

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

María Rosario BAUTISTA ZAMBRANA
Universidad de Málaga

Como citar este artículo:

BAUTISTA ZAMBRANA, María Rosario (2008) «La generación de lenguaje natural multilingüe: una nueva herramienta para el traductor», en PEGENAUTE, L.; DECESARIS, J.; TRICÁS, M. y BERNAL, E. [eds.] *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. La traducción del futuro: mediación lingüística y cultural en el siglo XXI. Barcelona 22-24 de marzo de 2007*. Barcelona: PPU. Vol. n.º 2, pp. 287-296. ISBN 978-84-477-1027-0. Versión electrónica disponible en la web de la AIETI:

<http://www.aieti.eu/pubs/actas/III/AIETI_3_MRBZ_Generacion.pdf>.



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

María Rosario Bautista Zambrana
Universidad de Málaga

1. Introducción

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).

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

* El presente trabajo ha sido realizado en el seno del proyecto *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* (ref. núm. HUM-892, 2006-2009. Proyecto de Excelencia, Junta de Andalucía).

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

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).

2. Conceptos básicos de la GLN

Son muchos los autores que han propuesto definiciones de generación de lenguaje natural. Algunas de las más relevantes son las formuladas por Hovy (1996), Reiter y Dale (2000: 1) y Aguayo *et al.* (2003). El primero afirma lo siguiente: «The area of study called natural language generation (NLG) investigates how computer programs can be made to produce high-quality natural language text from computer-internal representations of information». Para Reiter y Dale (2000: 1), los sistemas de GLN son aquellos que parten de algún tipo de representación no lingüística y se valen del conocimiento lingüístico y del dominio en cuestión para producir documentos, informes, explicaciones, mensajes de ayuda y otros tipos de textos. Por su parte, Aguayo *et al.* (2003: 22) consideran que:

[...] el resultado de la generación de lenguaje natural (GLN) es información textual en un lenguaje determinado partiendo de una base de conocimiento que contenga información bien estructurada, independiente del lenguaje, carente de ambigüedad y provista de cierta representación contextual.

De esta forma, como habíamos afirmado anteriormente, la GLN consiste en la producción de textos escritos en lengua natural a partir de una representación semántica abstracta del conocimiento, que se denomina interlengua. Así, un sistema de GLN toma en consideración información abstracta (generalmente no lingüística, inequívoca y bien estructurada), así como datos de la situación comunicativa, para producir un texto con una estructura coherente y unas expresiones lingüísticas adecuadas. En el caso de la generación multilingüe, el sistema es capaz de producir —de forma simultánea o secuencial— textos en varios idiomas a partir de la misma fuente abstracta.

Como se ha señalado, la entrada que recibe un sistema de generación es generalmente de naturaleza no lingüística (Dale, Di Eugenio y Scott 1998: 347): así, puede ser de carácter simbólico-conceptual (por ejemplo, procedente de una base de conocimiento) o numérico (por ejemplo, procedente de una base de datos de valores de bolsa). No obstante, existen algunos sistemas, como CourseViewGenerator (Barrutieta 2001), que sí cuentan con un *input* de carácter lingüístico.

Las modalidades o procesos de GLN se pueden clasificar atendiendo a su grado de sofisticación y a su poder expresivo. Eduard Hovy (1996) distingue cuatro modalidades de generación: sistemas de texto fijo, sistemas de plantilla, sistemas basados en frases y sistemas basados en características.

Los sistemas de texto fijo (en inglés, *canned text systems*), también llamados «textos enlatados», son los que se emplean en la mayoría de los programas informáticos y consisten en la presentación de cadenas invariables de palabras (mensajes de error, advertencias, etc.). Este método es fácil de aplicar, pero resulta insuficiente cuando necesitamos adaptar el texto a distintas situaciones.

Los sistemas de plantilla (*template systems*) suponen el siguiente nivel de sofisticación y se emplean para producir mensajes similares, entre los que existen sólo pequeñas variaciones. Esta modalidad se emplea principalmente para la generación de textos multioracionales de estructura bastante regular, como cartas tipo e informes comerciales, en los que sólo se pueden rellenar algunos campos de forma muy restringida.

Los sistemas basados en frases (*phrase-based systems*) emplean plantillas generalizadas, bien a nivel de oración, o bien a nivel de discurso. En estos sistemas, se selecciona un patrón de frase como nivel más alto del *input* (por ejemplo, [sujeto verbo objeto]); a continuación, cada elemento del patrón es sustituido y ampliado por un patrón frasal más específico que se corresponde con algún apartado del *input* (por ejemplo, [determinante nombre-núcleo adjetivos modificadores]). El proceso en cascada cesa cuando cada patrón frasal ha sido sustituido por una o más palabras. Los sistemas basados en frases pueden ser robustos, pero son difíciles de construir cuando superan un cierto tamaño; es por ello que este método se emplea sobre todo para generación de textos unioracionales.

