

PROPOSTA DE UM MODELO DE DICIONÁRIO TERMINOLÓGICO ONOMASIOLÓGICO

Eduardo Batista da Silva
(UEG/UnU Quirinópolis. PG/ Universidade
Estadual de São Paulo (UNESP/IBILCE)
eduardobatistadasilva@gmail.com

Maurizio Babini
Universidade Estadual de São Paulo (UNESP/ IBILCE)
maurizio@ibilce.unesp.br

RESUMO: O presente trabalho tem como objetivo principal aplicar o modelo de dicionário terminológico onomasiológico bilíngue, proposto por Babini (2001b), na elaboração de um dicionário eletrônico inglês-português e português-inglês dos termos fundamentais das Redes Neurais Artificiais (RNAs). Essa subárea da Inteligência Artificial foi escolhida devido à sua utilização nas mais variadas atividades tecnológicas. O dicionário onomasiológico caracteriza-se por permitir a busca de uma unidade lexical ou terminológica a partir de seu conteúdo semântico. 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 compõem o conceito de cada termo do dicionário.

PALAVRAS-CHAVES: Onomasiologia. Dicionário onomasiológico. Dicionário terminológico onomasiológico. Redes Neurais Artificiais.

ABSTRACT: The present paper aims at applying a model of bilingual onomasiological terminological dictionary, as proposed by Babini (2001b), for the development of an English-Portuguese and Portuguese-English electronic dictionary of the fundamental Artificial Neural Networks (ANN) terms. This subarea of Artificial Intelligence was chosen due to its use in several technological activities. The onomasiological dictionary is characterized by allowing searches of either lexical or terminological units from its semantic content. Our dictionary model allows two types of search:

semasiological and onomasiological. The onomasiological search is made possible by a set of semes or semantic traits that make up the concept of each term in the dictionary.

KEY WORDS: Onomasiology. Onomasiological Dictionary. Onomasiological Terminological Dictionary. Artificial Neural Networks

0. Introdução

O presente trabalho tem como objetivo principal desenvolver um protótipo de um dicionário eletrônico terminológico onomasiológico bilíngue inglês-português, português-inglês no domínio de Redes Neurais Artificiais (doravante, RNA). Nesse protótipo, serão utilizados trinta termos fundamentais de RNA para que a pesquisa possa ser efetuada tanto na direção português-inglês, quanto na direção inversa. O desenvolvimento da informática, especialmente das RNA, justifica uma pesquisa que tenha como resultado a elaboração de um dicionário terminológico, sendo de extrema importância para o trabalho de tradutores, terminólogos ou interessados pelo domínio em questão. A pesquisa prova ser relevante, também, por fornecer uma maior flexibilidade na busca por um termo quando já se tem a ideia do mesmo.

Para o desenvolvimento do protótipo de um dicionário eletrônico terminológico onomasiológico bilíngue inglês-português, português-inglês no domínio das Redes Neurais Artificiais, doravante DETOB, tornou-se fundamental o embasamento teórico da Teoria Comunicativa da Terminologia (Cabré, 1993). As principais referências teóricas que guiarão nossas buscas pelos equivalentes lexicais serão Felber (1987), Dubuc (1985), Alpizar-Castillo (1995), Cabré (1993), Berber-Sardinha (2004), Barros (2004) e Babini (2001a, 2001b). A Linguística de Corpus fornece os subsídios para o processamento e tratamento do corpus a ser explorado.

Para a realização do DETOB, constituímos dois *corpora* de especialidade, um em língua inglesa e um em língua portuguesa e a

coleta dos termos se deu com o auxílio do software *WordSmith Tools*. Para a constituição dos *corpora*, foram recolhidos textos de diversas fontes, em particular livros e artigos científicos sobre o assunto.

No próximo item, apresentamos uma breve introdução a respeito das Redes Neurais Artificiais.

