

EDUARDO BATISTA DA SILVA

PROPOSTA DE UM DICIONÁRIO ELETRÔNICO
TERMINOLÓGICO ONOMASIOLÓGICO BILÍNGUE INGLÊS-
PORTUGUÊS NO DOMÍNIO DAS REDES NEURAIAS
ARTIFICIAIS

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista, Câmpus de São José do Rio Preto, para obtenção do título de Mestre em Estudos Lingüísticos (Área de Concentração: Análise Lingüística)

Orientador: Prof. Dr. Maurizio Babini

São José do Rio Preto
2008

Silva, Eduardo Batista da.

Proposta de um dicionário terminológico onomasiológico bilingue inglês-português no domínio das redes neurais artificiais / Eduardo Batista da Silva. – São José do Rio Preto : [s.n.], 2009
143 f. : il. ; 30 cm.

Orientador: Maurizio Babini

Dissertação (mestrado) – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas

1. Terminologia - Dicionários. 2. Redes neurais (Computação) - Dicionários. 3. Dicionários onomasiológicos. 4. Dicionários bilingues – Português–Inglês. 5. Dicionários bilingues – Inglês–Português. 6. Dicionários eletrônicos. I. Babini, Maurizio. II Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III Título.

CDU – 81'374:004.02.26

Ficha catalográfica elaborada pela Biblioteca do IBILCE
Campus de São José do Rio Preto - UNESP

COMISSÃO JULGADORA

Titulares

Prof. Dr. Maurizio Babini - Orientador
Prof. Dr. Ricardo Baptista Madeira
Prof. Dr. Eli Nazareth Bechara

Suplentes

Prof. Dr. Evandro Silva Martins
Prof^a. Dr^a. Diva Cardoso de Camargo

EDUARDO BATISTA DA SILVA**PROPOSTA DE UM DICIONÁRIO ELETRÔNICO TERMINOLÓGICO ONOMASIOLÓGICO BILÍNGUE
INGLÊS-PORTUGUÊS NO DOMÍNIO DAS REDES NEURAS ARTIFICIAIS**

Dissertação apresentada para obtenção do título de Mestre em 1 de junho de 2009, área de Análise Lingüística junto ao Programa de Pós-Graduação em Estudos Lingüísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

BANCA EXAMINADORA

Prof^ª. Dr. Maurizio Babini
Professor Assistente Doutor
UNESP – São José do Rio Preto
Orientador

Prof^ª. Dr. Ricardo Baptista Madeira
Professor Doutor
UniFMU

Prof. Dr. Eli Nazareth Bechara
Professor Assistente Doutor
UNESP – São José do Rio Preto

São José do Rio Preto, 1 de junho de 2009

DEDICATÓRIA

Dedico esse trabalho
aos meus pais, pelo esforço em me educarem e sempre acreditarem em mim;
à Thaís, pela força e pelo apoio incondicional;
à Dona Fátima, pelos serviços reprográficos;
à Giovanna, que pode vir a se interessar pelas ciências da linguagem;
à Débora, a irmã que mais gosto;
ao grande amigo Artur, pelas dicas preciosas;
ao companheiro Celso, que apresentou-me à Lingüística de Corpus;
aos professores da ETEC Philadelpho Gouvêa Netto, pelo suporte informático;
aos amigos das repúblicas Curva de Rio, Saideira e Barca Furada;
ao Dr. Bruno e à Dra. Sílvia pelo auxílio na hora H;
à Profa. Lídia, pelo embasamento terminológico;
à Profa. Diva e ao Prof. Eli, pelas observações na banca de qualificação;
ao Professor Maurizio, pela orientação, pela paciência e pelos oportunos puxões de orelha;
a todos que contribuíram de alguma forma para a elaboração deste projeto.

SUMÁRIO

INTRODUÇÃO.....	10
1. FUNDAMENTAÇÃO TEÓRICA.....	15
<i>1.1. A Terminologia.....</i>	<i>15</i>
1.1.1. A lexicologia e a terminologia.....	19
1.1.2. O trabalho terminológico.....	21
1.1.3. A Teoria Comunicativa da Terminologia.....	23
<i>1.2. A Onomasiologia.....</i>	<i>24</i>
1.2.1. Os dicionários onomasiológicos.....	26
1.2.2. A questão da equivalência.....	31
1.2.3. Dicionários eletrônicos.....	33
<i>1.3. A lingüística computacional.....</i>	<i>38</i>
<i>1.4. A lingüística de corpus.....</i>	<i>41</i>
1.4.1. A ferramenta WordSmith Tools.....	43
1.4.2. A frequência de palavras.....	45
1.4.3. O corpus de especialidade.....	47
1.4.4. O etiquetamento.....	50
<i>1.5. O contraponto de Chomsky.....</i>	<i>51</i>
<i>1.6. A tradução e a terminologia.....</i>	<i>53</i>
<i>1.7. A tradução e os dicionários especializados.....</i>	<i>56</i>
<i>1.8. As redes neurais artificiais e suas aplicações.....</i>	<i>59</i>
2. METODOLOGIA DA PESQUISA.....	62
<i>2.1. Delimitação do domínio.....</i>	<i>62</i>
<i>2.2. Constituição do corpus.....</i>	<i>62</i>
<i>2.3. Comparação entre os corpora.....</i>	<i>63</i>
<i>2.4. Extração dos candidatos a termo.....</i>	<i>67</i>
<i>2.5. Seleção dos termos.....</i>	<i>69</i>
<i>2.6. A organização das fichas terminológicas.....</i>	<i>70</i>
<i>2.7. Ferramentas eletrônicas para elaboração de dicionários.....</i>	<i>76</i>
2.7.1. A elaboração do dicionário eletrônico no suporte CDS/ISIS.....	76
2.7.2. A elaboração do dicionário eletrônico no suporte Visual Basic.....	77
2.7.3. O gerenciador de banco de dados MySQL.....	80
3. O DICIONÁRIO ONOMASIOLÓGICO.....	82
<i>3.1. O sistema nocional em língua inglesa.....</i>	<i>82</i>
<i>3.2. O sistema nocional em língua portuguesa.....</i>	<i>83</i>
<i>3.3. O dicionário terminológico onomasiológico das RNAs.....</i>	<i>84</i>
4. ANÁLISE DOS DADOS.....	119
<i>4.1. Dificuldades relacionadas às buscas.....</i>	<i>129</i>
<i>4.2. A equivalência entre os termos.....</i>	<i>131</i>
5. CONSIDERAÇÕES FINAIS.....	134
REFERÊNCIAS BIBLIOGRÁFICAS.....	138

LISTA DE TABELAS

Tabela 1: Número de tokens, types e type/token ratio.....	64
Tabela 2: Número de tokens e types nos corpora de especialidade.....	64
Tabela 3: Variação do corpus de especialidade em relação aos corpora de referência....	66
Tabela 4: Primeiras 50 palavras-chaves no CR1 e CR2.....	68
Tabela 5: organização dos campos do modelo de dicionário onomasológico.....	71
Tabela 6: Termos fundamentais das RNAs.....	131

LISTA DE FIGURAS

Figura 1: Esquema de equivalência proposto por Felber.....	32
Figura 2: A lista de palavras do WST.....	45
Figura 3: Densidade lexical de 6 linguagens de especialidade.....	65
Figura 4: Ficha terminológica em inglês do MS Access XP.....	75
Figura 5: Tela do Winisis.....	77
Figura 6: Página desenvolvida em html puro.....	79
Figura 7: Página desenvolvida em PHP.....	80
Figura 8: Assimilação dos termos em língua portuguesa.....	119
Figura 9: Gênero gramatical dos termos.....	120
Figura 10: Composição dos termos.....	121
Figura 11: Ficha terminológica backpropagation model.....	125
Figura 12: Ficha terminológica modelo backpropagation.....	125
Figura 13: Ficha terminológica input layer.....	128
Figura 14: Ficha terminológica rede neural multicamada.....	130

SILVA, E. B. da. *Proposta de um dicionário eletrônico terminológico onomasiológico bilingüe inglês-português no domínio das Redes Neurais Artificiais*. São José do Rio Preto, 2009, 143 f. Dissertação (Mestrado em Estudos Lingüísticos – Área de concentração: Análise Lingüística). Instituto de Biociências, Letras e Ciências Exatas, Campus de São José do Rio Preto, Universidade Estadual Paulista

Resumo

O presente trabalho tem como objetivo principal aplicar o modelo de dicionário terminológico onomasiológico bilingüe, proposto por Babini (2001), na elaboração de um dicionário eletrônico inglês-português e português-inglês dos termos fundamentais das Redes Neurais Artificiais. O dicionário onomasiológico caracteriza-se por permitir a busca de uma unidade lexical ou terminológica a partir de seu conteúdo semântico. Constituímos um corpus de especialidade e um corpus de língua geral com o auxílio do programa computacional WordSmith Tools. A metodologia terminológica foi norteadada pela Teoria Comunicativa da Terminologia, que nos auxiliou nas reflexões quanto as diferenças terminológicas existentes entre os dois idiomas estudados. As principais referências metodológicas, que guiaram nossas buscas pelos equivalentes lexicais, foram Felber (1984), Dubuc (1985), Alpizar-Castillo (1995), Cabré (1993, 1999), Berber Sardinha (2001), Barros (2004), Babini (2001a, 2001b). Nosso modelo de dicionário permite dois tipos de pesquisa: semasiológica e onomasiológica. A busca onomasiológica é viabilizada pelo conjunto de semas ou traços semânticos que compõe o conceito de cada termo do dicionário. Foram utilizados o aplicativo MS Access, o gerenciador de banco de dados MySQL e a linguagem de programação PHP.

Palavras-chave: *Terminologia Bilingüe; Onomasiologia; Dicionário Eletrônico; Redes Neurais Artificiais.*

Abstract

The present thesis majorly aims at applying a bilingual onomasiological terminological dictionary model, proposed by Babini (2001), so as to develop an English-Portuguese and Portuguese-English electronic dictionary of the fundamental terms of the Artificial Neural Networks. The onomasiological dictionary allows the search of a lexical or terminological unit from its semantic content. We constituted a specialty corpus and a general language corpus with the aid of the computational program WordSmith Tools. The methodological approach was guided by the Communicative Terminology Theory, which allowed us to identify the existing differences among the languages studied. The main methodological basis, which guided our search for lexical equivalent, were Felber (1984), Dubuc (1985), Alpizar-Castillo (1995), Cabré (1993, 1999), Berber-Sardinha (2001), Barros (2004), Babini (2001a, 2001b). Our dictionary model allows two types of search: semasiological and onomasiological. The onomasiological search is made viable by a set of semantic traits that make up the concept of each term in the dictionary. The following computational resources were used: MS Access software, the MySQL database management system and the PHP programming language.

Keywords: *Bilingual Terminology; Onomasiology; Electronic Dictionary; Artificial Neural Networks*

INTRODUÇÃO

No mundo atual, os avanços constantes impulsionam a criação de novas tecnologias, especialmente na área da informática, as quais causam dificuldade de compreensão para indivíduos que não estão familiarizados com um domínio específico.

O conhecimento, seja ele qual for, é um meio de inclusão sócio-profissional indispensável no século XXI. Qualquer iniciativa que promova a difusão do conhecimento, ou, ainda, o acesso a esse bem, torna-se bem-vinda no meio acadêmico.

Vale salientar que o computador desempenha um papel de destaque nas novas tecnologias. Ele está presente em praticamente todas as áreas do saber, e as informações do mundo moderno transitam em algum momento por um computador.

O interesse em desenvolver esse projeto surgiu graças a um ciclo de palestras sobre o estudo do léxico, organizado no Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto. Cumpre ressaltar que as palestras foram ministradas por especialistas envolvidos em Lingüística de Corpus e Terminologia.

Esse trabalho é fruto da conjunção de áreas distintas do saber: a Terminologia, a Terminografia bilíngüe, a Lingüística de Corpus, a Onomasiologia e a Lingüística Computacional. Todas, sem exceção, podem recorrer ao auxílio do computador para fornecer e produzir resultados eficientes à comunidade.

O presente trabalho tem como objetivo principal desenvolver um protótipo de um dicionário eletrônico terminológico onomasiológico bilíngüe inglês-português, português-inglês no domínio de Redes Neurais Artificiais (RNAs). Pretendemos subsidiariamente verificar a aplicabilidade do modelo proposto por Babini (2001). Serão utilizados em nosso

protótipo trinta termos fundamentais de RNAs para que a pesquisa possa ser efetuada tanto na direção português-inglês quanto na direção inversa.

Com o desenvolvimento da informática, especialmente das RNAs, uma pesquisa que tenha como resultado a elaboração de um dicionário terminológico é mais uma fonte de pesquisa para o trabalho de tradutores, terminólogos ou interessados pelo domínio em questão.

As RNAs estão presentes em nosso cotidiano desde seu uso na identificação de placas de automóveis pelos radares fotográficos até no controle de estoque e tomadas de decisões automatizadas de grandes indústrias.

Nossa pesquisa propõe um modelo de dicionário, que, por sua vez, pode ser estendido a outras atividades da sociedade ou mesmo ao ensino de línguas estrangeiras. Em nosso caso, um dicionário onomasiológico proporciona uma maior flexibilidade na busca por um termo quando já se tem a idéia do mesmo. A agilidade conta como uma arma positiva na busca pela informação. Cumpre ressaltar que a busca semasiológica também será viabilizada no modelo eletrônico.

Para o desenvolvimento do protótipo de um dicionário eletrônico terminológico onomasiológico bilíngüe inglês-português, português-inglês no domínio de Redes Neurais Artificiais, doravante DETOB, tornou-se fundamental o embasamento teórico de distintas áreas do saber. Para desenvolver a fundamentação teórica do presente trabalho, no capítulo 1, recorreremos às contribuições da Terminologia, da Onomasiologia, da Linguística Computacional, dos Estudos da Tradução e das Redes Neurais Artificiais. Iniciamos a fundamentação teórica abordando alguns conceitos-chave da Terminologia. Optamos por seguir a Teoria Comunicativa da Terminologia (TCT) para auxiliar no tratamento dos termos, bem como na sua relação com os usuários de terminologia. A TCT permite a explicação e o entendimento funcional das diferenças, bem como das equivalências entre os termos. Nesse

sentido, o termo não é entendido como uma unidade isolada, mas, sim, reflete as características do meio no qual ocorre. As principais referências teóricas que guiarão nossas buscas pelos equivalentes lexicais, foram Felber (1984), Dubuc (1985), Alpizar-Castillo (1995), Cabré (1993, 1999), Berber Sardinha (2001, 2004), Barros (2004), Babini (2001a, 2001b).

Além da base teórica terminológica, recorreremos à Onomasiologia para demonstrar o princípio pelo qual o modelo de dicionário terminológico onomasiológico bilíngüe das Redes Neurais Artificiais seria regido. A Onomasiologia preocupa-se em examinar como uma idéia encontra expressão na palavra. Nosso modelo de dicionário possui, também, uma orientação onomasiológica, que pode revelar aspectos e relacionamentos lingüísticos despercebidos em uma consulta semasiológica. Para tornar essa modalidade de pesquisa possível, foi necessária a inserção de alguns traços semânticos pertencentes aos termos presentes no banco de dados. A partir de um ou mais traços semânticos, pode-se visualizar os termos que os compartilham.

Os preceitos da Lingüística de Corpus também foram de suma importância na construção do DETOB. A Lingüística de Corpus fornece os subsídios para o processamento e tratamento do corpus a ser explorado. Em primeiro lugar, o corpus em língua portuguesa foi construído a partir de obras fundamentais de RNAs, apresentações em formato PowerPoint, websites introdutórios sobre o assunto, artigos científicos, entre outros. Em língua inglesa, o corpus é constituído primordialmente de livros, artigos e textos introdutórios que abordam as RNAs. Houve a necessidade de digitalização de materiais com conteúdo relevante relacionado às RNAs. Posteriormente, todo o corpus foi convertido para formato de “texto simples”, para que pudesse ser explorado pela ferramenta de tratamento de texto WordSmith Tools.

O programa computacional WordSmith Tools foi escolhido por ser ágil na análise de milhões de palavras e fornecer resultados confiáveis úteis para a elaboração do DETOB. Um dicionário baseado em corpus apresenta maior autenticidade e representatividade de uso real

da língua. Os candidatos a termos serão elencados e selecionados pelo fator frequência de ocorrência. Em uma obra terminológica, essas características auxiliam o consulente a entender, ou ainda, contextualizar de forma adequada, o termo.

No capítulo 1.8 são apresentadas as RNAs, bem como suas aplicações. Essa subárea da Inteligência Artificial foi escolhida devido à sua utilização nas mais variadas atividades tecnológicas. As RNAs são sistemas inteligentes capazes de analisar determinado padrão e, após a aprendizagem, tomar decisões. Inúmeros setores de diferentes ramos da sociedade já contam com redes neurais artificiais para maximizar e obter resultados mais confiáveis, desde a área educacional até a área industrial. A tendência é que haja uma crescente aplicação das RNAs nas atividades humanas, principalmente pela sua capacidade de análise de milhões de dados em reduzido espaço de tempo.

No capítulo 2, apresentamos a metodologia da pesquisa. Em língua portuguesa, utilizamos dois corpora distintos: um corpus de especialidade das RNAs e outro corpus contendo obras de literatura brasileira. Em língua inglesa, o corpus foi constituído exclusivamente de material relacionado às RNAs. Obtivemos uma lista de palavras-chave realizando o cruzamento entre o corpus de especialidade das RNAs e um corpus de língua geral. A fim de justificar a escolha das palavras-chave que poderiam ser termos, efetuamos a comparação do corpus de especialidade das RNAs com outros corpus de especialidade. As palavras-chave que eram comuns em ambas as listas foram selecionadas porque poderiam ser encaradas como candidatas a termo.

Do capítulo 3 em diante, apresentamos o sistema nocional em línguas portuguesa e inglesa. Para armazenar todo o conteúdo terminológico do dicionário, será utilizado o programa de processamento de informação MS Access, pertencente ao pacote do Microsoft Office. Todo o banco de dados terminológico foi armazenado nesse programa. Devido à facilidade de visualização, as fichas terminológicas presentes nesse capítulo foram formatadas

para impressão pelo MS Access. No intuito de desenvolver uma interface de pesquisa mais amigável para o DETOB, houve a necessidade de utilização de outras duas ferramentas computacionais: o PHP e o MySQL. O primeiro é uma linguagem de programação voltada para a Internet ao passo que o MySQL é um gerenciador de banco de dados.

1. FUNDAMENTAÇÃO TEÓRICA

Haja vista o caráter interdisciplinar desse trabalho, faz-se necessária uma apresentação dos conceitos teóricos que nortearão nossas reflexões. Nesse capítulo, serão apresentados conceitos de áreas fundamentais para servir de base para o presente trabalho: a Terminologia, a Onomasiologia, a Lingüística Computacional, Estudos da Tradução e as Redes Neurais Artificiais.

1.1. A Terminologia

Ao se realizar uma apresentação da terminologia, é imperativo destacar suas diferentes acepções. A Terminologia entendida como uma disciplina propriamente dita é descrita como “estudo que tem por objeto a identificação e delimitação de conceitos próprios de uma arte, ciência, profissão etc., e a designação de cada um deles por um determinado vocábulo.” (HOUAISS, 2001). Quando utilizada para se referir às unidades terminológicas das línguas de especialidade,

[...] a palavra terminologia significa um ‘conjunto de palavras técnicas pertencentes a uma ciência, uma arte, um autor ou um grupo social’, como, por exemplo, a terminologia da medicina ou a terminologia usada pelos especialistas em computação.” (PAVEL e NOLET, 2002, p. xvii)

A Terminologia vem recebendo crescente atenção das ciências da linguagem por contribuir para o desenvolvimento, disseminação de informação e estabelecimento das áreas conhecidas como técnico-científicas. Barros (2006, p. 23) salienta que “com o acelerado desenvolvimento das ciências e das técnicas, as terminologias das áreas de especialidade ocupam cada vez mais terreno na comunicação, passam a integrar o universo léxico das línguas e, por conseguinte, ‘reivindicam’ maior espaço nos dicionários de língua geral.”

A constante criação de novas tecnologias e técnicas torna a Terminologia a partir da segunda metade do século XX, primordial para o desempenho satisfatório de atividades, como

por exemplo a tradução técnica, a elaboração de dicionários, glossários, entre outros. Vale salientar que a produção de material terminológico não é exclusividade da sociedade modernas. “No mundo oriental, desde a Idade Média trabalhos com designações de doenças, de órgãos do corpo humano e de plantas eram produzidos.” (ALVES, 2000, p. 262). No século XVIII, cientistas começam a mostrar a necessidade de denominação de conceitos nas várias áreas de especialidade. Índícios dessa necessidade são revelados nos trabalhos de Lavoisier na química, de Lineu na botânica e na zoologia. (idem). Na opinião de Barros (2004, p. 28), “as características da sociedade atual explicam o grande desenvolvimento da Terminologia no século XX, como disciplina científica que estuda as línguas (ou linguagens) de especialidade e o conjunto vocabular de campos específicos.”

A relevância dos estudos terminológicos mostra-se inegável, uma vez que não seria possível descrever aquilo que está à nossa volta sem lançar mão tanto das unidades lexicais da língua geral quanto das unidades das terminologias especializadas. A terminologia, não importa a área do conhecimento a que está relacionada, apresenta-se como uma parte funcional da língua geral. Seria errôneo afirmar que cada área de especialidade constitui uma linguagem independente. As unidades lexicais da língua geral e das línguas de especialidade estão interligadas. “Portanto, expressar o mundo sem Terminologia não é expressar o mundo completo.” (ALMEIDA, 2003, p. 221).

Qualquer tentativa de abordar a terminologia sem levar em consideração as línguas de especialidade esbarra na ausência de direcionamento. Para esse trabalho interessa a terminologia das RNAS. “A terminologia computacional adquiriu grande importância em todas as línguas devido à recente expansão dos computadores pessoais e da internet.”¹ (MEDINA, 2003, p. 315)

¹ *Computer terminology has acquired great importance in all languages due to the recent expansion of personal computers and the Internet.*

Para a definição da língua de especialidade, pode-se dizer que é a que é “utilizada para proporcionar uma comunicação sem ambigüidade numa área determinada do conhecimento ou da prática, com base num vocabulário e em usos lingüísticos específicos desse campo.” (PAVEL; NOLET, 2002, p. xvii)

A presença das unidades terminológicas nos dicionários de língua geral nem sempre é exaustiva. Pode-se perceber que sua inserção é importante dentro de um repertório que objetive certa extensividade ou completude.

Quanto mais percebemos a existência de línguas de especialidade nas rubricas dos dicionários, por exemplo (zool.), (naut.), (med.), (bibl.), que podem, naturalmente, apenas confirmar a crença popular de que não existem línguas especiais, mas apenas itens lexicais especiais. (SAGER; DUNGEWORTH; McDONALD, 1980, p. 2)²

Em língua portuguesa, o Novo Dicionário Aurélio século XXI em CD-ROM (2004) divide o conhecimento humano em aproximadamente 60 áreas do saber. Nas rubricas, o consulente dos mais diversos campos profissionais pode ter acesso ao significado de um termo. Nas notas introdutórias, lê-se que as definições que constam sob uma rubrica constituem um “subdicionário especializado na área delimitada pela rubrica”. A versão eletrônica do dicionário poderia viabilizar a busca de todos os termos do subdicionário, por exemplo, *Comp. Gráf.* (Computação Gráfica), mas não é possível acessar tal informação. Por sua vez, o dicionário eletrônico Houaiss traz por volta de 20 rubricas. A rubrica *Informática* abarca todas as possíveis subáreas.

Em língua inglesa, o *Cambridge International Dictionary of English* (1995) conta com o auxílio de uma equipe de consultores para os 70 domínios elencados no início da obra. No dicionário eletrônico *Collins COBUILD* (2003), as áreas de especialidades são teoricamente agrupadas como sendo *technical*. O dicionário explica que essa rubrica é utilizada

² *The closest we come to an acknowledgement of the existence of special languages is in the scope notes of dictionaries, e.g: (zool.), (naut.), (med.), (bibl.), which can, of course, only confirm the popular belief that there are no special languages but only special lexical items.*

principalmente quando se fala ou se escreve a respeito de objetos, eventos, ou processos de uma área de especialidade, como negócios, ciências, ou música. Seguindo esse pensamento, o item lexical “biotecnologia” poderia ser enquadrado como *technical*. Curiosamente também existem outras rubricas técnicas independentes como *business, computing, journalism, legal, literary, medical e military*.

Baseada em estudos anteriores, Biderman sugere que o número de áreas técnico-científicas possa ser bem maior. “Relativamente ao número de áreas técnico-científicas, [...] talvez totalize uns 250, 300 ou mais domínios. [...] A evolução da ciência [...] ampliou o número de domínios científicos, criando muitas especialidades novas.” (BIDERMAN, 2006, p. 37). Até o momento, o que se observa na práxis dicionarística é a tendência de condensação desse amplo espectro de áreas. De certa forma, a terminologia perde sua especificidade.

Mesmo assim, partindo-se do princípio de que “toda terminologia possui uma finalidade sócio-profissional e serve prioritariamente para exprimir saberes temáticos.” (BOULANGER, 1995, não-paginado), a presença de terminologia nos dicionários é inquestionável do ponto de vista cognitivo, social e profissional.

Nesse sentido, as línguas de especialidade constituem suas próprias unidades terminológicas, em função da ausência de unidades que sirvam para representar um conceito. Qualquer área do conhecimento especializado recorrerá ao chamado jargão. Com referência ao jargão e às línguas de especialidade, a terminologia da área médica ou jurídica parece ter recebido mais atenção no dicionário *Collins COBUILD*. Talvez pelo reconhecimento da importância desse domínio ou pelo prestígio que tais profissionais gozam na sociedade modernas.

Línguas de especialidade, ou mais precisamente línguas de domínios de especialidade, são geralmente vistas como meio de comunicação de especialistas de domínios altamente qualificados como engenheiros, médicos, advogados, etc. e são muitas vezes chamadas negativamente

de “jargão” [...]. (SAGER; DUNGEWORTH; McDONALD, 1980, p. 3)³

A utilização do chamado jargão proporciona uma amostragem do domínio em questão. Entre os profissionais de determinada área, é corriqueira a utilização de itens lexicais específicos daquela área de atuação. Com a preocupação em transmitir a mensagem com precisão, os especialistas recorrem aos termos. No domínio das RNAS, os especialistas entendem perfeitamente o que *perceptron learning rule* quer dizer e em quais situações o termo é usado. Dessa forma, quanto mais um indivíduo adentra pela língua de especialidade, maiores são as possibilidades de entrar em contato com um nível mais alto de complexidade.

Rey-Debove esclarece que,

A utilização do jargão seleciona, portanto, os leitores que têm o mesmo universo e a mesma terminologia, de tal forma que a comunicação seja a mais espontânea possível [...].⁴ (REY-DEBOVE, 1998, p. 146)

1.1.1. A lexicologia e a terminologia

A lexicologia é a ciência lingüística que descreve e estuda o léxico geral da língua e sua vertente aplicada é a lexicografia. Cabe a essa disciplina a elaboração de obras lexicográficas variadas, tais como dicionários de língua geral, dicionários de expressões idiomáticas, dicionários de sinônimos e antônimos, e outros.

A terminologia possui como objeto de estudo o léxico especializado, ou seja, os termos. Dentre as várias ramificações que uma pesquisa terminológica pode tomar, a orientação do estudo terminológico pode voltar-se ainda para a descrição da fraseologia

³ *Special languages, or more precisely special subject languages, are usually thought of as the means of expression of highly qualified subject specialists like engineers, physicians, lawyers, etc. and are often derogatively referred to as 'jargon' [...].*

especializada, das variantes sócio-terminológicas, dos neologismos terminológicos, etc. Krieger e Finatto (2004) ressaltam a diferença entre a lexicologia e a terminologia.

Lexicologia e Terminologia, embora aproximem-se, porquanto ambas constituem ciências do léxico, distinguem-se pela especificidade de seus objetos. A diferença entre estes, cabe ressaltar, não é outra senão a propriedade que possuem as unidades lexicais chamadas de termos de estruturas lingüísticas que, em sua dualidade sígnica, denominam e circunscrevem cognitivamente objetos, processos e conceituações pertinentes ao universo das ciências, das técnicas e das tecnologias; enquanto as palavras, realizando o mesmo processo denominativo e conceitual, cobrem toda a abrangência da realidade cognitiva e referencial apreendida e construída pelo homem. (KRIEGER ; FINATTO, 2004, p. 43)

À guisa de exemplificação, o item lexical “camada” encontra algumas definições no Novo dicionário Aurélio século XXI (2004). Dessas, existe uma definição para cada especialidade a seguir: a botânica, a geologia, a eletrônica, a arqueologia e a biologia. Para a física e a astronomia são duas. A geofísica tem três definições para *camada*. O domínio das RNAs não encontra definição, já que o termo *layer* não foi incluído nesse verbete. Em uma hipotética consulta, o consulente que procura descobrir o significado de *layer* nas RNAs, nesse caso, não teria suas dúvidas sanadas.

Esse é um exemplo de um item lexical que migrou da língua geral para uma língua de especialidade, das RNAs. O contrário também pode acontecer. Um item lexical pode também ser usado apenas dentro do domínio das RNAs, como é o caso de *backpropagation/retropropagação*.

