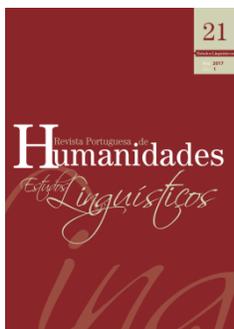


*Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.*

# Revista Portuguesa de Humanidades *Estudos Linguísticos*



This article appeared in *Revista Portuguesa de Humanidades* (2017, V. 21, 1 – *Estudos Linguísticos*) published by *Axioma – Publicações da Faculdade de Filosofia*. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

Authors requiring further information regarding *Revista Portuguesa de Humanidades* archiving and manuscript policies are encouraged to visit:

<http://rphumanidades.braga.ucp.pt/>

The copyright of this article belongs to *Aletheia – Associação Científica e Cultural*, such that any posterior publication will require the written permission of the President. For the use of any article or a part of it, the norms stipulated by the copyright law in vigor are applicable.



Revista Portuguesa de Humanidades  
Director Prof. Doutor Miguel Gonçalves

ALETHEIA - Associação Científica e Cultural  
Faculdade de Filosofia e Ciências Sociais  
Praça da Faculdade, 1  
4710-297 BRAGA  
Portugal  
aletheia.ffcs@braga.ucp.pt

# Lexicologia: mutações teórico-metodológicas no séc. XX

**CARLA ARAÚJO**

IPB

carla.araujo@ipb.pt

**CARLOS ASSUNÇÃO**

UTAD

cassunca@utad.pt

## Abstract

Lexicology, as the science of the lexicon of a language, is a recent discipline, having resulted, however, from a long tradition initiated by the thought of the philosophers of ancient Greece, who reflected on the origin of language and on the dialectic between thought and language and on how languages represent the world, but that was only formatted and validated as science in the second half of the twentieth century. The aim of this work is to demonstrate the main theoretical-methodological mutations of lexicology in the last half century, which led to its assertion as a science, and to describe some lexical ontologies that led to the development of databases for many languages in the world.

**Keywords:** Lexicon, Semantics, Database, Lexical Ontology.

## Introdução

Revisitaremos, neste texto, o rumo empreendido pela Lexicologia a partir da segunda metade do século XX, analisando alguns dos autores que consideramos mais marcantes e os pressupostos teórico-metodológicos que motivaram alterações teóricas e metodológicas na Lexicologia. Este constitui o objetivo primeiro a que nos propomos. Neste quadro, falaremos do reconhecimento da lexicologia como disciplina teórica e dos vários autores, associados a diversas escolas, que em muito contribuíram para essa afirmação: Jost Trier, Leo Weisgerber, Matoré, Quemada, Coseriu, Pottier, entre outros mais recentes.

O segundo escopo que procuraremos desenvolver está ligado com as redes ou *ontologias lexicais* que procuram cobrir o panorama lexical de uma língua. A ontologia lexical paradigmática é a WordNet (Fellbaum 1998), que também

é designada por WordNet de Princeton (WordNet.Pr). Seguiremos o conceito de Oliveira *et al.*, que definem a *ontologia lexical* de uma determinada língua como

Uma estrutura de conhecimento que relaciona itens lexicais (vulgo, palavras) de uma língua entre si, por relações que têm a ver com o significado desses mesmos itens; Uma estrutura que pretende abranger a língua toda e não conhecimento de um domínio em particular, ou seja, que não se encontre restrita a campos específicos (Oliveira *et al.* 2010: 77).

A realização do processamento computacional de uma língua exige aceder a recursos de ampla cobertura, como é o caso das *ontologias lexicais*, ou bases de dados lexicais, por um lado, ou, por outro lado, bases de conhecimento sobre o mundo<sup>1</sup>.

## 1. Mutações teóricas e metodológicas

No início do século passado, a estreita relação da lexicologia com a semântica decorre de uma origem e de um caminho traçado pelos comparatistas<sup>2</sup> e estreitado pelo estruturalismo.

A lexicologia só se emancipa como disciplina teórica apenas no final dos anos 1950, marcados por publicações como a de G. Matoré (*La méthode en Lexicologie*<sup>3</sup>), pelo congresso de 1957, realizado em Estrasburgo (*Lexicologie et lexicographie françaises et romanes*, 1960) e o início da publicação dos *Cahiers de lexicologie*, dirigidos por B. Quemada. Reveste-se de enorme importância distingui-la da lexicografia, disciplina aplicada que tem por objeto a elaboração de dicionários.

---

<sup>1</sup> A propósito da diferença e da relação entre ontologias e bases de dados lexicais, veja-se Hirst (2004), Dahlgren (1995) e Marcellino e Dias-da-Silva (2009).

<sup>2</sup> Mateus & Villalva (2007: 40): «Dos estudiosos comparatistas cujas obras ainda hoje são merecedoras de atenção, destacam-se Rasmus Rask (1787-1832), filólogo dinamarquês, e Franz Bopp (1791-1867), filósofo alemão, que estabeleceram princípios e métodos para o estudo comparado das línguas a partir da análise filológica de textos. A estes nomes deve acrescentar-se o de Wilhelm von Humboldt (1767-1835), linguista e político alemão que se interessou pela relação entre o homem e a linguagem [...]».

