência e Nanotecnologia e Biodiesel, o autor não considerou os empréstimos em sua análise.

O processo de prefixação, que não ocorre em nosso conjunto, está presente nas demais terminologias como um dos mais frequentes (Nanociência e Nanotecnologia, Biodiesel e Educação a Distância). A composição morfológica também é uma das mais frequentes no conjunto da Fisioterapia, da Nanociência e Nanotecnologia e do Biodiesel, enquanto que em nosso trabalho ocorreu apenas uma vez.

A semelhança em todos os trabalhos mais é a alta frequência da derivação sufixal. Trata-se do processo mais produtivo nas terminologias da Nanociência e Nanotecnologia e do biodiesel; e o segundo mais frequente em Fisioterapia e Educação a Distância e o terceiro em nosso conjunto terminológico. Nos trabalhos da Nanociência e Nanotecnologia, Biodiesel e Educação a Distância, os sufixos mais presentes são os mesmos: *-mento* e *-são*. A área da Fisioterapia apresenta nesse processo sua particularidade enquanto área médica, com a presença dos sufixos *-ite* e *-ose*. Já em nosso trabalho, o que ocorre mais de uma vez são *-ura*, *-inh(o/a)* e *-dão*.

O nosso segundo processo mais frequente é a lexicalização de sintagmas, que aparece como o mais frequente nas terminologias da Educação a Distância e Fisioterapia. A formação mais recorrente é N + ADJ em todas as terminologias descritas nesse trabalho, como “educação superior”, “articulação plana”, “podridão vermelha” e “mosca branca”.

A seguir apresentamos nossas conclusões retomando os objetivos iniciais e a proposta dessa pesquisa.

Conclusão

Como um dos objetivos de nosso trabalho é integrar uma base dados morfológicos de terminologias, realizamos uma leitura atenta de outros trabalhos semelhantes sobre outras terminologias, são elas: Nanociência e Nanotecnologia, Biodiesel, Educação a Distância e Fisioterapia. Pudemos verificar as particularidades de cada uma e ver em quais aspectos a nossa terminologia se diferencia e se assemelha a elas. A partir dessa comparação pudemos constatar que o principal diferencial do conjunto terminológico aqui descrito é o número elevado de importações. Também percebemos que os processos de sufixação e lexicalização sintagmática estão presentes em todas as terminologias aqui descritas como processos consideravelmente produtivos na formação de termos.

Em nosso conjunto terminológico os processos mais produtivos presentes na formação dos termos foram respectivamente: importação de palavras (ocorreu em todos os casos, ou seja, em todos os termos), lexicalização de sintagmas em (ocorreu em 15 termos) e sufixação (ocorreu em 10 termos), extensão semântica (ocorreu em 2 termos), derivação não afixal (ocorreu em 1 termo) e composição morfossintática (ocorreu em 1 termo). Lembrando que um mesmo termo pode conter mais de um processo, como é o caso de “raquitismo”, por exemplo, que é considerado resultado de um processo de sufixação (devido ao sufixo “-ismo”) e importação de palavras, pois “raquit(i)-“ é proveniente do grego *rakhítes*.

Também foi constatado após a análise que não há novos constituintes morfológicos, mas sim uma reutilização dos elementos que já se encontram disponíveis na língua portuguesa. No caso dos termos resultantes de sufixação encontramos sufixos que também circulam com frequência na língua geral como: -inho(a) e -ar. Tal fato reforça a escolha da TCT como teoria e metodologia em trabalhos terminológicos, pois ela defende que os termos são palavras da língua geral que são empregados em contextos de especialidade e a partir dessa inclusão contextual a palavra se torna um termo de determinado domínio do saber. Por isso faz sentido que os componentes morfológicos dos termos sejam, em sua maioria, os mesmos da língua geral. Podemos notar, dentre os trabalhos retomados que apenas no domínio da Fisioterapia existem componentes próprios da área médica e que não são comuns na língua geral, como: -ite e -ose.

