

Sistemas de traducción automática

Automatic translation systems

J. Gabriel Amores

En este artículo se abordan, de manera muy sucinta y general, algunas cuestiones relativas a los sistemas de traducción automática que han sido diseñados para traducir textos de una lengua a otra. En concreto, se describen características relativas al tipo de lenguaje en el que están escritos los textos, el grado de automatización que es posible en los sistemas de traducción automática y la integración y evaluación de los mismos en un entorno de trabajo.

This article provides a brief and general view to some of the issues involved in automatic translation designed to translate texts from one language into another. Specifically, the article describes characteristics in relation to language types used in texts, the degree of automatization these automatic translation systems provide for and the integration and assessment of automatic translation systems in work environments.

Con respecto a las características de los sistemas de traducción automática, independientemente del enfoque que se utilice (directo, transferencia o interlingua, por citar los más conocidos), se puede hacer una clasificación de los sistemas de traducción automática en función del tipo de texto o dominio para el que fueron diseñados.

Es muy conocido que el principal problema de la traducción automática no es de naturaleza informática sino lingüística. El lenguaje es masivamente ambiguo para un programa informático que trata de modelizar el léxico y gramática de una lengua, ya que tendrá en cuenta posibilidades de combinación sintáctica, y por tanto semántica, que a un hablante de dicha lengua no se le habrían planteado.

Para solucionar este problema se han adoptado numerosos enfoques, encaminados a restringir de algún modo o bien la sintaxis o bien el dominio semántico en que se escriben los textos. Alternativamente, se puede conseguir mayor flexibilidad mediante una selección previa del *glosario* que se ha de activar como preferente, en función del dominio al que pertenezca el texto, sin tener por ello que restringir su sintaxis. En los siguientes apartados nos ocuparemos de estas cuestiones.

Lenguaje controlado. Es aquel que fue producido teniendo en cuenta ciertas restricciones en cuanto a gramática, estilo y vocabulario. El objetivo de los lenguajes controlados es mejorar la coherencia, legibilidad, traducibilidad y recuperación de información en documentos de carácter técnico. En la actualidad hay cientos de empresas en todo el mundo que hacen uso de lenguajes controlados como medio para mejorar la legibilidad o facilitar la traducción de esos textos a otras lenguas.

El asunto no es trivial, ya que habitualmente las grandes empresas exportan sus productos a numerosos países y tienen que asegurarse que los manuales de mantenimiento serán entendidos adecuadamente por el personal técnico de cada país. Una traducción inadecuada puede llevar a desastres económicos, pero también humanos si se trata de instrumental médico, por ejemplo.

Uno de los primeros proyectos fue el *Caterpillar Fundamental English* en los años setenta, del cual se ha lanzado recientemente otra versión, el *Caterpillar Technical English*, que difiere sustancialmente del anterior, tanto en extensión como objetivos, ya que se propone facilitar la traducción de la documentación a otras 13 lenguas, ya sea por medio de sistemas de traducción automática o traductores profesionales. Actualmente, uno de los ejemplos más interesantes es el *AECMA Simplified English* (AECMA, 1995), un estándar de lenguaje controlado que ha sido adoptado por toda la industria aeroespacial para facilitar el uso de los manuales de mantenimiento a los hablantes no nativos de inglés. Otro ejemplo de lenguaje controlado es el desarrollado por la empresa Perkins (Perkins Approved Clear English, PACE).

Sublenguajes. Es aquel que se utiliza en un dominio o tipo de texto concreto, pero que surge de forma espontánea y no por restricciones impuestas, como en el caso de los lenguajes controlados. La ventaja que proporcionan los sublenguajes es similar a la de los lenguajes controlados, ya que tanto el vocabulario como la sintaxis son más reducidos que en los textos de propósito general. Con ello conseguimos reducir la complejidad del sistema de traducción automática y, por tanto, el problema de la traducción a unos límites manejables.

El mejor ejemplo del uso de los sublenguajes en traducción automática sigue siendo el *sistema Météo*

(Grimalia y Chandiox, 1992), que lleva traduciendo boletines meteorológicos del inglés al francés para la Oficina Canadiense de Meteorología desde hace más de dos décadas.

