

La generación de lenguaje natural multilingüe: una nueva herramienta para el traductor

María Rosario BAUTISTA ZAMBRANA
Universidad de Málaga

Resumen:

Las nuevas tecnologías se han convertido en un elemento imprescindible en el trabajo del traductor, dado que posibilitan una labor más rápida y eficaz. Esto supone, como apunta Alcina (2006), la automatización de gran parte de las tareas, como el envío y recepción de originales y traducciones a las agencias, el trabajo en equipo en entornos virtuales, la consulta de diccionarios y bases de datos terminológicas, la gestión de traducciones mediante programas de traducción asistida (memorias de traducción, alineadores de textos) y la familiarización con entornos de traducción automática. Muestra de la importancia de esta nueva realidad es la creación del Observatorio de Tecnologías de la Traducción,¹ una Red Temática financiada por el Ministerio de Educación y Ciencia (TIC2002-11705-E), cuyo objetivo reside en servir de catalizador para las diferentes iniciativas que se lleven a cabo en el ámbito de la traducción y la tecnología en España.

El Observatorio distingue nueve áreas principales dentro de las Tecnologías de la Traducción (TT): creación, gestión y técnicas de explotación de corpus, sistemas de memorias de traducción (SMT), localización de software, creación y gestión de bases de datos terminológicas, técnicas de pre- y post-edición, lenguajes controlados, estándares aplicados en la industria de la traducción, gestión de la documentación y traducción automática (TA). Como novedad, consideramos que a estas tecnologías podemos añadir una más reciente e innovadora: la generación de lenguaje natural (GLN), también denominada generación textual. Esta disciplina se centra en la producción de textos escritos en lengua natural a partir de una representación semántica abstracta del conocimiento, denominada *interlengua* (cf. Scott y Evans 1998, y Vander Linden 2000).

¹ Disponible en <http://www.uem.es/web/ott/> [Consulta: 2 de enero de 2007].

El auge de las Tecnologías de la Traducción ha tenido su consiguiente reflejo en el ámbito de la investigación: así, desde hace años se vienen llevando a cabo numerosos proyectos relacionados con el campo de la ingeniería lingüística y la traducción en Europa, Norteamérica y España. Sobre generación textual se han financiado numerosos proyectos europeos bajo los auspicios de los distintos programas marco de la Comunidad Europea, especialmente los últimos: Human Language Technologies (1998-2002), e-Content (2001-2004), Information Society Technologies (IST) —dentro del VI Programa Marco (2002-2006)— y e-Contentplus (2005-2008). Como ejemplo significativo de la importancia y novedad de la GLN como área de estudio, podemos citar a John Hutchins, uno de los grandes precursores de las TT y la TA, que en un reciente estudio consideraba la incipiente investigación sobre GLN como uno de los campos más prometedores en el ámbito de las TT y apuntaba a que se terminaría integrando en los sistemas de TA (Hutchins 2005: 3):

Users will want seamless integration of information retrieval, extraction and summarization systems with translation. Research has begun in such areas as cross-lingual information retrieval, multilingual summarization, multilingual text generation from databases, and so forth and, before many years, there may well be systems available on the market and the Internet.

En España, varias universidades han acometido proyectos sobre GLN. Concretamente, la Universidad de Málaga cuenta con una amplia experiencia en generación textual multilingüe, en la que lleva trabajando desde 1999, primero con el proyecto *Diseño de un tipologizador textual para la traducción automática de textos jurídicos (español/inglés/alemán/italiano/árabe)* (Dirección General de Investigación Científica y Técnica, ref. PB98-1399. Investigadora principal: G. Corpas), cuyo objetivo central fue la elaboración de un sistema de GLN basado en corpus para la producción automática de contratos de compraventa de bienes inmuebles (véanse los trabajos reunidos en Corpas Pastor (ed.) (2003); y, a continuación, con un proyecto sobre GLN multilingüe de contratos turísticos basada en corpus, denominado *TURICOR: Compilación de un corpus de contratos turísticos (alemán, español, inglés, italiano) para la generación textual multilingüe y la traducción jurídica* (Ministerio de Ciencia y Tecnología, ref.: BFF2003-04616, 2003-2006. Investigadora principal: G. Corpas). A su vez, gran parte de las líneas de trabajo anteriores se han continuado en el proyecto *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* (Junta de Andalucía, HUM-892).