Aubert resalta as semelhanças e particularidades da lexicologia e da terminologia da seguinte forma:

A terminologia, enquanto campo de investigação, entretém relações estreitas com a lexicologia, a lexicografia e a semântica, embora não se confunda com estas nem constitua simplesmente uma sub-área das mesmas, e seu estatuto de área de aplicação da lingüística e/ou da sociolingüística parece inquestionável. (AUBERT, 1996b, p. 13)

⁴ *L'usage du jargon sélectionne donc les lecteurs qui ont le même univers du discours et la même terminologie, de telle façon que la communication soit aisée [...]*

Barros (2006) alerta que a lexicologia e a terminologia lidam com o mesmo objeto de estudo, entretanto mantêm suas particularidades no tocante à metodologia e objetivos.

Terminologia, lexicologia e lexicografia têm como objeto de estudo a ‘palavra’. Embora trabalhem com a mesma “matéria-prima”, cada uma a recorta diferentemente, possui modelos teóricos e métodos de análise específicos, além de uma metalinguagem particular, o que garante a cada uma dessas ciências ou disciplinas uma identidade científica própria. (BARROS, 2006, p. 23)

1.1.2. O trabalho terminológico

A especificidade da nomenclatura das línguas de especialidade requer precisão no tocante à comunicação. Devido à necessidade em transmitir conhecimento sem duplo sentido ou má interpretação, as áreas técnico-científicas impulsionam a criação de novas palavras, nesse caso, novos termos, a todo momento. Geralmente, tal criação dá-se pelo fato de não haver um item lexical disponível para significar um conceito.

A intervenção da terminologia como conhecimento e como método específico é principalmente necessária na tecnologia, no direito e em todos os domínios onde existe circulação em larga escala de conhecimento. Nessas áreas, a função linguística, com ênfase na dimensão social, e a função sócio-cultural da terminologia é predominante. (REY, 1995, p. 58)⁵

Uma comunicação eficiente dentro de um domínio especializado aciona uma terminologia específica. “O conhecimento dos conceitos específicos e da terminologia utilizada em uma especialidade determinada é um precioso trunfo profissional.” (PAVEL; NOLET, 2002, p. xviii). Profissionais que compartilham os mesmos conhecimentos possivelmente utilizam sua terminologia de forma a maximizar sua interação. Rey-Debove

⁵ *The intervention of terminology as a body of knowledge and as a specific method is mainly needed in technology, in law and in all domains where there is large-scale circulation of knowledge. In these areas, the linguistic function, with emphasis on the social dimension, and the socio-cultural functions of terminology predominate.*

³ *Il est impossible de travailler dans une spécialité sans employer sa terminologie, et cet emploi est le signal d'un univers de discours.*

(1998, p. 146) salienta que “é impossível trabalhar em uma especialidade sem empregar sua terminologia, e esse emprego é sinal de um universo de discurso”⁶.

Uma vez que, na maioria das vezes, os textos técnico-científicos são redigidos por e para especialistas, costumam caracterizar-se por uma alta densidade terminológica. É uma maneira de interação profissional, bem como, de exclusão. Os indivíduos não-detentores dessa terminologia serão incapazes de compreender a mensagem. Isso deve-se ao fato de que os termos possuem a propriedade de transmitir e caracterizar conhecimento específico de mundo.

Cabré ressalta que,

Uma das características mais marcantes de um texto especializado é a presença de unidades terminológicas. Quanto maior o nível de especialização de um texto, maior sua densidade terminológica. Em geral, um texto altamente especializado mostra precisão, concisão e sistematicidade.⁷ (CABRÉ, 2000, p. 15)

Uma pesquisa terminológica, para alcançar êxito, necessitará identificar o jargão convencionalmente utilizado em uma ou mais áreas específicas. Cabré (2000, p. 14) destaca o procedimento a ser adotado em tal pesquisa afirmando que “um lingüista dedicado à descrição das unidades terminológicas deve buscá-las nas produções orais e escritas de especialistas. Chamamos texto especializado a totalidade das produções discursivas de caráter especializado⁸.” A busca por produções escritas é relativamente mais fácil. O pesquisador tem ao seu dispor programas computacionais de conversão de texto para digitalizar textos integralmente, a própria rede mundial de computadores, documentos prontos nos processadores de textos, etc. A busca por produções orais de especialistas é mais difícil.

⁷ *Une des caractéristiques les plus remarquables d'un texte spécialisé est la présence des unités terminologiques. Plus le niveau de spécialisation d'un texte est élevé, plus sa densité terminologique est grande. D'habitude, un texte hautement spécialisé montre précision, concision et systématique.*

⁸ *Un linguiste concerné par la description des unités terminologiques doit les chercher dans les productions orales et écrites des spécialistes. On appelle texte spécialisé la totalité de productions discursives de caractere spécialisé.*

Atualmente poucos são os materiais orais disponíveis para a pesquisa terminológica. Com efeito, o pesquisador pode realizar gravações com especialistas e transcrever posteriormente o conteúdo. A vantagem da combinação oral/escrita é que ela lança luz na variedade sociolingüística, ou seja, deixa claro que mesmo os especialistas de uma mesma área usam termos distintos para fazer referência ao mesmo conceito.

Na opinião de Corbeil (1999), os termos, no tratamento do corpus, devem ser elencados também pela sua variação sociolingüística.

Uma tendência recente em terminologia, designada sob o termo de socioterminologia, propõe-se a levar em conta diversos grupos de termos que gravitam ao redor de um termo padrão e de situá-los e relacioná-los de acordo com as circunstâncias de comunicação e grupos de usuários. Esta maneira de fazer teria a vantagem de fazer coexistir harmoniosamente diferentes termos para uma mesma noção evitando o ostracismo da exclusão.⁹ (CORBEIL, 1999, p. 77)

1.1.3. A Teoria Comunicativa da Terminologia

O presente trabalho de cunho terminológico segue a orientação metodológica da Teoria Comunicativa da Terminologia (doravante, TCT).

O antigo entendimento de que a terminologia seria capaz de eliminar as ambigüidades nos discursos técnico-científicos provou estar obliterado. Existe, hoje, uma visão que abarca contribuições de outras áreas do conhecimento. O postulado da TCT segue direção oposta aos princípios de Wüster, o criador da Teoria Geral da Terminologia. A TCT propõe um modelo teórico mais aberto. As unidades terminológicas são entendidas como unidades multidimensionais. Cabré (1999, p. 130) aponta que os termos são unidades lingüísticas, cognitivas e comunicativas. Dessa forma, as unidades terminológicas possuem o mesmo status das unidades léxicas da língua geral. Em outras palavras, os termos não são unidades

⁹ *Une tendance récente en terminologie, désignée sous le terme de socioterminologie, se propose de prendre en compte les diverses couches de termes qui gravitent autour d'un terme standard et de les situer par rapport à lui selon les circonstances de communication et les groupes d'utilisateurs. Cette manière de faire aurait l'avantage*

independentes que constituem um sistema próprio, mas, sim, unidades que se incorporam ao léxico de um falante enquanto adquire o papel de especialista pelo aprendizado de conhecimentos especializados. Seguindo essa orientação, o termo é entendido como tal a partir do momento em que é utilizado em contexto específico.

Cabré (1999) ressalta que, no trabalho terminológico, deve haver, por parte do pesquisador, o domínio de quatro habilidades. A primeira delas é o domínio do assunto estudado. Em segundo lugar, o conhecimento da língua que está sendo pesquisada. Na seqüência, as variações de registro devem ser levadas em consideração e, por último, a competência metodológica.

Os preceitos dessa teoria fornecem subsídios para a elaboração de dicionários terminológicos.

1.2. A Onomasiologia

A exploração onomasiológica do significado do termo proporciona perspectivas, até o momento, pouco ou nem um pouco exploradas em sua potencialidade pelos estudiosos do léxico. Essa negligência afasta os estudos onomasiológicos do grande público. Para tradutores, terminólogos, estudantes e profissionais de áreas especializadas, a onomasiologia pode oferecer precioso instrumento de pesquisa, estudo ou trabalho.

Babini (2006) apresenta a onomasiologia da seguinte forma:

O termo *onomasiologia* foi utilizado pela primeira vez em 1903 por A. Zauner, em um estudo sobre os nomes das partes do corpo humano nas línguas românicas [...]. A definição de *onomasiologia* proposta por Vittorio Bertoldi na *Enciclopédia Italiana di Scienze, Lettere ed Arti*, de 1935 é considerada como uma referência na matéria: Por onomasiologia entende-se um aspecto particular da pesquisa lingüística que, partindo de uma determinada idéia, examina as várias maneiras com as quais essa idéia encontrou expressão na palavra.

Assim, segundo Bertoldi, a onomasiologia (em lingüística) trata dos aspectos ligados ao processo da denominação (que vai da idéia ao signo). (BABINI, 2006, p. 38)

Dessa forma, o aspecto de extrema importância a ser salientado é que, ao se partir do conceito propriamente dito, procura-se a identificação do termo correspondente. Dentro da terminologia, em primeiro lugar aparece o conceito para depois se chegar ao termo. A abordagem onomasiológica, na prática terminológica, por exemplo, procura relacionar um conceito a um ou mais termos. É uma maneira complementar, e também eficiente, de se explorar o signo lingüístico do estruturalismo clássico. Via de regra, convencionou-se a prática da semasiologia, que parte do significante para o significado. “A Onomasiologia representa também uma das abordagens do signo saussuriano: sendo êle [sic] o resultado da associação da imagem acústica ao conceito, é a onomasiologia o estudo que busca atingir o signo partindo do conceito.” (CASTILHO; CARRATORE, 1967, p. 131).

Mesmo não pertencendo a modelos de análise recente, os estudos onomasiológicos carecem de maior atenção. Há mais de 40 anos, prevendo as diferentes possibilidades de pesquisa e mudança de atitude quanto ao entendimento da significação do mundo, e tendo por base a orientação onomasiológica, Baldinger afirma que,

É sobretudo a onomasiologia [...] que promete realmente resultados novos. Ela nos faz ver a estrutura lexical de uma só e mesma língua e possibilita a comparação entre diferentes línguas numa base estrutural. Pode-se mesmo afirmar que ela está revolucionando a lingüística comparada. (BALDINGER, 1966, p. 34)

No entanto, mesmo após tantas revoluções lingüísticas, a onomasiologia ainda carece de produções léxico-terminográficas. A tradição semasiológica, que busca um significado a partir de um significante, vem se perpetuando há séculos. Desse modo, o tradicionalismo arraigado nas sociedades perpetua a orientação semasiológica. Apesar de sua relevância, a orientação onomasiológica acaba sendo preterida. “Entre os modelos propostos para a

análise da estruturação do significado, um dos mais profícuos é a teoria dos campos semasiológico e onomasiológico.” (BIDERMAN, 2001, p. 199). A ausência de uma interação entre os dois campos impulsionou nosso trabalho no sentido de contribuir para o avanço das pesquisas de caráter onomasiológico.

1.2.1. Os dicionários onomasiológicos

Uma breve comparação entre o tradicional dicionário alfabético e o dicionário do tipo onomasiológico ilustra suas especificidades:

Se quero saber sobre as significações, o dicionário alfabético é prático, e o dicionário conceptual não, porque nele as acepções estão separadas umas das outras. Mas se quero saber sobre as designações, o dicionário conceptual é prático e o alfabético não, porque nele as designações para a mesma coisa, os ‘sinônimos’, estão separados uns dos outros. De fato, na macroestrutura do dicionário alfabético há uma infinidade de microestruturas semasiológicas (campos das significações), na macroestrutura do dicionário dividido conceptualmente estão incluídas uma infinidade de microestruturas onomasiológicas (campos das designações). (BALDINGER, 1970, p. 135-136)¹⁰

Para Bergenholtz e Tarp (1995, p. 198), “é inegável que o alfabeto seja fácil de lidar tanto para o lexicógrafo quanto para o usuário.”¹¹ Os consulentes já estão acostumados a recorrer a dicionários alfabéticos e possuem certa familiaridade no manuseio de tal obra. Sempre foi assim. O público não participa da escolha do dicionário, ou melhor, nem ao menos é informado que existem outras alternativas ao modelo tradicional. Uma observação atenta constatará que “os dicionários alfabéticos conquistaram o mercado. Esse fato não vai mudar

¹⁰ *Si pregunto por las significaciones, el diccionario alfabético es práctico, y el diccionario conceptual no, porque en éste las acepciones se separan unas de otras. Pero si pregunto por las designaciones, el diccionario conceptual es práctico y el alfabético no, porque aquí las designaciones para la misma cosa, los "sinónimos", están separados unos de otros. De hecho, en la macroestructura del diccionario alfabético hay una multitud de microestructuras semasiológicas (campos de significaciones), en la macroestructura del diccionario dividido conceptualmente se contiene una multitud de microestructuras onomasiológicas (campos de designaciones).*

¹¹ *It cannot be denied that the alphabet is easy to handle both for the lexicographer and for the user.*

em um futuro próximo. Uma consequência disso é que eles conquistaram também a atenção dos especialistas.”¹² (HÜLL, 1999, p. 21). Na outra ponta, mas igualmente importantes, os consulentes dificilmente se perguntam se existe alguma outra variedade de dicionário ou, ainda, dificilmente se questionam a respeito das possibilidades “alternativas” de consulta. No tocante a essa variante,

Existe, no entanto, uma tradição alternativa, na qual os compiladores usam grupos de tópicos em vez do alfabeto na base da organização. O alfabeto, com todas as suas virtudes, coloca animais e zoológicos, tios e tias distantes uns dos outros no esquema de coisas, enquanto que na mente humana tais palavras andam próximas. O dicionário alfabético tem uma lógica, mas não é a lógica do dia-a-dia. Em princípio, sente-se que as palavras devem ser definidas a partir da companhia que elas têm.¹³ (McARTHUR, 1981, p. vi)

As críticas aos dicionários alfabéticos abundam na literatura especializada porque na maioria das vezes ignoram a classificação sistemática. “[...] a menos que a lista de palavras seja preparada com base em uma classificação sistemática do campo do domínio em questão, a abordagem alfabética não garante uma cobertura completa do campo do domínio em questão.”¹⁴ (BERGENHOLTZ; TARP, 1995, p. 199). Teoricamente,

[...] as listas de palavras sistemáticas têm vantagens interessantes para o usuário. A macroestrutura fornece-lhe uma visão do domínio porque os termos relacionados estão em seqüência. A relação entre os termos logo torna-se clara. A macroestrutura, assim, auxilia o tradutor não-familiarizado na aquisição de conhecimento enciclopédico, que deve ser considerado um pré-requisito na tradução de textos de línguas de especialidade. Na lista de palavras sistemáticas, o usuário encontrará o termo que consultou no índice alfabético em seu contexto enciclopédico-sistemático. Ele pode, em uma consulta, conseguir até mesmo uma solução para mais de um problema de tradução, uma vez

¹² *Alphabetical dictionaries have conquered the market. This fact will not change in the near future. One consequence of this is that they have also conquered the attention of scholars.*

¹³ *There has, however, been an alternative tradition, in which compilers have used groups of topics instead of the alphabet as their basis for organization. The alphabet, with all its virtues, places animals and zoos, uncles and aunts far apart in its scheme of things, whereas in the human mind such words go close together. The alphabetical dictionary has a logic, but it is not the logic of everyday life. In principle, one feels, words should be defined in the company they usually keep.*

¹⁴ *[...] unless the word list is prepared on the basis of a systematic classification of the subject field in question, the alphabetic approach does not ensure full coverage of the subject field in question.*

que, naquele contexto, vários termos podem ocorrer, sendo relevantes para o seu próprio texto.¹⁵ (op.cit)

Conforme observado anteriormente,

Os dicionários tradicionais, organizados em ordem alfabética, permitem somente encontrar os significados das palavras (as idéias que elas representam) partindo de suas formas. O percurso seguido pelo leitor é, portanto, semasiológico ou do interpretante, percurso que é característico da maioria dos repertórios. O contrário, ou seja, o percurso onomasiológico ou do enunciador, não é muito freqüente nas obras lexicográficas. Como achar, então, uma palavra quando conhecemos somente sua “idéia”, seu “conceito”? Para dar solução a esse problema foram criados dicionários como os ideológicos e os analógicos. (BABINI, 2003, p. 66)

Tanto os dicionários ideológicos quanto os analógicos podem ser incluídos na categoria de dicionários onomasiológicos. “Os dicionários ideológicos e os analógicos foram concebidos para resolverem as dificuldades que o usuário enfrenta ao consultar um dicionário tradicional.” (BABINI, 2006, p. 40).

Na área técnica, a dificuldade advém principalmente no momento em que o indivíduo é detentor de um conceito, porém é incapaz de transformá-lo em termo. O mercado não dispõe de material de consulta que contemple tal empreendimento. Como resultado, a dúvida provavelmente será sanada após cansativa e minuciosa pesquisa.

No dicionário onomasiológico, algumas dificuldades são vencidas dentro de poucos segundos. “Como muitas palavras relacionadas semanticamente são apresentadas lado a lado, elas podem aumentar a capacidade de expressão e também o senso de lógica de seus

¹⁵ [...] systematic word lists have a number of interesting advantages to the user. The macrostructure provides him with an overview of the subject area, as related terms follow each other, and the relationship between the terms soon becomes clear. The macrostructure thus helps the layman translator acquire some of the encyclopedic knowledge that must be considered a prerequisite to the translation of LSP texts. In the systematic word list the user will find the term he has looked up in the alphabetic index in its systematic-encyclopedic context. He may even, in one lookup, get a solution to more than one translation problem, since in the surrounding context a number of terms may occur which are relevant to his own text.

usuários.”¹⁶ (HÜLL, 1999, p. 15). No ensino de terminologia, a orientação onomasiológica alcança êxito porque guia o consulente para o termo ou para os termos desejados. O capítulo 4 apresenta o funcionamento desse tipo de busca.

Sabe-se que “[...] a pesquisa terminológica temática propõe-se empreender o levantamento do vocabulário pertinente a uma determinada atividade ou técnica [...]” (AUBERT, 1996b, p. 57). Comercialmente, a publicação de dicionários específicos e onomasiológicos, visando atender um público restrito, acaba se mostrando, por vezes, inviável.

Por outro lado, infelizmente, a confecção de dicionários gerais da língua é planejada e executada por editoras comerciais que não estão dispostas a investir milhões em pesquisa para investigar a complexidade do léxico geral de uma língua e produzir um dicionário segundo critérios que poderíamos chamar de científicos. (BIDERMAN, 2006, p. 37)

No caso de dicionários terminológicos, a situação se complica ainda mais. O acesso à informação e ao termo é prejudicado pela falta de investimento nesse tipo de trabalho. “[...] em geral, dicionários são produtos comerciais. Em decorrência disso, acabam por seguir um padrão tradicionalmente aceito por consumidores [...]. (DIAS-DA-SILVA ; MORAES, 2003, p. 107). Quanto mais consumidores, mais investimentos são alocados para determinado modelo de estruturação e organização dos termos. Assim, a orientação semasiológica ganha destaque e o esforço da padronização contribui para gerar o apagamento, e o conseqüente ocultamento de outras apresentações dicionarísticas. No caso da apresentação de um dicionário técnico,

Os lexemas que denotam as partes de um navio podem ser dispostos na seqüência pela qual essas partes são construídas no estaleiro. Os lexemas que denotam pratos podem ser dados na seqüência de uma refeição farta. Os lexemas que denotam as partes do corpo humano são quase sempre dispostos em uma ordem que é sentida obviamente

¹⁶ *As many semantically related words are presented side by side, they can augment the power of expression and also the sense of logic of their users.*

como natural, que é da cabeça ao pé. Cada dicionário possui tais normas pragmáticas próprias.¹⁷ (HÜLL, 1999, p. 24)

À guisa de exemplificação, um outro dicionário que tratasse de itens relativos a veículos automotores, que disponibilizasse uma busca a partir da descrição de alguma parte ou característica do carro, seria de grande valia.

Caso se deseje conhecer a denominação do dispositivo fixado no chassi de um veículo de passeio e que serve de ponto de amarração/fixação de um reboque, pode-se definir os domínios pertinentes como compreendendo: veículos automotores, carros de passeio e reboques. (AUBERT, 1996b, p. 50)

O tratamento onomasiológico do termo pode enriquecer os estudos lingüísticos, uma vez que fornece subsídios para outras formas de entendimento da significação e do trabalho com a língua.

Babini destaca que

O problema que um dicionário onomasiológico deve resolver é exatamente o inverso daquele de um dicionário semasiológico: dada uma idéia (noção ou conceito), deve-se encontrar a unidade lexical ou o termo que a exprima. Em um dicionário semasiológico, o ponto de partida é o significante de um termo ou palavra; em um dicionário onomasiológico o ponto de partida é o significado. Assim, neste último tipo de obra lexicográfica deve-se encontrar um termo ou palavra desconhecida partindo do significado. (BABINI, 2006, p. 39)

Um dicionário que tinha uma “preocupação onomasiológica” foi o *Roget's Thesaurus*, publicado no ano de 1852. Seu sucesso deveu-se principalmente à sua organização, a saber: 1) estrutura da classificação hierárquica, 2) listagem das categorias (os verbetes), 3) índice com a listagem alfabética das palavras, e 4) números indicadores da categoria do verbete. Nem todos os dicionários onomasiológicos, porém, produzidos do século XIX em diante, seguem, necessariamente, as mesmas características do *Roget's Thesaurus*.

¹⁷ *Lexemes denoting the parts of a ship may be arranged in the sequence in which these parts are built in the shipyard. Lexemes denoting dishes may be given in the sequence of a huge meal. Lexemes denoting the parts of the human body are almost always arranged in an order which was obviously felt to be the natural one, that is, from head to foot. Each dictionary has such pragmatic norms of its own.*

1.2.2. A questão da equivalência

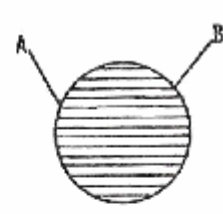
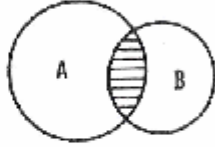
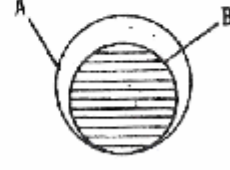

Na construção de um dicionário onomasiológico terminológico bilíngüe, faz-se necessário um criterioso trabalho de pesquisa terminológica, que ateste a equivalência ou não entre os termos envolvidos. Nesse sentido, dentro das RNAs, as diferentes maneiras de se enxergar a realidade podem influenciar o sentido dos termos: a língua portuguesa poderia contar com mais termos para representar determinado item; a língua inglesa poderia dispor de termos que, na língua portuguesa, não encontram termo correspondente; qualquer das línguas poderia contar com duas ou mais variantes que na outra língua não encontram a mesma diversidade. Cabré (1993) ressalta que a estruturação da realidade de uma língua de especialidade pode não coincidir em outra língua, especialmente em campos pouco estruturados.

Alpizar-Castillo (1995, p. 101) afirma que

a equivalência entre termos de idiomas diferentes situa-se em um diapasão de probabilidades que vai desde o total recobrimento do conteúdo do termo da língua A por um da língua B, até a total falta de equivalência, passando por uma variada gama de recobrimentos parciais.¹⁸

O dicionário onomasiológico terminológico bilíngüe contará com a apresentação dos traços semânticos dos termos, tanto em língua portuguesa quanto em língua inglesa. Essa peculiaridade permite a comparação entre os termos, bem como os contextos nos quais ocorrem. De modo a representar as possíveis relações de significado que pode haver entre os termos envolvidos em um trabalho de natureza técnico-científica, apresentamos o esquema originalmente desenvolvido por Felber (1987, p. 129):

Figura 1: Esquema de equivalência proposto por Felber.

Notions A et B	Comparaison	Compréhension
<p>1 Equivalence exacte des notions</p> <p>A = a₁, a₂, a₃ ... B = b₁, b₂, b₃ ...</p>	<p>A = B</p> <p>a₁=b₁ / a₂=b₂ a₃=b₃</p>	
<p>2 Intersection</p> <p>A = a₁, a₂, a₃, a₄ ... B = b₁, b₂, b₃, b₄ ...</p>	<p>A ∩ B</p> <p>a₁=b₁ b₂=a₂ a₃=b₃ a₄≠b₄ a₅≠b₅ etc.</p>	
<p>3 Supériorité</p> <p>A = a₁, a₂, a₃ ... B = b₁, b₂, b₃, b₄</p> <p>La compréhension de A est plus grande que celle de B. A possède par conséquent moins de caractères</p>	<p>A ⊃ B</p> <p>a₁=b₁ a₂=b₂ a₃=b₃ b₄ a₄= manque</p>	
<p>4 Non-équivalence des notions</p> <p>A = a₁, a₂, a₃ ... B = b₁, b₂, b₃ ...</p>	<p>A ≠ B</p> <p>a₁≠b₁ a₂≠b₂ a₃≠b₃ etc.</p>	

Pode-se inferir que, no caso 1, existe equivalência exata das noções e os termos partilham os mesmos traços semânticos nas duas línguas estudadas. Constata-se a equivalência total entre o termo A e o termo B. Na elaboração de um dicionário onomasiológico bilíngüe das RNAs, o fato de dois termos serem equivalentes diminui a possibilidade de problemas ou mal-entendidos na comunicação especializada. No caso 2, os termos possuem apenas alguns traços em comum. No caso 3, o termo A engloba mais características que o termo B, sendo superior a esse. No caso 4, não há equivalência entre os termos envolvidos pelo fato de seus traços semânticos serem distintos.

¹⁸ [...] la correspondencia entre términos de idiomas diferentes se situa en un diapasión de probabilidades que va desde el total recubrimiento del contenido del término de la lengua A por uno de la lengua B, hasta la total falta de equivalencias, pasando por una variada gama de recubrimientos parciales.

De acordo com Felber (1987), a equivalência total caracteriza-se quando o termo da língua A possui os mesmos traços semânticos que o termo da língua B. Assim sendo A=B. Seguindo a mesma orientação, Dubuc (1985, p. 55) afirma que “dois termos são equivalentes se possuem uma identidade completa de sentido e de usos dentro de um mesmo domínio de aplicação”.

1.2.3. Dicionários eletrônicos

“A união do dicionário e da informática perdura e traz frutos cada vez mais ricos e numerosos.”¹⁹ (GOUADEC, 1997, p. 24). O formato papel impede uma exploração adequada dos textos terminológicos, pois a manipulação dos dados é praticamente impossível.

Mesmo com todo o avanço da tecnologia computacional e dos recursos que podem ser utilizados para uma consulta rápida e eficiente, grande parte dos materiais terminológicos ainda esbarra no papel.

Os responsáveis pela maioria desses projetos escolheram difundir os dados já existentes nos dicionários impressos em papel e se acharam confrontados com a necessidade de convertê-los de maneira a que possam ser explorados a partir de uma interface de interrogação. De fato, a grande maioria dos dicionários disponíveis no mercado ainda é redigida em função do modo de apresentação herdado da tradição lexicográfica e destinado a facilitar a consulta sobre um suporte escrito²⁰. (VAN CAMPENHOULT, 2000, p. 5-6)

Existe uma discrepância entre a modernização dos procedimentos computacionais e a prática terminológica. São poucas as editoras que apostam no conteúdo baseado em suporte

¹⁹ *L'union du dictionnaire et de l'informatique perdure et porte des fruits de plus en plus riches et de plus en plus nombreux.*

²⁰ *Les responsables de la plupart de ces projets ont choisi de diffuser des données existant déjà dans des dictionnaires imprimés sur papier et se sont retrouvés confrontés à la nécessité de les convertir de manière à ce qu'on puisse les exploiter à l'aide d'une interface d'interrogation. En effet, la très grande majorité des dictionnaires disponibles sur le marché sont encore rédigés en fonction d'un mode de présentation hérité de la tradition lexicographique et destiné à faciliter la consultation sur un support écrit.*

eletrônico, como consulta online ou em CD-ROMs. Eles oferecem recursos indispensáveis para os usuários, que, por sua vez, apenas necessitam de pouca experiência com essas ferramentas para que sejam capazes de interagir.

Antes do advento do computador, é interessante notar os comentários a seguir, a respeito da idealização de uso do dicionário.

Os dicionários deveriam servir para todas as pesquisas que podem ser feitas sobre palavras, e como há nas palavras duas coisas essencialmente diferentes, a forma e a significação, um dicionário, para ser completo, deveria se dividir em duas partes distintas, nas quais uma pudesse servir para encontrar a significação e a outra para encontrar a forma. A primeira dessas partes foi feita mil vezes; é o dicionário da Academia, é o de Boiste, de Laveaux, de Bescherelle e de muitos outros. Mas a outra parte nunca foi feita, entretanto ela não é menos indispensável que a primeira, pois expressar todos os seus pensamentos é tão importante quanto poder compreender todos os pensamentos expressos (BOISSIÈRE, s.d., p. ii, apud BABINI, introdução, p. 148).

As digressões apresentadas acima podem ser facilmente executadas, na atualidade, frente às inúmeras linguagens de programação que a informática possui. O desenvolvimento de aplicativos que interagem com o usuário já se tornou uma realidade. Um dicionário eletrônico consegue explorar facilmente seu banco de dados para encontrar tanto a significação quanto a forma.

Mas que tipo de dicionário ou de base de dados seria mais eficaz? Todos os repertórios lexicográficos e terminográficos podem fornecer a definição e o significado de uma palavra ou de um termo, mas, como encontrar essa palavra ou esse termo, se não os conhecemos? Como chegar à expressão, se temos somente seu conceito ou parte deste? Os dicionários onomasiológicos são instrumentos importantes para o atendimento dessa necessidade. (BABINI, 2006, p. 38)

Ambas as formas de pesquisa provam ser frutíferas dentro da terminologia. Um dicionário eletrônico que seja acionado na orientação semasiológica, bem como na onomasiológica, adicionaria uma forma reveladora de tratamento das áreas de especialidade.