Los sistemas basados en características (*feature-based systems*) se encuentran entre los más sofisticados a día de hoy. Cada alternativa mínima de expresión se representa mediante una sola característica; por ejemplo, una oración es afirmativa o bien negativa; es interrogativa, imperativa o enunciativa; es de tiempo presente o pasado, etc. Cada oración queda especificada por un conjunto único de características. El proceso de generación tiene lugar mediante la recopilación progresiva de características para cada fragmento del *input*, hasta que la oración queda totalmente determinada o construida. La virtud de estos sistemas reside en la simplicidad de su concepción, ya que se le puede añadir cualquier distinción lingüística como característica. Por el contrario, el punto débil de estos sistemas radica en la dificultad de mantener interrelaciones de características y en el control de la selección de características (cuantas más haya, más complejo resultará el *input*).

La GLN puede ofrecer numerosas aplicaciones: el grupo de interés SIGGEN (*Special Interest Group in Natural Language Generation*)² menciona la generación de informes, la traducción automática y la creación de textos explicativos para sistemas basados en conocimiento; Bateman y Zock (2003: 286) añaden la generación de resúmenes y de documentación multilingüe y multimodal. Aguayo *et al.* (2003: 23), por su parte, afirman que los sistemas de GLN pueden servir para elaborar partes meteorológicas, visualizar el contenido de una base de datos, generar habla de forma automática, etc., y resaltan su aplicación en el campo de la traducción automática entre lenguajes naturales y la traducción a múltiples lenguas desde una fuente de representación. Así mismo, afirma Langkilde (2002: 1) que la GLN constituye una subtarea de numerosas

² Podemos encontrar más información en la URL <http://www.siggen.org/>. SIGGEN constituye el grupo de interés más importante dedicado al estudio de la GLN [Consulta: 2 de enero de 2007].

aplicaciones:

Such applications include machine translation, human-computer dialogue, summarization, report creation, automatic technical documentation, proof/decision explanation, customized instructions, item and event descriptions, question answering, tutorials, stories, and more.

3. La GLN multilingüe y su utilidad para la traducción

La generación textual multilingüe constituye una de las aplicaciones más interesantes de la GLN, por su capacidad de generar documentos —de forma simultánea o secuencial— en varios idiomas a partir de una única fuente de información: las elecciones conceptuales del usuario. Son numerosos los usos que puede adquirir en el marco de la globalización actual: por ejemplo, numerosas empresas multinacionales y organizaciones internacionales precisan de la producción constante de grandes cantidades de documentación; por su parte, el sector turístico viene experimentando desde hace algunos años un aumento considerable en el número de contrataciones por Internet, lo que implica en muchos casos la necesidad de creación de formularios, condiciones generales y contratos multilingües. La traducción tampoco permanece ajena al desarrollo de la GLN, pues se puede beneficiar de muchas de sus aplicaciones multilingües: de este modo, un sistema de generación puede convertirse en una valiosa fuente de textos paralelos para el traductor, que puede obtener no sólo documentos en varios idiomas, sino también seleccionar las distintas opciones conceptuales que desea que aparezcan en los respectivos textos. De este modo, el traductor selecciona en su propio idioma una serie de conceptos —por ejemplo, en el caso de un pronóstico del tiempo: temperaturas, velocidad del viento y humedad— y obtiene textos en dos o más idiomas que contengan oraciones relativas a los conceptos elegidos. Con ello se agiliza el proceso de documentación, tanto desde el punto de vista terminológico como enciclopédico. Una técnica que sigue esta filosofía es la propuesta por Scott y Evans (1998), *Symbolic Authoring*, que permite la especificación manual del contenido conceptual de un documento. A este sistema se le pueden añadir generadores adicionales para otras lenguas, de manera que un solo proceso de especificación de la entrada conceptual permite la generación de todas las variantes multilingües del documento, teniendo en cuenta además las convenciones lingüísticas y culturales de la lengua meta. Por tanto, esta técnica permite gestionar simultáneamente un documento en varias lenguas por medio de representaciones del contenido independientes de una lengua dada.