<sup>3</sup> Matoré (1953), seguindo ainda a corrente estruturalista, passa a considerar os aspetos sociais no estudo do léxico. Nesse sentido, a Lexicologia é encarada como uma disciplina sociológica. Na visão do autor, o léxico é concebido como um facto social. É o reflexo de uma sociedade, por isso, fora da mesma, a linguagem não encontra expressão.

En règle générale, dans les usages des spécialistes, comme dans les définitions des dictionnaires, la lexicologie est reconnue comme la “science du lexique” et englobe des approches aussi diverses que les théories linguistiques ou les méthodes dont elle se réclame: descriptive, historique, structurale, sociale, etc. La lexicographie, pour sa part, est “l’art et la science du lexicographe, auteur de dictionnaires (Quemada 1987: 232).

O reconhecimento da lexicologia como ciência do léxico não é pacífico. Desde o início, a lexicologia deparou-se com dois tipos de problemas que teve de resolver. O primeiro diz respeito à natureza do seu objeto de estudo: léxico/vocabulário; o segundo problema associado à sua relativa autonomia em relação às restantes disciplinas linguísticas. Niklas-Salminen (1997: 5), por exemplo, refere que, em sentido restrito, «[...] la lexicologie est considérée comme une branche de la sémantique qui a pour objet l’étude du sens des unités lexicales. Selon cette perspective, elle se confond, en grande partie, avec la sémantique structurale». Neste sentido, reveste-se de suma importância realçar, por um lado, o papel que adquire no âmbito da semântica lexical a teoria dos campos-semânticos influenciada, fundamentalmente, pelos estudos de Trier e, mais tarde, por Coseriu e, por outro, o modelo de análise sémica ou componencial. O estudo de Pottier sobre o campo lexical de assento exemplifica a combinação dos princípios subjacentes à análise sémica e à teoria dos campos (Baylon & Mignot 1995: 125, Tamba-Mecz 1998: 24). Deste modo, a lexicologia verá no dicionário o seu âmbito de análise e de ensaio para as suas metodologias e teorias.

No período temporal situado entre o final dos anos 50 e a década de 70 do século passado, os trabalhos em lexicologia evidenciam uma valorização da dimensão socio-semântica e sociolinguística do léxico (Tamba-Mecz 1998: 22-24).

Para assinalar que a própria organização das disciplinas linguísticas e o ponto de partida na descrição também podem divergir mediante a conceção que se preconiza, Coseriu (1980: 38) refere a corrente do neo-humboldtismo, representado nomeadamente na Alemanha por Jost Trier, criador da teoria do “campo semântico”, e Leo Weisgerber. Segundo esta corrente, a linguagem constituiu-se em dois estratos:

[...] o primeiro estrato representa a organização imediata do mundo por parte do homem, uma espécie de “intermundo” (em alemão: *Zwischenwelt*), responsável pela organização do mundo como tal mediante a linguagem, enquanto o segundo diz respeito ao falar sobre diversas situações do mundo, mas com os elementos dados neste intermundo. De um lado se terá então o léxico, que organiza de maneira imediata o mundo extralinguístico, de outro lado, a gramática; e a linguagem seria constituída de um estrato léxico, correspondente ao mundo enquanto conhecido e dominado pelo homem por meio da linguagem, e de outro estrato, gramatical,

correspondente à combinação desses elementos, isto é, do mundo já transformado em linguagem. A lexicologia deveria ser, portanto, a primeira disciplina linguística, enquanto estudo linguístico do modo e da ordem em que é organizado o mundo (Coseriu 1980: 38).

Coseriu prossegue a sua reflexão relativamente à corrente do neo-humboldtismo, no âmbito da conceção de lexicologia e do respetivo objeto de estudo, o léxico, mencionando que «as diferenças entre as línguas postas em relevo pelo neo-humboldtismo são sobretudo as lexicais, porque se supõe que a um léxico diferente corresponde uma maneira diferente de organizar, e, assim, de conceber a experiência do mundo exterior» (Coseriu 1980: 38).

Vilela, discípulo de Coseriu, afirma que a lexicologia se constitui nos modelos que nortearam a linguística contemporânea:

A lexicologia estuda as palavras de uma língua, em todos os seus aspetos [...] pode incluir a etimologia, a formação de palavras, a morfologia, a fonologia, a sintaxe, mas tem uma relação especial com a semântica. A lexicologia costuma ser definida como a ciência do léxico duma língua. Isto é, a lexicologia tem como objeto o relacionamento do léxico com os restantes subsistemas da língua, incidindo sobretudo na análise da estrutura interna do léxico, nas suas relações e inter-relações. [...]

Em resumo, entendemos e analisamos a lexicologia como semântica lexical. [...] Contudo, a lexicologia não tem como função inventariar todo o material armazenado ou incluído no léxico, mas sim fornecer os pressupostos teóricos e traçar as grandes linhas que coordenam o léxico duma língua. A sua função é apresentar as informações acerca das unidades lexicais necessárias à produção do discurso e caracterizar a estrutura interna do léxico, tanto no aspeto conteúdo, como no aspeto forma.