O usuário acessa a informação conforme seus objetivos, sem o empecilho do formato estanque. Biderman pensa que

O confronto de um campo onomasiológico com os campos semasiológicos afins demonstra que eles se interpenetram e se complementam. Portanto, a Onomasiologia e a Semasiologia constituem uma boa metodologia para o estudo da forma como se estrutura o Léxico de uma língua. (BIDERMAN, 2001, p. 201)

Ainda a respeito da semasiologia e da onomasiologia, Babini (2006, p. 41) preconiza que os dicionários, de modo geral, deveriam permitir encontrar tanto “as palavras pelas idéias” quanto “as idéias pelas palavras”.

Um dicionário terminológico onomasiológico bilíngüe inglês-português e português-inglês de RNAS pretende servir como uma base de consulta precisa. O dicionário em questão deverá ser baseado em corpus, que, por sua vez, é uma tendência na elaboração de dicionários, sejam gerais ou técnicos. Os termos serão arrolados, seguindo-se a metodologia da lingüística de corpus e da Teoria Comunicativa da Terminologia. Porém, “o grande problema com o dicionário bilíngüe é sua limitação no que diz respeito ao número de vocábulos arrolados e à má qualidade das definições apresentadas. (SCHMITZ, 2001, p. 163). O dicionário terminológico onomasiológico bilíngüe inglês-português e português-inglês de RNAs recorrerá, quando possível, às definições constantes do próprio corpus, a fim de garantir genuinidade ao texto e mínima interferência externas. “Este tipo de dicionário [bilíngüe] é fundamental para o desenvolvimento cultural e tecnológico do Brasil e deve ser visto como uma ferramenta para todos os profissionais.” (op. cit. p. 167)

No final dos anos 1960, Castilho e Carratore, vislumbram com otimismo o futuro da onomasiologia brasileira,

Falar em perspectivas para a onomasiologia, ou de qualquer outro ramo da Lingüística, com algumas exceções, no Brasil poderá parecer um gasto lugar-comum (tudo tem "perspectivas" em nosso país), mas a verdade é que se disséssemos que estamos na fase

inicial empregariamos um eufemismo descabido - na verdade, tudo está por fazer. Parece-nos, entretanto, que não faltam as bases que podem permitir a edificação de trabalhos sólidos e valiosos [...]. (CASTILHO ; CARRATORE, 1967, p. 140)

Babini acredita que um dicionário onomasiológico eletrônico seja de grande valia por tornar viável uma consulta interativa, de forma completa.

Para que todos os percursos onomasiológicos que repertoriamos e os que desejamos criar sejam possíveis, acreditamos que seja necessário realizar este novo modelo de dicionário onomasiológico, sob forma de base de dados, com a ajuda de software (informático). Seria muito difícil realizar pesquisas onomasiológicas a partir dos semas num dicionário impresso, sobretudo, quando for combiná-los para encontrar a unidade lexical ou o termo pesquisado. (BABINI, no prelo, p. 220)

A dificuldade em se fazer esse tipo de consulta é menos eficiente em um dicionário em papel. Um dicionário eletrônico terminológico onomasiológico tenciona sanar, em partes, o déficit de conteúdo terminológico em suporte eletrônico. A partir do momento em que uma obra terminológica está disponível eletronicamente, faz-se interessante uma ampla opção de busca de informação, de forma rápida e prática. Barros discorre sobre os dicionários inteligentes:

A evolução das obras lexicográficas e terminográficas conduziu à criação de uma nova expressão (e conceito): a de dicionários inteligentes. Uma das funções mais importantes de que dispõem as obras consideradas desse tipo é a de busca sofisticada dos dados. Esta se dá por meio de filtros, que utilizam critérios como, por exemplo, categorias gramaticais, usos, frequência, palavras correlatas e outros. (BARROS, 2004, p. 277)

Quanto à disponibilização dos dicionários terminológicos no final da década de 1990, Gouadec revela que

[...] o mercado de dicionário cresce a um ritmo de 40% ao ano, e os dicionários em suporte eletrônico diversos representam unicamente 1% do mercado. No ano 2000, a tiragem dos dicionários técnicos se

situará ao redor de 3000 exemplares, na razão de 1500 em suporte papel e 1500 em suporte em CD-ROM.²¹ (GOUADEC, 1997, p. 25)

A realidade mais de dez anos depois parece não garantir número suficiente de exemplares em suporte em CD-ROM. A tendência atual é a publicação do dicionário juntamente com o CD-ROM. Há casos isolados nos quais a compra do CD-ROM separadamente é possível. No entanto, dicionário eletrônico terminológico permanece raro.

Auger, no final dos anos 1980, prevê a popularização de dicionários em forma eletrônica e acrescenta a utilização do hipertexto na pesquisa.

De um outro ponto de vista, poderíamos sonhar em utilizar da mesma forma os dicionários em sua forma eletrônica (todos os dicionários importantes existentes dessa forma antes de serem publicados), esses “hipertextos” tratados pelos programas de indexação poderiam se revelar ajudas preciosas para a terminólogo (pesquisa de contextos, redação de definições, etc)²² (AUGER, 1989, p. 454)

Os recursos que os aplicativos em suporte eletrônico podem executar acabam surpreendendo por conta das perspectivas e pelo constante avanço no modo de consulta.

No intuito de desenvolver aplicativos cada vez melhores que serão executados em computadores pessoais, a representatividade do computador nessa tarefa é destacada por Bergenholtz e Tarp (1995, p. 31) da seguinte forma: “é rápido, tanto em buscar informações quanto em executar operações. É consistente e capaz de desempenhar operações complicadas em materiais complicados com grande precisão.”²³

²¹ [...], le marché du dictionnaire croit au rythme de 40 % par an, et les dictionnaires sur supports électroniques divers représentent uniquement 1 % du marché. En l' an 2000, le tirage courant des dictionnaires techniques se situera autour de 3000 exemplaires, à raison de 1500 sur support papier et 1500 sur CD-ROM.

²² D'un autre point de vue, on pourrait songer à utiliser de la même façon les dictionnaires sous leur forme électronique (tous les dictionnaires importants existent sous cette forme avant d'être publiés), ces "hypertextes" traités par des logiciels d'indexation pourraient se révéler être de précieux aides pour le terminologue (recherches de contextes, rédaction de définitions etc.).

²³ It is fast, both at retrieving material and performing operations on it. It is consistent and capable of performing complicated operations on complicated material with great precision.

É interessante notar que Landau (1989, p. 272) destaca que é “inconcebível que qualquer grande dicionário possa ser levado a cabo hoje sem pensar em armazenar o material no computador.”²⁴

No caso da informática, mais especificamente das redes neurais artificiais, que estão em constante desenvolvimento,

O número de termos criados para expressar novos conceitos relacionados a esse campo está constantemente crescendo em inglês, uma vez que novas máquinas e programas são inventados continuamente e distribuídos mundialmente. Como resultado, os países com línguas que não o inglês têm que lutar regularmente com os termos computacionais para achar uma versão popular e adequada em seu estoque de palavras, o que não é fácil fazer.²⁵ (MEDINA, 2003, p. 317-318)

Berber Sardinha (2008) considera que um dicionário onomasiológico baseado em corpus possui maior abrangência e maior quantidade de dados. Nesse sentido, a lingüística computacional pode oferecer sua contribuição para que a pesquisa onomasiológica ganhe mais agilidade.

1.3. A lingüística computacional

Othero e Menuzzi (2005, p. 12) afirmam que a “[...] lingüística computacional é a área da lingüística que se ocupa do tratamento computacional da linguagem para diversas finalidades práticas.”. Costuma-se dividir a lingüística computacional em duas subáreas: a lingüística de corpus e o processamento de linguagem natural (PLN).

Graças aos avanços da informática, especialmente após os anos 1980, diversas áreas do conhecimento puderam observar e ao mesmo tempo manipular seus objetos de estudo sob

²⁴ *It is inconceivable that any major dictionary could be undertaken today without planning to store the material in a computer.*

²⁵ *The number of terms created to express new concepts related to this field is constantly increasing in English, since new machines and programs are invented continuously and distributed worldwide. As a result, countries with languages other than English have to struggle regularly with computer terms in order to find a suitable and popular version in their word stock, which is not easy to do [...].*

a perspectiva da chamada lingüística computacional, de fato, inovadora e empírica. A possibilidade de manipulação da língua pelo computador agiliza a realização de análises lingüísticas, sejam elas quais forem. Com o aporte da informática, atesta-se que a pesquisa de viés lingüístico é incrementada sobremaneira. Assim, ao demonstrar como a língua se comporta em determinadas situações reais de uso, o lingüista encontra meios de comprovar ou refutar hipóteses. Dentre inúmeras outras aplicações, pode ainda utilizar os recursos da lingüística computacional para fazer a revisão ortográfica ou mesmo gramatical de textos.

O sucesso do software moderno no processamento de língua natural impressiona nossa imaginação. Programas para correção ortográfica e gramatical, busca de informação em bancos de dados, e tradução de uma língua natural em outra, entre outros, são vendidos mundialmente em milhões de cópias, hoje em dia. ²⁶ (BOLSHAKOV; GELBUKH, 2004, p. 5)

Apesar de todas as vantagens que a lingüística computacional proporciona, “[...] se perguntarmos à maioria dos cidadãos se eles acham que a informática está presente nas áreas vistas como menos tecnológicas, como as da esfera das Ciências Humanas, e Letras, em particular, a maioria das respostas, com quase certeza, seria negativa.” (BERBER SARDINHA, 2005a, p. 8). Deduz-se que ainda não está estabelecida na consciência do grande público a participação extremamente positiva dos recursos da informática nos estudos lingüísticos. A noção que a maioria dos cidadãos possui é de que o universo das letras pode prescindir de tecnologia em seus mais diversos segmentos. É comum encontrar pesquisadores devotados aos estudos lingüísticos ou literários abstraindo-se da tecnologia por ora disponível. A falta de interesse aumenta significativamente quando entra em questão o uso de estatística. São poucas as pesquisas que se aventuram na utilização da estatística, seja para ilustrar dados principais ou secundários. Alguns dizem que as técnicas estatísticas não estão relacionadas à lingüística. Outros parecem acreditar que a estatística encontra abrigo apenas nas ciências

exatas. “Entre a comunidade lingüística, os métodos estatísticos [...] são em grande parte ignorados ou rejeitados devido à falta de treinamento, medo e aversão também.”²⁷ (GOMEZ, 2002, p. 234).

Interessa neste trabalho abarcar a lingüística computacional e a terminologia. Conforme discussão anterior, os benefícios trazidos pela aliança da tecnologia e da lingüística, especialmente da terminologia, proporcionaram, no caso dos estudos do léxico, a comprovação de dúvidas, a visualização de fatos lingüísticos e produção de material baseado em corpus.

A terminologia e a informática não são independentes uma da outra. Desde 1960, em países desenvolvidos e com grande tradição em pesquisa terminológica, os dois campos de estudo estão ligados de forma a facilitar o armazenamento e a difusão de dados terminológicos [...]. (ALMEIDA *et alii*, 2006, p. 42)

Percebe-se que a informática está crescentemente imbricada na terminologia de forma a torná-la eficiente aos fins que se presta. Pelo mundo todo, existe a preocupação em facilitar o acesso a conteúdo terminológico, com o uso de ferramentas computacionais, bem como da lingüística computacional. A criação de bancos terminológicos informatizados passa a ser uma necessidade nas sociedades modernas.

A importância da lingüística computacional para a terminologia e para as áreas afins é explicitada por Barros:

A lingüística computacional, disciplina científica multidisciplinar que se caracteriza pela confluência de várias áreas, como a lingüística, a inteligência artificial, a computação e a lógica computacional, cria instrumentos de composição, extração, análise, manipulação e processamento de dados lingüísticos em texto, ferramentas eletrônicas que auxiliam enormemente o trabalho dos terminólogos. (BARROS, 2006, p. 25)

²⁶ *The success of modern software for natural language processing impresses our imagination. Programs for orthography and grammar correction, information retrieval from document databases, and translation from one natural language into another, among others, are sold worldwide in millions of copies nowadays.*

Levando-se em consideração as discussões acima, é possível encontrar respostas satisfatórias para as duas questões a seguir. “Para que propósitos precisamos desenvolver a lingüística computacional? Que resultados práticos ela fornece para a sociedade?”²⁸ (BOLSHAKOV; GELBUKH, 2004, p.53)

1.4. A lingüística de corpus

A lingüística de corpus é um ramo da lingüística computacional que vem ganhando atenção nos meios acadêmicos. No Brasil, principalmente na região sul e sudeste, existem diversas instituições voltadas aos estudos baseados em corpus. Em resumo, “a Lingüística de Corpus ocupa-se da exploração de corpora, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística.” (BERBER-SARDINHA, 2004, p. 3).

A definição mais completa de corpus, segundo Berber-Sardinha, pode ser a seguinte:

Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (BERBER SARDINHA, 2004, p. 18).

A pesquisa lingüística contemporânea é grandemente influenciada pela lingüística de corpus. Muitos dos resultados obtidos com a lingüística de corpus têm impacto direto nos estudos da tradução, no ensino de línguas estrangeiras e na produção de obras lexicográficas e terminográficas.

Nos estudos da tradução, a lingüística de corpus coloca em evidência o estilo do autor/tradutor, suas escolhas lexicais, a variação lexical da obra original e da traduzida, etc.

²⁷ *Among the linguistic community, statistical methods [...] are mostly ignored or avoided because of the lack of training, fear and dislike too.*

Para o ensino de línguas estrangeiras, a lingüística de corpus recorre à frequência de palavras, às colocações e também ao corpus de aprendizes (*learner's corpus*) para destacar pontos importantes que apresentam dificuldades ao aprendiz. Dois importantes corpora merecem destaque: O *Bank of English*, por exemplo, conta com mais de 500 milhões de ocorrências. O *Longman Corpus Network* registra mais de 300 milhões de ocorrências. Corpora da ordem de milhões de ocorrências retratam a língua em movimento, tal qual é usada pelos falantes de determinada comunidade lingüística. Biber (2000) aponta que uma das grandes vantagens de uma abordagem baseada em corpus é que ela proporciona um alcance e uma fidedignidade antes impossível

No tocante à terminologia, seria imprescindível recorrer à lingüística de corpus para o desenvolvimento de materiais de orientação terminológica. No caso de um dicionário de uma língua de especialidade, extenso e confiável, um corpus técnico-científico mostra-se importante. De uma forma ou outra, “a principal qualidade de um dicionário é sua exaustividade.”²⁹ (HOMBROUCK, s.d., p. 101). Tal exaustividade pode ser lograda se a obra em questão for baseada em corpus. Tanto os termos quanto seus conceitos podem ser analisados e processados pelo corpus. Evidentemente, a manipulação dos dados textuais armazenados em bancos de dados para essa finalidade requer uma boa dose de atenção e certa habilidade por parte do terminólogo.

Sem dúvida, a lingüística de corpus transformou a maneira pela qual a língua é concebida.

“[...] até recentemente muitas investigações de uso da linguagem eram impraticáveis ou simplesmente impossíveis. A abordagem baseada em corpus, no entanto, fornece um meio de manipular grandes quantidades de língua e de manter o controle de muitos fatores contextuais ao mesmo tempo. Isso, portanto, abriu caminho

²⁸ *For what purposes do we need to develop computational linguistics? What practical results does it provide for society?*

²⁹ *La principal qualité d'un dictionnaire est son exhaustivité.*

para uma gama de investigações de uso da língua.”³⁰ (BIBER; CONRAD; REPPEN, 1998, p. 3)

O campo de atuação da lingüística de corpus parece crescer incessantemente. Além de enriquecer o conhecimento humano a respeito da linguagem em uso, passa a ser uma forte aliada na produção de programas computacionais que processam a linguagem. No início dos anos 1990, Leech observa que “[...] a pesquisa baseada em corpus realmente mostra êxito, não apenas como um reconhecido paradigma para investigação lingüística, mas como contribuição essencial para o desenvolvimento de software de processamento de língua natural.”³¹ (LEECH, 1991, p. 21). Trata-se de uma tendência do mundo contemporâneo o fato de a lingüística de corpus e a PLN unirem-se para garantir materiais educativos ou profissionais de qualidade. Hoje, o desenvolvimento de software de processamento de língua natural começa a popularizar-se.

O software para tratamento de textos WordSmith Tools foi escolhido para processar e analisar o corpus de RNAS.

1.4.1. A ferramenta WordSmith Tools

Nas palavras de Mike Scott, o próprio criador do software, o programa WordSmith Tools (doravante, WST) é “[...] uma suíte de programas integrados para observar como as palavras se comportam nos textos.”³² (SCOTT, 2007, p. 2).

³⁰ [...] until recently many investigations of language use were unfeasible or simply impossible. The corpus-based approach, however, provides a means of handling large amounts of language and keeping track of many contextual factors at the same time. It therefore has opened the way to a multitude of new investigations of language use.

³¹ [...] corpus-based research has truly taken off, not only as an acknowledged paradigm for linguistic investigation but as a key contribution to the development of natural language processing software.

³² [...] an integrated suite of programs for looking at how words behave in texts.

Os recursos mais importantes do programa WST são o *WordList*, o *KeyWords* e o *Concord*. A função *WordList* possibilita, em poucos segundos, a ordenação da lista de palavras de um corpus de duas maneiras: por ordem alfabética e por ordem de frequência. Nos resultados estatísticos obtidos pelo WST, o conceito de *token* e *type* chama a atenção. *Token* é o número de vezes que determinada palavra foi identificada no corpus. *Type* é a própria palavra considerada isoladamente. Por exemplo, cada palavra é um *type*, cada ocorrência dela no corpus é um *token*. Pelo *Concord*, qualquer palavra da lista é visualizada tal como ocorre no corpus, dentro do texto. Finalmente, para acionar o *KeyWords* o usuário precisará dispor de um corpus de referência, além do corpus já investigado. Dessa forma, as palavras-chave do corpus investigado em questão serão listadas após comparadas com o corpus de referência, que deve ser mais abrangente. Berber Sardinha estimou “a quantidade média de palavras-chave numa coletânea de 40 corpora de estudos variados como sendo da ordem de 1472, com alguns corpora ultrapassando a marca de 3000 palavras-chave.” (BERBER SARDINHA, 2005b, p. 238).

Esse conhecimento auxilia o pesquisador no direcionamento de sua investigação. Se as palavras-chave são identificadas, elas podem fazer parte de material pedagógico, lexicográfico, terminológico, de divulgação profissional, ou serem destinadas a qualquer outra aplicação.

Com relação ao uso do WST, pode dizer que,

As ferramentas [*WordList*, *Concord* e *KeyWords*] são usadas pela Oxford University Press na realização do próprio trabalho lexicográfico que prepara dicionários, por professores e estudantes de idiomas, e por pesquisadores que investigam padrões linguísticos em muitas línguas diferentes em muitos países pelo mundo todo.³³ (idem)

³³ *The tools [WordList, Concord and KeyWords] are used by Oxford University Press for their own lexicographic work in preparing dictionaries, by language teachers and students, and by researchers investigating language patterns in lots of different languages in many countries world-wide.*

1.4.2. A frequência de palavras

“O critério quantitativo que vem imediatamente à cabeça, ao se falar de textos, é evidentemente o de frequência de palavras (ou de termos).”³⁴ (THOIRON; SERANT, 1989, p. 435). Pode-se afirmar que a frequência de palavras é um recurso que pode demonstrar a relevância de determinado termo em dada área de especialidade. A título de exemplificação, o termo *network* tem frequência altíssima no corpus das RNAs. Ocupa a décima posição de palavras mais frequentes e é o substantivo mais usado nesse domínio.

Figura 2: a lista de palavras do WST

N	Word	Freq.	%	Lemmas
1	THE	94.330	6,73	
2	OF	46.665	3,33	
3	A	29.479	2,10	
4	AND	27.363	1,95	
5	TO	25.177	1,79	
6	IN	24.461	1,74	
7	IS	23.121	1,65	
8	FOR	14.499	1,03	
9	THAT	10.714	0,76	
10	NETWORK	9.576	0,68	
11	BE	9.363	0,67	
12	NEURAL	9.354	0,67	
13	ARE	9.192	0,66	
14	WITH	8.942	0,64	
15	THIS	8.563	0,61	
16	AS	7.843	0,56	
17	BY	7.604	0,54	
18	NETWORKS	6.700	0,48	
19	ON	6.427	0,46	
20	WE	6.137	0,44	

Por outro lado, no corpus de língua geral compilado pela *American Heritage Society*, o termo *network* não figura entre as três mil palavras mais comuns. De fato, a frequência de uso dos termos na língua geral é baixa ou, muitas vezes, inexistente. Para que um termo “raro” possa ocorrer apenas uma vez, o corpus pode ter que abrigar mais de um milhão de *tokens*. Constata-se que, na sua essência, “[...] o vocabulário técnico-científico tem frequência de uso muito baixa na língua como um todo, sendo usado apenas no âmbito de cada língua de especialidade pelos profissionais dessa área [...]”.(BIDERMAN, 2006, p. 36). Entende-se,

³⁴ *Le critère quantitatif qui vient immédiatement à l'esprit, s'agissant de textes, est évidemment celui de la fréquence des mots (ou des termes).*

assim, o motivo pelo qual a devida seleção de textos específicos de campos de domínio mostra-se um dos procedimentos fundamentais no trabalho terminológico profissional.

Nessa seleção de termos constantes do corpus, a frequência é um recurso que faz com que o dicionário que tenha base na lingüística de corpus forneça as realizações reais da língua. “O fato de o dicionário fundamentar-se em corpus autêntico da língua como ela é realmente utilizada pelos falantes confere a essa obra uma autoridade indiscutível.” (BIDERMAN, 2003, p. 181). A devida abonação de termos, tomando-se como critério a frequência deles na área de especialidade, levará à criação de produtos terminológicos confiáveis.

É sabido que

O fenômeno da repetição é frequentemente evocado no tocante aos textos científicos. [...]. Dizemos, de fato, que um texto é repetitivo quando encontramos, em intervalos próximos, diversas ocorrências dos mesmos vocábulos. É então o conceito de topografia que aparece³⁵. (THOIRON; SERANT, 1989, p. 435)

Além de vislumbrar dados relevantes para a pesquisa lingüística, o trabalho de cunho estatístico permite saber em “[...] qual medida uma abordagem quantitativa de um texto de especialidade pode auxiliar a colocar em evidência os elementos nos quais o conteúdo informacional deve ser considerado como importante.”³⁶ (idem).

Cabe ao terminólogo a seleção dos textos que serão usados na constituição do corpus. Os textos, necessariamente, precisam ser separados em função de seu campo de domínio. Se as informações terminológicas forem retiradas de textos originais autênticos, tem-se a certeza de que elas são representativas do contexto no qual ocorrem. Dessa forma, a tarefa de cálculo de frequência de palavras é tornada possível.

³⁵ *Le phénomène de la répétition est d'ailleurs souvent évoqué à propos des textes scientifiques. [...]. On dit bien, en effet, qu'un texte est répétitif quand on y retrouve, à intervalles rapprochés, plusieurs occurrences des mêmes vocables. C'est donc le concept de topographie qui réapparaît.*

³⁶ *[...] quelle mesure une approche quantitative d'un texte de spécialité peut aider à mettre évidence les éléments dont le contenu informationnel doit être considéré comme important.*

Sinclair (2004, p. 23 apud SCOTT, 2007, p. 79) esclarece que “[...] deveríamos aceitar que uma grande parte de nosso comportamento lingüístico é subliminar, e que, portanto, podemos encontrar muitas surpresas.”³⁷. A frequência de palavras corrige justamente “o comportamento lingüístico subliminar”. O que entra em questão passa a ser o que realmente mostra-se importante e não o que se acredita que deva ser importante.

De acordo com Barros (2007), o uso e a frequência constituem características relevantes na pesquisa terminológica.

O uso é, enfim, um critério muito importante na identificação de termos, visto que nele se leva em conta a estabilidade da relação entre a unidade lexical e o conceito especializado. [...]. A frequência de uso constitui uma pista para a identificação de uma unidade terminológica. (BARROS, 2007, p. 50)

1.4.3. O corpus de especialidade

No final dos anos 1970, Houaiss, renomado lexicógrafo, tinha o desejo de incluir em seu dicionário toda a terminologia possível. O objetivo seria trazer para o dicionário todo, ou quase todo, o universo do conhecimento.

Um dos meus objetivos, no futuro, seria exatamente esse [computar proporcionalmente os vocabulários de cada universo do conhecimento]. A matéria não só é computável como também perfeitamente capitulável. Mas a tarefa demandaria, claro, uma equipe de sessenta a oitenta pessoas trabalhando durante quinze anos e, a partir daí, umas dez, quinze pessoas cuidando da manutenção e da atualização. (HOUAISS, 1978)

Atualmente, se esse trabalho fosse levado a cabo, renderia alguns anos de pesquisa. Seria provavelmente mais ágil e eficiente do ponto de vista da não-necessidade de intervenção total do ser humano na análise e processamento de imensa quantidade de dados. Já existem algumas iniciativas que ensejam tal empreitada, diga-se, de passagem, a da editora *Longman, Oxford, Collins*, e outras.

³⁷ [...] we should accept that a large part of our linguistic behaviour is subliminal, and that therefore we may find a lot of surprises.

Para computar proporcionalmente os vocabulários de cada universo do conhecimento, uma pesquisa terminológica exaustiva seria necessária. Nesse aspecto, Barros (2004) ressalta que é importante estabelecer com exatidão os domínios, como sendo um dos passos iniciais da pesquisa. “[...] a precisa delimitação do domínio é importante para o sucesso da sistematização da nomenclatura, para o equilíbrio desta, assim como de seu tratamento de uma maneira aprofundada e homogênea.” (BARROS, 2004, p. 119).

Sobre esse assunto, Cabré cita a mudança na sistemática do trabalho terminológico e as possibilidades de investigações. A presença de recursos computacionais, computadores mais potentes e velozes, facilidade de constituição de corpus e manipulação de dados conseqüentemente influenciam diretamente o resultado do estudo terminológico.

De fato, hoje em dia, a grande quantidade de textos digitalizados que se pode conseguir via Internet permite a criação de corpora de todas as temáticas. A aplicação a esses corpora de ferramentas de tratamento automático permite localizar e extrair a terminologia que contém e armazená-la em bases terminológicas. Os sistemas automáticos de extração de informação permitem também recuperar dos textos os contextos mais adequados, fragmentos de definição, variantes denominativas, e termos relacionados que sejam sinônimos, hiperônimos ou co-hipônimos do termo de entrada. Todas estas ferramentas, adequadamente articuladas em uma mesma plataforma, constituem estações de trabalho integradas que facilitam em grande medida o processo terminográfico de elaboração de glossários especializados.³⁸ (CABRÉ, 2005, p.6)

Não raro, um corpus de especialidade é mais completo que um corpus de língua geral. Mais completo no sentido de possuir uma densidade terminológica relativamente maior. A constituição de um corpus especializado e de dimensões extensas requer qualidade e representatividade dos textos selecionados. “[...] é razoável supor que um corpus que é

³⁸ *En efecto, hoy en día, la gran cantidad de textos digitalizados que pueden conseguirse vía Internet permiten la creación de corpora de todas las temáticas. La aplicación a estos corpora de herramientas de tratamiento automático permite localizar y extraer la terminología que contienen y almacenarla en bases terminológicas. Los sistemas automáticos de extracción de información permiten también recuperar de los textos los contextos más adecuados, fragmentos de definición, variantes denominativas, y términos relacionados que sean sinónimos, hiperónimos o cohipónimos del término de entrada. Todas estas herramientas, adecuadamente articuladas en una misma plataforma, constituyen estaciones de trabajo integradas que facilitan en gran medida el proceso terminográfico de elaboración de glosarios especializados.*

especializado dentro de um certo domínio terá uma maior concentração de vocabulário que um corpus de maior amplitude [...].”³⁹ (SINCLAIR, 2004). Como o corpus de especialidade concentra o vocabulário especializado, existe uma probabilidade maior de se encontrar as realizações mais importantes daquele domínio. O corpus de especialidade das RNAs possui mais de um milhão de palavras e foi constituído de forma a permitir a identificação dos termos fundamentais dessa área do conhecimento. No entanto, Almeida (2006) acredita que um corpus médio, porém mais balanceado e diversificado, é muito mais representativo, pois constitui uma amostra mais fiel das possibilidades comunicativas de determinado domínio.

De acordo com os cinco níveis de categorização do tamanho de um corpus, proposto por Berber Sardinha (2000), um corpus com menos de 80 mil ocorrências é um corpus pequeno. O corpus pequeno-médio possui entre 80 mil e 250 mil ocorrências. De 250 mil a 1 milhão, é considerado médio. De 1 milhão a 10 milhões de ocorrências pode ser denominado um corpus médio-grande. Acima dessa quantidade, grande. A constituição de um corpus com esses requisitos exige do pesquisador bastante rigor, tanto na seleção quanto na análise.

Aubert (1996b) sugere uma metodologia para o levantamento de termos para um trabalho terminológico. Não cita, no entanto, se o inventário dos termos essenciais é assistido por computador. A lingüística de corpus não é mencionada nessa passagem.

[...] a constituição de um inventário dos termos essenciais à especialidade em questão, na média não ultrapassando 300 ou 400 termos; no segundo, será preciso prever o levantamento de por volta de 2.500 termos. O levantamento básico proporciona resultados mais rápidos, mas coloca a questão, talvez problemática para o iniciante na área de conhecimento em pauta, de distinguir termos essenciais e termos acessórios. O levantamento exaustivo, inversamente, demanda mais tempo, mas coloca menos exigências para a seleção dos termos, uma vez consolidada a grade conceptual. (AUBERT, 1996b, p. 59)

Se o iniciante da área em pauta tivesse conhecimentos da lingüística de corpus e tivesse, a seu dispor, um corpus representativo, não encontraria muita dificuldade em

³⁹ [...] it is reasonable to suppose that a corpus that is specialised within a certain subject area will have a greater concentration of vocabulary than a broad-ranging corpus [...].

distinguir os termos essenciais daqueles menos freqüentes, em determinada área do conhecimento.

