Abs-256

Irene Renau and Araceli Alonso (Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona)

Using corpus pattern analysis for the Spanish Learner's Dictionary DAELE (Diccionario de aprendizaje del español como lengua extranjera)

Learner's dictionaries for non-native speakers are based on the assumption that users look up words in the dictionary not only for decoding texts, but also for encoding. In this sense, more grammatical information is included as the main target users, who do not have full command of the grammar of the language, need information about the use of words in a specific context. Corpus linguistics has brought a light for the building-up of dictionaries in the English lexicographical tradition, and specifically, in the elaboration of learner's dictionaries. Spanish lexicographical tradition, on the contrary, has dedicated itself mainly to elaborating monolingual dictionaries for native speakers, and as a result, currently there are neither many learners' dictionaries nor many Spanish learners' dictionaries which take into account the use of words in texts.

Corpus-driven methodologies have been applied for English learners' dictionaries for the selection of headwords, for the organisation of data and for establishing the uses of a given word that are to be included in the dictionary (Atkins & Rundell 2008; Rundell 2002). One of these methodologies has been developed by Patrick Hanks for the creation of a Pattern Dictionary of English Verbs (Hanks 2004, 2005). Corpus Pattern Analysis (CPA) is a corpus-driven method to establish syntactic and semantic verbal patterns. It is based on the Theory of Norms and Exploitations - TNE (Hanks 2004, forthcoming) which establishes that there are norms for a given word which can be exploited to create specific uses in a given context. TNE and CPA establish a theoretical and practical framework for "mapping meaning onto words in text". Although CPA has been developed for English language, it is also being applied to other languages such as English, Italian or Spanish.

In previous studies (Alonso 2009; Alonso & Renau, forthcoming), it has been stated that CPA is not only a useful technique for studying the main common patterns of a word, but also for determining terminological uses of a given word in a specific context. Actually, CPA is being analysed for being applied in the case of the Spanish Learners' Dictionary DAELE (Bernal & Renau 2010; http://www.iula.upf.edu/rec/daele/).

In the present study, the use of CPA applied to Spanish language is shown by analysing different Spanish verbs. Several entries of Spanish verbs created by using CPA are compared to the same entries created without using this methodology. As it is well known, corpus analysis implies a thoroughly study of occurrences of a word. If the process of going through the occurrences is not systematized, the lexicographer may make mistakes at the time of interpreting the data. Our main objectives are, in one hand, to illustrate in which way CPA contributes to the systematization of this process and improves the way of writing dictionary entries and, on the other hand, to show the potential of CPA as a unique method for being applied to several languages, specifically, to Romance languages.

References:

Alonso, A. (unpublished, 2009) Características del léxico del medio ambiente en español y pautas de representación en el diccionario general. [PhD dissertation]. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Alonso, A. & Renau, I. (forthcoming). "Los verbos españoles y el Corpus Pattern Analysis de Hanks: primera propuesta de adaptación". IV Congreso Internacional de Lexicografía Hispánica. Tarragona (Spain), 20-22 September 2010. Tarragona: Universitat Rovira i Virgili.

Atkins, B. & Rundell, M. (2008). The Oxford Guide to Practical Lexicography. Oxford: Oxford University Press.

Bernal, E. & Renau, I. (2010). "¿Lo que necesitan es lo que encuentran? Reflexiones a propósito de la representación de los verbos en los diccionarios de aprendizaje de español". In A. Dykstra, Schoonheim, T., (eds.). Proceedings of the XIV Euralex International Congress. Ljowert (Nederlands): Fryske Akademy, pp. 484-496.

Hanks, P. (2004). "The Syntagmatics of Metaphor and Idioms". International Journal of Lexicography, Volume 17, Number 3. Oxford: Oxford University Press. 245-274.

Hanks, P. (forthcoming). Lexical Analysis: Norms and Exploitations. Cambridge: MIT Press.

Hanks, P. & Pustejovsky, J. (2005). "A Pattern Dictionary for Natural Language Processing". Revue Francaise de linguistique appliquée, 10:2. 63-82.

Rundell, M. (2002). "Good old fashioned lexicography: Human judgement and the limits of automation". In Corréard, M. H. (ed.) Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins. Grenoble, France: EURALEX. 138-155.