1. As Redes Neurais Artificiais

As RNA, conhecidas também como *redes conexionistas* e *sistema de processamento paralelo e distribuído*, são uma subárea da Inteligência Artificial, e relembram em sua estrutura, de alguma maneira, o funcionamento do cérebro humano. Os primeiros trabalhos sobre as RNAs começaram em 1943, com Warren McCulloch (psiquiatra e neuroanatomista) e Walter Pitts (matemático), que realizaram o primeiro modelo de neurônio artificial, tendo como ponto de partida os estudos sobre a estrutura do cérebro humano e o funcionamento dos neurônios biológicos. Os elementos constitutivos de uma rede neural artificial são chamados de neurônios artificiais e são dispostos em uma ou mais camadas, sendo interligados por inúmeras conexões (BABINI; MARRANGHELLO, 2007). As RNAs são desenvolvidas para que tenham a capacidade de agir como o cérebro humano ao desempenhar uma determinada atividade, tanto na organização quanto no funcionamento. Existe, portanto, uma ampla gama de disciplinas que podem receber contribuições desse domínio. Apesar de serem concebidas por especialistas em informática e programação, Haykin (1999) afirma que são um campo multidisciplinar com raízes nas neurociências, matemática, estatística, física, ciência da computação e engenharia. Dentre as possíveis aplicações das RNAs, temos inúmeros exemplos, em áreas muito diferentes entres si. Podemos citar como aplicações, o reconhecimento de faces, a identificação de placas de carros, o reconhecimento de caracteres por meio de um software *OCR* e a compressão de dados em um computador.

2. A macroestrutura e a microestrutura do DETOB

Para a aplicação do modelo de dicionário terminológico onomasiológico bilíngue proposto por Babini (2001b), é necessário recorrer a uma ferramenta computacional. Para a realização do DETOB, optamos pela utilização do *Visual Basic*, uma linguagem de programação distribuída pela Microsoft, que é parte integrante do pacote Microsoft Visual Studio. O *Visual Basic* possui uma interface que permite a customização dos aplicativos criados, fato esse que o torna extremamente útil para a realização de nosso dicionário.

Uma vez escolhida a ferramenta computacional, e depois de termos efetuados as diferentes etapas do trabalho terminológico, procedemos a determinação da macroestrutura e da microestrutura de nosso dicionário. A macroestrutura do dicionário se divide em duas partes, uma para a língua inglesa e uma para a língua portuguesa, contendo a primeira o sistema nocional e os verbetes dos termos em língua inglesa, e a segunda, o sistema nocional e os termos em língua portuguesa. A microestrutura dos verbetes em língua inglesa é a mesma dos verbetes em português e contém oito campos: símbolo, área (e subárea), termo, categoria gramatical, outras designações, contexto, definição e semema:

Símbolo	O primeiro dos campos é o símbolo de classificação, que indica a posição que o termo ocupa no sistema conceptual.
Área	Esse campo contém duas informações, a área e a subárea dos termos de nosso dicionário, isto é Inteligência Artificial e Redes Neurais Artificiais.
Termo	Esse campo contém a entrada do verbete.
Outras designações	Aqui, se houver, são listados quase-sinônimos, siglas e outras variantes terminológicas.
Categoria gramatical	Esse campo registra a categoria gramatical do termo: em língua portuguesa, observa-se se o termo é substantivo masculino ou feminino; em língua inglesa, se é substantivo contável ou incontável.
Contexto	Esse campo traz um dos contextos onde o termo foi encontrado.
Definição	Aqui é colocada a definição do termo.
Semema	O campo semema traz os traços semânticos de cada termo.

A seguir, apresentamos um exemplo de verbete em língua inglesa e um em língua portuguesa:

Symbol	2.1.
Area	Artificial Intelligence
Term	hidden layer
Other designations	-
Grammar Category	n
Context	The output of the hidden layer consists therefore of normalized numbers.
Definition	The hidden layer is meant to recode (or to provide a representation for) the inputs.
Sememe	<output> <output of the hidden layer> <normalized numbers> <to recode> <that recodes> <representation for the inputs> <representation> <inputs> <processing> <characteristic extractors> <pattern recognition> <external input>

Símbolo	2.1.
Área	Inteligência Artificial
Termo	camada oculta
Outras designações	-
Categoria gramatical	sf
Contexto	É na camada oculta onde é feita a maior parte do processamento, através das conexões ponderadas; podem ser consideradas como extratoras de características.
Definição	A camada oculta serve para intervir entre a entrada externa e a saída da rede de maneira útil.
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>

Como podemos ver, o campo que apresenta o semema contém os traços semânticos dos termos, nas respectivas línguas. Veremos, mais adiante, como efetuar buscas de tipo onomasiológico a partir dos traços semânticos. O preenchimento desse campo foi o mais trabalhoso, tendo em vista que, para facilitar as buscas de tipo onomasiológico, optamos pela inclusão do maior número possível de traços semânticos. Nesse sentido, os traços semânticos foram combinados e apresentados de diferentes maneiras, isto é: **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>.

No próximo item, apresentamos alguns exemplos de buscas onomasiológicas realizáveis no DETOB.

3. Buscas de tipo onomasiológico

Conforme demonstrado nos exemplos acima, o campo semema traz os traços semânticos de cada termo, o que permite, em nosso dicionário, efetuar buscas onomasiológicas. Apresentamos, a seguir, alguns exemplos de busca de tipo onomasiológico em língua portuguesa e em língua inglesa.