Um levantamento dos termos fundamentais de áreas de especialidade resulta em um maior conhecimento dos mistérios da linguagem humana, bem como sua capacidade de transformação e adaptação.

1.4.4. O etiquetamento

O etiquetamento dos termos é um procedimento que permite saber, dentre outros detalhes, qual o número de substantivos, adjetivos ou verbos em determinada língua de especialidade. Para se fazer o etiquetamento do corpus, é preciso um etiquetador que “serve para inserir automaticamente, no corpus, códigos que indicam a classe gramatical de cada palavra.” (BERBER-SARDINHA, 2004, p. 113).

Por meio do recurso do etiquetamento, pode-se, por exemplo, identificar quais são as colocações mais comuns em um corpus.

No domínio da terminologia computacional, [...] cada vez mais trabalhos destacam a importância da estruturação da terminologia. Essa estruturação precisa descobrir ligações, notadamente semânticas, entre as unidades terminológicas e de eventualmente etiquetá-las.⁴⁰ (CLAVEAU & L’HOMME, 2005, p. 1)

Essa modalidade de estudo terminológico pode ser útil, por exemplo, no ensino de terminologia, ou seja, os aprendizes podem ter contato com os verbos mais comuns de determinada área, visualizando quais são as palavras que co-ocorrem mais comumente.

⁴⁰ *Dans le domaine de la terminologie computationnelle, [...], de plus en plus de travaux soulignent l'importance de la structuration de terminologie. Cette structuration nécessite de découvrir les liens, notamment sémantiques, entre les unités terminologiques et d'éventuellement les étiqueter.*

1.5. O contraponto de Chomsky

Chomsky nunca acreditou que a lingüística de corpus ou estudos baseados em corpus pudessem ser eficientes e contribuir para o conhecimento da língua. Nem a frequência, muito menos o sistema probabilístico são encarados como algo a ser levado em consideração. Chomsky descarta a manipulação de grandes quantidades de texto. Prefere levar em consideração a intuição que o falante tem da língua, juntamente com suas abstrações. As elucubrações de Chomsky a respeito da lingüística de corpus não são exclusividade da era da tecnologia. Nos primórdios da lingüística computacional, quando já se idealizava a manipulação do corpus pelo computador, Chomsky opinava que “qualquer corpus natural será distorcido. Algumas frases não ocorrerão porque são óbvias, outras porque são falsas, outras ainda porque são indelicadas. O corpus, se natural, será distorcido de tal forma que a descrição seria nada mais que uma mera lista.”⁴¹ (CHOMSKY, 1962, p. 159, apud LEECH, 1991, p. 9).

A visão do corpus como uma mera lista acabou por restringir o avanço dos estudos baseados em corpus em instituições que se pronunciavam seguidoras do gerativismo chomskyano. Pelo forte apelo que as teorias de Chomsky detinham nessa época, a lingüística de corpus deixou de ser estudada em todo o seu potencial em algumas partes do mundo. Especialmente na Europa, a lingüística de corpus recebeu guarida, e lá conseguiu se desenvolver fortemente, como na Inglaterra e na França, dois países de tradição nos estudos de corpora.

Chomsky acrescenta, ainda, que

A lingüística de corpus não significa nada. [...] suponha que a física ou a química decida que em vez de basear-se em experimentos, o que farão é filmar as coisas que acontecem no mundo e coletarão gigantescas filmagens de tudo que está acontecendo. Talvez criem

⁴¹ *Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list.*

algumas generalizações ou insights. Bem, como se sabe, a ciência não faz assim. [...] Não conseguirão muito apoio no departamento de química, física ou biologia. [...] Meu julgamento, se quer saber, é que aprendemos mais sobre a linguagem ao seguir o método padrão das ciências. O método padrão das ciências é não acumular grandes volumes de dados não-analisados e tentar tirar algumas generalizações a partir deles. As ciências modernas, pelo menos desde Galileu, têm sido surpreendentemente diferentes.⁴² (CHOMSKY, 2004, p. 97)

Contrariamente ao que Chomsky afirma, não se acumulam grandes volumes de dados não-analisados e não se tenta tirar algumas generalizações a partir deles. Os dados são analisados sob a perspectiva da terminologia, sob a metodologia da lingüística de corpus. O computador processa os dados de modo infinitamente superior aos humanos, sem substituí-los de forma alguma na ciência lingüística. “[...] o computador é uma boa máquina para ajudar os humanos a identificarem padrões e tendências. [...] Uma ferramenta ajuda na execução de um trabalho, ela não executa o trabalho para você.”⁴³ (SCOTT, 1997). São os pesquisadores que fazem o trabalho após o auxílio da máquina em indicar tendências.

Scott (2008) acredita que os estudos terminológicos ficam empobrecidos sem o aporte da lingüística de corpus. O mesmo pesquisador da universidade de Liverpool afirma que a lingüística de corpus é uma ferramenta útil. “Se é útil, por que não usá-la?” A mesma opinião é compartilhada por Berber Sardinha (2008), segundo o qual, se a pesquisa terminológica exigir análise de muitos dados, então a pesquisa fica empobrecida.

⁴² *Corpus linguistics doesn't mean anything. [...] suppose physics and chemistry decide that instead of relying on experiments, what they're going to do is take videotapes of things happening in the world and they'll collect huge videotapes of everything that's happening and from that maybe they'll come up with some generalizations or insights. Well, you know, sciences don't do this. [...] They're not going to get much support in the chemistry or physics or biology department. [...] My judgment, if you like, is that we learn more about language by following the standard method of the sciences. The standard method of the sciences is not to accumulate huge masses of unanalyzed data and to try to draw some generalization from them. The modern sciences, at least since Galileo, have been strikingly different.*

⁴³ *[...] the computer is a good device for helping humans to spot patterns and trends. [...] A tool helps you to do your job, it doesn't do your job for you.*

1.6. A tradução e a terminologia

Um tradutor que tenha em mãos um texto de uma área de especialidade qualquer necessita estar familiarizado com os termos empregados nela. Caso contrário, correrá o risco de não transmitir a mensagem satisfatoriamente. Dubuc (1974) apresenta um estudo realizado para comparar a terminologia estatística utilizada em uma tradução com os dados de uma pesquisa terminológica original. Dos 137 casos de expressões pesquisadas, houve apenas 27 casos de concordância completa, o que deixa um saldo de 110 casos de disparidade.

O nível de disparidade foi de mais de 80% dos casos. Essa situação serve para ilustrar como a falta de estudos baseados em corpus pode prejudicar a atividade tradutória especializada. Se houvesse uma pesquisa no banco de dados ou no corpus terminológico, o nível de disparidade cairia drasticamente a um dígito ou alcançaria o nível zero.

Em se tratando de uma língua de civilização com antiga tradição literária e escrita o léxico atinge dimensões enormes, o que dificulta o trabalho de recolha, seleção e tratamento dos dados vocabulares. Assim sendo, os dicionaristas saudaram o advento e a banalização do uso do computador em sua práxis como uma ferramenta *sui-generis* e quase miraculosa. (BIDERMAN, 2003, p. 179)

Fica claro que o corpus a que o tradutor especializado recorrerá será um corpus especial, até mesmo porque os “[...] ‘termos’ das línguas de especialidade não chegam a cobrir 2%” de um corpus de língua geral.⁴⁴ (PICOCHÉ, 2002, p. 108). A tarefa de incluir toda a terminologia de todas as áreas de especialidade demandaria um dicionário com milhões de páginas e milhares de volumes. “Em tese, somente o vocabulário ‘técnico’ de base de grandes áreas especializadas é incluído em um DLG [Dicionário de língua geral]. Isto é, somente aparecem os termos especializados mais representativos da área [...]” (KRIEGER e FINATTO, 2004, p. 131-132). Deduz-se que os dicionários gerais não são voltados para

⁴⁴ [...] les “termes” des langues de spécialités n'en couvre que 2%.

indivíduos que precisam de informações a respeito de uma área técnica específica, ainda que exibam alguns termos.

A tradução de textos especializados nos mais diversos setores da atividade humana cresce incessantemente. A busca por fontes confiáveis de consulta torna-se, cada vez mais, um pré-requisito para o bom desempenho da atividade profissional do tradutor.

Assim, a tradução e a terminologia compartilham uma relação de relativa importância. Ambas as disciplinas contribuem para a transmissão do conhecimento, seja ele técnico ou não. O tradutor recorre ao material elaborado por terminólogos, ao passo que os terminólogos recorrem ao material que pode ter sido produzido pelo tradutor. Sob a perspectiva de Van Hoof, “na verdade, a profissão de terminologista deriva da profissão de tradutor.” (VAN HOOFF, 1998, p. 253). A propósito da terminologia e da tradução, Cabré adiciona que

A terminología e a tradução surgiram as duas da prática, da necessidade de expressar um pensamento especializado ou de resolver um problema de compreensão. A terminologia como atividade consciente surgiu do interesse dos estudiosos em estarem de acordo com a fixação dos conceitos e denominações de suas respectivas ciências, sobretudo as ciências naturais. [...]. A tradução nasceu devido à necessidade de facilitar a compreensão entre línguas distintas, uma necessidade comunicativa evidente.⁴⁵ (CABRÉ, 1999, p. 178-179)

O contato entre as línguas, especialmente entre línguas de especialidade, impele o tradutor a adentrar os domínios da terminologia bilíngüe. De certa forma, a habilidade em lidar com o léxico especializado e com a terminologia bilíngüe resulta em uma tradução que reflete o uso real de uma área técnica. “Assim, no caso da tradução de textos especializados, é

⁴⁵ *La terminología y la traducción surgieron las dos de la práctica, de la necesidad de expresar un pensamiento especializado o de resolver un problema de comprensión. La terminología como actividad consciente surgió del interés de los científicos para ponerse de acuerdo en la fijación de los conceptos y denominaciones de sus respectivas ciencias, sobre todo, las ciencias naturales. [...]. La traducción nació ante la necesidad de facilitar la comprensión entre lenguas distintas, una necesidad comunicativa evidente.*

fundamental ao tradutor saber identificar e delimitar os termos presentes no texto sobre o qual trabalha.” (BARROS, 2007, p. 42)

Krieger e Finatto explicam que

A primeira motivação do encontro que direciona a Tradução para a Terminologia relaciona-se ao fato de que os termos técnico-científicos são elementos-chave, nódulos cognitivos, dos textos especializados. É em relação a essa tipologia textual que se efetua a chamada tradução técnica ou especializada. (KRIEGER; FINATTO, 2004, p. 66)

Fica evidente que o tradutor, especialmente o profissional, não pode prescindir das contribuições que a terminologia é capaz de fornecer. Caso contrário, a possibilidade de haver problemas no texto traduzido é muito alta, prejudicando a qualidade da mensagem. Barros (2006, p. 23) recomenda que “uma boa tradução não deve apenas expressar o mesmo conteúdo que o texto de partida, mas fazê-lo com as formas que um falante nativo da língua de chegada utilizaria”. A naturalidade no emprego das terminologias advém de uma pesquisa terminológica minuciosa. Ainda hoje, inúmeros artigos e textos traduzidos trazem incoerências ou termos que podem ser denominados como verdadeiros “corpos estranhos”. De fato, são a consequência da falta de uma pesquisa terminológica profunda ou mesmo confiável na área de especialidade em questão. Portanto, causa pouco estranhamento a presença de “pseudo-termos” nas traduções de áreas de especialidade, sejam elas quais forem. Infelizmente, “[...] é freqüente, encontrarmos na prática, traduções técnicas eivadas de erros provenientes da ignorância da língua de origem e da intuição de seus hábitos de expressão.” (MAILLOT, 1975, p. 98). Ángel Vega (1996), ao abordar a mesma problemática tradutória, afirma que as falhas, chamadas gramaticais, não são as mais problemáticas. “Foi provado que grande parte dos erros de tradução se devem, deixando de lado as más-interpretações das estruturas sintáticas, ao fato de não haver informações léxicas

ou terminológicas suficientes, ao fato de não terem sido consultadas.”⁴⁶ (ÁNGEL VEGA, 1996, p. 69). Dentre as várias estratégias e procedimentos a que o tradutor de textos especializados pode recorrer, uma delas consiste em identificar os termos presentes nos textos. O levantamento das unidades terminológicas exige um preparo prévio em terminologia. Barros (2007, p. 51) ilustra esse conceito da seguinte maneira:

A prática tradutória de textos de áreas técnicas, científicas ou especializadas exige que o profissional da tradução tenha conhecimentos básicos de Terminologia para que possa realizar bem sua tarefa. O primeiro passo é saber identificar as unidades terminológicas existentes no texto sobre o qual se debruça. (BARROS, 2007, p. 51)

A terminologia pode ser entendida como um subsídio para que a tradução técnico-científica seja desenvolvida de maneira satisfatória. Frente ao que foi exposto nos parágrafos acima, depreende-se que “a maior colaboração que a disciplina terminológica pode oferecer aos tradutores é a de auxiliá-los a compreender a natureza, o estatuto, a constituição e o funcionamento dos termos técnico-científicos” (KRIEGER; FINATTO, 2004, p.70).

1.7. A tradução e os dicionários especializados

Os autores concordam que a maior parte dos consulentes dos dicionários especializados, mono ou bilíngües, são indivíduos com pouca familiaridade com o domínio em questão. Assim sendo, estão à procura de dados terminológicos por vezes essenciais para sanar uma dúvida, semântica ou ortográfica, ou mesmo dar continuidade a uma tradução. “De fato, os dicionários especializados são freqüentemente consultados por um público pouco ou nada especializado no domínio.”⁴⁷ (BARROS, 1999, p. 73). Os tradutores se enquadram na

⁴⁶ Se ha demostrado que gran parte de los errores de traducción se deben, dejando aparte las malinterpretaciones de las estructuras sintácticas, bien al hecho de no disponer de la información léxica o terminológica suficiente, bien al hecho de no haberlas consultado.

⁴⁷ *En effet, les dictionnaires spécialisés sont très souvent consultés par un public moins ou pas de tout spécialisé dans le domaine.*

⁴ *[...] les vocabulaires son rarissimes dans les spécialités émergentes.*

categoria daqueles que detêm pouco ou nenhum conhecimento do domínio no qual está pesquisando. Com exceção dos profissionais de tradução que acabaram por se especializar em um campo, há aqueles que se deparam freqüentemente com novos desafios na rotina de sua atividade. Dessa forma,

[...] os tradutores profissionais apresentam-se como um dos principais grupos de usuários finais dos produtos da pesquisa terminológica (glossários, dicionários técnicos, bases de dados terminológicos, etc.). Nessa perspectiva, cabe aos estudos terminológicos gerarem esses produtos, com base em metodologias adequadas, proporcionando recursos confiáveis para o exercício da tarefa tradutória. (AUBERT, 1996b, p. 14)

Na mesma linha de pensamento, com relação aos dicionários técnicos, “pelo que se pode observar, é o tradutor o profissional que aparece como um beneficiário mais direto ou evidente.” (KRIEGER; FINATTO, 2004, p. 124). “Afinal, o dicionário não é um instrumento essencial para o trabalho do tradutor? Os dicionários de mais de uma língua são a primeira coisa que vem à mente.” (VAN HOOFF, 1998, p. 241). Tal associação tradutor-dicionário integra um dos itens pertencentes ao conceito estereotipado do que é ser tradutor.

Quando necessitam consultar dicionários terminológicos, na maior parte das vezes, os tradutores utilizam repertórios bilíngües. “[...] os vocabulários são raríssimos nas especialidades emergentes.”⁴⁸ (PAVEL, 1989, p. 346). É o caso das RNAs que reclamam material apropriado que descreva apropriadamente sua especialidade.

Um dicionário ou glossário que abone os termos das redes neurais artificiais, por exemplo, ainda não existe no mercado, nem em papel, muito menos em versão eletrônica. O tradutor que lida com a terminologia das RNAs possivelmente terá sua tarefa dificultada pelo fato de as RNAs serem relativamente novas dentro do domínio da inteligência artificial.

“De fato, a pesquisa documentária e terminológica ocupa mais de 40% do tempo destinado à tradução [...]”⁴⁹ (SADER FEGHALI, s/d, p. 11). No caso do tradutor das RNAs, o tempo dedicado à pesquisa terminológica será consideravelmente maior. Na ausência de recursos terminológicos que tratem das RNAs, sua tarefa será praticamente intuitiva e passível de inexatidões. “E, nessa *utilização* são obrigados, solitariamente, a encontrar soluções de manejo terminológico [...]” (KRIEGER, 2006, p. 195).

Deve-se levar em consideração que, para o tradutor,

“As dificuldades aumentam porque nem sempre os profissionais da tradução podem contar com boas obras de referência especializada, bi e/ou multilíngües. O acesso a esse tipo de obra é de grande utilidade, porquanto repertórios temáticos já sistematizados contribuem positivamente para o ato de documentar-se, componente que integra o rol de competências exigidas do tradutor especializado [...]” (KRIEGER, 2006, p. 194)

Pode-se afirmar que, quando não existem fontes de consulta na língua portuguesa, os tradutores são impelidos a criarem soluções para transpor suas barreiras terminológicas. Para que essas soluções sejam aceitáveis, é recomendável que os tradutores possuam outros conhecimentos lingüísticos concernentes à formação de palavras, prefixação e sufixação, porque

[...] negligenciar as peculiaridades lingüístico-terminológicas de uma determinada comunidade profissional pode trazer como efeito traduções mal-sucedidas, que não atendem às expectativas dos leitores dessa comunidade. Mas, infelizmente, é inegável a carência de instrumentos de referência em português brasileiro, o que obriga tradutores a atuarem como verdadeiros "pesquisadores-exploradores" das linguagens especializadas. (KRIEGER; FINATTO, 2004, p. 179)

No que diz respeito ao trabalho terminológico bilíngüe, o pesquisador não pode prescindir de uma adequada pesquisa terminológica que leve em consideração os traços semânticos dos termos envolvidos.

⁴⁹ *En effet, même si la recherche documentaire et terminologique accapare parfois plus de 40% du temps consacré à la traduction [...].*

1.8. As Redes Neurais Artificiais e suas aplicações

As Redes Neurais Artificiais são uma subárea da Inteligência Artificial. Da mesma forma que o cérebro humano, as Redes Neurais Artificiais são construídas para que funcionem em conjunto. Cada neurônio é responsável por processar determinada informação. As RNAs podem ser definidas como sendo estruturas inteligentes capazes de tomar decisões. Após minuciosa análise de padrões e dados previamente inseridos no sistema, as RNAs, conseguem “aprender” os padrões – com ou sem supervisão – e, baseadas em cálculos probabilísticos, apresentam respostas às tarefas mais variadas. São comumente utilizadas no reconhecimento de padrões. De acordo com Babini e Marranghello (2007, p. 15-16), “reconhecer padrões é uma das atividades mais recorrentes e mais importantes da vida do ser humano, que a pratica constantemente, e que lhe permitiu uma relativa rapidez na evolução [...]”. Em outras palavras, “as Redes Neurais Artificiais (RNAs) são ferramentas de Inteligência Artificial que possuem a capacidade de se adaptar e de aprender a realizar uma certa tarefa, ou comportamento, a partir de um conjunto de exemplos dados.” (OSÓRIO; BITTENCOURT, 2000, p. 2). Como visto há pouco, as RNAs são desenvolvidas para que tenham a capacidade de agir como o cérebro humano, tanto na organização quanto no funcionamento, ao desempenharem uma determinada atividade. Existe uma ampla gama de disciplinas que podem receber contribuições desse domínio. Apesar de serem concebidas por especialistas em informática e programação, outras áreas do conhecimento vislumbram seu potencial. Haykins (1999, p. 812) afirma que as Redes Neurais Artificiais constituem “um campo multidisciplinar com raízes nas neurociências, matemática, estatística, física, ciência da computação e engenharia [...]”⁵⁰

⁵⁰ *Neural networks represent a multidisciplinary subject with roots in the neurosciences, mathematics, statistics, physics, computer science, and engineering [..].*

A capacidade de generalizações que as RNAs possuem é considerada promissora pelos especialistas. Pelo método da tentativa e erro, as RNAs aprendem certas regras que foram pré-programadas e, a partir disso, executam a tomada de decisões baseadas em dados variados. Por meio da solução de problemas, as RNAs demonstram que se caracterizam pelo sucesso na capacidade de extrair e fornecer informações.

Para melhor entendimento das RNAs,

RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que computam determinadas funções matemáticas (normalmente não lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O funcionamento destas redes é inspirado em uma estrutura física concebida pela natureza: o cérebro humano. (BRAGA; CARVALHO; LUDERMIR, 1998, p. 1)

Muitas tarefas que envolvem inteligência ou reconhecimento de padrões são extremamente difíceis para o ser humano. É preciso contar com o dispêndio de inúmeras horas e paciência, além da alta possibilidade de erro. As RNAs estão à disposição para serem orientadas a fornecer previsões, classificações e tomada de decisões e é justamente para que essas características sejam possíveis que “[...] procura-se construir um computador que tenha circuitos modelando os circuitos cerebrais e espera-se ver um comportamento inteligente emergindo, aprendendo novas tarefas, errando, fazendo generalizações e descobertas [...]” (BARRETO, 2002, p. 6).

Há uma gigantesca quantidade de aplicações para as RNAs com taxas de êxito por volta de 90%. “Mesmo não sendo perfeitas nas suas previsões, as redes superam todos os outros métodos [...] permitindo que se realizem previsões sobre as estimativas de retornos com certa confiabilidade nos resultados.” (MELLO, 2004, p. 41). Um indivíduo, mesmo com auxílio de planilhas eletrônicas, não é capaz de analisar, por exemplo, todas as cotações da

bolsa de valores dos últimos anos para chegar à conclusão de que as ações de uma empresa podem ganhar uma valorização acima da média, nos próximos meses.

O reconhecimento de faces humanas é outra aplicação que beneficia os sistemas de segurança. Na engenharia de tráfego, as RNAs são utilizadas para a identificação de placas de carros. Em caso de multa, roubo ou qualquer outra finalidade, essa identificação é dinâmica. O controle de abertura e fechamento dos faróis nas ruas e avenidas também pode ser acionado tomando como responsáveis as RNAs, através da tomada de decisões, sem intervenção humana. Câmeras indicam que uma via, naquele momento, tem um fluxo mais pesado que outra. Se a via perpendicular, estiver livre ou com um número reduzido de veículos, então, terá o farol fechado nessa via e liberado na outra para que não haja congestionamento.

A compressão de dados no computador recorre a processos similares, uma vez que é necessária a transformação de dados. Um texto com milhares de palavras que precisa ser compactado dispõe de seqüência de caracteres que são repetidas ao longo do documento. Essas seqüências de duas, três ou mais caracteres são transformadas em códigos menores. Assim, é viabilizada a diminuição do tamanho original do arquivo.

Dentre as diversas utilidades, pode-se destacar:

- análise de radiografias, diagnóstico de doenças, identificação de células, simulação de funções cerebrais;
- caracterização de rochas, prospecção mineral, sensoriamento remoto;
- detecção de alvos, classificação de sinais de radar;
- efeitos especiais, animação;
- visão computacional, controle de manipuladores, análise de situações;
- simulação de processos químicos, processamento de sinal, diagnóstico de falhas;
- análise de texturas e fases. (SILVA, 2005, p. 2)

As aplicações ainda encontram ramificações como na psicologia, na detecção de ataques a redes de computadores, na assistência médica, atuando como instrumento intelectual e dedutivo.

2. METODOLOGIA DA PESQUISA

Neste capítulo, apresentamos o método de constituição do corpus de especialidade das RNAs, a comparação entre os corpora de distintas áreas do saber, a extração dos itens lexicais candidatos a termo, a seleção dos termos, a organização das fichas terminológicas e as ferramentas eletrônicas para elaboração de dicionários.

2.1 Delimitação do domínio

As RNAs pertencem ao domínio da Inteligência Artificial, que é uma área de pesquisa da ciência da computação dedicada a buscar métodos ou dispositivos computacionais que possuam ou simulem a capacidade humana de resolver problemas, pensar ou, de forma ampla, ser inteligente. Apenas recentemente, com o surgimento do computador moderno, é que a inteligência artificial pôde se estabelecer como ciência integral. Desde então, seu desenvolvimento tem extrapolado os clássicos programas de xadrez ou de conversão e envolvido áreas como visão computacional, análise e síntese da voz, lógica difusa, redes neurais artificiais e muitas outras.

2.2 Constituição do corpus

O processo de constituição do corpus teve início com a seleção dos livros, artigos, tutoriais, palestras, websites, manuais, apresentações em formato PowerPoint e textos introdutórios sobre RNAs. Todo o conteúdo textual estava em língua inglesa e foi digitalizado para que pudesse ser posteriormente analisado pelo programa computacional WordSmith Tools. Para facilitar a manipulação dos dados pelo software, todos os textos tiveram que ser convertidos para o formato “texto simples”, com a extensão txt. Nessa conversão, as fórmulas, as tabelas, as figuras e os gráficos não puderam ser aproveitados. As

legendas, no entanto, foram aproveitadas e constam no corpus. Todos os documentos selecionados abordavam algum aspecto das RNAs e, obrigatoriamente, deveriam apresentar terminologia que servisse aos propósitos da pesquisa. Após processamento do corpus pelo WST, os dados do corpus de especialidade obtidos ao longo dessa fase inicial são os seguintes: tokens: 1.457.442; types: 35.842.

Em língua portuguesa, constituímos um corpus de especialidade das RNAs com 129.083 tokens e 9.671 types. O objetivo desse corpus é comprovar a existência do termo em língua portuguesa.

2.3. Comparação entre os corpora

Sinclair (2004) compara um corpus da ciência da computação com o corpus de língua geral LOB (London/Oslo/Bergden). Ambos com um milhão de tokens. Os resultados indicam que a quantidade de types é 40% menor no texto de especialidade analisado. Isso demonstra que o corpus de especialidade possui menos variedade lexical, concentrando um maior número de termos. A fim de confirmar o postulado de Sinclair, no qual a variação vocabular é bem menor no texto especializado que no geral, comparamos nosso corpus especializado das RNAs com dois corpora de referência, CR1 e CR2. O corpus de referência CR1 constituía-se de 33 obras literárias originalmente escritas em língua inglesa e o corpus de referência CR2 apresentava diversos gêneros textuais também em língua inglesa. A utilização de dois corpora de referência pode proporcionar resultados mais confiáveis. O corpus de literatura, o CR1, abrange o repertório lexical da língua geral. Por sua vez, o corpus CR2, mais balanceado, pode confirmar ou refutar as suspeitas geradas pela análise do CR1.

Tabela 1. Número de tokens, types e type/token ratio

Corpora	Tokens	Types	Type/Token Ratio
RNAs	1.457.442	35.842	2,46
CR1	4.243.943	65.213	1,53
CR2	6.815.483	119.260	1,74

Em sua pesquisa, Sinclair (2004) faz a comparação entre um corpus da ciência da computação e um corpus de referência de língua geral. Para confirmar se qualquer corpus de especialidade caracteriza-se como sendo 40% menor em relação a um corpus de referência, procedemos com a comparação de cinco corpora com o corpus de referência CR1 e CR2. Os cinco corpora de especialidade são pertencentes a Engenharia de Alimentos (EA), Genética (Gen), Matemática (Mat), Microbiologia (Mic) e Zoologia (Zoo). Dessa forma, viabilizamos uma abordagem isenta de pressuposições.

Tabela 2: Número de tokens e types nos corpora de especialidade

Corpora	Tokens	Types	Type/Token Ratio
EA	1.467.078	45.723	3,12
Gen	2.163.003	49.494	2,29
Mat	3.158.610	41.674	1,32
Mic	2.031.679	59.035	2,91
Zoo	3.334.408	58.653	1,76

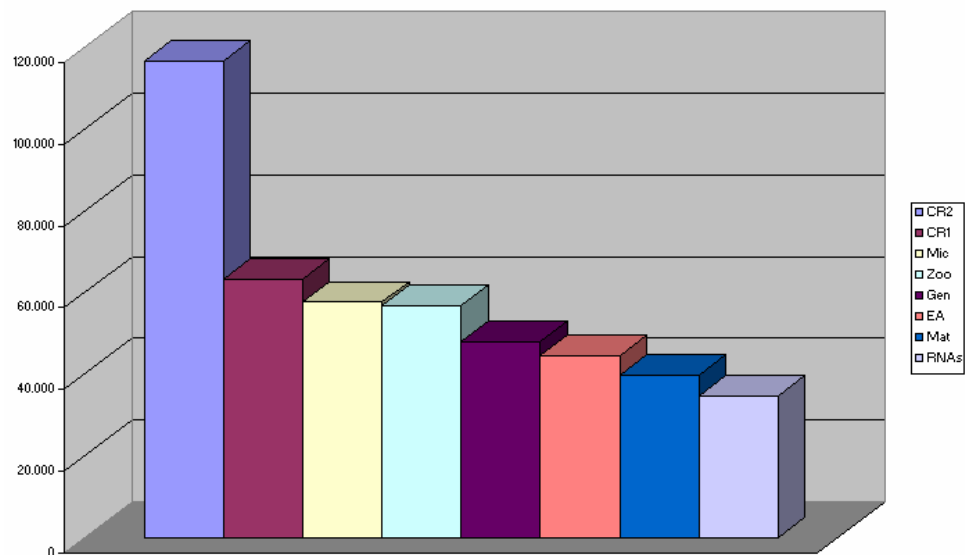
Causou estranhamento o corpus da Engenharia de Alimentos possuir uma razão type/token quase duas vezes maior que o corpus da Zoologia. Mais uma vez, o corpus de especialidade parece caracterizar-se como possuidor de maior variação lexical que os corpora de referência.