Percebemos que, com a decisão de reduzir o conjunto terminológico original, conforme a sugestão dada pela banca no exame de qualificação houve um empobrecimento no que se refere à variedade dos processos de formação de palavras. O conjunto terminológico

anterior apresentava termos mais abrangentes, como os relativos a sintomas e métodos de tratamento, e não apenas referentes a nomes de pragas e doenças. Ao optarmos por retirá-los da lista a ser descrita, o número de termos reduziu consideravelmente (de 941 para 135). Comparando com os outros trabalhos, esse número é baixo e isso poderia ocasionar um questionamento sobre a validade e confiabilidade desse trabalho. Entretanto, podemos afirmar que, durante a fase de compilação de *corpus*, foram buscadas fontes seguras e tudo o que encontramos de confiável referente às pragas e doenças da cana-de-açúcar foi inserido no trabalho. O número de termos também é reduzido em relação aos demais, pois o domínio é mais restrito. Fisioterapia, por exemplo, é um campo de saber muito amplo, diferentemente de pragas e doenças da cana-de-açúcar.

A descrição morfológica desse campo do saber é um elemento importante na composição de uma base de dados morfológicos a partir de descrições de terminologias, pois traz uma área que, diferentemente das quatro aqui detalhadas, não apresenta variedade de processos produtivos. O que podemos concluir a partir disso que é de grande importância que mais terminologias sejam descritas morfológicamente, para verificarmos se existem mais diferenças contrastivas com ocorreu no caso das pragas e doenças da cana-de-açúcar.

Acreditamos ter concretizado os objetivos propostos no início do trabalho que são: a análise morfológica do conjunto terminológico de 135 termos pertencentes ao domínio das pragas e doenças da cana-de-açúcar, apresentar esses dados de forma que seja possível uma futura implementação computacional e oferecer conhecimento sobre a morfologia derivacional do português em terminologias.

Como mostramos na Introdução, é necessário que sejam realizadas pesquisas que visem o controle das pragas e doenças da cana-de-açúcar e para que isso ocorra sem nenhum tipo de ruído de comunicação entre os especialistas do domínio é necessário que o conhecimento sobre o tema esteja sempre renovado e organizado, por isso o trabalho do terminólogo e do linguista é importante.

Para pesquisa futura, acreditamos que seria enriquecedor trabalhar com *corpus* de caráter menos científico e mais informal como entrevistas com agricultores, por exemplo. Assim poderia ser realizada uma comparação para constatar se os termos mais frequentes são semelhantes ao conjunto terminológico dessa pesquisa e quais podem ser acrescentados, assim como verificar se os processos mais produtivos de formação morfológica seguem a mesma tendência.

Esperamos que nosso trabalho possa vir a integrar e enriquecer bases de dados linguísticos e terminológicos, contribuindo na construção e aprimoramento de sistemas híbridos de extração automática de termos.