A unidade básica da lexicologia é a palavra, a que atribuímos a definição dada por Pottier a “lexie”, ou seja, a “unité de comportement syntaxique” (1967: 17) ou “unité fonctionnelle mémorisée en compétence” (1974: 326) (Vilela 1994: 9-10).

Alain Rey, nos anos 70 do século passado, denunciava o caráter transdisciplinar da ciência do léxico (Rey 1977: 169): «La vocation transdisciplinaire “science carrefour” des disciplines anthropologiques et domaine essentiel de l’activité pragmatique sur le langage, la lexicologie est dominée méthodologiquement par la linguistique. Pourtant son objet dépasse et englobe celui de la linguistique».

No sentido de reclamar uma delimitação não ambígua para o objeto da lexicologia, Delessalle & Gary-Prieur (1976: 6) afirmam o seguinte: «[...] ce que la lexicologie atteint, au travers des champs et des textes, c’est parfois un – ou des – vocabulaire (on note d’ailleurs des hésitations significatives entre ce terme et celui de lexique), ce sont des éléments de discours des éléments de société; ce n’est pas du lexique».

Ao longo da década de oitenta do século XX, a lexicologia absorve uma nova dinâmica na diversificação de áreas disciplinares e na criação de novos domínios de investigação, como consequência da renovação teórica operada em linguística.

Como exemplo, realce-se em semântica a afirmação da semântica cognitiva ao longo dos anos 1980, bem como a emergência de correntes neo-estruturalistas, como por exemplo, a semântica interpretativa de Rastier (1987). Os novos modelos teóricos e princípios metodológicos inspirados numa perspetiva sintática como o léxico-gramática de M. Gross ou numa perspetiva sintático-semântica como as *classes d'objet* (G. Gross) ou ainda o modelo teórico defendido por Pierre e Danièle Corbin vão também refletir-se na lexicologia.

O desenvolvimento de investigações centradas na vertente sociossemântica e sociolinguística do léxico focaliza-se, essencialmente, no estudo das relações paradigmáticas entre as unidades lexicais estruturadas a partir de campos léxico-semânticos e no estudo das relações sintagmáticas entre unidades lexicais organizadas a partir de fragmentos discursivos.

Como se refere em Rey-Debove (1998), consideramos que o objeto de estudo da lexicologia se situa entre o sistema, o uso e o discurso.

Com a publicação, em 1995, da obra de Igor Mel'čuk e Alain Polguère, intitulada *Lexicologie Explicative et Combinatoire*, demonstra-se a evolução crescente na vertente da lexicologia descritiva e, em simultâneo, confirma-se o desenvolvimento da lexicologia experimental.

*L'Introduction à la lexicologie explicative et combinatoire*, ou ILEC, peut à maintes occasions surprendre ou même choquer par son rigorisme extreme et sa sécheresse formaliste. En effet, ni en lexicologie ni en lexicographie, on n'a l'habitude de rencontrer des constructions quasi mathématiques, avec des systèmes de définitions logiques et tout leur appareillage de déductions. Même si plusieurs disciplines de la linguistique moderne, surtout la syntaxe et la sémantique, ont déjà franchi le pas depuis des années, il n'en est pas ainsi en lexicologie/lexicographie. Nous croyons que l'ILEC est un des premiers ouvrage qui tentent d'implanter une approche logique et formelle dans l'étude des mots. Nous nous sommes donné une tâche bien spécifique: proposer des methods de description rigoureuse, formelle et exhaustive du lexique.

Cette orientation impose inévitablement des cheminements rigides, des formulations catégoriques pouvant même agacer ou des analyses excessivement tranchées des matériaux linguistique. Dans le domaine du lexique, il est habituel de proceder avec précaution, de proposer des descriptions approximatives et nuances, en les situant de façon continue et en laissant un rôle vital au context et à l'intuition des locuteurs. Tout en reconnaissant le caractère naturel de cette façon de faire, nous avons apté pour une autre approche, celle de la prise de décisions discrètes (dans le sens mathématique) et absolues. Cette tendance s'intègre d'ailleurs parfaitement

dans l'ensemble des tendances engendrées par la révolution informatique, ce qui justifie notre tâche et confirme sa justesse (Mel'čuk *et al.* 1995: 9).

A lexicologia experimental pode definir-se como um ramo da lexicologia aplicada à dicionarística e à lexicografia. Neste sentido, o dicionário é concebido como campo de ensaio para testar teorias e metodologias.

Igor Mel'čuk apresenta a obra *L'Introduction à la lexicologie explicative et combinatoire* como uma teorização do *Dictionnaire Explicativo e Combinatório do Francês*, DEC, que se inscreve nesta nova linha de investigação. «L'ILEC est en fait une theorization du DEC; nous avons essayé de généraliser et de systématiser notre expérience lexicologique/lexicographique et, en même temps, d'établir des objectifs idéalisés à atteindre. En un mot, nous voulons proposer un modèle dictionnaire à discuter, à développer et à imiter» (Mel'čuk *et al.* 1995: 10).