Percebemos que o tamanho dos corpora estava influenciando o cálculo da razão type/token. Desconfiamos que a razão type/token fosse deficiente na comparação de corpora com dimensões distintas. Tínhamos corpus até duas vezes maior que outros.

Gómez (2002) tece algumas considerações sobre a razão type/token e sugere que esse cálculo seja substituído pelo valor K (K-value). A fórmula para obtenção desse valor é a seguinte: $\text{type} = K\sqrt{\text{token}}$. A margem do erro amostral é de $\pm 5\%$.

Seguimos a mesma metodologia de Gómez para realizar um levantamento da densidade terminológica de corpora distintos, com tamanhos variados. Pela contagem do K-value, recorreremos aos seis corpora de especialidade, já mencionados anteriormente - Engenharia de Alimentos (EA), Genética (Gen), Matemática (Mat), Microbiologia (Mic), Zoologia (Zoo) e Redes Neurais Artificiais (RNAs). Apesar do fato de cada corpus totalizar um número diferente de tokens e types, é possível vislumbrar quais deles possuem uma densidade ou riqueza lexical mais pronunciada, independentemente do tamanho. A Figura 3, diferentemente da tabela 2, exemplifica essas especificidades ao revelar os valores K.

Figura 3: Densidade lexical de 6 linguagens de especialidade



Depreende-se que, no domínio das Redes Neurais Artificiais, a densidade terminológica é maior. Isso se explica pelo fato de as RNAs serem uma ciência lógica, da qual há maior uso de frases diretas e pouca necessidade de recursos metafóricos ou jogos de

linguagem para transmitirem a mensagem. Os corpora de referência encontram-se em posição de destaque, graças à sua riqueza lexical.

Uma análise mais acurada entre os corpora leva a crer que outros cálculos estatísticos devem ser utilizados pelos pesquisadores, quando estiverem lidando com corpora de dimensões distintas. Suas características intrínsecas podem induzir a conclusões equivocadas.

O novo cálculo permite uma nova visualização da densidade lexical dos corpora em questão. Vale salientar que, confirmando o postulado de Sinclair (2004), os corpora de especialidade, por ora analisados, são, em geral, 42,74% menores no que concerne à quantidade de types quando comparados a um dado corpus de referência. Com os resultados fornecidos pelo valor K, podemos calcular a diferença no tamanho dos corpora.

Tabela 3: Variação do corpus de especialidade em relação aos corpora de referência

Corpora	CR1	CR2
EA	29,52	61,84
Gen	23,96	58,83
Mat	36,83	65,8
Mic	8,78	50,61
Zoo	10,02	51,28
RNASs	45,16	70,3

Como nosso interesse está voltado especialmente para o corpus das RNAs, não discutiremos os demais domínios. Os dados obtidos mostram que o corpus das RNAs é 45,16% menor que o corpus de referência CR1, aproximando-se do valor apresentado por Sinclair. Por outro lado, quando o corpus das RNAs é confrontado com o corpus de referência CR2, a diferença mostra-se elevada. Levando em conta a diferença entre o corpus de referência CR1 e o corpus de referência CR2, o corpus das RNAs mostra-se, em média, 57,73% menor.

Nosso próximo passo foi a extração dos candidatos a termo.

2.4. Extração dos candidatos a termo

Após estudar uma coletânea de 40 corpora, Berber Sardinha (2005) estima que a quantidade média de palavras-chave seria da ordem de 1.472. Alguns corpora alcançam 3000 palavras-chave. Em um corpus de especialidade, as palavras-chave constituem peças fundamentais, já que são representantes do domínio ao qual pertencem. A frequência com que elas ocorrem nos textos especializados (e pouco ou raramente no corpus de referência) justifica a escolha dessas palavras pela importância das mesmas.

Dentre outras fundamentações teórico-metodológicas, baseamo-nos também em um recente estudo de Paiva e Cardoso (2007, p. 434) para nortear a manipulação e a seleção da terminologia a ser analisada.

A extração dos candidatos a termo deu-se pelo aporte da lingüística de corpus. Foi necessária a utilização de dois corpora de referência – CR1 e CR2 – para viabilizar a extração das palavras-chave candidatas a termo. A utilização de dois corpora explica-se pelo fato de proporcionarem uma maior fiabilidade e redução do erro amostral.

Em primeiro lugar, foi gerada uma lista das palavras mais frequentes do corpus de especialidade das RNAs (CE). Em seguida, a lista do corpus CR1 e CR2 foi criada.

Pode-se afirmar que os termos são palavras-chave nos textos de onde foram retirados. Dessa forma, um cruzamento da lista de palavras do CE com a lista de palavras do CR1 pode identificar quais são as ocorrências que possuem uma frequência maior no CE que no CR1. Isso indica que são importantes no CE.

Em seguida, houve a comparação da lista do CE com a lista do CR2.

A ferramenta *Keywords* do WST geralmente elenca 500 types. Fizemos um recorte para selecionar os 200 primeiros types.

Tabela 4: Primeiras 50 palavras-chaves no CR1 e CR2

CE x CR1	
1	NETWORK
2	NEURAL
3	NETWORKS
4	X
5	INPUT
6	LEARNING
7	FUNCTION
8	K
9	TRAINING
10	DATA
11	N
12	E
13	OUTPUT
14	T
15	IS
16	S
17	J
18	F
19	LAYER
20	R
21	ALGORITHM
22	W
23	USING
24	UNITS
25	WEIGHTS
26	B
27	ERROR
28	P
29	FIGURE
30	SYSTEM
31	VECTOR
32	PROBLEM
33	Y
34	MODEL
35	SYSTEMS
36	NUMBER
37	UNIT
38	CAN
39	VALUES
40	HIDDEN
41	D
42	M
43	PATTERN
44	USED
45	FUNCTIONS
46	WEIGHT
47	G
48	EQUATION
49	ALGORITHMS
50	C

CE x CR2	
1	NEURAL
2	NETWORK
3	NETWORKS
4	INPUT
5	TRAINING
6	OUTPUT
7	LEARNING
8	FUNCTION
9	K
10	WEIGHTS
11	ALGORITHM
12	LAYER
13	ERROR
14	HIDDEN
15	UNITS
16	DATA
17	T
18	X
19	FIGURE
20	FUZZY
21	ALGORITHMS
22	VECTOR
23	LINEAR
24	RECURRENT
25	VALUES
26	NEURON
27	EQUATION
28	MODEL
29	PATTERN
30	INPUTS
31	SET
32	MATRIX
33	W
34	GRADIENT
35	UNIT
36	NEURONS
37	NONLINEAR
38	WEIGHT
39	OPTIMIZATION
40	BACKPROPAGATION
41	MODELS
42	NODE
43	ROJAS
44	USING
45	MULTILAYER
46	SPRINGER
47	PROBLEM
48	TRAINED
49	NET
50	PROBABILITY

As duas listas resultantes apresentaram disparidade com relação à ordem de importância das palavras-chave. Decidimos selecionar aquelas que apareceram em ambas as listas de palavras-chave. Obtivemos 153 types em comum nas duas listas de palavras-chave. Os demais types só ocorreram uma vez em cada lista.

Acreditamos que para a elaboração do dicionário eletrônico terminológico onomasiológico bilíngüe das RNAs, 30 termos são suficientes para testar sua funcionalidade.

Optamos por analisar os candidatos a termos pela ordem de frequência. Percebemos que, entre as dez primeiras palavras, no cruzamento do CE com o CR1, excluindo-se as duas letras, temos 7 substantivos. No cruzamento do CE com o CR2, observamos 8 substantivos. O primeiro substantivo candidato a termo encontrado como freqüente foi “network”. Para certificarmos se “network” aparecia isoladamente ou acompanhado, geramos a concordância pelo concordanceador do WST. Logo depois, selecionamos a opção “clusters”. Esse recurso desempenha o trabalho exaustivo de leitura de todas as ocorrências de “network”, para somente então iniciar a contagem das palavras que comumente apresentam associação. Observamos os clusters, que são os aglomerados de palavras, ou melhor, as palavras que ocorrem juntas.

Foram selecionados os candidatos a termo que ocorreram 30 ou mais vezes no corpus. Grande parte deles era de termos compostos. Se são aglomerados de palavras comuns nesse domínio, possivelmente caracterizam-se como termos.

Conseqüentemente, selecionamos “neural network” como um dos candidatos a termo.

2.5. Seleção dos termos

Com a ajuda de dois profissionais pesquisadores do curso de Ciência da Computação do Instituto de Biociências Letras e Ciências Exatas da Universidade Estadual Paulista e outro em Engenharia Elétrica do mesmo Instituto, foi possível a seleção dos termos que entrariam

no dicionário eletrônico terminológico. Cada especialista recebeu separadamente uma lista com 30 itens candidatos a termo. Foi solicitado que indicassem quais realmente faziam parte do vocabulário fundamental das RNASs, de acordo com sua experiência. Os itens mais escolhidos foram separados para serem incluídos no dicionário.

Somente após a análise lingüístico-terminológica e a validação dos especialistas, os candidatos a termo receberam o referido estatuto. Ao término dessa etapa, chegamos ao número de 30 termos propriamente ditos.

2.6. A organização das fichas terminológicas

Em posse dos termos fundamentais e do sistema conceptual das RNASs, iniciamos a preparação das fichas terminológicas. As fichas terminológicas foram armazenadas em um banco de dados a partir do aplicativo MS Access XP.

Por tratar-se de um programa integrante do pacote Microsoft Office Professional, é um sistema acessível à maioria dos pesquisadores lingüísticos. O MS Access pode ser descrito como sendo um sistema relacional de administração de banco de dados, que prescinde de conhecimentos avançados em programação. O resultado final do trabalho nesse aplicativo pode ser exportado, disponibilizado em rede intranet ou publicado em meio virtual.

O suporte eletrônico torna a manipulação das informações mais ágeis e menos propensas a erros. Outrossim, a investigação e o acesso aos dados em uma base computacional são vantagens para os consulentes por duas razões: gerenciamento e compartilhamento de informações.

A elaboração das fichas terminológicas eletrônicas tomou por base o modelo proposto por Aubert (1996).

Um de nossos propósitos era tornar a consulta amigável ao usuário. Foi por esse motivo que todos os campos de busca estão presentes na mesma tela: na parte superior, os

campos com informações em língua inglesa e, na parte inferior, os mesmos campos com informações em língua portuguesa.

Cada ficha foi formatada de modo a apresentar 11 campos. São eles: símbolo, área, subárea, termo, categoria gramatical, outras designações, contexto, definição, semema e fonte. Tanto a ficha terminológica em inglês quanto aquela em português possuem o mesmo número de campos.

Tabela 5: organização dos campos do modelo de dicionário onomasiológico

símbolo	O primeiro dos campos é o símbolo. Aqui, apresenta-se a numeração que o termo ocupa no sistema conceptual. Pela observação desse símbolo, compreende-se a relação que os termos mantêm entre si. A noção de hipônimo e hiperônimo pode apresentar-se mais clara com esse recurso.
área	O campo seguinte é denominado área. Esse campo será sempre preenchido como “Inteligência Artificial”. Ao lado do campo área encontra-se a subárea, que, nesse caso, é a das Redes Neurais Artificiais.
termo	O campo destinado ao termo permite a busca inteligente e interativa. Quando o consulente digita o termo “network”, vai receber como resposta uma gama de opções com todos os termos que trazem “network” em sua composição. A título de exemplificação, dois termos poderiam ser de interesse nessa consulta, tais como “artificial neural network” e “Hopfield network”. Na seqüência, ao selecionar qualquer termo retornado da consulta, pode-se visualizar os respectivos campos com as devidas

	informações terminológicas. Caso a procura seja realizada em língua portuguesa, o sistema de busca funciona da mesma forma, sempre fornecendo opções e posteriormente os detalhes procurados.
outras designações	A variante terminológica, se houver, será apresentada nesse campo. Quando houver uma sigla, ela também será incluída no mesmo espaço.
categoria gramatical	No campo categoria gramatical, em língua portuguesa, observa-se se o termo é substantivo masculino ou feminino. No mesmo campo, em língua inglesa, há a indicação da categoria do termo, se é substantivo contável ou incontável.
contexto	O campo contexto traz exemplos reais de uso. Em língua portuguesa, serão apresentados excertos de produções de especialistas, material introdutório, guias, tutoriais, artigos científicos, textos traduzidos, etc. Além da consulta ao corpus de especialidade das RNAs, utilizamos basicamente duas obras de referência na área: <i>Redes Neurais: princípios e prática</i> , de Simon Haykins (1999) (em língua inglesa e sua respectiva tradução) e <i>Redes Neurais em Delphi</i> , de Luciano Medeiros (2003). Em língua inglesa, a consulta foi feita quase que exclusivamente com o corpus de especialidade das RNAs em língua inglesa.
definição	No campo definição, inserimos e, muitas vezes, adaptamos, a definição encontrada no próprio corpus ou nas obras em papel. A

	<p>consulta ao <i>Cambridge Advanced Learner's Dictionary</i>, em sua versão eletrônica, tornou-se indispensável, quando não havia definições dentro do corpus. Isso ocorria, principalmente, quando o termo era oriundo da língua geral. Por exemplo, não encontramos a definição de “input” no corpus de especialidade das RNAs. Foi necessária a consulta e o uso da definição presente no <i>Cambridge Advanced Learner's Dictionary</i>.</p>
<p>semema</p>	<p>O campo semema traz os traços semânticos que permitem a busca onomasiológica propriamente dita. Para se saber qual é a rede neural mais popular, basta digitar <most popular neural network>, no campo semema (em língua inglesa). Ou ainda, <rede neural mais popular>, na ficha terminológica em língua portuguesa. Se existir tal rede, o consulente vai ter acesso ao termo procurado.</p> <p>No preenchimento do campo relacionado ao termo “Kalman filter”, encontram-se os seguintes traços semânticos: <type of training> <supervised training> <Kalman filter> <Kalman> <filter> <extended Kalman filter> <algorithms> <to locally linearize the nonlinear system> <to linearize the nonlinear system> <that linearizes the nonlinear system> <current state estimate> <state estimate> <estimate> <linearized system> <linearized> <system> <approximate state distribution> <recurrent network> <state distribution> <distribution> <Gaussian> <perceptron> <multilayer perceptron> <weights> <node>.</p> <p>No preenchimento do campo relacionado ao termo “extended</p>

	<p>Kalman filter”, que é hipônimo de “Kalman filter”, podem ser visualizados todos os descritores pertencentes ao hiperônimo mais aqueles pertencentes ao hipônimo: <type of training> <supervised training> <Kalman filter> <Kalman> <filter> <extended Kalman filter> <algorithms> <to locally linearize the nonlinear system> <to linearize the nonlinear system> <that linearizes the nonlinear system> <current state estimate> <state estimate> <estimate> <linearized system> <linearized> <system> <approximate state distribution> <recurrent network> <state distribution> <distribution> <Gaussian> <perceptron> <multilayer perceptron> <weights> <node> <minimum mean-square> <estimator> <linear dynamical system> <digital computer> <recursive structure> <recursive> <structure> <input> <training> <iteration> <learning process> <learning> <process>.</p>
--	---

A seguir, a ficha terminológica eletrônica, que apresenta a seguinte disposição dos campos.

Figura 4: Ficha terminológica em inglês do MS Access XP

symbol:	1.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	
term:	other designations:	
artificial neural network	ANN	
context:		
A neural network is a circuit composed of a very large number of simple processing elements that are neurally based.		
definition:		
According to Zurada (1992), Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge.		
sememe		
<type of network> <artificial neural network> <network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <input layer> <ANN>		
source:		
corpus de especialidade das RNAs		

Conquanto a consulta tradicional termo-conceito continue disponível, o dinamismo e a inovação dessa ficha terminológica eletrônica caracteriza-se pela possibilidade de consulta reversa, conceito-termo. Ou seja, a partir do conhecimento de alguns traços semânticos (descritores) do termo chega-se ao termo ou termos relacionados.

Na ficha terminológica eletrônica, os descritores funcionam bem para detalhar o conceitos. Dubuc (1985) define os descritores como elementos reveladores do conceito contido no contexto.

Durante a consulta à ficha terminológica eletrônica, grande parte do sucesso da busca pelo termo depende dos hipônimos e dos hiperônimos. Diz-se que são hipônimos os vocábulos com sentido mais específico em relação ao de um outro mais geral. Por exemplo,

“filtro de Kalman estendido” é hipônimo de “filtro de Kalman”. “Filtro de Kalman” é hiperônimo de “filtro de Kalman estendido”

Para enriquecer a busca no MS Access, incluímos cinco fichas remissivas no dicionário eletrônico terminológico onomasiológico das RNAs. No caso das abreviações, as fichas remissivas podem guiar o consulente diretamente ao termo desejado. Isso faz com que a consulta se torne mais dinâmica. Das cinco abreviações selecionadas, três são usadas pelos especialistas indistintamente, seja em língua inglesa ou portuguesa: TDNN, HMM e RBF. As abreviações em inglês ANN e EKF são modificadas para atender a sintaxe da língua portuguesa, a saber RNA e FKE. No modelo de dicionário eletrônico terminológico onomasiológico das RNAs, essa busca pode ser feita diretamente pelo termo ou pelo semema.

2.7. Ferramentas eletrônicas para elaboração de dicionários

A tecnologia disponível atualmente possibilita a elaboração de dicionários por meio de uma série de aplicativos. Decidimos apresentar alguns dos que apresentam resultados satisfatórios nos estudos lingüísticos: dois gerenciadores de banco de dados – CDS/ISIS e o MySQL – e uma linguagem de programação, denominada PHP.

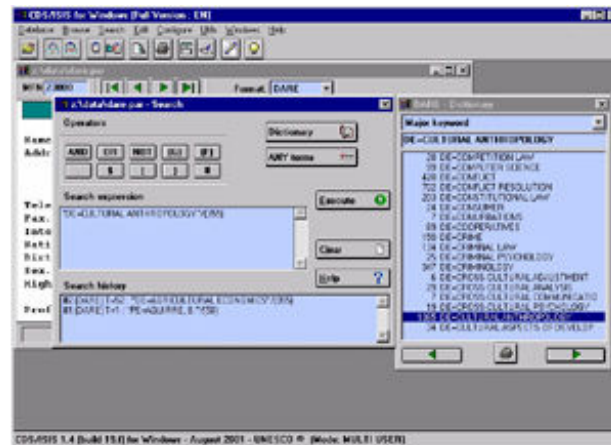
2.7.1 A elaboração do dicionário eletrônico no suporte CDS/ISIS

Para a aplicação do modelo de dicionário terminológico onomasiológico bilíngüe proposto por Babini (2001), pode-se recorrer a um aplicativo para o sistema operacional Windows chamado CDS/ISIS – também conhecido como Winisis. Trata-se de um aplicativo gratuito distribuído pela Unesco.

O Winisis permite o desenvolvimento e o gerenciamento de banco de dados não-numéricos, ou seja, o aplicativo requer um banco de dados textual. Seu ponto de destaque é a possibilidade de inserção de campos com tamanhos variados, conseqüentemente permitindo a

otimização de cada campo de entrada de informação. Após a criação de cada entrada, pode-se modificar, corrigir e deletar os arquivos.

Figura 5: Tela do Winisis



Para a viabilização do dicionário onomasiológico terminológico bilíngüe das RNAs, tal aplicativo dispõe de um sofisticado sistema de busca, sendo capaz de recuperar dados pelo conteúdo.

Ranzani (2006) desenvolveu um dicionário terminológico onomasiológico nesse suporte eletrônico. Os resultados obtidos foram satisfatórios.

A fim de avaliar uma plataforma distinta, optamos por desenvolver nosso dicionário terminológico onomasiológico com uma outra interface, testando, conseqüentemente, sua viabilidade para os estudos terminológicos.

2.7.2 A elaboração do dicionário eletrônico com o PHP

O PHP, Hypertext Preprocessor, também é uma linguagem de programação de código aberto que pode ser obtido gratuitamente do site do fabricante. É uma linguagem embutida no código do html para tornar as páginas mais dinâmicas. Desse modo, o PHP é responsável por dinamizar as páginas na Internet, ou seja, possibilita o desenvolvimento de sites realmente dinâmicos. Dominando essa linguagem, pode-se transformar sites estáticos, feitos em html

puro, em sites interativos, utilizando todas as técnicas de programação que essa linguagem oferece. Além disso, o PHP é uma linguagem de fácil utilização com diversos bancos de dados. Ainda que o PHP possibilite que uma página web tenha fácil acesso a determinado banco de dados, faz-se necessária certa habilidade na elaboração dos comandos próprios dessa linguagem. A possibilidade de adaptar-se às necessidades do pesquisador mostra-se como uma vantagem. Para a aplicação do modelo de dicionário terminológico onomasiológico bilíngüe proposto por Babini (2001), recorreremos a essa linguagem de programação, além de associá-la a um gerenciador de banco de dados, chamado MySQL.

Conforme demonstrado na Figura 6, a página inicial desenvolvida em html puro foi dividida em duas colunas: a primeira permite a consulta diretamente em língua inglesa e a segunda, em língua portuguesa. Na primeira linha de entrada de dados, faz-se a consulta ao termo. Ao passo que, na segunda linha de entrada de dados, pode-se pesquisar a partir do semema.

Caso o usuário efetue a entrada de dados de forma errônea e deseje limpar os campos de busca, basta pressionar o botão “Clean /Limpar” para que esse procedimento elimine todos os caracteres digitados em qualquer dos campos.

O botão “Search/Busca” desempenha a tarefa de chamar outra página que fará a conexão com o banco de dados e efetuará a procura do termo ou semema nesse banco. Quando o resultado da busca é positivo, o usuário escolhe o termo ou semema pesquisado.

Figura 6: Página desenvolvida em html puro



The image shows a web page titled "DICTIONARY DEMO / DICIONÁRIO DEMO". It features a search interface with two columns of input fields. The left column has fields labeled "Term:" and "Sememe:". The right column has fields labeled "Termo:" and "Semema:". Below the input fields, there are two buttons: "Clean / Limpar" and "Search / Procurar". The entire page has a solid blue background.

A Figura 7 representa a ficha terminológica desenvolvida na linguagem de programação PHP. Além de aceitar alterações no estilo da fonte, na cor do fundo de tela, no tamanho dos campos, o layout da página pode ser alterado a qualquer momento. O botão “Nova Consulta” leva o consulente à primeira página de consulta, já apresentada na Figura 6.

A página com os dados terminológicos baseada em PHP é melhor visualizada pelo navegador Mozilla Firefox, versão 3.0.9 ou superior. Para tanto, a tecla F11 do teclado deve ser pressionada, e, dessa forma, toda a página será expandida para ocupar a tela inteira.

Figura 7: Página desenvolvida em PHP

TERMINOLOGICAL FORM / FICHA TERMINOLÓGICA

Símbolo: <input type="text"/> Área: <input type="text"/> Subárea: <input type="text"/> Term: <input type="text"/> Noun: <input type="text"/> Other Designations: <input type="text"/> Context: <input type="text"/> <input type="radio"/> Haykins <input type="radio"/> Corpus (Port.) <input type="radio"/> Corpus (Eng.) <input type="radio"/> Medeiros <input type="radio"/> Babini/Marranghello <input type="radio"/> Cambridge Definition: <input type="text"/> <input type="radio"/> Haykins <input type="radio"/> Corpus (Port.) <input type="radio"/> Corpus (Eng.) <input type="radio"/> Medeiros <input type="radio"/> Babini/Marranghello <input type="radio"/> Cambridge Sememe: <input type="text"/>	Área: <input type="text"/> Subárea: <input type="text"/> Termo: <input type="text"/> Categoria Gramatical: <input type="text"/> Outras Designações: <input type="text"/> Contexto: <input type="text"/> <input type="radio"/> Haykins <input type="radio"/> Corpus (Port.) <input type="radio"/> Corpus (Eng.) <input type="radio"/> Medeiros <input type="radio"/> Babini/Marranghello <input type="radio"/> Cambridge Definição: <input type="text"/> <input type="radio"/> Haykins <input type="radio"/> Corpus (Port.) <input type="radio"/> Corpus (Eng.) <input type="radio"/> Medeiros <input type="radio"/> Babini/Marranghello <input type="radio"/> Cambridge Semema: <input type="text"/>
--	--

2.7.3 O gerenciador de banco de dados MySQL

O MySQL é um sistema de gerenciamento de banco de dados gratuito que utiliza a linguagem SQL (Structured Query Language) como interface. O banco de dados pode ser definido como uma coleção de dados estruturados. O MySQL é atualmente um dos bancos de dados mais populares do mundo. Com ele, para adicionar, acessar e processar dados armazenados em um banco de dados de um computador, é preciso um sistema de gerenciamento de bancos de dados como o Servidor MySQL. Como os computadores são ágeis para lidar com grandes quantidades de dados, o gerenciamento de bancos de dados funciona como um recurso importante na elaboração de um dicionário, por exemplo.

O MySQL é um banco de dados relacional que armazena dados em tabelas separadas, em vez de colocar todos os dados em um só local. Isso proporciona velocidade e flexibilidade.

O servidor de banco de dados MySQL é rápido, confiável e fácil de usar. Sua conectividade, velocidade e segurança fazem com que o MySQL seja adaptável para acessar bancos de dados na Internet.

3. O DICIONÁRIO ONOMASIOLÓGICO

Nesse capítulo, apresentaremos o dicionário onomasiológico eletrônico baseado nas fichas terminológicas; a seguir, a organização do sistema nocional elaborada a partir da frequência dos termos mais comuns. No subcapítulo 4.1, apresentaremos o sistema nocional em língua inglesa, e no subcapítulo 4.2, o sistema nocional em língua inglesa. As fichas terminológicas eletrônicas dos termos fundamentais das RNAs são apresentadas no subcapítulo 4.3.

