

## **Dicionário-Aberto: Construção semiautomática de uma funcionalidade codificadora**

Alberto Simões, Álvaro Iriarte S., José João Almeida

27 Congrès International de Linguistique et de Philologie Romanes  
Section 5. Lexicologie, phraséologie, lexicographie.

### **Sobre o Dicionário Aberto**

O *Dicionário Aberto* (DA, <http://www.dicionario-aberto.net>) iniciou em Junho de 2005, como a transcrição dos dois volumes do *Novo Dicionário da Língua Portuguesa*, de Cândido de Figueiredo, de 1913, transcrição essa ao abrigo do *Projeto Gutenberg*.

A transcrição, na grafia original, terminou em 2010. Desde então o dicionário tem sido sujeito a um conjunto de alterações, como a transformação para um formato standard (TEI - Text Encoding Initiative - <http://www.tei-c.org/>), a validação da notação usada, e a modernização da ortografia. Ao mesmo tempo, a aplicação web também tem evoluído, incorporando diversos tipos de pesquisa (pela cabeça do verbete, por afixo ou ainda a pesquisa reversa), uma *Application Programmer Interface* (API) para pesquisa remota, entre outras funcionalidades. Mais recentemente incorpora-se um sistema semiautomático para a extração de hiperónimos/hipónimos e merónimos/holónimos.

### **O dicionário codificador**

Um dicionário codificador ou de produção é uma ferramenta pensada para ajudar o utilizador a produzir textos numa língua. Um dicionário deste tipo fornece abundante informação morfossintática, léxico-semântica e pragmática, necessária para ajudar à codificação linguística, uma vez que não dispomos de estratégias de tipo textual ou pragmático que nos permitam inferir o significado de determinada palavra ou combinação lexical, tal como acontece na atividade descodificadora.

Nos séculos XIX e XX foram elaborados alguns dicionários ideológicos (ou onomasiológicos, ou analógicos ou conceptuais) que permitiam ao leitor pesquisar uma ideia ou um conceito numa lista de descritores, semelhante a um *thesaurus*, ou numa lista estruturada de conceitos ordenados por assuntos, que conduziam o leitor à palavra supostamente procurada.

## **Construção de uma faceta codificadora sobre o Dicionário Aberto**

As possibilidades de pesquisa reversa transformam os dicionários eletrônicos em bases de dados conceptuais ou ideológicas, que poderíamos também chamar dicionários onomasiológicos, ideológicos, analógicos ou conceptuais.

Mas o nosso propósito vai para além de num dicionário onomasiológico tradicional. Não se pretende que o utilizador procure a informação numa espécie de ontologia ordenada por matérias, que contém uma série de hiperónimos ou categorias que permitem, hipoteticamente, chegar à palavra procurada (usando a ontologia apenas como uma estrutura classificativa).

O que se pretende é que, utilizando as funcionalidades de pesquisa reversa do Dicionário Aberto, o utilizador possa procurar unidades lexicais relacionadas (sinónimos, quase-sinónimos, hiperónimos, hipónimos, merónimos, holónimos, coocorrentes, etc.) a partir de um conjunto de palavras. Por exemplo, que com base nas palavras “*rijo*” e “*metal*” se consiga encontrar termos como “*temperar*”.

Para que isto seja possível, será necessário que o sistema de pesquisa não use apenas os termos introduzidos pelo utilizador e as respetivas ocorrências nas definições, mas também uma estrutura semântica, que providencie relações léxico-conceptuais com os termos introduzidos. Estas relações serão cruzadas e serão calculadas medidas de proximidade, sendo deste modo possível apresentar um conjunto de resultados ordenados por (uma métrica de) relevância.

## **Referências**

- [1] Simões, A., & Farinha, R. (2011). Dicionário Aberto: Um novo recurso para PLN. *Vice-Versa*, 16, 159–171.
- [2] Simões, A., Iriarte Sanromán, A., & Almeida, J. J. (2012). Dicionário-aberto – a source of resources for the portuguese language processing. In Caseli, H., Villavicencio, A. Teixeira, A., & Perdigão, F. (eds.) *Computational Processing of the Portuguese Language, Lecture Notes for Artificial Intelligence*, 7243: 121–127.
- [3] Leal, J. P., Rodrigues, V., & Queirós, R. (2012). Computing Semantic Relatedness using DBpedia. In A. Simões, R. Queirós, & D. C. da Cruz (ed.), *'SLATE'*, *Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik* (pp. 133-147).
- [4] Oliveira, H. G., Pérez, L. A., Costa, H. & Gomes, P (2011). Uma rede léxico-semântica de grandes dimensões para o português, extraída a partir de dicionários electrónicos. In A. Simões, X. Gómez Guinovart, & J. J. Almeida (ed.) (2011) *Linguamática* (3) 2., 23-38.
- [5] Mendes, P. N., Jakob, P. N., & Bizer, C. (2012) DBpedia for NLP: A Multilingual Cross-domain Knowledge Base. In Calzolari et al (ed.), *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2012* (21–27). Istanbul, Turkey.