

Sobre unidades de traducción y su reconocimiento automático

Patricia Fernández Carrelo (*patty.fil@gmail.com*)
Universidad de Deusto (Bilbao, Spain)

PAPER

El desarrollo alcanzado en las últimas décadas en herramientas y técnicas de procesamiento del lenguaje natural aplicadas al campo de la traducción ha supuesto grandes avances en la realización automática de este complejo ejercicio lingüístico. No obstante, quedan aún espacios en los que la labor del lingüista resulta de vital importancia para la obtención de resultados de calidad, especialmente en la caracterización de las unidades en las que se ha de descomponer el texto dentro del proceso traductológico.

Tal y como recoge Abaitua (1997), "Toury (1995) sugiere que, en la traducción de un texto, 'lo ideal sería traducir de una tacada el texto completo'. Sin embargo, a no ser de que se trate de un texto muy breve, esto es humanamente inviable y el traductor necesita proceder por partes." Así, mientras el traductor humano pone en juego sus saberes lingüísticos y conocimientos del mundo para llevar a cabo este proceso de segmentación, las herramientas automáticas necesitan de formalizaciones lingüísticas que les permitan discriminar las unidades de traducción de manera objetiva.

Es indudable, por tanto, el interés de aquellas secuencias de palabras que actúan como un mismo bloque en determinados niveles lingüísticos y que, en el campo de la traducción, han de tratarse como una unidad única. Estas unidades, entendidas como combinaciones lexicalizadas de palabras que establecen dependencias morfosintácticas entre ellas, suponen la unidad base de estudio en el presente proyecto. Partimos del supuesto de que las lexías complejas o unidades poliléxicas (*multiword units*), se sitúan en un nivel superior al de las colocaciones, en cuanto a grado de lexicalización se refiere, pero sin llegar a alcanzar la categoría de locuciones, giros idiomáticos o fórmulas, elementos discursivos de mayor complejidad y fijación en la lengua.

Así, la presente comunicación pretende asomarse a las distintas tipologías existentes para la clasificación de estos elementos morfoléxicos, junto a algunas de las herramientas existentes para su procesamiento automático en el par de lenguas euskara-castellano. Estas dos lenguas, muy lejanas tipológicamente pero cercanas geográficamente, se encuentran aún en proceso de análisis y desarrollo en el ámbito de la traducción automática, y es cada vez mayor la aparición de herramientas orientadas a esta tarea, por lo que la recopilación de estos datos puede resultar de gran interés.

Para ello, se efectuará un recorrido por la amplia tradición teórica recogida en Abaitua (2001) y obras posteriores, como Hurtado Albir (2001), a la vez que se indagará en los procedimientos automáticos para el reconocimiento de este tipo de unidades de traducción. Puesto que en la base de estos procesos se sitúan los analizadores morfosintácticos, se partirá del estudio de las aportaciones del Grupo IXA para el euskara (fundamentalmente MORFEUS y EUSLEM), así como de otras herramientas disponibles para el castellano: FreeLing y SVMTool, tecnologías del Centro de investigación TALP (Universitat Politècnica de Catalunya y Centre de Llenguatge i Computació, Universitat de Barcelona); CLiC (Universitat de Barcelona); FLAPE y SCOGEME (Universidad de Las Palmas) o DATALEXICA (Empresa Bitext), entre otros. No obstante, no se profundizará en la caracterización de estas herramientas por no tratarse de objeto de estudio de este trabajo.

Dado que las unidades que analizamos (denominadas en euskera *hitz anitzeko unitate lexikalak* -HAUL-), han sido ya tratadas desde la perspectiva monolingüe, pero no entendidas como unidad de traducción, se atenderá a un enfoque multilingüe en la revisión de las numerosas tipologías existentes para la clasificación de estas locuciones léxicas, basadas en criterios que van desde la composicionalidad semántica hasta la disposición sintáctica de los elementos.

Bibliografía

- Abaitua, J. 2001. *Memorias de traducción en TMX compartidas por internet. Tradumática 0*. <<http://www.fti.uab.es/tradumatica/revista/articles/jabaitua/art.htm>>
- Abaitua, J. 1997. *Segmentos y unidades de traducción*. <<http://paginaspersonales.deusto.es/abaitua/konzeptu/12uutts.htm>> (April 2006)
- Aduriz I., I. Aldezabal, I. Alegria, et al. 1996. 'Del analizador morfológico al etiquetador /lematizador: unidades léxicas complejas y desambiguación'. *Procesamiento del Lenguaje Natural (SEPLN)* 19: 90-100, Sevilla.
- Ahmad, K. y M. Rogers. 2001. 'Corpus Linguistics and Terminology Extraction'. Wright, S.E. y Budin, G. (comp.) *Handbook of Terminology Management* 2: 725-760.
- Alegria I., O. Ansa, X. Artola, et al. 2004. 'Representation and Treatment of Multiword Expressions in Basque'. *ACL'04 Workshop on MWEs*, Barcelona. <<http://ixa.si.ehu.es/Ixa/Argitalpenak/Artikuluak/1084865372/publikoak/acl2004.pdf>> (April 2006)
- Alonso Ramos, M. 1994-1995. 'Hacia una definición del concepto de colocación: de J. R. Firth a I. Mel'cuk'. *Revista de Lexicografía* 1: 9-28.
- Bennett, P. 1994. 'Translation units in human and machine'. *Babel* 40:12-20.
- Canals, R., A. Garrido, M. Guardiola, et al. 2000. 'Herramientas para la construcción de sistemas de traducción automática: aplicación al par castellano-catalán'. *IV Congreso de Lingüística General*. <<http://www.internostrum.com/docum/articad.pdf>> (April 2006)
- Corpas Pastor, G. 1996. *Manual de fraseología española*. Madrid: Gredos.
- Daille, B. 1996. *Study and implementation of combined techniques for automatic extraction of terminology*. <<http://acl.ldc.upenn.edu/W/W94/W94-0104.pdf>> (April 2006)
- Hurtado Albir, A. 2001. *Traducción y traductología. Introducción a la traductología*. Madrid: Cátedra.
- ESPEUS (Prototipo de traductor automático español-euskara). Eusko Jaurlaritzza-Kultura Saila. <<http://www.kultura.ejgv.euskadi.net/>> (April 2006)
- Fillmore, C., R. Grishman, N. Ide, et al. 2002. 'Towards best practice for multiword expressions in computational lexicons'. *Procedures of the 3rd International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas. <<http://www.humaniora.sdu.dk/ifki/ontoquery/readings/lenci/Calzolarie tal-02.pdf>> (April 2006)
- Melby, A. K. 1995. *The possibility of language. A discussion of the nature of language with implications for human and machine translation*. John Benjamins.
- Toury, G. 1995. *Descriptive Translation Studies and beyond*. John Benjamins.
- Vázquez-Ayora, G. 1977. *Introducción a la traductología*. Georgetown: University Press.
- Villavicencio, A. et al. 2004. 'Lexical encoding of MWEs'. *Proceedings of the 2nd ACL Workshop on Multiword Expressions: Integrating Processing (MWE-2004)*. Barcelona. <<http://acl.ldc.upenn.edu/acl2004/mwe/pdf/villavicencio.pdf>> (April 2006)
- Vinay, J.P. y J. Darbelnet. 1958. *Stylistique comparée du français et l'anglais*. Paris: Didier.