3.1 O sistema nocional em língua inglesa

1 TYPE OF NETWORK

- 1.1. artificial neural network
 - 1.1.1. feedforward neural network
 - 1.1.2. backpropagation model
 - 1.1.3. multilayer neural network
 - 1.1.4. time delay neural network
- 1.2. hidden Markov models
- 1.3. Hopfield network (Hopfield model)

2 NEURON ORGANIZATION

- 2. 1. hidden layer
 - 2. 1.1. single hidden layer
 - 2.1.2. unidade oculta
- 2. 2. input
 - 2. 2. 1. input layer
 - 2. 2. 2. input pattern
 - 2.2.3. input signal
- 2. 3. node
- 2. 4. perceptron

3 ALGORITHMS AND FUNCTIONS

- 3.1. mean squared error
- 3.2. prediction error
- 3.3. local minimum
- 3.4. global minimum
- 3.5. error function
- 3.6. activation function
- 3.7. fuzzy logic
- 3.8. linear combination
- 3.9. learning rate
- 3.10. nonlinear dynamic system
- 3.11. threshold function

3.12. radial basis function

4 TYPE OF TRAINING

4.1. SUPERVISED TRAINING

4. 1.1. Kalman filter

4.1.1.1. extended Kalman filter

4.1. 2. perceptron learning rule

4.2 UNSUPERVISED TRAINING

4.2.1. competitive learning

5 PROCESS

5.1. adaptive resonance theory

5.2. data mining

3.2. O sistema nocional em língua portuguesa

1 TIPO DE REDE

1.1. rede neural artificial

1.1.1. rede neural alimentada adiante

1.1.2. modelo backpropagation

1.1.3. rede neural multicamadas

1.1.4. rede neural com atraso de tempo

1.2. modelo oculto de Markov

1.3. rede de Hopfield

2 ORGANIZAÇÃO DO NEURÔNIO

2. 1. camada oculta

2. 1.1. camada oculta única

2.1.2. unidade oculta

2. 2. entrada

2. 2. 1. camada de entrada

2. 2. 2. padrão de entrada

2. 2. 3. sinal de entrada

2. 3. nó

2. 4. perceptron

3 ALGORITMOS E FUNÇÕES

3.1. erro quadrado médio

3.2. predição de erro

3.3. mínimo local

3.4. mínimo global

3.5. função de erro

3.6. função de ativação

3.7. lógica fuzzy

3.8. combinação linear

3.9. taxa de aprendizado

3.10. sistema dinâmico não-linear

- 3.11. função limiar
- 3.12. função de base radial

4 TIPO DE TREINAMENTO

- 4.1. TREINAMENTO SUPERVISIONADO
 - 4.1.1. filtro de Kalman
 - 4.1.1.1. filtro de Kalman estendido
 - 4.1.2. regra de aprendizado do perceptron
- 4.2. TREINAMENTO NÃO-SUPERVISIONADO
 - 4.2.1. aprendizado competitivo

5 PROCESSO

- 6.1. teoria da ressonância adaptiva
- 6.2. mineração de dados

3.3 O dicionário terminológico onomasiológico das RNAs

Pela facilidade de impressão, as fichas terminológicas dos termos fundamentais das RNAs foram incluídas com o mesmo layout fornecido pela opção relatório do MS Access. Na parte superior, encontram-se informações terminológicas em língua inglesa e, na parte inferior, as informações terminológicas em língua portuguesa. No entanto, a fim de demonstrar o funcionamento do modelo de dicionário eletrônico, todas as fichas foram exportadas para o gerenciador de banco de dados MySQL. Nessa versão eletrônica, todas as informações terminológicas em língua portuguesa foram inseridas do lado esquerdo da ficha terminológica. Do lado direito, encontram-se as informações em língua inglesa.

symbol:	1.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	
term:	other designations:	
artificial neural network	ANN	
context:		
A neural network is a circuit composed of a very large number of simple processing elements that are neurally based.		
definition:		
According to Zurada (1992), Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge.		
sememe		
<type of network> <artificial neural network> <network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <input layer> <ANN>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
rede neural artificial		
contexto:		
A rede neural se assemelha ao cérebro porque o conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.		
definição:		
É uma máquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou função de interesse.		
semema:		
<tipo de rede> <redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistemas neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <RNA>		
fonte:		
corpus de especialidade das RNAs		

symbol:	1.1.2	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
feedforward neural network	FNN	
context:		
Use a feedforward neural network having 20 hyperbolic tangent activation (set = 1) units in its hidden layer, feeding into a linear output unit.		
definition:		
The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.		
sememe		
<type of network> <artificial neural network> <network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <input layer> <feedforward neural network> <feedforward> <type of neural network> <tangent activation units> <activation units> <hidden layer> <layer> <linear output unit> <linear output> <unit> <simplest type of artificial neural network> <simplest> <type of artificial network> <information moves in one direction> <one direction> <input nodes> <hidden nodes> <output nodes> <no cycles> <no loops> <network> <one way> <different layers> <FNN>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
rede neural feedforward		
contexto:		
Nas redes alimentadas adiante (feedforward), os neurônios estão dispostos em camadas conectadas por pesos unidirecionais na direção entrada-saída, ou seja, as conexões ocorrem apenas entre camadas diferentes e subsequentes. (2)		
definição:		
Uma rede em camadas com neurônios organizados. Na forma mais simples, de uma rede em camadas, temos uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de neurônios, mas não vice-versa. (1)		
semema:		
<tipo de rede> <redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistema neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <rede alimentada adiante> <rede neural feedforward> <feedforward> <tipo de rede> <rede neural> <unidades de ativação tangentes> <unidades de ativação> <camada oculta> <camada> <unidade de saída linear> <saída linear> <unidade> <tipo mais simples de rede neural> <mais simples> <tipo de rede artificial> <informação se move em uma direção> <uma direção>		
fonte:		
Haykins; corpus de especialidade das RNAs		

symbol:	1.1.3	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
backpropagation model		
context:		
Among them, the backpropagation model has been most widely used in terms of practical applications.		
definition:		
Backpropagation minimizes output error in a feed-forward network. The major application categories of this model are classification, prediction, and control.		
sememe		
<type of network> <artificial neural network> <network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <input layer> <backpropagation model> <backpropagation> <type of neural network> <practical applications> <feed-forward network> <classification> <prediction> <control> <pattern classification> <pattern> <training> <supervised learning> <multiple layers> <error> <hidden layers> <character recognition> <recognition>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
modelo backpropagation		
contexto:		
O perceptron teve sua validade provada no modelo backpropagation , que possibilitou a implementação da terceira camada necessária para o aprendizado do XOR.		
definição:		
Este modelo é utilizado, principalmente, para classificação de padrões, tais como em aplicações de reconhecimento de caracteres, reconhecimento de sinais em geral, treinamento de robôs e composição musical. Também chamado de retropropagação, pois consiste em um modelo de aprendizado supervisionado de múltiplas camadas, cujo erro obtido na saída é transferido às camadas escondidas, caracterizando-se como uma rede acíclica.		
semema:		
<tipo de rede> <redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistema neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <modelo backpropagation> <modelo de retropropagação> <tipo de rede neural> <backpropagation> <aplicações práticas> <rede alimentada adiante> <classificação> <predição> <controle> <classificação de padrões> <padrões> <treinamento> <aprendizado supervisionado> <camadas múltiplas> <erro> <camada oculta> <reconhecimento de caracteres>		
fonte:		
corpus de especialidade das RNAs		

symbol:	1.1.4	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
multilayer neural network		
context:		
The neural chip structure corresponds to the results of structure and adjustment algorithm development of multilayer neural networks (in the case of commonly used neural computers) and neural network solution algorithms (in the case of problem-oriented and special-purpose neural computers).		
definition:		
Multilayer neural networks are undoubtedly the most popular networks used in applications.		
sememe		
<artificial neural network> <network> <type of network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <knowledge> <input layer> <output layer> <multilayer neural network> <type of neural network> <multilayer> <multilayer perceptron> <knowledge> <most popular network> <most popular>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
rede neural multicamada		
contexto:		
As redes multicamadas podem usar um grande número de técnicas de aprendizado, sendo que a mais popular é a propagação reversa.		
definição:		
As redes neurais multicamadas são arquiteturas onde os neurônios são organizados em duas ou mais camadas de processamento, já que sempre vão existir pelo menos uma camada de entrada e uma camada de saída. Estas arquiteturas são as mais frequentemente encontradas na literatura referente a redes neurais artificiais.		
semema:		
<redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistemas neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <camada de saída> <redes mais populares> <mais populares> <rede mais popular> <mais popular>		
fonte:		
corpus de especialidade das RNAs		

symbol:	1.1.5	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
time delay neural network	TDNN	
context:		
MLPs can also be architecturally constrained in various ways, for instance by limiting their connectivity to geometrically local areas, or by limiting the values of the weights, or tying different weights together. One type of constrained MLP which is especially relevant to this thesis is the Time Delay Neural Network.		
definition:		
The time-delay neural network (TDNN) model combined with a genetic algorithm is proposed and used to predict the nonlinear relationship and to analyze specific characteristics.		
sememe		
<artificial neural network> <network> <type of network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <TDNN> <to predict the nonlinear relationship> <time delay> <time> <delay> <time processing> <feedforward> <hidden neurons> <hidden> <neurons>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
rede neural de atrasos de tempo		
contexto:		
Na avaliação de desempenho envolvendo o uso de dados de teste de três locutores, a TDNN alcançou um escore médio de reconhecimento de 98,5 por cento.		
definição:		
Utiliza atrasos de tempo para realizar processamento temporal. É uma rede alimentada adiante cujos neurônios ocultos e neurônios de saída são replicados através do tempo (1)		
semema:		
<redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistemas neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <prever a relação não-linear> <atraso de tempo> <tempo> <atraso> <processamento temporal> <rede feedforward> <neurônios ocultos> <oculto> <neurônios> <TDNN>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
1.2		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
hidden Markov models	HMM	
context:		
<p>Researchers have become rather pragmatic and combine the best features of neural modeling with traditional algorithms or with other statistical approaches, like Hidden Markov Models, which we will briefly review.</p>		
definition:		
<p>Hidden Markov Models (HMM) form an important special type of recurrent network. Any system capable of assuming one of n different states at time t.</p>		
sememe		
<p><type of network> <special type of recurrent network><type of recurrent network> <recurrent network> <network> <system> <different states> <assume different states> <that can assume different states> <system capable of> <neural modeling> <traditional algorithms> <Hidden Markov Models> <stochastic process> <Markov chain> <output symbols> <probabilistic> <all possible symbols> <probabilistic output symbols> <HMM></p>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
modelos ocultos de Markov		
contexto:		
<p>Basicamente, um modelo oculto de Markov representa um processo estocástico gerado por uma cadeia de Markov subjacente, e um conjunto de distribuições de observações associadas com seus estados ocultos.</p>		
definição:		
<p>Um modelo oculto de Markov é similar a uma cadeia de Markov porque em ambos os casos a transição de um estado para outro é probabilística. No modelo oculto de Markov, os símbolos de saída são também probabilísticos, resultando que, em cada estado, todos os símbolos são possíveis.</p>		
semema:		
<p><tipo de rede> <tipo especial de rede recorrente> <tipo de rede recorrente> <rede recorrente> <redes> <sistema> <estados diferentes> <assumir estados diferentes> <que pode assumir estados diferentes> <sistema capaz de> <modelagem neural> <algoritmo tradicional> <modelos ocultos de Markov> <processo estocástico> <cadeia de Markov> <símbolos de saída> <probabilístico> <todos os símbolos possíveis> <símbolos de saída probabilísticos> <HMM></p>		
fonte:		
corpus de especialidade das RNAs		

symbol:	1.3	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
Hopfield model	~ network	
context:		
The Hopfield Type Network is a multiple-loop feedback neural computation system. The neurons in this network are connected to all other neurons except to themselves, that is there are no self-feedbacks in the network.		
definition:		
The Hopfield network can be considered a data structure that makes possible the verification of the found solution. The Hopfield network can be considered a data structure that makes possible the verification of the found solution.		
sememe		
<Hopfield model> <Hopfield> <model> <type of network> <network> <data structure> <verification of the found solution> <verification> <found solution> <continuous mode> <discreet mode> <neurons> <set of neurons> <delay units> <multiple loops> <multiple-loop feedback neural computation system> <neural computation system> <computational system> <neurons> <network> <connected to all other neurons> <connected> <self-feedbacks>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
modelo de Hopfield	rede de ~	
contexto:		
A rede de Hopfield pode ser operada em um modo contínuo ou em um modo discreto, dependendo do modelo adotado para descrever os neurônios.		
definição:		
A rede (modelo) de Hopfield consiste de um conjunto de neurônios e um conjunto correspondente de atrasos unitários, formando um sistema realimentado de múltiplos laços.		
semema:		
<modelo de Hopfield> <Hopfield> <modelo> <tipo de rede> <rede> <estrutura de dados> <verificação da solução encontrada> <verificação> <solução encontrada> <modo contínuo> <modo discreto> <neurônios> <conjunto de neurônios> <atrasos unitários> <laços múltiplos> <sistema computacional realimentado de múltiplos laços> <sistema computacional neural> <sistema computacional> <neurônios> <rede> <conectado a outros neurônios> <conectado> <auto alimentação>		
fonte:		
corpus de especialidade das RNAs		

symbol:	2.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
hidden layer		
context:		
The output of the hidden layer consists therefore of normalized numbers.		
definition:		
The hidden layer learns to recode (or to provide a representation for) the inputs.		
sememe		
<neuron organization> <output> <neuron organization> <output of the hidden layer> <normalized numbers> <to recode> <that recodes> <representation for the inputs> <representation> <inputs> <processing> <characteristic extractors> <pattern recognition> <external input>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
camada oculta		
contexto:		
Camadas Ocultas: onde é feita a maior parte do processamento, através das conexões ponderadas; podem ser consideradas como extratoras de características. (2)		
definição:		
A função da camada oculta é intervir entre a entrada externa e a saída da rede de maneira útil (1)		
semema:		
<organização do neurônio> <saída> <saída da camada oculta> <números normalizados> <para recodificar> <que recodifica> <representação das entradas> <representação> <entradas> <processamento> <extratores de características> <reconhecimento de padrões> <entrada externa>		
fonte:		
Haykins; corpus de especialidade das RNAs		

symbol:	2.1.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
single hidden layer	SHL	
context:		
When the size of the training set is much larger than the dimension of the input exemplars, then networks with two or more hidden layers may require substantially fewer units than networks with a single hidden layer.		
definition:		
The single hidden layer learns to recode (or to provide a representation for) the inputs.		
sememe		
<output> <neuron organization> <output of the hidden layer> <normalized numbers> <to recode> <that recodes> <representation for the inputs> <representation> <inputs> <processing> <characteristic extractors> <pattern recognition> <external input> <neural networks> <neuron organization> <networks> <single hidden layer> <hidden layers> <to recode> <that recodes> <representation> <perceptron> <computational node> <node> <global perspective> <feedforward network> <decision regions> <input> <backpropagation> <SHL>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
camada oculta única		
contexto:		
O teorema de aproximação universal dita que toda função contínua que mapeia intervalos de números reais de entrada a algum intervalo de números reais de saída pode ser arbitrariamente aproximada com precisão por um perceptron multicamadas com camada oculta única. (2)		
definição:		
A camada oculta única possui nó computacional. Com ela, a rede adquire uma perspectiva global apesar de sua conectividade. (1)		
semema:		
<saída> <saída da camada oculta> <números normalizados> <to recodificar> <que recodifica> <representação das entradas> <representação> <entradas> <processamento> <extratores de características> <reconhecimento de padrões> <entrada externa> <redes neurais> <redes> <camada oculta única> <camada oculta> <recodificar> <que recodifica> <representação> <perceptron> <nó computacional> <nó> <perspectiva global> <rede alimentada adiante> <rede feedforward> <regiões de decisão> <entrada> <retropropagação>		
fonte:		
Haykins; corpus de especialidade das RNAs		

symbol:	2.1.2	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
hidden unit	hidden nod	
context:		
If you do not provide enough hidden units, the neural network will not have the computational power to learn some complex nonlinear functions.		
definition:		
neurons, nodes		
sememe		
<neuron organization> <hidden unit> <hidden> <unit> <units> <hidden node> <hidden nodes> <linear functions> <neuron> <node> <neurons> <nodes> <value> <input> <output>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
unidade oculta	nó oculto	
contexto:		
Uma unidade de entrada envia seu valor de ativação para cada uma das unidades ocultas com as quais está conectada. Cada unidade oculta calcula seu próprio valor de ativação, dependendo dos valores de ativação que recebe das unidades de entrada.		
definição:		
Neurônios montados em um padrão de conexões dentro de uma rede neural.		
semema:		
<organização do neurônio> <unidade oculta> <oculto> <oculta> <unidade> <unidades> <nó oculto> <nós ocultos> <funções lineares> <neurônio> <nó> <neurônios> <nós> <valor> <entrada> <saída>		
fonte:		
corpus de especialidade das RNAs		

symbol:	2.2	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
input		
context:		
The first component of our system is a filter that receives as input a 20x20 pixel region of the image, and generates an output ranging from 1 to -1, signifying the presence or absence of a face, respectively.		
definition:		
information that is put into a system or machine so that it can operate.		
sememe		
<neuron organization> <information that is put into a system> <neuron organization> <information that is put into a machine> <information> <system> <machine> <neuron> <set of values> <values>		
source:		
corpus de especialidade das RNAs; Cambridge		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
entrada		
contexto:		
A função básica de um neurônio é somar as entradas e retornar uma saída, caso esse valor seja maior que o valor de soma (threshold).		
definição:		
Uma série de valores que podem ser aplicados sobre um neurônio.		
semema:		
<organização do neurônio> <informação que é inserida em um sistema> <informação que é inserida em uma máquina> <informação> <sistema> <máquina> <neurônio> <conjunto de valores> <valores>		
fonte:		
corpus de especialidade das RNAs		

symbol:	2.2.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
input layer		
context:		
The input layer is the only layer that does not contain transfer functions.		
definition:		
There are networks in which there is an input layer consisting of nodes that simply accept the input values and successive layers of nodes that are neurons as depicted in Figure 1.		
sememe		
<neuron organization> <information that is put into a system> <neuron organization> <information that is put into a machine> <information> <system> <machine> <neuron> <set of values> <values> <neural networks> <networks> <input layer> <neuron organization> <layer> <input> <output> <transfer functions> <no transfer functions> <nodes> <neurons> <bias> <special neuron> <adaptation> <patterns>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
camada de entrada		
contexto:		
A camada de entrada possui um neurônio especial chamado de "bias" e serve para aumentar os graus de liberdade, permitindo uma melhor adaptação, por parte da rede neural, ao conhecimento a ela fornecido.		
definição:		
Usualmente, as camadas são classificadas em três grupos. É na camada de entrada que os padrões são apresentados à rede.		
semema:		
<organização do neurônio> <informação que é inserida em um sistema> <informação que é inserida em uma máquina> <informação> <sistema> <máquina> <neurônio> <conjunto de valores> <valores> <redes neurais> <redes> <camada de entrada> <camada> <entrada> <saída> <funções de transferência> <sem funções de transferência> <nós> <neurônios> <bias> <neurônio especial> <adaptação> <padrões>		
fonte:		
corpus de especialidade das RNAs		

symbol:	2.2.2	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
input pattern		
context:		
In example 6.6, only four input patterns were stored using a Hopfield network with 19 x 19 nodes. The weights become densest where input patterns are most common. (1)		
definition:		
arrangement; any regularly repeated arrangement, especially a design made from repeated lines, shapes or colours on a surface. (3)		
sememe		
<neuron organization> <information that is put into a system> <neuron organization> <information that is put into a machine> <information> <system> <machine> <neuron> <set of values> <values> <neural networks> <networks> <input layer> <layer> <input> <output> <nodes> <neurons> <bias> <adaptation> <patterns> <weights> <repeated arrangement> <repeated> <arrangement> <behavior> <learning>		
source:		
Haykins; corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
padrão de entrada		
contexto:		
O aprendizado supervisionado utiliza um agente externo que indica à rede um comportamento bom ou ruim de acordo com o padrão de entrada. (2)		
definição:		
uma amostra ou conjunto de amostra seguindo um certo critério preestabelecido ou não. (4)		
semema:		
<organização do neurônio> <informação que é inserida em um sistema> <informação que é inserida em uma máquina> <informação> <sistema> <máquina> <neurônio> <conjunto de valores> <valores> <redes neurais> <redes> <camada de entrada> <camada> <entrada> <saída> <nós> <neurônios> <bias> <adaptação> <padrões> <pesos> <organização repetida> <repetida> <organização> <comportamento> <aprendizado>		
fonte:		
corpus de especialidade das RNAs; Medeiros		

symbol:		
2.2.3		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
input signal		
context:		
Strength of an input signal is determined by the strength of the connection denoted by a weight. Higher the connection weight, larger the contribution of the corresponding input signal to neuron firing.		
definition:		
An action which gives information or an order to a system or machine so that it can operate.		
sememe		
<neuron organization> <information that is put into a system> <neuron organization> <information that is put into a machine> <information> <system> <machine> <neuron> <set of values> <values> <input signal> <input> <signal <input layer> <processing units> <nodes> <neurons> <input vectors> <hidden layers> <hidden> <layers>		
source:		
corpus de especialidade das RNAs; Cambridge		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
sinal de entrada		
contexto:		
O sinal de entrada se propaga para a frente através da rede camada por camada. (1)		
definição:		
O sinal de entrada é caracterizando como um comando, que será pré-processado para otimizar conteúdo do vetor e, em seguida, passá-lo à rede neural.		
semema:		
<organização do neurônio> <informação que é inserida em um sistema> <informação que é inserida em uma máquina> <informação> <sistema> <máquina> <neurônio> <conjunto de valores> <valores> <sinal de entrada> <entrada> <sinal> <camada de entrada> <unidades de processamento> <nós> <neurônios> <vetores de entrada> <camadas ocultas> <oculto> <camadas>		
fonte:		
Haykins; corpus de especialidade das RNAs		

symbol:		
2.3		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
node		
context:		
Nodes of input layer represent input fields or values of categorical input fields. In the figure, there are five input nodes: "Female", "Male", "Income", "Age", and "Education".		
definition:		
In communication networks, a node (Latin nodus, 'knot') is an active electronic device that is attached to a network, and is capable of sending, receiving, or forwarding. A node is a connection point, either a redistribution point or a communication endpoint (some terminal equipment).		
sememe		
<node> <input layer> <output layer> <input> <output> <neuron organization> <values> <active electronic device> <network> <connection point> <redistribution point> <communication endpoint> <connection> <processing units>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
nó		
contexto:		
Cada nó então passa seu dado valor para as conexões que saem dele e em cada conexão o valor é multiplicado por um valor associado a essa conexão.		
definição:		
Nós são unidades de processamento.		
semema:		
<nó> <camada de entrada> <camada de saída> <entrada> <saída> <organização do neurônio> <valores> <dispositivo eletrônico ativo> <rede> <ponto de conexão> <ponto de redistribuição> <terminal de comunicação> <conexão> <unidades de processamento>		
fonte:		
corpus de especialidade das RNAs		

symbol:	2.4	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
perceptron		
context:		
Although the perceptron initially seemed promising, it was eventually proved that perceptrons could not be trained to recognise many classes of patterns.		
definition:		
The perceptron is a program that learn concepts, i.e. it can learn to respond with True (1) or False (0) for inputs we present to it, by repeatedly "studying" examples presented to it.		
sememe		
<perceptron> <program> <neuron organization> <learning> <>true or false> <examples> <classes of patterns> <concepts> <weights> <sets of weights> <network> <input> <neural network> <input vector>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
perceptron		
contexto:		
Durante o processo de treinamento do Perceptron, busca-se encontrar um conjunto de pesos que defina uma reta que separe as diferentes classes, de forma que a Rede classifique corretamente as entradas apresentadas.		
definição:		
O perceptron é uma rede neural cujos pesos e inclinações podem ser treinados para produzir um vetor alvo que, quando apresentamos, tem que corresponder ao vetor de entrada.		
semema:		
<perceptron> <programa> <organização do neurônio> <aprendizado> <verdadeiro ou falso> <exemplo> <classes de padrões> <conceito> <pesos> <conjunto de pesos> <rede> <entrada> <redes neurais> <vetor de entrada>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.1		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
mean squared error	MSE	
context:		
In other words, network development is seen as the task of minimizing a combination of the mean squared error and network complexity, with network complexity being a function of the effective number of parameters in a network.		
definition:		
The average or mean squared error (MSE) or the root mean squared error (RMS) are good measures of the prediction accuracy. When training is just started and the neural network weights have been randomized, the RMS error is usually quite high.		
sememe		
<algorithms and functions> <algorithms> <functions> <mean squared error> <squared error> <algorithm> <function> <mean> <parameters> <network> <average> <training> <prediction accuracy> <weights> <neural networks> <MSE> <high> <low> <input> <values> <estimate> <MSE>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
erro quadrado médio	EQM	
contexto:		
A calibração consiste em alterar os dados de entrada para se obter um erro quadrado médio menor possível.		
definição:		
Erro Quadrado Médio é a soma dos quadrados dos desvios entre os valores do estimador (variável) e o parâmetro que ele se propõe a estimar.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <padrão> <peso> <erro quadrado médio> <erro quadrado> <médio> <função> <parâmetros> <rede> <média> <treinamento> <precisão de previsão> <pesos> <redes neurais> <MSE> <EQM> <alto> <baixo> <entrada> <valores> <estimativa>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.10		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
nonlinear dynamic system		
context:		
In nonlinear dynamical systems, the transition from self-organizing to chaotic behavior is sometimes referred to as a phase transition (or, more specifically, as an order-disorder transition).		
definition:		
Nonlinear dynamical systems can exhibit a completely unpredictable behavior, which might seem to be random. This seemingly unpredictable behavior has been called chaos.		
sememe		
<algorithms and functions> <algorithms> <functions> <nonlinear dynamical systems> <nonlinear> <dynamic systems> <self-organizing behavior> <chaotic behavior> <behavior> <order-disorder transition> <unpredictable behavior> <random> <chaos> <random behavior>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical:
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
sistema dinâmico não-linear		
contexto:		
A característica principal dos sistemas dinâmicos não-lineares é a aleatoriedade, ou o movimento ou comportamento aleatório, ou caótico.		
definição:		
Um sistema dinâmico não-linear é um sistema não-determinista, no qual as implicações dos seus integrantes individualmente são aleatórias e não-previsíveis.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <sistema dinâmico não-linear> <não-linear> <dinâmico não-linear> <comportamento auto-organizável> <comportamento caótico> <comportamento> <transição ordem-desordem> <comportamento imprevisível> <aleatório> <caos> <comportamento aleatório>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.11		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
threshold function		
context:		
The signals from other neurons are summed together and compared against a threshold to determine if the neuron shall excite ("fire").		
definition:		
The level or point at which you start to experience something, or at which something starts to happen:		
sememe		
<algorithms and functions> <algorithms> <functions> <threshold function> <threshold> <function> <neuron> <signal> <pattern> <error>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
função de limiar	função lim	
contexto:		
Os resultados mostram que quanto maior o número de padrões utilizados para o treinamento da rede, mais difícil para a rede atingir um limiar de erro desejado.		
definição:		
A função de limiar (threshold) é uma das mais usadas na construção de redes neurais. (5)		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função de limiar> <limiar> <função> <neurônio> <signal> <padrão> <erro>		
fonte:		
corpus de especialidade das RNAs; Babini/Marranghe		

symbol:	3.12	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
radial basis function	RBF	
context:		
A radial basis function network is a neural network approached by viewing the design as a curve-fitting (approximation) problem in a high dimensional space.		
definition:		
A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin.		
sememe		
<algorithms and functions> <algorithms> <functions> <radial basis function> <radial> <basis> <function> <value> <input> <pattern> <weights> <RBF>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
função de base radial		
contexto:		
A camada escondida utiliza funções de base radial para classificar os dados de entrada e a camada de saída consiste de dez centros, cujos pesos são treinados por retropropagação.		
definição:		
O uso de funções de base radial proporciona deformações suavizadas com um comportamento facilmente controlável. Da sua definição, temos que RBF's são funções suaves e contínuas que realizam interpolações entre dados fornecidos, produzindo resultados contínuos.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função de base radial> <radial> <base> <função> <valor> <entrada> <padrão> <peso>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.2		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
prediction error		
context:		
This rather naive approach failed because the force of discrimination is proportional to the prediction error of the network, such that negative training is weakest for the most confusable model, and becomes stronger for all models that have already been pushed away.		
definition:		
Thus the prediction error is normally distributed with mean zero and some variance. This allows us to express the conditional probability density. This function is also known as the likelihood of w with respect to observation d_i . The maximum likelihood estimate of w is that value of w which maximizes the probability of occurrence of observation d_i for input x_i .		
sememe		
<algorithms and functions> <algorithms> <functions> <prediction error> <mean zero> <variance> <conditional probability density> <likelihood of w > <maximum likelihood> <value of w > <probability of occurrence> <observation d_i > <naive approach> <force of discrimination> <negative training> <models> <estimate> <to estimate> <sets of input in the neural network> <sets of input> <neural network> <input layer>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
erro de predição		
contexto:		
Uma maneira de evitar a polarização é estimar o modelo do ruído. Esta estimação pode ser feita através da inclusão do erro de predição no conjunto de sinais de entrada da rede neural		
definição:		
Diferença entre o valor previsto e o valor observado.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <erro de predição> <média zero> <variância> <densidade de probabilidade condicional> <probabilidade de w > <probabilidade máxima> <valor de w > <probabilidade de ocorrência> <observação d_i > <abordagem inocente> <força de discriminação> <treinamento negativo> <modelos> <estimativa> <estimar> <conjuntos de entrada na rede neural> <conjunto de entrada> <rede neural> <camada de entrada>		
fonte:		
corpus de especialidade das RNAs		

symbol:	3.3	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
local minimum		
context:		
As can be seen from the diagrams, the dynamic of the Hopfield model is always the same: the energy of the system eventually reaches a local minimum and the state of the network can no longer change.		
definition:		
the smallest amount or number allowed or possible.		
sememe		
<algorithms and functions> <algorithms> <functions> <local minimum> <local> <minimum> <critical point> <The lowest point in a particular section of a graph> <lowest point> <graph> <smallest amount> <smallest> <amount> <algorithm> <global> <learning> <value> <function>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
mínimo local		
contexto:		
Esta taxa é um parâmetro de uso opcional, de valor positivo menor do que 1, cuja utilização visa imprimir uma dinâmica no treinamento tal que, eventualmente, possibilite que o algoritmo livre-se de mínimos locais durante o processo.		
definição:		
Assim que um mínimo é encontrado, seja global ou local, o aprendizado cessa. Se a rede alcançar um mínimo local, do seu ponto de vista limitado, todas as direções em sua volta representam valores maiores que o alcançado e, conseqüentemente, a convergência para o mínimo global não é atingida.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <mínimo local> <local> <mínimo> <ponto crítico> <o menor ponto em uma seção específica de um grafo> <menor ponto> <grafo> <menor quantidade> <menor> <quantidade> <algoritmo> <global> <aprendizado> <valor>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.4		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
global minimum		
context:		
This strategy will always lead to the global minimum when y is a function of a single variable. The standard genetic algorithm will be used to obtain the global minimum of this function.		
definition:		
The global minimum can be reached by gradient descent on the noisy function.		
sememe		
<algorithms and functions> <algorithms> <functions> <local minimum> <local> <minimum> <critical point> <The lowest point in a particular section of a graph> <lowest point> <graph> <smallest amount> <smallest> <amount> <algorithm> <global> <learning> <the lowest point over the entire domain of a function> <entire domain> <domain>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
mínimo global		
contexto:		
O objetivo geral é encontrar o peso que minimize o erro, fazendo com que a rede, da melhor maneira possível, execute o mapeamento de todos os exemplos para os valores de alvo. Tenta-se chegar iterativamente ao mínimo global.		
definição:		
Seja f uma função definida sobre um conjunto S . O valor mínimo (mínimo global) para f sobre o conjunto S é um número real m , denotado por $m = \min \{ f(x) : x \text{ em } S \}$ significando que, para todo x em S , temos que $m < f(x)$ e, além disso, existe um ponto d em S tal que $f(d)=m$. O ponto $x=d$ é ponto de mínimo (global) e m é o valor mínimo para f sobre o conjunto S .		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <mínimo local> <local> <mínimo> <ponto crítico> <o menor ponto em uma seção específica de um grafo> <menor ponto> <grafo> <menor quantidade> <menor> <quantidade> <algoritmo> <global> <aprendizado> <valor> <função> <o menor ponto dentro do domínio da função> <domínio todo> <domínio>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.5		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
error function		
context:		
The backpropagation algorithm is used to find a local minimum of the error function. The network is initialized with randomly chosen weights.		
definition:		
the error function (also called the Gauss error function) is a special function (non-elementary) of sigmoid shape which occurs in probability, statistics, materials science, and partial differential equations.		
sememe		
<algorithms and functions> <algorithms> <functions> <error function> <error> <weight> <Gauss error function> <Gauss> <sigmoid shape> <sigmoid> <shape> <probability> <output>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
função de erro		
contexto:		
Se uma rede neural é treinada usando a função de erro de entropia cruzada e se a saída dessa rede neural tem uma forma sigmoideal não-linear, então as saídas serão estimativas de uma probabilidade posterior real de uma classe.		
definição:		
A função erro foi criada para poder calcular a integral da distribuição normal, sendo definida por uma integral que não pode ser expressa por funções elementares.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <função de erro> <erro> <função> <algoritmo> <peso> <erro de função de Gauss> <Gauss> <formato de sigmóide> <sigmóide> <formato> <probabilidade> <saída>		
fonte:		
corpus de especialidade das RNAs		