O DEC é definido pelos seus autores como um dicionário de lexicólogo ou ainda um dicionário teórico e um modelo idealizado, que é acessível apenas a especialistas. Este dicionário não é um dicionário de lexicógrafo, ou seja, não é um dicionário de língua francesa concebido para ser utilizado pelo público em geral. Trata-se de um modelo de dicionário que preconiza uma teoria linguística: a teoria *Sens-Texte*<sup>4</sup>.

Nous avons résolument bâti notre approche sur la théorie linguistique Sens-Texte, qui sert de charpente et de fil directeur et nous fait aboutir à un modèle de dictionnaire, qui est un dictionnaire théorique – non pas le sens qu'il n'a pas d'existence concrète, mais dans l'acception qu'il répond à une certaine vision théorique de la réalité linguistique, que son pouvoir descriptif est maximalisé, et donc que la justesse de la présentation des faits de langue rend justice à la théorie sous-jacente. Un tel dictionnaire correspond, nous en sommes convaincus, à un dictionnaire idéalisé (Mel'čuk *et al.* 1995: 5-6).

No âmbito da lexicologia, Mel'čuk, situando-se no quadro da Teoria Linguística *Sens-Texte*, assume uma atitude inovadora:

Avec cette mise en relief de l'importance capitale du lexique dans la langue, on comprend que nous voulons placer l'étude du lexique, c'est-à-dire la lexicologie, au cœur de la linguistique théorique. Cette attitude n'est pas très habituelle non plus. À la fin du XX siècle, la lexicologie a encore peu d'importance et dans les manuels et dans l'enseignement de la linguistique (Mel'čuk *et al.* 1995: 17).

---

<sup>4</sup> A propósito da Teoria Sens-Texte, veja-se Brito (2010: 99-114).

Com esta mudança de paradigma, o âmbito dos estudos lexicais evidenciou alterações decorrentes de uma nova conceção do léxico, isto é, o léxico de uma língua não é atualmente perspetivado como um mero repositório das unidades lexicais e das suas idiossincrasias, configurando uma vertente da língua dissociada da gramática, mas antes como uma parte da gramática. Os estudos do léxico passaram também a integrar o objetivo de compreender a forma de estruturação do conhecimento lexical, adquirido e processado a nível mental, passando a linguística a conceber o léxico pluridimensionalmente. Consequentemente, a apreensão do conhecimento lexical necessita da descrição da própria gramática das unidades lexicais.

A reflexão levada a cabo pela psicolinguística permite-nos entender a constituição, a aquisição, bem como a forma de processamento do saber lexical. Os avanços da neurobiologia e das ciências cognitivas revestem-se de suma importância para a compreensão do conceito de léxico mental, enquanto módulo interativo e paralelo situado entre as estruturas concetuais e as linguísticas.

O léxico passou, então, a ser perspetivado como uma rede de itens lexicais, os quais se encontram associados por nexos semânticos e concetuais, imprescindíveis para a estruturação do léxico. Neste sentido, surgiu o conceito de *WordNet*, que nos permite navegar através das palavras: cada palavra cria uma rede de outras palavras e de outros conceitos.<sup>5</sup>

## 2. Modelos de Ontologia Lexical

### 2.1. *WordNet*

A *WordNet.Pr* é uma ontologia lexical para a língua inglesa, que visa representar o modo como o ser humano processa o vocabulário. Encontra-se dispo-

---

<sup>5</sup> O conceito de *WordNet* decorre da designação de uma base de dados de palavras, concebida para o Inglês sob a direção de George A. Miller (Miller 1990: 235-244) e constituída por palavras (nomes, verbos, adjetivos, advérbios) agrupadas por relações semânticas de base cognitiva, cada uma expressando um conceito. Visando desenvolver uma *WordNet* de ampla cobertura que incluísse todas as variedades do Português (Angola, Brasil, Cabo Verde, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste), realizou-se o projeto *WordNet.PT*<sup>global</sup> - *Rede Léxico-Conceptual das variedades do Português*, que foi financiado pelo Instituto Camões no biénio 2009-2010. Na sequência do projeto *WordNet* para as variedades do Português, que permitiu a codificação numa rede léxico-concetuual de todas as variedades do Português, o projeto *WordNet.PT*global tem como resultado a disponibilização de uma rede de 10 000 conceitos, integrando expressões das categorias nominal, verbal e adjetival, com as

nível *on-line*, em <http://wordnet.princeton.edu>. A sua estrutura mais básica diz respeito a um grupo de sinónimos, isto é, um conjunto de palavras que podem possuir o mesmo significado e ser utilizadas para representar o mesmo conceito, num determinado contexto. Através de ligações, é criada uma rede semântica na WordNet.Pr, que corresponde a relações semânticas, entre os nós, correspondentes aos grupos de sinónimos. Na WordNet.Pr, verifica-se uma nítida distinção entre nós que são substantivos, verbos, adjetivos, advérbios ou palavras gramaticais.