Referências

- ALVES, I. M. *Neologismo: criação lexical*. São Paulo: Ática, 2004.
- ALMEIDA, G. M. B. *Neologismos na Informática: natureza e deriva de um vocabulário*. Dissertação (Mestrado em Linguística e Língua Portuguesa). Araraquara: Faculdade de Ciências e Letras, Universidade Estadual Paulista, 1995.
- _____. Fazer Terminologia é fazer Linguística. In: PERNA, C. L., DELGADO, H. K. FINATTO, M. J. (Orgs.). *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa*. 1. ed. Porto Alegre: Editora da PUCRS, 2010, p. 72-90. (vol. 1).
- _____. O percurso da Terminologia: de atividade prática à consolidação de uma disciplina autônoma. *Tradterm*, São Paulo, v. 9, 2003, p. 211-222.
- _____. A Teoria Comunicativa da Terminologia e a sua prática. *Alfa* (IBILCE/UNESP), v. 50, 2006, p. 81-97.
- _____. Terminologia: o que é e como se faz. In: GONÇALVES, A. V.; GÓIS, M. L. S. (Orgs.). *Ciências da Linguagem: O Fazer Científico?* Campinas: Mercado de Letras, 2012, p. 197-230. (vol. 1).
- ALMEIDA, G. M. B.; OLIVEIRA, L. H. M. Terminology and computational linguistics: new praxes in terminography. *Cahiers de Lexicologie*, Paris, v. 101, 2012, p. 139-153.
- ALMEIDA, G. M. B.; OLIVEIRA, L. H. M.; ALUÍSIO, S. M. A Terminologia na era da Informática. *Ciência e Cultura (SBPC)*, Campinas, v. 58, n. 2, 2006, p. 42-45.
- AMBIENTE COLABORATIVO WEB DE GESTÃO TERMINOLÓGICA (E-TERMOS). Disponível em: <www.etermos.cnptia.embrapa.br>. Acesso em: 17 set. 2015.
- ARONOFF, M.; FUDEMAN, K. *What is Morphology?* Oxford: Blackwell Publishing, 2005.
- BARROS, L. A. *Curso básico de Terminologia*. São Paulo: Editora da Universidade de São Paulo, 2004.
- BECHARA, E. *Moderna Gramática Portuguesa*. 37. ed. Rio de Janeiro: Editora Lucerna, 1999.
- BIDERMAN, M. T. C. *Teoria linguística (Teoria lexical e linguística computacional)*. São Paulo: Martins Fontes, 2001.
- BIODIESEL BR. *PróAlcool – Programa Brasileiro de Álcool*. 2006. Disponível em: <<http://www.biodieselbr.com/proalcool/pro-alcool/programa-etanol.htm>>. Acesso em: 18 set. 2015.
- BLOOMFIELD, L. *Language*. Nova Iorque: Henry Holt, 1933.

BORBA, F. S. *Introdução aos estudos linguísticos*. 11. ed. Campinas: Pontes, 1991.

BRASIL. Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. Disponível em <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_53_711200516718.html>. Acesso em: 19 ago. 2015.

BRASIL. Ministério da Agricultura. *Cana-de-açúcar*. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/cana-de-acucar>>. Acesso em: 11 abr. 2016.

CABRAL, L. S. *Introdução à linguística*. Porto Alegre: Editora Globo, 1979.

CABRÉ, M. T. Theories of terminology: their description, prescription and explanation. *Terminology*, v. 9, n. 2, p. 163-200, 2003.

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

_____. *Bases de dados morfológicos de terminologias do português do Brasil. Descrição e análise morfológica com vistas à disponibilização on-line*. Dissertação (Mestrado em Linguística). Universidade Federal de São Carlos, 2012.

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

CORREIA, M. Bases digitais lexicais na União Europeia. In: SIMPÓSIO DE LEXICOLOGIA, LEXICOGRAFIA E TERMINOLOGIA, 1994, Araraquara. *Anais...* Araraquara: 1994, p. 1-37. Disponível em: <<http://www.iltec.pt/pdf/wpapers/1994-mcorreia-bdigitais.pdf>>. Acesso em: 10 jun. 2015.

CORREIA, M.; ALMEIDA, G. *Neologia em Português*. São Paulo: Parábola, 2012.

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

DIAS-DA-SILVA, B. C. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Tese (Doutorado em Letras). Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara: 1996, 272 p.

DIAS-DA-SILVA, B. C.; MONTILHA, G.; RINO, L. H. M.; SPECIA, L.; NUNES, M. das G. V.; Oliveira JR., O. N.; MARTINS, R. T.; PARDO, T. A. S. *Introdução ao Processamento das Línguas Naturais e Algumas Aplicações*. NILC. NÚCLEO INTERINSTITUCIONAL DE LINGÜÍSTICA COMPUTACIONAL. NILC-TR-07-10, agosto 2007.

DI FELIPPO, A.; DIAS-DA-SILVA, B. C. Dos olhares sobre o léxico: diferenças e semelhanças. In: LONGO, B. N. O.; DIAS-DA-SILVA, B. C.; BIDERMAN, M. T. C. (Orgs.). *A construção de dicionários e bases de conhecimento lexical*. 1. ed. Araraquara: Laboratório Editorial (FCL/UNESP), 2006, p. 169-185. (vol. 9).