symbol:	3.6	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
activation function		
context:		
Activation functions for the hidden units are needed to introduce nonlinearity into the network. Without nonlinearity, hidden units would not make nets more powerful than just plain perceptrons (which do not have any hidden units, just input and output units)		
definition:		
In computational networks, the activation function of a node defines the output of that node given an input or set of inputs. A standard computer chip circuit can be seen as a digital network of activation functions that can be "ON" (1) or "OFF" (0), depending on input.		
sememe		
<algorithms and functions> <algorithms> <functions> <network> <output> <input> <algorithm> <set of inputs> <hidden units> <output of a neuron> <output of a node> <node> <activate a node> <activate a neuron> <function to activate a node> <function to activate a neuron> <function that activates a node> <function that activates a neuron> <finite value> <induced local field> <induced> <local field> <on> <off> <nonlinearity> <powerful nets>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf.
termo:	outras designações:	
função de ativação		
contexto:		
A função de ativação é também referida como função restritiva, já que restringe (limita) o intervalo permissível de amplitude do sinal de saída a um valor finito.		
definição:		
A função de ativação define a saída de um neurônio em termos do campo local induzido v.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <rede> <saída> <algoritmo> <conjunto de entradas> <unidades ocultas> <saída de um neurônio> <saída de um nó> <nó> <ativar um nó> <ativar um neurônio> <função para ativar um nó> <função para ativar um neurônio> <função que ativa um nó> <função que ativa um neurônio> <valor finito> <campo local induzido> <induzido> <campo local> <on> <off> <não-linearidade> <redes poderosas>		
fonte:		
corpus de especialidade das RNAs		

symbol:	3.7	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
fuzzy logic		
context:		
Conversely, for certain problems a fuzzy system can be used as a front-end or an aid of a rough set system. For example, some condition attribute values may be expressed in terms of fuzziness, which may be dealt with fuzzy logic.		
definition:		
Fuzzy logic offers an interpretation model for the analysis of neural networks.		
sememe		
<algorithms and functions> <algorithms> <functions> <fuzzy logic> <fuzzy> <logic> <fuzzy system> <system> <set system> <values> <fuzziness> <interpretation model> <model> <analysis of neural networks> <analysis> <neural networks> <intelligent control> <neuro-fuzzy approach> <neuro-fuzzy> <approach> <neuro-fuzzy hybridization> <hybridization> <hybrid intelligent system> <intelligent system> <uncertainty>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
lógica fuzzy		
contexto:		
Tendo estabelecido definições precisas para Redes Neurais e Redes Neurais "Fuzzy", nós podemos definir diferentes tipos de redes.		
definição:		
A lógica fuzzy deve ser vista mais como uma área de pesquisa sobre tratamento da incerteza, ou uma família de modelos matemáticos dedicados ao tratamento da incerteza, do que uma lógica propriamente dita.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <lógica fuzzy> <lógica difusa> <lógica> <algoritmo> <função> <sistema fuzzy> <sistema difuso> <sistema> <sistema de conjunto> <valores> <difusão> <modelo de interpretação> <modelo> <análise de redes neurais> <análise> <redes neurais> <controle inteligente> <abordagem neuro-difusa> <neuro-difusa> <abordagem> <hibridização neuro-difusa> <hibridização> <sistema inteligente híbrido> <sistema inteligente> <incerteza>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.8		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
linear combination		
context:		
The Traditional Time Series Prediction analyzes historic data and attempts to approximate future values of a time series as a linear combination of these historic data.		
definition:		
An expression of first order, composed of the sums and differences of elements with coefficients in a field. It is a concept central to linear algebra and related fields of mathematics.		
sememe		
<algorithms and functions> <algorithms> <functions> <linear combination> <linear> <combination> <sums> <differences> <elements> <coefficient> <field> <linear algebra> <mathematics> <elements with coefficients in a field> <vector> <finite> <finite sum>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
combinação linear		
contexto:		
Sabemos que todo vetor u do plano se escreve de maneira única como combinação linear de dois vetores v e w linearmente independentes. O objetivo deste applet é observar esta propriedade.		
definição:		
Uma combinação linear de um conjunto S de vetores de um espaço vetorial V sobre um corpo K que é uma soma finita.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <combinação linear> <linear> <combinação> <somas> <diferenças> <elementos> <coeficiente> <campo> <álgebra linear> <matemática> <elementos com coeficientes em um campo> <vetor> <finito> <soma finita> <algoritmo> <função>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
3.9		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
learning rate		
context:		
Assume a learning rate = 0.01, and start with random weight vectors distributed uniformly in the region $ w_1 \leq 10$ and $ w_2 \leq 10$.		
definition:		
the speed at which a machine or system learns, or the amount or number of times it happens or changes in a particular period. (3)		
sememe		
<algorithms and functions> <algorithms> <functions> <learning rate> <learning> <rate> <learning speed> <speed> <machine learns> <system learns> <weight> <processing units> <values> <high> <low> <input>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
taxa de aprendizado, taxa de aprendizagem		
contexto:		
Outro parâmetro que interfere na convergência para uma solução é a taxa de aprendizado. Sua influência está relacionada à mudanças nos pesos entre as conexões das unidades de processamento.		
definição:		
Regula a velocidade na qual os valores são ajustados. Uma taxa de aprendizado maior vai fazer com que os valores se ajustem muito mais rápido.		
semema:		
<algoritmos e funções> <algoritmos> <funções> <função> <valor> <entrada> <taxa de aprendizado> <taxa de aprendizagem> <aprendizado> <aprendizagem> <taxa> <velocidade de aprendizado> <velocidade de aprendizagem> <velocidade> <máquina aprende> <sistema aprende> <peso> <unidade de processamento> <valores> <alto> <baixo> <entrada>		
fonte:		
corpus de especialidade das RNAs		

symbol:	4.1.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
Kalman filter	KF	
context:		
A key property of the Kalman filter is that it is the minimum mean-square (variance) estimator of the state of a linear dynamical system.		
definition:		
The basic Kalman filter is a linear, discrete-time, finite-dimensional system, which is endowed with a recursive structure that makes a digital computer well suited for its implementation.		
sememe		
<type of training> <supervised training> <Kalman filter> <Kalman> <filter> <extended Kalman filter> <algorithms> <to locally linearize the nonlinear system> <to linearize the nonlinear system> <that linearizes the nonlinear system> <current state estimate> <state estimate> <estimate> <linearized system> <linearized> <system> <approximate state distribution> <recurrent network> <state distribution> <distribution> <Gaussian> <perceptron> <multilayer perceptron> <weights> <node> <KF>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
filtro de Kalman		
contexto:		
Fazendo isso, a derivação do filtro de Kalman é simplificada por expressar a estimativa do estado no tempo i , dado o conjunto de inovações.		
definição:		
Fornece uma utilização eficiente da informação contida nos dados de entrada. Utiliza recursivamente a informação contida nos dados de treinamento voltando à primeira iteração do processo de aprendizagem.		
semema:		
<tipo de treinamento> <treinamento supervisionado> <filtro de Kalman> <Kalman> <filtro> <filtro de Kalman estendido> <algoritmo> <linearizar localmente o sistema não-linear> <linearizar o sistema não-linear> <que lineariza o sistema não-linear> <estimativa do estado da corrente> <estimativa do estado> <estimativa> <sistema linearizado> <linearizado> <sistema> <distribuição do estado aproximado> <rede recorrente> <distribuição do estado> <distribuição> <Gaussiano> <perceptron> <perceptron multicamadas> <pesos> <nó> <FK>		
fonte:		
corpus de especialidade das RNAs		

symbol:	4.1.1.1	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
extended Kalman filter	EKF	
context:		
The derivation of the extended Kalman filter hinges on linearization of the nonlinear state-space model on the assumption that deviation from linearity is of first order.		
definition:		
Extended Kalman filters (EKF) are algorithms used to locally linearize the nonlinear system about the current state estimate so that applying the Kalman filter to the linearized system the approximate state distribution remains Gaussian.		
sememe		
<type of training> <supervised training> <Kalman filter> <Kalman> <filter> <extended Kalman filter> <algorithms> <to locally linearize the nonlinear system> <to linearize the nonlinear system> <that linearizes the nonlinear system> <current state estimate> <state estimate> <estimate> <linearized system> <linearized> <system> <approximate state distribution> <recurrent network> <state distribution> <distribution> <Gaussian> <perceptron> <multilayer perceptron> <weights> <node> <minimum mean-square> <estimator> <linear dynamical system> <digital computer> <recursive structure> <recursive> <structure> <input> <training> <iteration> <learning process> <learning> <process> <EKF>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical:
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
filtro de Kalman estendido		
contexto:		
O filtro de Kalman estendido torna-se adequado às exigências de uma planificação particular e dos recursos computacionais disponíveis.		
definição:		
Uma rede recorrente construída em torno de um perceptron de múltiplas camadas com w pesos sinápticos e p nós de saída.		
semema:		
<filtro de Kalman> <Kalman> <filtro> <filtro de Kalman estendido> <algoritmo> <linearizar localmente o sistema não-linear> <linearizar o sistema não-linear> <que lineariza o sistema não-linear> <estimativa do estado da corrente> <estimativa do estado> <estimativa> <sistema linearizado> <linearizado> <sistema> <distribuição do estado aproximado> <rede recorrente> <distribuição do estado> <distribuição> <Gaussiano> <perceptron> <perceptron multicamadas> <pesos> <nó> <quadrado médio mínimo> <estimador> <sistema dinâmico linear> <computador digital> <estrutura recursiva> <recursiva> <estrutura> <entrada> <treinamento> <iteração> <processo de aprendizagem> <aprendizagem> <processo> <FKE>		
fonte:		
corpus de especialidade das RNAs		

symbol:	4.1.2	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
perceptron learning rule		
context:		
The perceptron learning rule generated great interest due to its ability to generalize from its training vectors and work with randomly distributed connections. It is especially suited for simple problems in pattern classification.		
definition:		
The perceptron is a single layer neural network whose weights and biases could be trained to produce a correct target vector when presented with the corresponding input vector. The training technique used is called the perceptron learning rule.		
sememe		
<type of learning> <supervised learning> <perceptron learning rule> <learning rule> <training vectors> <randomly distributed connections> <pattern classification> <weights> <target vector> <input vector> <input> <training technique> <bias> <training>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
regra de aprendizado do perceptron,		
contexto:		
Por outro lado, os pesos e as inclinações são atualizados usando as regras de aprendizado do perceptron.		
definição:		
A regra de aprendizado do perceptron consiste em pesos e inclinações que podem ser treinados para produzir um vetor alvo. A regra de aprendizado tem sido provar que existe uma solução em tempo finito, se a solução existir.		
semema:		
<tipo de aprendizado> <tipo de aprendizagem> <regras de aprendizado do perceptron> <regras de aprendizagem do perceptron> <regra de aprendizado> <regra de aprendizagem> <vetor de treinamento> <conexões de distribuição randômica> <classificação de padrões> <pesos> <vetor alvo> <vetor de entrada> <entrada> <técnica de treinamento> <bias>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
4.2.1		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
competitive learning		
context:		
Because the weights are modified in response to the inputs, rather than in response to desired outputs, competitive learning is called unsupervised learning, to distinguish it from the supervised learning of Perceptrons, Adalines and Backpropagation.		
definition:		
Competitive learning is a rule based on the idea that only one neuron from a given iteration in a given layer will fire at a time. Weights are adjusted such that only one neuron in a layer, for instance the output layer, fire. Competitive learning is useful for classification of input patterns into a discrete set of output classes.		
sememe		
<type of learning> <unsupervised training> <competitive learning> <input> <unsupervised learning> <rule> <iteration> <weights> <layer> <classification of input patterns> <input patterns> <algorithms> <data clustering> <training> <input layer>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
aprendizado competitivo, aprendizagem competitiva		
contexto:		
Os algoritmos utilizam a aprendizagem competitiva através do Mapa Auto-organizável de Kohonen em uma arquitetura de rede 2 x 2 e uma arquitetura de rede 5 x 5.		
definição:		
As redes de aprendizagem competitiva estão baseadas no processo competitivo de aprendizagem entre suas unidades. Uma das principais aplicações destas redes é a descoberta de agrupamentos de dados.		
semema:		
<tipo de aprendizagem> <tipo de aprendizado> <aprendizado competitivo> <aprendizagem competitiva> <entrada> <aprendizado não-supervisionado> <aprendizagem não-supervisionada> <<regra> <iteração> <pesos> <camada> <classificação do padrão de entrada> <padrões de entrada> <algoritmos> <agrupamento de dados> <camada de entrada>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
5.1		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	C
term:	other designations:	
adaptive resonance theory	ART	
context:		
Common activation functions include sigmoid, hyperbolic tangent, signum, and Gaussian. adaptive resonance theory (ART)		
definition:		
Adaptive resonance theory (ART) networks are a family of recurrent networks that can be used for clustering. A recurrent neural network that has input and output layers joined by two sets of connection weights.		
sememe		
<process> <activation function> <activation> <function> <adaptive resonance theory> <networks> <recurrent networks> <recurrent> <input layer> <output layer> <layer> <sets of connection weights> <connection weights> <weights> <statistical tool> <clustering> <self-organization> <bottom-up recognizing> <bottom-up> <top-down recognition> <theory> <ART> <input>		
source:		
Corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
teoria de ressonância adaptiva		
contexto:		
Uma rede neural artificial baseada na teoria da ressonância adaptativa (ART) foi usada como ferramentas estatística não-paramétrica e não-linear de forma a definir e comparar agrupamentos segundo os achados neuropsicológicos.		
definição:		
Trata-se de um princípio de auto-organização. A teoria envolve uma camada de reconhecimento de baixo para cima (bottom-up) e uma camada generativa de cima para baixo (top-down). (1)		
semema:		
<processo> <função de ativação> <ativação> <função> <teoria de ressonância adaptiva> <redes> <redes recorrentes> <recorrentes> <camada de entrada> <camada de saída> <camada> <conjunto de pesos conectados> <pesos conectados> <pesos> <ferramenta estatística> <aglomeração> <auto-organização> <reconhecimento de baixo para cima> <de baixo para cima> <reconhecimento de cima para baixo> <teoria> <entrada>		
fonte:		
corpus de especialidade das RNAs		

symbol:		
5.2		
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
data mining		
context:		
As more data is gathered, with the amount of data doubling every three years, data mining is becoming an increasingly important tool to transform this data into information.		
definition:		
Data mining is the process of extracting hidden patterns from large amounts of data.		
sememe		
<process> <data mining> <data> <mining> <information> <hidden pattern> <amounts of data> <learning algorithms> <neural network> <pattern> <extraction> <association rules> <tool> <process>		
source:		
corpus de especialidade das RNAs		

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
mineração de dados		
contexto:		
A mineração de dados é formada por um conjunto de ferramentas e técnicas que, através do uso de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatística, são capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento.		
definição:		
É o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando, assim, novos subconjuntos de dados.		
semema:		
<processo> <mineração de dados> <dados> <mineração> <informação> <padrão oculto> <quantidade de dados> <algoritmos de aprendizagem> <rede neural> <padrão> <extração> <regras de associação> <ferramenta>		
fonte:		
corpus de especialidade das RNAs		

4. ANÁLISE DOS DADOS

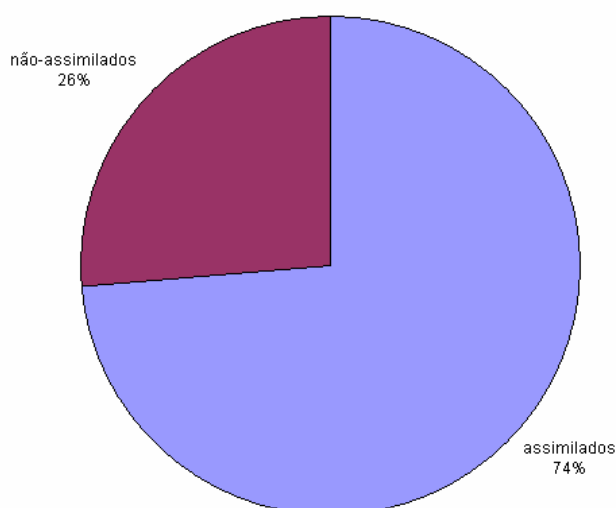
Conforme a Figura 8, dentre os 34 termos fundamentais das RNAs, 26% não foram totalmente adaptados para a língua portuguesa, a saber:

- rede neural feedforward;
- modelo backpropagation;
- modelos ocultos de Markov;
- modelo de Hopfield;
- perceptron;
- lógica fuzzy;
- filtro de Kalman;
- filtro de Kalman estendido;
- regras de aprendizado do perceptron.

Desses, quase 50% possuem nomes próprios em sua concatenação. Os nomes próprios identificados são Markov, Hopfield e Kalman.

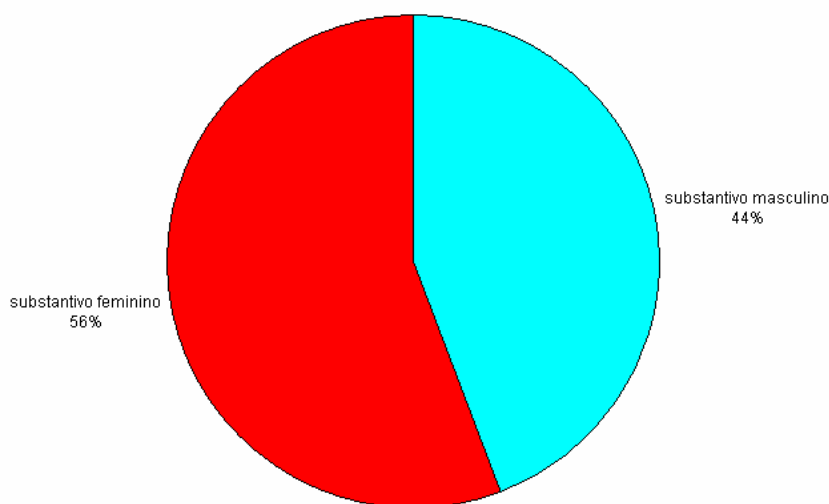
Quanto ao aportuguesamento dos termos, percebe-se que 74% dos termos das RNAs foram adaptados para a língua portuguesa, recorrendo a itens lexicais já existentes na língua, como por exemplo, “camada oculta” e “mineração de dados”.

Figura 8: assimilação dos termos em língua portuguesa



No que diz respeito ao gênero gramatical dos termos das RNAs em língua portuguesa, é possível notar a presença um pouco maior de substantivos femininos. A diferença entre os termos que são substantivos masculinos e femininos é da ordem de 12% na amostra estudada. Dos termos levantados pela análise lexical, 56% estão na categoria de substantivo feminino e, o restante, 44%, na categoria de substantivo masculino. De fato, a frequência de substantivos relacionados às funções e redes contribuiu para que fosse observada uma predominância de substantivos femininos.

Figura 9: Gênero gramatical dos termos

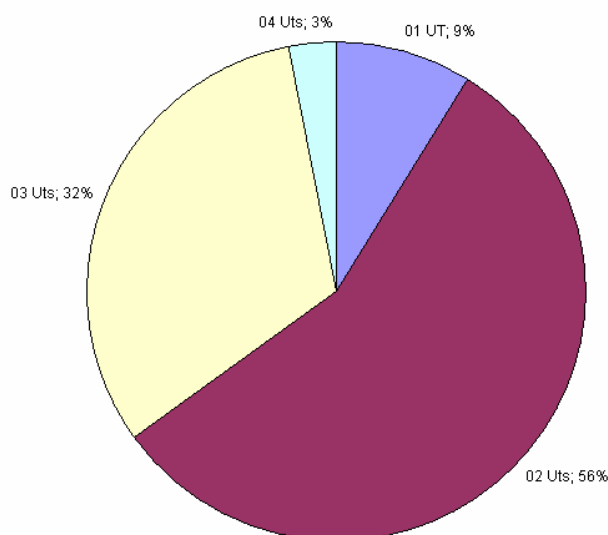


Notamos que o termo “learning rule” mantém dois equivalentes em língua portuguesa: “regra de aprendizado” e “regra de aprendizagem”. Por esse motivo, tivemos que incluir “aprendizado” e “aprendizagem” na ficha em língua portuguesa todas as vezes que “learning” ocorria em língua inglesa. Tanto “regra de aprendizado” quanto “regra de aprendizagem” ocorrem com frequência semelhante no corpus de especialidade das RNAs.

A manipulação do corpus viabilizada pela ferramenta computacional WordSmith Tools identificou as ocorrências mais comuns. Posteriormente, o mesmo aplicativo permitiu

visualizar os “clusters” presentes no corpus. Os “clusters” servem como um indício de unidades lexicais que ocorrem com mais frequência. Em um corpus de especialidade, esses aglomerados podem configurar termos compostos. Nossa análise dos termos fundamentais das RNAs indica que menos de 10% dos termos mais frequentes são unidades terminológicas compostas por um só termo. Há predominância de termos compostos, principalmente de termos constituídos de duas unidades terminológicas. Dos termos presentes nesse levantamento, 90% estão estruturados entre dois e quatro unidades terminológicas (UTs).

Figura 10: Composição dos termos



Com relação às definições, todas elas, apresentadas na ficha terminológica, foram extraídas do corpus de especialidade das RNAs. Muitas vezes, tanto em língua portuguesa quanto em língua inglesa, o fato de haver o verbo “é”/“is” ou “são”/“are”, logo à direita do termo, já era um indício de definição. Quando o termo era de uso muito comum, não encontramos definição. Nesse caso, recorremos ao dicionário de língua geral para que a ficha não ficasse incompleta.

A ausência de definição aconteceu, por exemplo, com o termo “input”. O dicionário *Cambridge Advanced Learner’s Dictionary* traz a seguinte definição: “something such as

energy, money or information that is put into a system, organization or machine so that it can operate”.

Na ficha, optamos por remover algumas partes dessa definição a fim de caracterizar o termo no meio em que ele ocorreu. Caso houvesse a inclusão do item lexical “dinheiro”, a definição perderia sua orientação informática. A definição editada na ficha foi a seguinte: “information that is put into a system or machine so that it can operate”. A edição de tal conceito esclarece o termo em questão de forma adequada aos propósitos dos consulentes.

Optamos por realizar algumas correções quando as definições retiradas do corpus de especialidade das RNAs apresentavam algum tipo de problema. No que diz respeito às falhas identificadas, as mais comuns foram letras repetidas ou ausentes, a ausência de vírgula, e falta de concordância verbal. Essas interferências no conteúdo lingüístico do corpus foram necessárias, principalmente em língua portuguesa.

No que concerne ao contexto, o processo de preenchimento do “campo contexto” revelou-se mais ágil. Os termos eram recolhidos tal qual apareciam no corpus. Em língua portuguesa, foi preciso corrigir algumas imperfeições ortográficas e, também, incluir a concordância verbal de acordo com a norma culta. Em língua inglesa, não houve alterações.

Após a observação da linha de concordância do WST, escolhíamos uma ocorrência e, com o botão direito do mouse, selecionávamos a opção “View Text”. Com isso, era possível observar o contexto de onde o termo advinha. Era comum eliminarmos determinada frase contextual, por haver uma série longa de símbolos e códigos numéricos que acabariam dificultando o entendimento do contexto.

Como o corpus sofreu alguns cortes, as fórmulas, equações e gráficos não puderam ser utilizados. Conseqüentemente, quando o termo estava presente próximo ao local do corpus de onde havíamos efetuado uma limpeza, optávamos por buscar um outro termo.

De certa forma, a magnitude do corpus em língua inglesa facilitou a escolha do exemplo que faria parte do “campo contexto”. Devido à multiplicidade de ocorrências, o processo de seleção do contexto foi dinâmico.

Quanto às fichas terminológicas, foram formatadas de modo a permitirem a visualização completa dos dados pertinentes ao domínio estudado. Na apresentação baseada no MS Access, na parte superior, encontram-se informações a respeito dos termos em língua inglesa. Na parte inferior, as informações estão em língua portuguesa. Na página desenvolvida na linguagem de programação PHP, no lado esquerdo estão os dados em língua portuguesa e, no lado direito, os dados em língua inglesa.

Uma das conveniências dessa ficha terminológica é a possibilidade da busca onomasiológica. A busca é dinamizada por disponibilizar consultas baseadas no(s) traço(s) semântico(s). Para realizar tal procura, basta digitar o semema no campo apropriado. À guisa de exemplificação, caso o consulente deseje descobrir quais são os tipos de redes existentes, receberá uma lista com todos os termos que possuem “tipo de rede” como traço semântico.

O resultado da pesquisa mostraria os seguintes termos:

- rede neural artificial
- rede neural alimentada adiante
- modelo backpropagation
- rede neural multicamada
- rede neural com atraso de tempo
- modelo oculto de Markov
- rede de Hopfield

Todos os termos elencados trazem o traço semântico “tipo de rede” no campo semema. Após a seleção de um dos termos elencados, podem-se descobrir outros detalhes, tais como o contexto real de uso do termo, sua definição e demais sememas. Além disso, a ficha permite a visualização dos mesmos dados em língua inglesa para que a pesquisa também

possa ser executada nessa língua. A procura pelo traço semântico “type of network” retornaria os termos em língua inglesa:

- artificial neural network
- feedforward neural network
- backpropagation model
- multilayer neural network
- time delay neural network
- hidden Markov models
- Hopfield network

Podem-se consultar todos os termos encontrados, dado que todos estão disponíveis na base de dados e contêm informações úteis para a pesquisa terminológica bilíngüe. Escolheremos o termo “backpropagation model” para apresentar suas especificidades. A seguir, as duas divisões da ficha terminológica para o termo “backpropagation model” e “modelo backpropagation”.

Figura 11: Ficha terminológica backpropagation model

symbol:	1.1.3	
area:	subarea:	word class:
Artificial Intelligence	Artificial Neural Networks	U
term:	other designations:	
backpropagation model		
context:		
Among them, the backpropagation model has been most widely used in terms of practical applications.		
definition:		
Backpropagation minimizes output error in a feed-forward network. The major application categories of this model are classification, prediction, and control.		
sememe		
<type of network> <artificial neural network> <network> <neural network> <circuit> <simple processing elements> <neurally based> <artificial neural systems> <neural systems> <can acquire, store, and utilize experiential knowledge> <brain> <neuron> <learning process> <learning> <process> <machine> <input> <output> <weight> <perceptron> <layer> <multilayer> <multilayer perceptron> <knowledge> <input layer> <backpropagation model> <backpropagation> <type of neural network> <practical applications> <feed-forward network> <classification> <prediction> <control> <pattern classification> <pattern> <training> <supervised learning> <multiple layers> <error> <hidden layers> <character recognition> <recognition>		
source:		
corpus de especialidade das RNAs		

Figura 12: Ficha terminológica modelo backpropagation

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sm
termo:	outras designações:	
modelo backpropagation		
contexto:		
O perceptron teve sua validade provada no modelo backpropagation , que possibilitou a implementação da terceira camada necessária para o aprendizado do XOR.		
definição:		
Este modelo é utilizado, principalmente, para classificação de padrões, tais como em aplicações de reconhecimento de caracteres, reconhecimento de sinais em geral, treinamento de robôs e composição musical. Também chamado de retropropagação, pois consiste em um modelo de aprendizado supervisionado de múltiplas camadas, cujo erro obtido na saída é transferido às camadas escondidas, caracterizando-se como uma rede acíclica.		
semema:		
<tipo de rede> <redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistema neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <modelo backpropagation> <modelo de retropropagação> <tipo de rede neural> <backpropagation> <aplicações práticas> <rede alimentada adiante> <classificação> <predição> <controle> <classificação de padrões> <padrões> <treinamento> <aprendizado supervisionado> <camadas múltiplas> <erro> <camada oculta> <reconhecimento de caracteres>		
fonte:		
corpus de especialidade das RNAs		

A busca onomasiológica em língua inglesa também permite a visualização de todos os termos que possuem traços semânticos mais genéricos, por exemplo, “input”. Por ser um dos termos mais comuns identificado no corpus, sua presença é atestada nos traços semânticos de diversos termos. Para descobrir quais os termos que apresentam “input” no campo semema, é necessário iniciar a pesquisa diretamente no campo semema. O resultado, nesse caso, será mais abrangente e retornará os seguintes termos:

- artificial neural network
- feedforward neural network
- backpropagation model
- multilayer neural network
- time delay neural network
- hidden layer
- single hidden layer
- unidade oculta
- input
- input layer
- input pattern
- input signal
- node
- perceptron
- mean squared error
- activation function
- learning rate
- radial basis function
- perceptron learning rule
- competitive learning
- adaptive resonance theory

Em língua portuguesa, o resultado obtido para a pesquisa “entrada” no campo semema será o seguinte:

- rede neural artificial
- rede neural alimentada adiante
- modelo backpropagation
- rede neural multicamadas
- rede neural com atraso de tempo
- camada oculta
- camada oculta única
- unidade oculta
- entrada
- camada de entrada
- padrão de entrada
- sinal de entrada
- nó
- perceptron
- erro quadrado médio
- função de erro
- função de ativação
- taxa de aprendizado
- função de base radial
- regra de aprendizado do perceptron
- aprendizado competitivo
- teoria da ressonância adaptiva

A pesquisa dinâmica não é esgotada apenas pela procura onomasiológica. A construção da ficha terminológica foi desenvolvida de modo a permitir a consulta direta do termo. Caso o consulente tenha em mente qual o termo que deseja consultar, o simples preenchimento do termo em questão no campo apropriado, em inglês ou em português, retornará a ficha correspondente.