Son muchos los autores que resaltan las amplias posibilidades de la generación multilingüe. Es el caso, por ejemplo, de Lavid (2005), Androutsopoulos, Kallonis y Karkaletsis (2005) y Hartley y Paris (1997). Lavid (2005) ha trabajado en el desarrollo de recursos gramaticales para la generación contrastiva inglés-español. Androutsopoulos, Kallonis y Karkaletsis (2005) han desarrollado un generador multilingüe de descripciones de objetos a partir de ontologías. Por su parte, Hartley y Paris (1997) crearon, junto con otros investigadores, el sistema DRAFTER, que conciben como una herramienta de ayuda tanto para redactores como para traductores: (Hartley y Paris 1997: 109)

We argue that (...) multilingual generation provides the appropriate technology, shifting attention to an even earlier stage in the authoring process, that of specifying the semantics of the text to be produced. We describe a prototype system which exploits this technology to meet the expressed needs of authors and translators by supporting them in the drafting of multilingual instructions. We suggest that, in the future, a single platform to support multilingual documentation

should integrate translation-oriented tools and generation-based tools to be employed as appropriate by different types of users (translators and authors) in different circumstances.

Como observamos anteriormente al mencionar los usos de la GLN, el vínculo entre ésta y la traducción automática es algo que resaltan numerosos autores, puesto que la tecnología aplicada en la GLN se aproxima bastante a los últimos desarrollos de la traducción automática —los sistemas de traducción automática basados en ejemplos (*Example-based Machine Translation-EBMT*). No obstante, existen notables diferencias entre ambos sistemas, dado que en la GLN no se parte de un texto origen para llegar a un texto meta, sino que se producen textos originales en una o varias lenguas a partir de especificaciones abstractas; y tampoco se requiere una fase de análisis previa, pues la información lingüística y demás especificaciones provienen de sistemas de bases de datos relacionales (SBDR) ya elaboradas y estructuradas (cf. Aguayo *et al.*, 2003: 23). Por otro lado, se han construido numerosos sistemas de TA a partir del concepto de interlengua o representación subyacente de la GLN. Ejemplos de ello son los trabajos de Dorr (1993) (quien creó el sistema UNITRAN), Wilcock (1993) y Lavoie *et al.* (2000). En general, los métodos para llevar a cabo la generación de texto en dos o más idiomas no difieren mucho de los empleados para la generación monolingüe, sobre todo si las lenguas en cuestión están en mayor o menor medida emparentadas. La idea básica de la mayoría de los sistemas multilingües reside en aprovechar la misma información abstracta de entrada para generar después textos en distintos idiomas; en este sentido, son la lexicalización —la elección de palabras y estructuras sintácticas concretas— y la realización gramatical —el proceso de transformación de las expresiones abstractas en texto real— los campos más estudiados respecto a la generación multilingüe, ya que son los más implicados en el proceso de elegir las expresiones adecuadas para cada lengua. Dos enfoques destacados de GLN multilingüe son el propuesto por Stede (1999) y el basado en KPML.

Stede (1999) propone un sistema multilingüe de generación de instrucciones técnicas en inglés y alemán basado en el proyecto TECHDOC (Rösner y Stede, 1994). El autor considera la generación multilingüe como una extensión del problema de la producción de paráfrasis monolingües; es decir, para él se trata de decir lo mismo con distintas palabras, independientemente de que esas palabras pertenezcan a un mismo idioma o a varios. Dentro de la arquitectura del sistema propuesto por Stede podemos destacar dos elementos de gran relevancia para la generación multilingüe: una ontología de conceptos sobre el dominio en cuestión, válida para las dos lenguas implicadas; y un lexicón encargado de establecer correspondencias unívocas entre las representaciones conceptuales subyacentes y las representaciones semánticas oracionales en cada una de las lenguas.

El sistema KPML (Komet-Penman Multilingual) (Bateman, 1997; Reiter y Dale, 2000)³ es uno de los componentes de realización lingüística más conocidos y empleados dentro de la GLN. Se basa en la gramática sistémico-funcional de Halliday (1985) y su objetivo principal es facilitar la generación multilingüe de manera que un solo *input* o especificación sintagmática puedan realizarse —es decir, convertirse en texto real— en distintas lenguas mediante el simple cambio de la gramática activa en la aplicación. Una característica importante del sistema es que cuenta con una ontología de tipos semánticos, orientada lingüísticamente, que se denomina *Upper Model* (Bateman *et al.*, 1990).

³ Véase también la página oficial de KPML, mantenida por John Bateman, dentro del sitio web de la Universidad de Bremen: <http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/README.html> [Consulta: 15 de diciembre 2006].