FERREIRA, A. B. H. *Novo dicionário eletrônico Aurélio versão 6.0*. 4. ed. Curitiba: Positivo, 2009.

GIANOTI, A. C. *Descrição e análise morfológica da terminologia do domínio da “Educação à Distância”*. Trabalho de Conclusão de Curso (Bacharelado em Linguística). Departamento de Letras, Universidade Federal de São Carlos: São Carlos, 2013.

GUILBERT, Louis. *La créativité lexicale*. Paris: Larousse, 1975.

INSTITUTO ANTÔNIO HOUAISS. *Dicionário eletrônico Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Banco de dados agregado*. Disponível em: <<http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=1612&z=t&o=1&i=P>>. Acesso em: 10 mar. 2013.

MARTÍNEZ, S. M. Estructuración conceptual y formalización terminográfica de frasesmas en el subdominio de la oncología. In: *Estudios de Lingüística del Español (ELiEs)*, v. 19, 2003. Red Iris/Universidad Autónoma de Barcelona. Disponível em: <<http://elies.rediris.es/elies19/cap13222.html>>. Acesso em: 3 abr. 2016.

MATTOS, D. F. *Descrição e análise morfológica da terminologia da Fisioterapia: subsídios para organização de uma base de dados morfológicos de terminologias do português do Brasil*. Dissertação (Mestrado em Linguística). Universidade Federal de São Carlos, 2013.

OLIVEIRA, C. C. *A terminologia das pragas da cana-de-açúcar*. Trabalho de Conclusão de Curso. Universidade Federal de São Carlos: São Carlos, 2013, 185 p.

OLIVEIRA, L. H. M. *e-Termos: um ambiente colaborativo web para a criação de produtos terminológicos*. Tese (Doutorado em Computação). Instituto de Ciência Matemática e Computação. Universidade de São Paulo: São Carlos, 2006.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA ALIMENTAÇÃO E AGRICULTURA (FAO). Secretaria da Convenção Internacional para a Proteção dos Vegetais. Normas Internacionais para Medidas Fitossanitárias. *NIMF nº 5: Glossário de Termos Fitossanitários*. FAO, 2009. Disponível em: <http://www.arp.cnptia.embrapa.br/documentos-orientadores/NIMF_05_2009_PT.pdf>. Acesso em: 20 maio 2015.

LAROCCA, M. N. C. *Manual de Morfologia do Português*. 3. ed. Campinas: Pontes, 2004.

MATORÉ, G. *La méthode en lexicologie*. Paris: Marcel Didier, 1953.

RONDEAU, G. Problems and Methods of Terminological Neology (Neonymy). *Infoterm*, München, n. 6, 1983.

ROCHA, L. C. *Estruturas Morfológicas do Português*. São Paulo: WMF Martins Fontes, 2008.

ROSSETTO, R.; SANTIAGO, A. D. *Pragas*. AGÊNCIA EMBRAPA DE INFORMAÇÃO TECNOLÓGICA (AGEITEC). *Árvore do conhecimento cana-de-açúcar*.

Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_53_711200516718.html>. Acesso em: 11 abr. 2016.

SAUSSURE, F. *Curso de Linguística Geral*. Organizado por Charles Bally e Alberte Sechehaye, com a colaboração de Albert Riedlinger. 4. ed. São Paulo: Cultrix, 1972.