Na sequência do sucesso da WordNet.Pr, o modelo da mesma inspirou a representação de ontologias lexicais em outras línguas, como por exemplo, as wordnets concebidas para as línguas patentes no projeto EuroWordNet (Vossen 1997; Vossen 1998).

Relativamente às wordnets criadas para a língua portuguesa, destacam-se a WordNet.PT, para a variante do português europeu, e a WordNet.BR, para a variante do português do Brasil. Para aceder às mesmas, temos o endereço <http://cvc.instituto-camoes.pt:8080/wordnet/index.jsp>.

Destacamos também o projeto MultiWordNet (Pianta, Bentivogli & Girardi 2002), que pretendeu, de igual modo, alinhar diversas wordnets com a WordNet.Pr. Embora este projeto fosse inspirado pelo EuroWordNet, o objetivo foi criar novas wordnets nas quais fosse mantida a maioria dos nós e das relações estabelecidas na WordNet.Pr. Neste sentido, na MultiWordNet, encontram-se alinhadas com a WordNet.Pr wordnets para o português, acessíveis em <http://mwnpt.di.fc.ul.pt/>, wordnets para o latim, para o italiano, o espanhol, o romeno e o hebraico.

O TeP, *Thesaurus Eletrónico de Português*, foi desenvolvido em 2000-2001 pela equipa liderada por Bento Dias-da-Silva (Dias-da-Silva & Moraes 2003; Dias-da-Silva *et al.* 2000). Este projeto tinha em vista a elaboração de um recurso que fosse possível utilizar como *thesaurus* de um processador de texto, no sentido de auxiliar os respetivos utilizadores a descobrirem palavras distintas que lhes permitissem expressar as suas ideias. A versão 2.0 do TeP surgiu em 2008 (Maziero *et al.* 2008). Diz respeito à mesma base de dados, mas encontra-se disponível na Web<sup>6</sup>, com uma interface que faculta o acesso interativo a esses dados. Os dados encontram-se representados de modo semelhante à WordNet,

---

correspondentes lexicalizações de todas as variedades do Português, associados a glosas e integrados numa rede de mais de 40 000 relações de vários tipos, e apresentando etiquetas de registo para todas as variantes codificadas. O Centro de Linguística de Lisboa tem a WordNet disponível em <http://www.clul.ul.pt/wnglobal/index.jsp>.

<sup>6</sup> <http://www.nilc.icmc.usp.br/tep2/busca.php>

tendo em vista uma evolução nesse sentido, o que se veio a verificar através do projeto WordNet.BR.

Segundo Santos *et al.* (2010: 686-687), contrastando com a metodologia utilizada no MindNet ou no PAPEL, a construção do TeP implicou um trabalho principalmente manual, relativamente a um *corpus* de referência formado por vários dicionários disponíveis eletronicamente, mas pensados unicamente para a utilização humana. Os dados estão disponíveis em ficheiro de texto, com uma estrutura simples, constando em um grupo de sinónimos por linha, como a seguir se mostra:

11519. [Substantivo] {aberta, saída, solução} <16854>, em que o primeiro campo representa o identificador do nó e o segundo a sua categoria gramatical. O conjunto representado entre chavetas é o grupo propriamente dito e o identificador que pode surgir no final consiste numa referência para um nó com um sentido oposto (antónimo) ao corrente (Santos *et al.* 2010: 686-687).

Em relação ao TeP, a autora refere ainda que

Existem 19.885 destes registos, dos quais 4.312 se encontram relacionados pela relação de antonímia. (Ou seja, existem 2.156 relações de antonímia entre nós.) O número médio de palavras por nó é 3,8, enquanto que a moda é 2. Constata-se que 85% dos nós têm menos de seis palavras, enquanto que o maior nó compreende 53 palavras.

Se fizermos o exercício de desmultiplicar estes grupos em triplos, seja de sinonímia, seja de antonímia, ignorando a informação relativa à categoria gramatical, constatamos que o TeP 2.0 representa 202.514 relações de sinonímia e 49.141 relações de antonímia, envolvendo um total de 44.325 palavras distintas (com 75.713 sentidos diferentes) (Santos *et al.* 2010: 687).

## 2.2. MindNet

A MindNet, enquanto recurso que pode ser encarado como uma ontologia lexical, representa uma metodologia constituída por um conjunto de instrumentos para adquirir, estruturar, aceder e explorar, de forma automática, informação semântica integrada em texto, definindo um amplo âmbito de relações semânticas. Segundo Oliveira & Gomes (2007: 8),