4. GenTur, un sistema de GLN multilingüe para contratos turísticos

Las investigaciones llevadas a cabo en los proyectos TURICOR y HUM-892 de la Universidad de Málaga han dado lugar a la creación del sistema GenTur, un generador multilingüe de contratos turísticos, capaz de generar secuencialmente documentos en cuatro idiomas: español, inglés, alemán e italiano. En concreto, la primera versión del programa se ha centrado en la producción de condiciones generales del contrato de viaje combinado, en las cuatro lenguas indicadas.

El primer paso ha consistido en estudiar los textos que posteriormente se van a generar. Para ello, se han analizado detalladamente las características lingüísticas, textuales y discursivas de los contratos alojados en el corpus *Turicor*. A partir de este análisis se han elaborado cuatro prototipos textuales del contrato de viaje combinado, uno para cada idioma implicado; éstos consisten, como explica Corpas Pastor (2003b: 51), en «textos preparados artificialmente en soporte informático, elaborados según la superestructura y otros elementos macro- y microestructurales característicos» del documento en cuestión.

Una vez creados los prototipos textuales correspondientes al documento (en nuestro caso, el contrato de viaje combinado) que deseamos generar, se ha procedido a construir una interlengua, denominada XGTLING, para cada uno de los idiomas del proyecto. Esta formalización tiene lugar por medio de un editor que permite representar, por medio de etiquetas basadas en XML, la superestructura, la macroestructura y la microestructura del contrato en cuestión. Así, encontramos etiquetas como *clause*, para introducir una cláusula, *concept*, para insertar un concepto, o *tf*, para introducir una forma textual. Así mismo, el editor posibilita el establecimiento de variables, a fin de que el usuario final pueda introducir nombres reales, direcciones, cifras, etc.

Resulta interesante destacar que la interlengua utilizada sigue un enfoque basado en plantillas, ya que constituye la forma más cómoda y funcional de expresar un contenido multioracional, el de los contratos de viaje combinado, que es bastante regular y prefijado y que se presta a relativamente pocas variaciones. Este método basado en plantillas presenta una gran flexibilidad y es fácilmente adaptable a la inclusión de nuevas lenguas para la posterior generación de documentos. Sistemas muy conocidos como TEXT, TAILOR, ANA y ELIZA lo han implementado.

Una vez que los prototipos textuales han quedado formalizados con la ayuda del editor —con lo cual han tomado forma etiquetada, apta para su procesamiento computacional—, el programa está listo para generar. Es entonces cuando el usuario final pasa a desempeñar un papel muy importante, ya que es el encargado de decidir, por medio de la interfaz de GenTur, qué partes del contrato van a aparecer en los textos finales, qué valor van a tomar las variables, qué formas textuales se van a seleccionar, etc. Al final obtenemos, de forma secuencial, cuatro contratos de viaje combinado, en formato HTML, que contienen las cláusulas, conceptos y formas textuales elegidas por el usuario.

5. Conclusión

La interlengua XGTLING y el sistema GenTur constituyen una buena muestra de lo que se puede llegar a conseguir con la GLN multilingüe: se trata de un sistema fácil de usar, en el que el usuario final sólo tiene que seleccionar los conceptos y formas textuales deseados, y que permite obtener textos equivalentes —desde el punto de vista conceptual— en varias lenguas. En este sentido, consideramos que la GLN multilingüe puede convertirse en una herramienta de gran utilidad para la producción de textos paralelos para el traductor y permitir así un proceso de documentación más rápido y

fiable. Además, GenTur se puede utilizar para la generación de cualquier tipo de documento, ya que la interlengua XGTLING se puede adaptar a cualquier texto e idioma, siempre que previamente se hayan construido prototipos textuales adecuados del tipo de documento en cuestión.