Apêndice

Tabela com a descrição morfológica dos termos do domínio
“pragas e doenças da cana-de-açúcar”

Termo	Observações	Frequência no corpus	Classificação
AMARELECIMENTO	amarelecer + -mento [amarelecer: amarelo + -ecer [amarelo: lat. hispanico amarellus] + [-ecer terminação de v. latinos] = amarelecer] + [-mento: lat. mentu]	105	Derivação sufixal e Importação de palavras
ACIDOVORAX AVENAE		11	Importação de palavras
AGROBACTERIUM TUMEFACIENS		32	Importação de palavras
AMARANTHUS HYBRIDUS		36	Importação de palavras
AMARELINHO	amarelo + -inho [amarelo: lat. hispanico amarellus] + [-inho: suf. lat.]	12	Derivação sufixal e Importação de palavras
ANTHONOMUS GRANDIS		18	Importação de palavras
ANTICARSIA GEMMATALIS		9	Importação de palavras
ASPERGILLUS NIDULANS		34	Importação de palavras
ASPERGILLUS TRICHODERMA		46	Importação de palavras
BACILLUS		187	Importação de palavras
BIDENS PILOSA		28	Importação de palavras

BORBOLETA		17	Importação de palavras
BOTRYTIS CINEREA		12	Importação de palavras
BOUGAINVILLEA SPECTABILIS		17	Importação de palavras
BRACHIARIA DECUMBENS		64	Importação de palavras
BRACONIDAE		36	Importação de palavras
BROCA GIGANTE DA CANA-DE-AÇÚCAR	broca: catalão. broca 'saliente, pontiagudo', prov. de orig. céltica	598	Lexicalização sintagmática e importação de palavras
CAMUNDONGO	umbundo. Okamundongo	9	Importação de palavras
CARVÃO	lat. <i>Carbonis</i>	154	Importação de palavras
CASTNIA		16	Importação de palavras
CASTNIIDAE		16	Importação de palavras
CERATITIS CAPITATA		18	Importação de palavras
CERATOCYSTIS PARADOXA		34	Importação de palavras
CERCOPIIDAE		174	Importação de palavras
CIGARRINHA DA FOLHA	cigarrinha: cigarra + -inha [cigarra: or. Controversa + -inha: suf. Lat], folha: lat. <i>Folium</i>	15	Derivação sufixal, Importação de palavras, Lexicalização sintagmática
CIGARRINHA DA RAIZ	cigarrinha: cigarra + -inha [cigarra: or. Controversa + -inha:	645	Derivação sufixal, Importação de palavras,

	suf. Lat], raiz: lat. <i>Radix</i>		Lexicalização sintagmática
COLEOPTERA		224	Importação de palavras
COLLETOTRICHUM FALCATUM		9	Importação de palavras
COLÔNIA	lat. <i>Colonia</i>	229	Importação de palavras
COMMELINA BENGHALENSIS		11	Importação de palavras
CORAÇÃO MORTO	coração: or. Contr. Posv. do lat. Cor, <i>cordis</i> ; morto: lat. <i>Mortuus</i>	15	Lexicalização sintagmática e Importação de palavras
CORDA DE VIOLA	corda: lat. <i>Chorda</i> , viola: lat. Gen. <i>Viola</i> , do lat. <i>Viola</i>	19	Lexicalização sintagmática, Extensão semântica e Importação de palavras
COSMOPOLITES SORDIDUS		31	Importação de palavras
COTESIA FLAVIPES		70	Importação de palavras
CUPIM	tupi: <i>kupi'i</i>	35	Importação de palavras
CURCULIONIDAE		160	Importação de palavras
CURCULIONÍDEO	lat. Cient. fam. <i>Curculionidae</i>	61	Importação de palavras
CYPERUS ROTUNDUS		9	Importação de palavras
DERMAPTERA		9	Importação de palavras
DIABROTICA SPECIOSA		9	Importação de palavras
DIATRAEA		299	Importação de