Para a construção do MindNet foi utilizado o parser PLN utilizado na verificação gramatical do Microsoft Word 97 que se revelou essencial e que é também a base para o crescimento do MindNet. Este parser produz árvores sintáticas e formas lógicas sobre as quais são aplicadas regras para a extração de relações semânticas. A forma de encontrar semelhanças entre palavras combina a medição da semelhança substi-

tucional e também na procura de clusters de palavras geralmente relacionadas. Há também um procedimento que, a partir das relações encontradas procura inferir novas relações. A maior parte da desambiguação de sentidos é feita com o auxílio de informação morfológica e sintática obtida na altura do parsing. Outras formas de desambiguação passam pela informação acerca do domínio de cada palavra (presente nos dicionário) e também na análise de relações entre as várias palavras. No MindNet existe a distinção entre as quatro classes gramaticais abertas, mas as relações não estão restringidas a nenhuma em particular. Existe um interface web para o MindNet, o MNEX (MindNet Explorer), criado com o objetivo de facilitar as buscas feitas à estrutura e às relações do MindNet [VKSM05]. O utilizador insere duas palavras e o MNEX mostra todos os caminhos de relações semânticas entre essas duas palavras.<sup>7</sup>

Desta forma, a metodologia da MindNet permite que para cada palavra definida exista, além da informação típica num dicionário, por exemplo, informações gramaticais, um conjunto de registos associados aos diversos sentidos que a palavra pode assumir.

Por sua vez, para cada sentido, além da definição, juntam-se ligações a outras entradas, de forma que cada ligação tem um tipo correspondente a uma relação gramatical, por exemplo, sujeito típico, predicado típico, ou semântica, por exemplo, sinónimo, hiperónimo, parte, causa, finalidade, maneira. Estas relações são conseguidas com base na aplicação de regras sobre árvores sintático-semânticas, realizadas por um analisador sintático de ampla cobertura. Cada relação produzida possui um peso atribuído em conformidade com a sua saliência.

### 2.3. FrameNet

A FrameNet (Baker, Fillmore & Lowe 1998) consiste numa rede semântica baseada no conceito de enquadramentos (do inglês, *frames*) (Fillmore 1982). Nesta representação, cada enquadramento procede à descrição de um objeto, um evento ou um estado, correspondente a um conceito e é possível relacioná-lo com outros enquadramentos, partindo de um conjunto de relações semânticas, por exemplo, herança, sub-frame, causador, utiliza<sup>8</sup>. No âmbito do português, existe um projeto representativo deste modelo de recurso<sup>9</sup>, o FrameNet Brasil<sup>10</sup>, que tem como objetivo desenvolver, com base na Semântica de Frames e na Gramática das

<sup>7</sup> Citamos o texto de fonte *on-line*: <http://linguateca.dei.uc.pt/papel/relatorio1.pdf>.

<sup>8</sup> Os exemplos são de Oliveira *et al.* (2010: 80).

<sup>9</sup> Veja-se Salomão (2009) e Afonso (2009).

<sup>10</sup> A este propósito, consulte-se <http://www.framenetbr.ufjf.br/>.

Construções, recursos lexicais e sintáticos para o português do Brasil que estejam disponíveis *on-line* e que possam ser usados no desenvolvimento de soluções em Linguística Computacional e no Processamento de Linguagem Natural.

Oliveira e Gomes (2007: 10) apresentam-nos algumas diferenças entre WordNet e FrameNet:

Algumas diferenças entre o FrameNet e o WordNet [REP+06]:

- O FrameNet contém vários exemplos anotados de cada unidade lexical ou significado. Os exemplos consistem em texto natural, retirado de corpora.
- A análise contida no FrameNet é realizada enquadramento por enquadramento e não por lema.
- A rede de relações do WordNet é feita entre os seus nós, enquanto a rede de relações do FrameNet é feita entre os enquadramentos semânticos. O FrameNet contém uma rede de relações entre frames. De seguida apresentam-se as mais importantes:
  - Inheritance (herança): A típica relação is-a (hiponímia). Exemplo: Revenge (vingança) herda de Rewards\_and\_punishments (recompensas e punições).
  - Using: a frame “filha” pressupõe a frame pai. Exemplo: Speed (velocidade) utiliza Motion (movimento).
  - Subframe: a frame filha\_ é um subevento da frame pai, que será um evento mais complexo. Exemplo: Arrest (detenção), Arraignment (acusação), Trial (julgamento) e Sentencing (sentencição) utilizam Criminal\_process (processo criminal).
  - Perspective on: a frame “filha” representa uma perspectiva da frame pai. Exemplo: Hiring (contratar) e Get\_a\_job (arranjar emprego) perspectivam a frame Employment\_start (início de um emprego) respectivamente do ponto de vista do empregador e do empregado. (Oliveira & Gomes 2007: 10)

Moreira e Salomão (2012) apontam algumas limitações à FrameNet em detrimento da WordNet:

A FrameNet possui o mérito de promover o entendimento de um lexema a partir das cenas em que é empregado. Essa base é um rico repositório de informações semânticas com um nível de detalhamento fino (nível fino de granularidade). Porém, a FrameNet possui algumas limitações, que dificultam seu uso em certas aplicações. Uma das limitações da FrameNet é a falta de formalização. A FrameNet é armazenada em uma base textual semiestruturada em HTML (*HyperText Markup Language*), o que dificulta seu uso por sistemas computacionais. Existe uma formalização em OWL DL1 (SCHEFFCZYK, 2006), mas sem dispor de uma axiomatização, ou seja, foram listados em OWL-DL os tipos de elementos que ocorrem na FrameNet, mas não foram definidas, por meio de axiomas, as restrições e as propriedades associadas a esses tipos. Outra queixa comum a respeito da FrameNet é a sua baixa cobertura lexical, inferior à da *WordNet* (12000 lexemas contra 155000) (Moreira & Salomão 2012: 498).