Bibliografía

- Aguayo Maldonado, A., J. L. Caro Herrero, G. Corpas Pastor, I. Gómez Gallego y A. J. Guevara Plaza (2003). «Traducción automática y generación textual: herramientas, grupos y proyectos de investigación». En G. Corpas Pastor (ed.). *Recursos documentales y tecnológicos para la traducción del discurso jurídico (español, alemán, inglés, italiano, árabe)*. Granada: Comares. 1-32.
- Alcina, A. (2006). *Curso de Postgrado TECNOLOC'06*. [Castellón:] Universitat Jaume I. En línea en http://tecnolettra.uji.es/es/?page_id=9 [Consulta: 10 de diciembre de 2006].
- Androutsopoulos, I., S. Kallonis y V. Karkaletsis (2005). «Exploiting OWL Ontologies in the Multilingual Generation of Object Descriptions». En G. Wilcock, K. Jokinen, C. Mellish y E. Reiter (eds.). *Proceedings of the 10th European Workshop on Natural Language Generation (ENLG 2005)*. Aberdeen: [s. l.]. 150-155. En línea en http://www.aueb.gr/users/ion/docs/enlg2005_paper.pdf [Consulta: 19 de diciembre de 2006].
- Barrutieta, G. (2001). Generador inteligente de documentos de formación. En línea en <http://sapiens.ya.com/cyberformacion/Generador.pdf> [Consulta: 26 de diciembre de 2006].
- Bateman, J. (1997). «Enabling technology for multilingual natural language generation: the KPML development environment». *Journal of Natural Language Engineering* 3 (1). 15-55.
- Bateman, J. y M. Zock, M. (2003). «Natural Language Generation». En R. Mitkov (ed.). *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press. 284-304.
- Bateman, J., R. Kasper, J. D. Moore y R. A. Whitney. (1990). «A general organization of knowledge for natural language processing: the Penman Upper Model». Informe técnico inédito. Marina del Rey: Information Sciences Institute.
- Corpas Pastor, G. (ed.). (2003a). *Recursos documentales y tecnológicos para la traducción del discurso jurídico (español, alemán, inglés, italiano, árabe)*. Granada: Comares.
- Corpas Pastor, G. (2003b). «Diseño de un *tipologizador* para la traducción jurídica: del corpus al prototipo textual». En G. Corpas Pastor (ed.). *Recursos documentales y tecnológicos para la traducción del discurso jurídico (español, alemán, inglés, italiano, árabe)*. Granada: Comares. 33-58.
- Dale, R., B. Di Eugenio y D. Scott. (1998). «Introduction to the Special Issue on Natural Language Generation». *Computational Linguistics*. 24 (3). 345-353. En línea en: <http://www.cs.mu.oz.au/acl/J/J98/J98-3001.pdf> [Consulta: 16 de diciembre de 2006].
- Dorr, B. (1993). «Interlingual Machine Translation: A Parameterized Approach». *Artificial Intelligence* 63 (1&2). 429-492.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Londres: Edward Arnold.
- Hartley, A. y C. Paris (1997). «Multilingual Document Production From Support for Translating to Support for Authoring». *Machine Translation* 12. 109-129.
- Hovy, E. (1996). «Overview (of Language Generation)». En R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen y V. Zue (eds.). *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press. En línea en <http://cslu.cse.ogi.edu/HLTSurvey/ch4node3.html#SECTION41> [Consulta: 15 de diciembre de 2006].
- Hutchins, J. (2005). «Computer-based Translation in Europe and North America, and its Future

- Prospects». En *JAPIO 20th anniversary*. Toquio: Japan Patent Information Organization. 156-160. En línea en <http://ourworld.compuserve.com/homepages/WJHutchins/JAPIO-2005.pdf> [Consulta: 15 de diciembre de 2006].
- Langkilde, I. (2002). «An Empirical Verification of Coverage and Correctness for a General-Purpose Sentence Generator». Nueva York: [s. l.]. 17-24. En *Proceedings of the International Natural Language Generation Conference*. En línea en <http://faculty.cs.byu.edu/~irenelg/papers/verifcc.ps.gz> [Consulta: 18 de diciembre de 2006].
- Lavid, J. (2005). *Lenguaje y nuevas tecnologías: nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Cátedra.
- Lavoie, B., R. Kittredge, T. Korelsky y O. Rambow (2000). «A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Processing». En *Proceedings of ANLP/NAACL 2000*. San Francisco: Morgan Kaufmann. 60-67.
- Reiter, E. y R. Dale (2000). *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Rösner, D. y M. Stede (1994). «Generating multilingual documents from a knowledge base: the TECHDOC project». En *Proceedings of the 15th Conference on Computational Linguistics (COLING)*. Kyoto [s. l.]. 1, 339-343.
- Scott, D. y R. Evans (1998). «Multilingual Document Management Without Translation: Using natural language generation in the Multilingual Information Society». *Elsnews* 7/1. 2-3. En línea en <http://mcs.open.ac.uk/ds5473/publications/IEE-paper.htm> [Consulta 15 de diciembre de 2006].
- Stede, M. (1999). *Lexical semantics and knowledge representation in multilingual text generation*. Boston: Kluwer Academic Publishers.
- Vander Linden, K. (2000). «Natural Language Generation». En D. Jurafsky y J. H. Martin (eds.). *Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River: Prentice Hall. 763-798.
- Wilcock, G. (1993). *Interactive Japanese-European Text Generation. An approach to multilingual export translation based on Systemic Functional Grammar*. Manchester: University of Manchester. [Tesis de máster.] En línea en <http://www.ling.helsinki.fi/~gwilcock/Pubs/1993/MScThesis.pdf> [Consulta: 19 de diciembre de 2006].