SACCHARALIS			palavras
DIGITARIA HORIZONTALIS		17	Importação de palavras
DIPTERA		44	Importação de palavras
DORU LUTEIPES		9	Importação de palavras
DROSOPHILIDAE		10	Importação de palavras
ESCALDADURA DAS FOLHAS	escaldadura: escalda(r) + -ad(o) + -ura [escaldar: lat. <i>excaldare</i>] + [-ado: suf. lat.] + [-ura: suf. lat.] = escaldadura + folhas: folha: lat. <i>folium</i>	29	Derivação sufixal, Importação de palavras, Lexicalização sintagmática
ESCARABÉÍDEO	lat. Cien. Fam. <i>Scarabaeidae</i> , lat. <i>Scarabeus</i>	16	Importação de palavras
ESCHERICHIA COLI		36	Importação de palavras
ESTRIA VERMELHA	estria: lat. <i>Striga</i> ; vermelha: vermelho: lat. <i>Vermiculus</i>	18	Lexicalização sintagmática, Importação de palavras
FERRUGEM ALARANJADA DA CANA-DE-AÇÚCAR	ferrugem: lat. <i>Ferrugo</i> ; alaranjada: alaranjado: part. De alaranjar; cana (lat. <i>Canna</i>) e açúcar (ár. <i>As-sukkar</i>)	347	Derivação sufixal, Lexicalização sintagmática, Importação de palavras
FORMIGA	lat. <i>Formica</i> , ae	43	Importação de palavras
FUSARIUM MONILIFORME		71	Importação de palavras
FUSARIUM OXYSPORUM		79	Importação de palavras

FUSARIUM VERTICILLIOIDES		69	Importação de palavras
GELECHIIDAE		11	Importação de palavras
GLIOCLADIUM VIRENS		55	Importação de palavras
GORGULHO	lat. Gurgulio e curculio	30	Importação de palavras
HELICOVERPA ZEA		9	Importação de palavras
HEMIPTERA		121	Importação de palavras
HETEROTERMES TENUIS		14	Importação de palavras
HOMOPTERA		36	Importação de palavras
HYMENOPTERA		83	Importação de palavras
HYPONEUMA TALTULA		14	Importação de palavras
ISOPTERA		17	Importação de palavras
LAGARTA	prov. Lat. <i>Lacarta</i>	630	Importação de palavras
LARVA	lat. <i>Larva</i>	334	Importação de palavras
LEIFSONIA XYLI		142	Importação de palavras
LEPIDOPTERA		363	Importação de palavras
LEVEDURA	levedo + -ura [levedo: lat. <i>Levitus</i>] + [-ura: suf. Lat.]	274	Derivação sufixal, Importação de palavras
LEVIS VAURIE		31	Importação de palavras
LUTEOVÍRUS	lute(i)-: el. Comp. Lat.	19	Importação de

	+ vírus: lat. <i>Virus</i>		palavras
LYMANTRIIDAE		11	Importação de palavras
LYMATRIA		16	Importação de palavras
MAGNAPORTHE GRISEA		11	Importação de palavras
MAHANARVA FIMBRIOLATA		240	Importação de palavras
MAHANARVA POSTICATA		9	Importação de palavras
MARIPOSA	esp. <i>Mariposa</i>	88	Importação de palavras
METAMASIUS HEMIPTERUS		12	Importação de palavras
METARHIZIUM ANISOPLIAE		231	Importação de palavras
MIGDOLUS FRYANUS		8	Importação de palavras
MOSAICO	it. <i>Mosaico</i>	100	Importação de palavras
MOSCA	lat. <i>Musca</i>	54	Importação de palavras
MOSCA BRANCA	mosca: lat. <i>Musca</i> ; branca: branco: germ. <i>Blanck</i>	11	Lexicalização sintagmática, Importação de palavras
MUCUNA-PRETA	lat. Cien. Gen. <i>Mucuna</i> + preta: preto: <i>prettu</i>	10	Lexicalização sintagmática, Importação de palavras
NOCTUIDAE		69	Importação de palavras
NOMURAEA RILEYI		9	Importação de palavras
NOSEMA		375	Importação de palavras