## Conclusão

No período temporal situado entre o final dos anos 50 e a década de 70 do século passado, os trabalhos em lexicologia evidenciam uma valorização da dimensão sociosemântica e sociolinguística do léxico.

A lexicologia, que estuda cientificamente o léxico de uma língua, ou seja, o repositório das palavras e de todas as suas propriedades, em meados do século XX, orienta-se pelos modelos estruturalistas que determinam os princípios de análise do léxico.

Decorrente da investigação nas diversas disciplinas da linguística contemporânea, acedemos, atualmente, a um conhecimento mais cabal dos diversos tipos de saberes associados às unidades lexicais, saberes esses que contemplam a sua forma oral e escrita, a sua estrutura interna, a flexão e as relações com palavras da mesma família, os respetivos significados, o seu alcance denominativo, as relações com outras unidades no plano sintático, as suas combinatórias, bem como as relações semânticas estabelecidas com outras palavras.

A nova dinâmica sentida em lexicologia nas últimas décadas do século XX resulta da emergência de novos paradigmas em semântica, da afirmação da pragmática, da sociolinguística, da linguística cognitiva e da psicolinguística, bem como da linguística computacional.

Nesta linha, o PLN (Processamento Automático das Línguas Naturais), desde muito cedo, anunciou a indispensabilidade de recursos de ampla cobertura para poder passar de protótipos para sistemas reais. Por isso, um dos caminhos para atingir esse objetivo era usar repositórios vastos já criados e que permitissem, de certo modo, compensar a necessidade de grande trabalho na produção de recursos. Nesse sentido, procedeu-se ao uso de dicionários e de outros materiais lexicográficos como base para a estruturação de conhecimento. As ontologias lexicais constituem o resultado desta modificação ou melhoria de materiais lexicográficos, tendo em vista o uso no processamento da língua, isto é, as ontologias lexicais são as estruturas representativas desse conhecimento.

## Referências

- Afonso, Susana. (2009): Uma FrameNet para o português. Apresentação na Escola de Verão Belinda Maia (Edv 2009), Porto, Portugal. Disponível para descarregar em <http://www.linguateca.pt/Repositorio/AfonsoFrameNetEdV2009.pdf>.
- Baker, Collin F., Charles J. Fillmore & John B. Lowe (1998): The Berkeley FrameNet Project. In: *Proceedings of the 17th International Conference on Computational linguistics*. Morristown, NJ, EUA: Association for Computational Linguistics, 86-90.

- Baylon, Ch. & X. Mignot (1995): *Semantique du language. Initiation*. Paris: Nathan.
- Brito, A. M. (2010): Gramáticas descritivas do Português. In: Cornelia Doll, Sybille Große, Christine Hundt & Axel Schönberger (orgs.), *De arte grammatica: Festschrift für Eberhard Gärtner zu seinem 65 Geburtstag*. Frankfurt am Main: Valentia, 39-55.
- Corbin, Danielle & Pierre Corbin (1991): Un traitement unifié du suffixe –ier(e). *Lexique* 10: 61-145.
- Coseriu, E. (1980): *Lições de Linguística Geral*. Tradução do Prof. Evanildo Bechara. Rio de Janeiro: Ao Livro Técnico.
- Dahlgren, Kathleen (1995): A linguistic ontology. *International Journal Human Computer Studies* 43 (5-6): 809-818.
- Delessale, S. & Gary-Prieur (1976): Le Lexique, entre la lexicologie et l'hypothèse lexicaliste. *Langue Française* 30: 4-33.
- Dias-da-Silva, B. C. & H. R. Moraes (2003): A construção de um thesaurus eletrônico para o português do Brasil. *ALFA*, Vol. 47, N. 2: 101-115. Disponível para descarregar em [http://www.nilc.icmc.usp.br/tep2/Dias-da-Silva\\_et\\_al.ALFA.2003.pdf](http://www.nilc.icmc.usp.br/tep2/Dias-da-Silva_et_al.ALFA.2003.pdf).
- Dias-da-Silva, B. C., H. R. Moraes, M. F. Oliveira, R. Hasegawa, D. A. Amorim, C. Paschoalino & A. C. Nascimento (2000): Construção de um thesaurus eletrônico para o português do Brasil. *Processamento computacional do português escrito e falado* (PROPOR), Vol. 4: 1-10. Disponível para descarregar em [http://www.nilc.icmc.usp.br/tep2/Dias-da-Silva\\_et\\_al.PROPOR.2000.pdf](http://www.nilc.icmc.usp.br/tep2/Dias-da-Silva_et_al.PROPOR.2000.pdf).
- Fellbaum, Christiane (1998): *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.
- Fillmore, Charles J. (1982): Frame semantics. In: Linguistic Society of Korea (eds), *Linguistics in the morning calm*. Seoul, Coreia do Sul: Hanshin Publishing Co., 111-137.
- Gross, Gaston (1989): *Les constructions converses du français*. Genève, Paris: Droz, Coll. Langue et Cultures n.º 22.
- Gross, Maurice (1975): *Méthodes en syntaxe*. Paris: Hermann.
- Hirst, Graeme (2004): Ontology and the léxicon. In: Steffen Staab & Rudi Studer (eds), *Handbook on Ontologies*. Germany: Springer, 209-230.
- Marcellino, Erasmo Roberto & Bento Dias-da-Silva (2009): Sistematização linguístico-computacional do léxico do domínio conceitual Indústria do Bordado de Ibitinga. In: *The 7th Brazilian Symposium in Information and Human Language Technology (STIL 2009)*, 8-11 de Setembro, 2009.
- Mateus, Maria Helena Mira & Alina Villalva (2007): *Linguística*. Lisboa: Editorial Caminho.
- Matoré, George (1953): *La Méthode en Léxicologie*. Paris: Didier.
- Mazieiro, E. G., T. A. S. Pardo, A. Di Felippo & B. C. Dias-da-Silva (2008): A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. *VI Workshop em tecnologia da informação e da linguagem humana* (TIL): 390-392. Disponível para descarregar em <http://www.nilc.icmc.usp.br/tep2/TIL2008-MazieiroEtAl.pdf>.
- Mel'čuk I., A. Clas & A. Polguère (1995): *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve/Paris: Éd. Duculot/AUPELF UREF.
- Miller G. A., R. Beckwith, C. D. Fellbaum, D. Gross & K. Miller (1990): WordNet: An online lexical database. *Int. J. Lexicograph*, 3, 4: 235-244.
- Moreira, Alexandra & Maria Margarida Martins Salomão (2012): Análise ontológica aplicada ao desenvolvimento de frames. *Alfa*, 56 (2): 491-521. Disponível para descarregar em <http://www.ufjf.br/framenetbr/publicacoes/>.
- Niklas Salminen, Aino (1997): *La Lexicologie*. Paris: Armand Colin.