Por exemplo, o consulente já conhece, em língua portuguesa, o termo “camada de entrada”. No entanto, precisa descobrir o termo em língua inglesa. Basta digitar “camada de

entrada” na ficha em língua portuguesa e finalizar a busca. Nesse caso, o resultado retornado será unitário, isto é, apenas da ficha do termo em inglês, e não em português.

Figura 13: Ficha terminológica input layer

symbol:	2.2.1	
area:	Artificial Intelligence	subarea: Artificial Neural Networks
		word class: C
term:	input layer	other designations:
context:	The input layer is the only layer that does not contain transfer functions.	
definition:	There are networks in which there is an input layer consisting of nodes that simply accept the input values and successive layers of nodes that are neurons as depicted in Figure 1.	
sememe	<neuron organization> <information that is put into a system> <neuron organization> <information that is put into a machine> <information> <system> <machine> <neuron> <set of values> <values> <neural networks> <networks> <input layer> <neuron organization> <layer> <input> <output> <transfer functions> <no transfer functions> <nodes> <neurons> <bias> <special neuron> <adaptation> <patterns>	
source:	corpus de especialidade das RNAs	

Alternativamente, a mesma busca “camada de entrada” pelo sema, retornaria todos os termos que partilham do mesmo traço semântico. O resultado da pesquisa seria o seguinte:

- rede neural artificial
- rede neural multicamada
- camada de entrada
- padrão de entrada
- nó
- erro de predição
- aprendizado competitivo
- teoria da ressonância adaptiva

Para demonstrar a funcionalidade da busca onomasiológica pelo traço semântico/sema, a inserção de mais traços retornará resultados cada vez mais precisos. Tomando como ponto

de partida os termos identificados acima, em língua portuguesa, é possível inserir na busca mais de um traço semântico. Uma possível consulta onomasiológica poderia incluir dois traços:

<camada de entrada> + <ferramenta estatística>

A identificação dos dois traços semânticos ocorreu no campo semema do termo “teoria de ressonância adaptiva”. Constam no campo semema os seguintes traços:

<processo> <função de ativação> <ativação> <função> <teoria de ressonância adaptiva> <redes> <redes recorrentes> <recorrentes> <**camada de entrada**> <camada de saída> <camada> <conjunto de pesos conectados> <pesos conectados> <pesos> <**ferramenta estatística**> <aglomeração> <auto-organização> <reconhecimento de baixo para cima> <de baixo para cima> <reconhecimento de cima para baixo> <teoria>

A ordem dos traços semânticos dentro do campo semema não influencia o sucesso na busca onomasiológica porque a simples presença do traço pesquisado já permite a identificação do termo que o possui.

O entendimento da terminologia das RNAs ganha uma nova perspectiva e uma nova abordagem com este tipo de estudo.

4.1. Dificuldades relacionadas às buscas

A despeito dos benefícios mencionados anteriormente, a maior dificuldade enfrentada no desenho das fichas terminológicas foi a inclusão do maior número possível de traços semânticos. Esse processo mostrou-se laborioso, tendendo a “pecar” pela incompletude.

Figura 14: Ficha terminológica da rede neural multicamada

área:	subárea:	categoria gramatical
Inteligência Artificial	Redes Neurais Artificiais	sf
termo:	outras designações:	
rede neural multicamada		
contexto:		
As redes multicamadas podem usar um grande número de técnicas de aprendizado, sendo que a mais popular é a propagação reversa.		
definição:		
As redes neurais multicamadas são arquiteturas onde os neurônios são organizados em duas ou mais camadas de processamento, já que sempre vão existir pelo menos uma camada de entrada e uma camada de saída. Estas arquiteturas são as mais freqüentemente encontradas na literatura referente a redes neurais artificiais.		
semema:		
<redes neurais artificiais> <rede> <rede neural> <circuito> <elementos simples de processamento> <baseados em neurônios> <sistemas neurais artificiais> <sistemas neurais> <podem adquirir, armazenar e utilizar conhecimento experimental> <cérebro> <neurônio> <processo de aprendizagem> <aprendizado> <processo> <máquina> <entrada> <saída> <peso> <perceptron> <camada> <multicamada> <perceptron multicamadas> <conhecimento> <camada de entrada> <camada de saída> <redes mais populares> <mais populares> <rede mais popular> <mais popular>		
fonte:		
corpus de especialidade das RNAs		

O preenchimento do “campo semema” do termo “rede neural multicamada” provavelmente será enriquecido com a inclusão e derivação do maior número de descritores. As múltiplas variações e combinações que podem ser realizadas acabam por dificultar esse trabalho. A saber,

- **singular e plural:** <rede neural> e <redes neurais>; <rede> e <redes>; <elemento simples de processamento> e <elementos simples de processamento>, etc.
- **conjugação dos verbos:** <baseados em neurônios>; <baseado em neurônios>; <que se baseiam em neurônios>; <que se baseia em neurônios>
- **preposições:** <com aprendizado>; <do aprendizado> <sem aprendizado> <para o aprendizado>; <de aprendizado>

Além dos itens supra-mencionados, os artigos, as diversas designações do termo e os verbos a ele relacionados, requerem bastante atenção.

Cada sinal gráfico – letras, números, caracteres especiais, espaço em branco, entre outros – é entendido pelos sistemas computacionais como tendo um código binário correspondente e independente. Desse modo, a seqüência gráfica do termo “rede neural” é diferente de “redes neurais” na interpretação binária, ainda que ambas façam referência à mesma realidade. Somada aos espaços em branco a seqüência de caracteres é uma questão fulcral para que o sistema reconheça a seqüência de caracteres digitada no campo de busca. Em outras palavras, a ausência de espaço em branco ou até mesmo de um caracter a mais pode impedir que a consulta seja realizada.

4.2. A equivalência entre os termos

A partir da classificação elaborada por Felber (1987), todos os termos em língua inglesa no domínio das Redes Neurais Artificiais encontraram equivalentes totais em língua portuguesa, o que pode ser atestado pela tabela abaixo:

Tabela 6: Termos fundamentais das RNAs

TERMOS FUNDAMENTAIS DAS RNAs		
#	termo em inglês	termo em português
01	activation function	função de ativação
02	adaptive resonance theory	teoria de ressonância adaptiva
03	artificial neural network	rede neural artificial
04	backpropagation model	modelo backpropagation
05	competitive learning	aprendizado competitivo
06	data mining	mineração de dados
07	error function	função de erro
08	extended Kalman filter	filtro de Kalman estendido
09	feedforward neural network	rede neural feedforward
10	fuzzy logic	lógica fuzzy
11	global minimum	mínimo global
12	hidden layer	camada oculta
13	hidden Markov models	modelos ocultos de Markov

TERMOS FUNDAMENTAIS DAS RNAs		
#	termo em inglês	termo em português
14	hidden unit	unidade oculta
15	Hopfield model	modelo de Hopfield
16	input	entrada
17	input layer	camada de entrada
18	input pattern	padrão de entrada
19	input signal	sinal de entrada
20	Kalman filter	filtro de Kalman
21	learning rate	taxa de aprendizado
22	linear combination	combinação linear
23	local minimum	mínimo local
24	mean squared error	erro quadrado médio
25	multilayer neural network	rede neural multicamada
26	node	nó
27	nonlinear dynamic system	sistema dinâmico não-linear
28	perceptron	perceptron
29	perceptron learning rule	regras de aprendizado do perceptron
30	prediction error	erro de predição
31	radial basis function	função de base radial
32	single hidden layer	camada oculta única
33	threshold function	função de limiar
34	time delay neural network	rede neural de atrasos de tempo

A equivalência total pode ter relação com o fato de que grande parte da produção brasileira nessa área advém de traduções da língua inglesa. As obras de referência das RNAs encontram-se basicamente nos Estados Unidos e na Inglaterra. Dessa forma, os tradutores seriam os responsáveis pela inserção e adaptação desses termos no Brasil. Vale salientar que existem poucas obras consideradas cânones, quando o assunto abordado são as RNAs. Apesar de haver em língua portuguesa algumas obras de ótima qualidade, que não adentraram o mercado editorial por meio de tradução, as principais obras de referência ainda são as traduções oriundas da língua inglesa.

Nesse levantamento dos termos fundamentais das RNAs, assistido pela lingüística de corpus, identificamos a equivalência total em 100% dos termos português-inglês e inglês-português.

O termo em língua inglesa “learning rate” encontra dois equivalentes em língua portuguesa. Uma consulta ao corpus de especialidade das RNAs em língua portuguesa revela que os equivalentes são “taxa de aprendizado” e “taxa de aprendizagem”. O WST indica que, no corpus de especialidade das RNAs, “aprendizado” ocorre 317 vezes e “aprendizagem”, 230. Como nosso interesse era descobrir se “taxa de aprendizado” ou “taxa de aprendizagem” era o termo preferencial, geramos a concordância de cada um e depois recorremos ao recurso clusters. O termo “taxa de aprendizado” ocorre 22 vezes e “taxa de aprendizagem” ocorre 24 vezes. Nesse caso, a diferença não permite decidir qual deles é o termo preferencial. A partir de uma pesquisa pelo motor de busca Google, que encontra 1.630 registros para “taxa de aprendizado” e 1.490 para “taxa de aprendizagem”, optamos por incluir “taxa de aprendizado” como termo preferencial.

No domínio das RNAs, a diferença na visão de mundo de diferentes culturas não alterou os traços semânticos dos termos, muito menos a equivalência entre os termos. Uma vez que as RNAs encontram-se na área das ciências exatas, a necessidade de precisão dos termos das RNAs, no que diz respeito à descrição de seus componentes, processos, organização, etc, gerou a correspondência entre os termos das línguas inglesa e portuguesa.

Ranzani (2006), ao estudar os termos fundamentais italiano-português da imigração, descobriu que a equivalência, tal como descrita anteriormente por Felber (1987) e Dubuc (1985), ocorreu em apenas 30% dos termos analisados. Por abordar o domínio do direito internacional, a subjetividade aliada às peculiaridades sociais das culturas demonstra como o trabalho terminológico não pode prescindir de um cuidadoso trabalho lingüístico.

5. CONSIDERAÇÕES FINAIS

Nos capítulos iniciais deste trabalho, apresentamos conceitos básicos a respeito da terminologia, bem como suas imbricações em outros domínios. Conforme exposto nos capítulos anteriores, as RNAs foram o domínio selecionado para este trabalho porque estão ganhando cada vez mais importância no cotidiano. Esses sistemas inteligentes são capazes de tomar decisões após terem aprendido determinado padrão. Já são amplamente utilizadas em diversos setores da sociedade onde há a necessidade de análise de milhões de informações e tomadas de decisões. O ser humano não consegue realizar a mesma tarefa com os mesmos resultados.

Toda a fundamentação teórica do trabalho terminológico foi norteada pela Teoria Comunicativa da Terminologia, desenvolvida por Cabré nos anos 1990. Para que pudéssemos manipular os dados terminológicos de forma adequada, primeiramente digitalizamos uma série de livros, artigos, guias, manuais, tutorias, páginas da Internet, etc. O segundo passo foi promover a limpeza do corpus, a fim de corrigir algumas imperfeições e remover caracteres ininteligíveis. O programa computacional de análise lingüístico-estatística WordSmith Tools, versão 3, revelou ser útil em nossa investigação terminológica. Em língua inglesa, constituímos um corpus de especialidade das RNAs com 1.457.442 tokens. A existência dos termos foi confirmada no corpus de especialidade em língua portuguesa com 129.083 tokens. Após o levantamento dos termos fundamentais das RNAs em língua inglesa presentes em nosso corpus de especialidade, o passo seguinte foi cruzá-lo com outros corpora: um literário e seis outros de especialidade. Desse cruzamento, obtivemos as palavras-chave do domínio das RNAs. Identificamos os clusters mais comuns e procedemos com a pesquisa terminológica propriamente dita. Percebemos, a partir da análise da Figura 1, que a densidade terminológica das RNAs é maior que todos os demais corpora analisados. Isso se explica pelo

fato de as RNAs serem uma ciência exata, da qual há maior uso de frases diretas e pouca necessidade de recursos metafóricos ou jogos de linguagem para se transmitir a mensagem.

Dentre os termos analisados, identificamos 100% de equivalência total. Diferentemente dos achados de outros estudos, os termos analisados nesse trabalho não sofreram influência do modo de enxergar a realidade das culturas. Isso leva a crer que, no domínio das ciências exatas, a subjetividade cede lugar à objetividade conceitual, havendo menos espaço para adaptações. Nesse caso, um termo em língua inglesa vai encontrar equivalente em língua portuguesa. Não existem diferentes recortes dessa realidade. No domínio das RNAs, identificamos correspondência entre os termos das línguas inglesa e portuguesa.

Quanto às ferramentas e aplicativos computacionais, acreditamos que os estudos lingüísticos podem ser enriquecidos pelas possibilidades de pesquisa, obtendo resultados satisfatórios. Além da ferramenta computacional WST, no desenvolvimento de nosso trabalho terminológico, outros aplicativos nos auxiliaram tanto no armazenamento quanto na estruturação de fichas terminológicas. Em um primeiro momento, o MS Access foi utilizado para armazenar todo o banco de dados terminológico. Contou como ponto positivo nessa pesquisa a facilidade de utilização, bem como o layout amigável ao usuário que o MS Access oferece. Com poucas horas de preparação, quer o pesquisador quer o terminólogo são capazes de adquirir autonomia na construção de um banco de dados terminológico. A impressão das fichas terminológicas elaboradas nessa base é uma tarefa relativamente simples, com diversas opções de apresentação.

Apesar da facilidade em armazenar e apresentar o banco de dados terminológico, percebemos que o MS Access não dispunha de um sistema de busca e relacionamento de informações sofisticado que estivesse à altura de nossas necessidades. Como o objetivo do DETOB era permitir duas modalidades de busca – a semasiológica e a onomasiológica –

tivemos que recorrer a duas outras ferramentas computacionais, a saber, o PHP e o MySQL. O PHP é uma linguagem de programação voltada para Internet. Seu objetivo é tornar a página de um site dinâmica, com entrada e validação dos dados e estruturas condicionais. Trata-se de uma linguagem de programação simples que suporta diversos bancos de dados. O MySQL, por sua vez, é um gerenciador de banco de dados. Foi escolhido por possuir recursos mais sofisticados que o MS Access, nas consultas termo-conceito (semasiológico) e conceito-termo (onomasiológico). A fim de recuperar todo o banco de dados, para viabilizar as consultas, foi necessário exportar os dados terminológicos do MS Access e importá-los no MySQL. O funcionamento do PHP e do MySQL apresentou desempenho satisfatório nessa proposta de um dicionário terminológico.

No momento, existem algumas áreas como a fonética, a sintaxe, a lexicografia, a terminografia, o ensino de língua, e outras, que começam a voltar sua atenção para uma associação com a tecnologia. Em geral, percebe-se que ainda não existe a devida exploração do uso de tecnologia disponível para o desenvolvimento de aplicativos voltados para os estudos lingüísticos.

Acreditamos que os recursos eletrônicos e computacionais enriquecem as pesquisas lingüísticas. No caso do DETOB, esse aporte positivo se dá não apenas pela riqueza de informações disponibilizadas ao mesmo tempo e de forma dinâmica, mas também pelo recurso de busca inteligente. O acesso à informação torna-se facilitado pela disposição das entradas nas fichas terminológicas. Apesar do viés onomasiológico desta pesquisa, a fim de evitar a restrição de possibilidades de busca, foram disponibilizadas tanto a pesquisa onomasiológica quanto a semasiológica. Vale salientar que, devido ao foco deste trabalho, destacamos a pesquisa onomasiológica na consulta às fichas terminológicas.

A aplicação de Babini (2001) à proposta de um dicionário terminológico onomasiológico bilíngüe das RNAs pôde ser levada a cabo com resultados confiáveis. Trata-

se de um modelo viável para ser elaborado em suporte eletrônico. Tanto o consulente monolíngüe quanto o consulente bilíngüe terão facilidade em manusear o DETOB porque a consulta pode ser executada em dois idiomas e o resultado final – a ficha terminológica – foi elaborada de modo a apresentar informações em línguas inglesa e portuguesa.

Apesar das dificuldades encontradas, a busca onomasiológica contribuiu para o melhor entendimento desse domínio. Por meio da consulta onomasiológica, é possível entender a relação que existe entre os termos que compartilham do(s) mesmo(s) traço(s) semântico(s). A maior dificuldade encontrada no desenvolvimento de um dicionário terminológico onomasiológico foi o preenchimento do “campo semema”. O trabalho de inserção dos traços semânticos é difícil pela quantidade de sememas que devem estar presentes, pois, quanto maior o número de sememas, maior será a possibilidade de êxito de uma busca onomasiológica.

Nossa intenção é alargar os horizontes das pesquisas de orientação onomasiológica, terminológica e computacional, trazendo a discussão aos meios acadêmicos. Com essa pesquisa, esperamos incentivar o desenvolvimento de dicionários em outras áreas do conhecimento, contribuindo para que sejam implementados em suporte eletrônico, facilitando, assim, sua divulgação.

Outrossim, pretendemos demonstrar que os estudos lingüísticos, especialmente as pesquisas terminológicas, podem contar com o suporte informático para desenvolver suas aplicações de forma ágil e prática. Essa associação entre áreas interdisciplinares tende a gerar avanços na forma de entender, bem como no modo de estudar, os fenômenos lingüísticos. Pretendemos, ainda, dar continuidade aos estudos onomasiológicos e computacionais iniciados aqui, em um trabalho maior, em nível de doutorado.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, G.M.B. O percurso da terminologia: de atividade prática à consolidação de uma disciplina autônoma. **TradTerm**, n.9. São Paulo, FFLCH, 2003, p. 211-222

_____.; OLIVEIRA, L.H.M.; ALUÍSIO, S.M. A terminologia na era da informática. **Ciência e Cultura**, v. 58, n. 2, São Paulo, abr/jun 2006, p. 42-44.

_____. A teoria comunicativa da terminologia e a sua prática. **Alfa**, São Paulo, v.50, n.2, 2006b, p.85-101.

ALPÍZAR-CASTILLO, R. **¿Como hacer un diccionario científico-técnico?** Buenos Aires: Editorial Memphis, 1995.

AUBERT, F.H. **Introdução à metodologia da pesquisa terminológica bilíngüe**. São Paulo: Humanitas/FFLCH, 1996b, 100p. (Cadernos de Terminologia, n. 2)

AUGER, P. La terminotique et les industries de la langue. **Méta** XXXIV, 1989, p. 451-455.

BABINI, M. Dicionário analógico do padre Carlos Spitzer: uma versão do *Thesaurus* de Roget? **Cadernos de tradução**, n. xii. Florianópolis: Pós-graduação em estudos da tradução, 2003/2, p. 65-73.

_____. Do conceito à palavra: os dicionários onomasiológicos. **Ciência e Cultura**, v. 58, n. 2, São Paulo, abr/jun 2006, p. 38-41.

BARRETO, J.M. **Introdução às Redes Neurais Artificiais**. Laboratório de Conexionismo e Ciências Cognitivas UFSC - Departamento de Informática e de Estatística. Florianópolis, 2002.

BALDINGER, K. **Semasiologia e onomasiologia**. Trad. Ataliba T. de Castilho. **Alfa**. n. 9. Marília, 1966, p. 7-36.

_____. **Teoria semántica**: hacia una semántica modernas. Madrid: Alcalá, 1970.

BARROS, L.A. **Curso básico de terminologia**. São Paulo: Editora da Universidade de São Paulo, 2004. 285 p.

_____. Aspectos epistemológicos e perspectivas científicas da terminologia. **Ciência e Cultura**, v. 58, n. 2, São Paulo, abr/jun 2006, p. 22-26.

_____. **Conhecimentos de terminologia geral para a prática tradutória**. São José do Rio Preto, SP: NovaGraf, 2007.

BIBER, D. Investigating Language Use Through Corpus-Based Analyses of Association Patterns. In: BARLOW, M.; KEMMER, S. **Usage Based Models of Language**. Stanford: CSLI, 2000.

BERBER SARDINHA, T. Lingüística de corpus: histórico e problemática. **Delta**, v. 16, n. 2, p. 323-367. 2000.

_____. **Lingüística de corpus**. São Paulo: Manole, 2004.

_____. **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005a.

_____. Como encontrar as palavras-chave mais importantes de um corpus com WordSmith Tools. **Delta**, Campinas, v. 2, n. 21, p. 237-250, 2005b.

BERGENHOLTZ, H.; TARP, S. (ed.). **Manual of Specialized Lexicography**. Amsterdam: Philadelphia: John Benjamins, 1995.

BIDERMAN, M. T. **Teoria Lingüística**: teoria lexical e computacional. 2a. ed. São Paulo: Martins Fontes, 2001.

_____. Corpora informatizados e dicionários. In: DIAS-DA-SILVA, B.C. (org). **Todas as trilhas**: perfilando pesquisas e projetos. Araraquara: Laboratório Editorial/FCL/UNESP; São Paulo: Cultura Acadêmica, 2003, p. 179-193.

_____. O conhecimento, a terminologia e o dicionário. **Ciência e Cultura**, v. 58, n. 2, São Paulo, abr/jun 2006, p. 35-37.

BOLSHAKOV, I.A.; GELBUKH, A. **Computational Linguistics**: Models, Resources, Applications. México, DF: Universidad Nacional Autónoma de México, 2004.

BOULANGER, J.-C. Alguns componentes lingüísticos no ensino da terminologia. **Ciência e Cultura**, v. 24, n. 3, Brasília, 1995.

BRAGA, A de; CARVALHO, AP. de L. F. de; LUDERMIR, T.E. Fundamentos de redes neurais artificiais. Rio de Janeiro: NCE/UFRJ, 1998.

CABRÉ, M.T. Terminologie et linguistique: la théorie des portes. **Terminologies Nouvelles**. Bruxelles: Agence de la francophonie et Communauté française de Belgique, n. 21, juin, 2000, p. 10-15.

CASTILHO, A.T.B.; CARRATORE, E. del. A onomasiologia no léxico e na sintaxe. In: **Alfa**, n. 11, Marília, 1967, p. 129-149.

CHOMSKY, N. The master and his performance: An interview with Noam Chomsky. Entrevistador József. Andor. In: **Intercultural Pragmatics**, v. 1, n. 1, 2004, p. 93-111

CLAVEAU, V.; L'HOMME, M.-C. Apprentissage par analogie pour la structuration de terminologie – utilisation comparée de ressources endogènes et exogènes. [s.l], 2005.

CORBEIL, J.-C. Le plurilinguisme terminologique. In: **Terminologia y modelos culturales**. Barcelona: IULA, 1999, p. 75-85.

DIAS-DA-SILVA, B.C.; MORAES, H.R. A construção de um thesaurus eletrônico para o português do Brasil. **Alfa**. São Paulo. 47(2), 2003, p. 101-115.

DUBUC, R. Les textes traduits et leur utilisation en terminologie. **Meta** XIX, n. 4, 1972, p. 205-208.

_____. **Manuel pratique de terminologie**. Montreal: Linguatex, 1985.

FELBER, H. **Manuel de terminologie**. Paris: Unesco/Infoterm, 1987.

FERREIRA, A. B. H. **Novo dicionário Aurélio século XXI**: dicionário eletrônico versão 5.0. Curitiba: Positivo, 2004. 1 CD-ROM.

FINATTO, M.J.B. Corpora para reconhecimento de terminologias e de linguagens técnico-científicas: o desafio da interação de bases de dados. **Estudos Lingüísticos** XXXV, 2006, p.146-155.

GÓMEZ, P. C. Do we need statistics when we have linguistics? **Delta**, Campinas, v. 2, n. 18, p. 233-271, 2002.

GOUADEC, D. Dictionnaires terminologiques: l'impact des nouvelles technologies. **Méta** XLII, 1997, p. 24-32.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**, 2nd ed. Delhi: Pearson Education, 1999.

HOMBROUCK, É. Étude comparée de dictionnaires électroniques et banques de termes. **Terminologies Nouvelles**. Bruxelles: Agence de la francophonie et Communauté française de Belgique, n.15?, p. 90-103.

HOUAISS, A. Na pista da língua. Entrevistadores Humberto Werneck e Zuenir Ventura. Veja, São Paulo, 30 ago. 1978. In: **ALMANAQUE Abril**: 2003, São Paulo: Abril, s/d. 1 CD-ROM.

HÜLL, W. The Onomasiological Approach. In: _____. **800-1700: The Topical Dictionaries**. Cambridge: OUP, 1999, p. 3-27.

LEECH, G. (1991). The State of the Art in Corpus Linguistics. In: AIJMER, K.; ALTENBERG, B (eds.). **English Corpus Linguistics**, London: Longman, 1991, p. 8-29.

KRIEGER, M.G.; FINATTO, M.J.B. **Introdução à terminologia**. São Paulo: Contexto, 2004. 223p.

_____. Do ensino de terminologia para tradutores: diretrizes básicas. Disponível em: <<http://www.cadernos.ufsc.br/online/cadernos17/maria%20da%20graca%20krieger.pdf>>. Acesso em: 01 abr. 2007.

McARTHUR, T. Preface. **Longman Lexicon of Contemporary English**. Essex: Longman, 1981.

MEDINA, J.B. Conversion in English computer terminology. **Target**. Amsterdam: Philadelphia: John Benjamins, v. 15, n.2, 2003, p. 317-336.

MELLO, M. T de. Aplicação de redes neurais artificiais no processo de precificação de ações. Universidade Federal de Pelotas - Instituto de Física e Matemática - Curso de Bacharelado em Ciência da Computação - Trabalho de Conclusão apresentado como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação. Pelotas, outubro de 2004

OSÓRIO, F.; BITTENCOURT, J. R. Sistemas inteligentes baseados em Redes Neurais Artificiais aplicados ao processamento de imagens. In: **I WORKSHOP DE INTELIGÊNCIA ARTIFICIAL**, 2000, Santa Cruz do Sul: UNISC/Departamento de Informática.

OTHERO, G.A.; MENUZZI, S.M. **Linguística computacional: teoria e prática**. São Paulo: Parábola, 2005.

PAIVA, P.T.P.; CARDOSO, D.C. de. Estudos de tradução baseados em corpus e linguística de corpus: levantamento de termos médicos para a elaboração de um glossário bilíngüe. In: **Estudos Linguísticos: XXXVI(2)**, mai/ago., 2007, p. 428-436.

PAVEL, S.; NOLET, D. **Manual de terminologia**. Trad. Enilde Faulstich. Quebec: Travaux publics et Services gouvernementaux. 2002.

RANZANI, E. **Dicionário terminológico onomasiológico bilíngüe italiano-português: busca de equivalentes e tratamento dos dados em português**. Dissertação de mestrado apresentada ao programa de pós-graduação em Estudos Linguísticos da Unesp, campus de São José do Rio Preto, 2005.

REY, A. **Essays on Terminology**. Translated and edited by Juan C. Sager. Philadelphia: John Benjamins, 1995.

REY-DEBOVE, J. Le sens des mots. In: _____. **La linguistique du signe: une approche sémiotique du langage**. Paris: Armand Colin, 1998, p. 90-150.

SADER FEGHALI, L. Pour une terminologie traductionnelle tridimensionnelle et personnalisée. Disponível em: <<http://www.certa.usj.edu.lb/alkimiya/linafeghalifr.pdf>>. Acesso em: 01 abr. 2007.

SAGER, J.C.; DUNGWORTH, D.; McDONALD, P.F. **English Special Languages: principles and practice in science and technology**. Wiesbaden: Brandstetter, 1980.

SCOTT, M. **WordSmith Tools: Version 3**. Oxford: Oxford University Press, 1999.

_____. John Sinclair (1933-2007). Language Pioneer and Explorer. In: **Language Awareness**, v. 16, n. 2, 2007, p-79-80.

_____. **Re: online teaching (fwd)**. Lista de discussão da universidade de Virginia. Disponível em: <http://lists.village.virginia.edu/lists_archive/Humanist/v10/0342.html>. Acesso em: 30 nov. 2007

_____. **Oxford WordSmith Tools: version 4.0**. Oxford, 2007.

SCHMITZ, J.R. A problemática dos dicionários bilíngües. In: OLIVEIRA, A.M.P.P. de; ISQUERDO, A.N. (Org). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: Ed. UFMS, 2001, p. 161-170.

SILVA, R. R. da. **Reconhecimento de imagens digitais utilizando redes neurais artificiais**. Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do curso de Ciência da Computação para obtenção do título de Bacharel. Lavras, 2005.

SINCLAIR, J. Corpus and text: basic principles. In: WYNNE, M. (Ed.) **Developing linguistic corpora**: a guide to good practice. 2004. Disponível em: <<http://ahds.ac.uk/linguistic-corpora/>>. Acesso em: 7 dez. 2007.

THOIRON, P.; SERANT, D. La topographie des termes. **Méta** XXXIV, 1989, p. 435-442.

VAN CAMPENHOULT, M. Les voies actuelles en terminologie et terminotique. **En bons termes**. Paris. 1998, p. 1-10.

Autorizo a reprodução xerográfica para fins de pesquisa.

São José do Rio Preto, ____/____/____

Assinatura