ORTHOPTERA		9	Importação de palavras
OSTRINIA NUBILALIS		19	Importação de palavras
PANTOEA AGGLOMERANS		51	Importação de palavras
PENICILLIUM		40	Importação de palavras
PICÃO-PRETO	picão: pica + -ão [pica: regr. De picar] + [-ão: suf. Lat.]; preto: lat. Prettu	22	Derivação não-afixal, Derivação sufixal, Lexicalização sintagmática, Importação de palavras
PLUTELLA XYLOSTELLA		10	Importação de palavras
POACEAE		9	Importação de palavras
PODRIDÃO ABACAXI	podridão: podre + -i- + -dão [podre: lat. <i>Putris</i>] + [-i-: vogal de ligação] + [-dão: suf. Lat.]; abacaxi: tupi <i>iwaka'ti</i>	209	Derivação sufixal, Importação de palavras, Lexicalização sintagmática
PODRIDÃO VERMELHA	podre + -i- + -dão [podre: lat. <i>Putris</i>] + [-i-: vogal de ligação] + [-dão: suf. Lat.]; vermelha: vermelho: lat. <i>Vermiculus</i>	111	Derivação sufixal, Importação de palavras, Lexicalização sintagmática
POTYVIRUS		62	Importação de palavras
PROSAPIA BICINCTA		9	Importação de palavras
PSEUDOMONAS		98	Importação de

FLUORESCENS			palavras
PSEUDOMONAS SYRINGAE		76	Importação de palavras
PUCCINIA KUEHNII		119	Importação de palavras
PUCCINIA MELANOCEPHALA		117	Importação de palavras
PYRALIDAE		43	Importação de palavras
PYRAUSTA		29	Importação de palavras
PYTHIUM ULTIMUM		36	Importação de palavras
RAQUITISMO	raquit(i)- + -ismo [raquit(i)-: gr. <i>Rakhítes</i> + -ismo: suf. Gr.]	161	Derivação sufixal, Importação de palavras
RHIZOCTONIA SOLANI		20	Importação de palavras
RHYNCHOPHORUS PALMARUM		17	Importação de palavras
SACCHARALIS LARVAE		12	Importação de palavras
SACCHAROMYCES CEREVISIAE		28	Importação de palavras
SCARABAEIDAE		17	Importação de palavras
SERRATIA MARCESCENS		9	Importação de palavras
SOLENOPSIS		14	Importação de palavras
SORGHUM BICOLOR	bicolor: lat. <i>Bicolor</i> , <i>oris</i>	10	Lexicalização sintagmática, Importação de palavras
SPHENOPHORUS LEVIS		121	Importação de palavras

SPITTLEBUG MAHANARVA		11	Importação de palavras
SPODOPTERA FRUGIPERDA		32	Importação de palavras
SPODOPTERA LITURA		28	Importação de palavras
SPORISORIUM SCITAMINEUM		19	Importação de palavras
STEINERNEMA BRAZILENSE		9	Importação de palavras
STEINERNEMATIDAE		10	Importação de palavras
STIZOLOBIUM ATERRIMUM		33	Importação de palavras
TELCHIN LICUS		69	Importação de palavras
TEPHRTIDAE		20	Importação de palavras
TORTRICIDAE		12	Importação de palavras
TRAÇA	segundo Nascentes, regr. traçar	28	Derivação não-afixal
TRICHODERMA HARZIANUM		50	Importação de palavras
TRICHODERMA VIRENS		41	Importação de palavras
TRICHOGRAMMA GALLOI		121	Importação de palavras
USTILAGO SCITAMINEA		74	Importação de palavras
VESPA	lat. <i>vespa,ae</i>	41	Importação de palavras
VÍRUS DA GRANULOSE	vírus: lat. <i>Vírus</i> , granulose: grânulo + - ose [grânulo: lat. <i>Granulu</i>] + [-ose: suf.	15	Derivação sufixal, Lexicalização sintagmática, Importação de

	Gr.]		palavras
--	------	--	----------

VÍRUS DO MOSAICO	vírus: lat. <i>Vírus</i> , mosaico: it. <i>Mosaico</i>	31	Lexicalização sintagmática, Importação de palavras
XANTHOMONAS ALBILINEANS		285	Importação de palavras
XYLELLA FASTIDIOSA		37	Importação de palavras