- Oliveira, Hugo Gonçalves, Paulo Gomes & Diana Santos (2007): *PAPEL - Trabalho relacionado e relações semânticas em recursos semelhantes*. Coimbra: Departamento de Engenharia Informática, FCTUC, CISUC. Disponível para descarregar em <http://linguateca.dei.uc.pt/papel/relatorio1.pdf>.
- Oliveira, Hugo Gonçalves, Diana Santos & Paulo Gomes (2010) : *Extração de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação*. Disponível para descarregar em <http://www.lbd.dcc.ufmg.br/bdbcomp/servlet/Trabalho?id=9186>.
- Pianta, Emanuele, Luisa Bentivogli & Christian Girardi (2002): MultiWordNet: developing an aligned multilingual database. In: *Proceedings of the First International Conference on Global WordNet*, Mysore, India, January 21-25. Disponível para descarregar em MultiWordNet: developing an aligned multilingual database”
- Pottier, B. (1973): *Le Langage*. Paris: Denoel.
- \_\_\_\_\_ (1992): *Sémantique générale*. Paris: Presses Universitaires de France.
- Quemada, B. (1987): Notes sur lexicographie et dictionnaire. *Cahiers de Lexicologie* 51-2: 229-242.
- Rastier, F. (1987): *Sémantique interprétative*. Paris: Presses Universitaires de France.
- Rey-Debove, J. (1998): *La Linguistique du signe. Une approche sémiotique du langage*. Paris: Armand Colin.
- Rey, A. (1977): *Le Lexique – images et modèles*. Paris: Armand Colin.
- Salomão, Maria M. M. (2009): Framenet Brasil: Um trabalho em progresso. *Calidoscópio* 7 (3): 171-182.
- Santos, Diana, Anabela Barreiro, Cláudia Freitas, Hugo Gonçalves Oliveira, José Carlos Medeiros, Luís Costa, Paulo Gomes & Rosário Silva (2010): Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL. In: A. M. Brito, F. Silva, J. Veloso & A. Fiéis (eds.), *Textos seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*, 681-700. Disponível para descarregar em <http://www.apl.org.pt/apl-actas/xxvi-encontro-nacional-da-associacao-portuguesa-de-linguistica.html>.
- Tamba-Mecz, I. (1998): *La sémantique*. Paris: Presse Universitaires de France.
- Trier, Jost (1931): *Der deutsche Wortschatz im Sinnbezirk des Verstandes*. Heidelberg: C. Winter.
- Vilela, Mário (1994): *Estudos de lexicologia do português*. Coimbra: Almedina.
- Vossen, Piek (1998): *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.
- Vossen, Piek (1997): Eurowordnet: a multilingual database for information retrieval. In: *Proceedings of the DELOS workshop on Cross-Language Information Retrieval*, Zúrique, Suíça, 5-7 de Março, 1997.
- Weisgerber, Leo (1973): *Die sprachliche Gestaltung der Welt*. Düsseldorf: Schwann.