Para saber como se chama em língua portuguesa um <tipo de rede recorrente> que <pode assumir diferentes estados>, o usuário digitará essas informações no sistema de busca do dicionário e terá como resposta: *1.2 modelo oculto de Markov*. Para encontrar em língua inglesa o mesmo tipo de rede, o usuário poderá digitar as mesmas informações nessa língua, a saber <type of recurrent network> e <that can assume different states>, obtendo como resultado: *1.2. hidden Markov models*.

A lista de todos os traços semânticos dos termos do dicionário foi indexada no aplicativo em *Visual Basic*, não sendo necessário

digitar todo o traço semântico. É possível escolher esses traços a partir de uma lista, disponível no momento da busca.

No DETOB, é possível também efetuar buscas um pouco mais genéricas. Caso o consulente deseje descobrir quais são os tipos de redes neurais contidas no dicionário, poderá efetuar uma busca a partir do traço semântico <tipo de rede>, ou alternativamente, <tipos de rede>; e, como resposta à sua consulta, visualizará uma lista com todos os termos que possuem esses traços semânticos em seu semema, ou seja, *rede neural artificial*, *rede neural alimentada adiante*, *modelo backpropagation*, *rede neural multicamadas*, *rede neural com atraso de tempo*, *modelo oculto de Markov*, *rede de Hopfield*. Após a seleção de um dos termos elencados, o usuário pode acessar o verbete correspondente. O software apresenta na mesma tela, também, se houver, o verbete do termo equivalente em língua inglesa.

Essa busca poderia ter sido feita a partir da língua inglesa. A consulta pelo traço semântico “type of network” retornaria os seguintes termos: *artificial neural network*; *feedforward neural network*; *backpropagation model*; *multilayer neural network*; *time delay neural network*; *hidden Markov models* e *Hopfield network*.

Apresentamos neste item somente alguns exemplos de busca de tipo onomasiológico. Vale ressaltar que o dicionário oferece outras possibilidades de pesquisa ao consulente. Além das modalidades de buscas apresentadas, é possível efetuar as tradicionais buscas de tipo semasiológico, que não apresentamos nesse trabalho.

4. Considerações finais

O objetivo de nosso trabalho era a aplicação de modelo de dicionário onomasiológico bilíngue proposto por Babini (2001b) no domínio das RNAs. Para isso, desenvolvemos um protótipo de dicionário, chamado DETOB, a partir de um conjunto de termos em língua portuguesa e em língua inglesa. Para podermos efetuar buscas de tipo onomasiológico, tanto os verbetes em língua portuguesa, quanto os verbetes em língua inglesa contêm o campo semema, no

qual são listados os traços semânticos de cada termo. Como podemos ver nos exemplos listados no item 3, o DETOB permite vários tipos de buscas onomasiológicas, fornecendo o termo ou os termos desejados a partir dos traços semânticos contidos no semema. A elaboração do campo semema foi, contudo, muito trabalhosa, tendo em vista a metodologia utilizada. Esperamos, com esse trabalho, ter trazido nossa contribuição às pesquisas de cunho onomasiológico e à elaboração de novos modelos de dicionários eletrônicos.

REFERÊNCIAS

- ALPÍZAR-CASTILLO, R. *¿Como hacer um diccionario científico-técnico?* Buenos Aires: Editorial Memphis, 1995.
- AUBERT, F.H. *Introdução à metodologia da pesquisa terminológica bilingüe*. São Paulo: Humanitas/FFLCH, 1996. (Cadernos de Terminologia, n. 2)
- BABINI, M. *Onomasiologie et dictionnaires onomasiologiques*. São José do Rio Preto: Beatriz, 2001a.
- . *Proposition d'un nouveau modèle de dictionnaire terminologique onomasiologique*. São José do Rio Preto: Beatriz, 2001b.
- ; MARRANGHELLO, N. *Introdução às Redes Neurais Artificiais*. São Paulo: Cultura Acadêmica; São José do Rio Preto: Laboratório Editorial do Ibilce/Unesp, 2007. (Coleções Brochuras)
- BARROS, L. A. *Curso básico de terminologia*. São Paulo: Editora da Universidade de São Paulo, 2004.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.
- CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Editorial Antártida/Empúries, 1993.
- DUBUC, R. *Manuel pratique de terminologie*. Montreal: Linguattech, 1985.
- FELBER, H. *Manuel de terminologie*. Paris: Unesco/Infoterm, 1987.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed. Delhi: Pearson Education, 1999.