El problema habitual del enfoque basado en sublenguajes es que difícilmente se encuentra un sublenguaje tan bien caracterizado como el de los partes meteorológicos. Por el contrario, la mayoría de los textos pertenecientes a un dominio suelen contener lenguaje figurado, vocabulario perteneciente a distintos dominios, etc. Quizá por esta razón se ha abandonado en gran medida este enfoque frente al desarrollo que están experimentando los lenguajes controlados.

Lenguaje especializado o de dominio general. Como hemos visto, tanto los lenguajes controlados como los sublenguajes simplifican enormemente el problema general de la traducción automática y lo asemejan a la práctica real en traducción, ya que la mayoría de los traductores profesionales se especializan en determinadas áreas: medicina, economía, aeronáutica, etc.

Asimismo, los lenguajes controlados y sublenguajes parecen ser la solución óptima para muchas empresas con un gran volumen interno de producción de documentación técnica que ha de ser traducida a otras lenguas en un período muy breve.

Sin embargo, desde el punto de vista del diseño de sistemas de traducción automática, tanto el enfoque de sublenguajes como el de lenguajes controlados suponen una gran limitación. Dichos sistemas son difícilmente reutilizables para otros dominios, donde la sintaxis y el vocabulario difieren de aquél para el que se diseñó originalmente el sistema.

Como alternativa a este problema, la mayoría de los sistemas de traducción automática comerciales suelen optar por diseñar un sistema de carácter general, sin pensar en ninguna aplicación concreta, y después ofrecer, por una cantidad adicional, glosarios especializados según las necesidades del cliente. En el momento de la traducción, se ofrece al usuario la posibilidad de jerarquizar u ordenar la secuencia en la que el sistema ha de buscar las palabras en los diccionarios, para así dar prioridad a los glosarios especializados sobre el diccionario de carácter general.

De manera similar, se puede indicar al principio del proceso de traducción a qué dominio pertenece el texto que se va a traducir, para que el sistema elija por sí mismo los glosarios apropiados. Como alternativa a esta indicación «manual» del dominio al que pertenece el texto, sería posible también que el programa de traducción incluyera un algoritmo que resuelva la ambigüedad semántica que presenten las palabras del texto (Ide y Véronis, 1998). Este enfoque es el más prometedor, ya que puede adaptarse al posible cambio de dominio que se produce dentro del texto en función de su contexto más inmediato, sin tomar una decisión *a priori* para todo el texto como hacen los enfoques anteriores.

Sistemas de traducción automática con ayuda del traductor frente a sistemas de traducción con ayuda del ordenador. Independientemente del tipo de texto que se vaya a traducir, el traductor cuenta con una amplia gama de productos para llevar a cabo su trabajo. Son productos que se encuentran en algún lugar entre los casos extremos del traductor profesional que no hace uso alguno de ordenadores para su trabajo y el programa informático que realiza todo tipo de traducciones de manera totalmente automática. Habitualmente se hacen dos distinciones, dependiendo del grado de automatización del proceso de traducción y también de si el control del proceso de traducción lo lleva el traductor o el programa informático. En el primer caso se habla de *traducción humana con ayuda del ordenador* y en el segundo, de *traducción automática con ayuda del traductor*. En el primer caso, el traductor profesional controla el proceso y hace uso de distintos programas informáticos o bases de datos informatizadas según las va necesitando. Las herramientas pueden afectar a todo el proceso de traducción. Por ejemplo, en la recepción, preparación y entrega de los documentos: recepción y entrega por correo electrónico, escaneo y formateo de los documentos finales, etc. Durante el proceso de traducción se pueden utilizar un procesador de textos, diccionarios y glosarios terminológicos accesibles en Internet o en CD, memorias de traducción, corrector ortográfico, etc. En el segundo, el programa informático realiza la traducción y el traductor profesional puede intervenir antes (preedición), durante (sistemas interactivos) o posteriormente (postedición) para asegurar o mejorar la calidad del producto final. La preedición consiste en *adelantarse* a los posibles fallos que puede cometer el sistema y facilitarle el trabajo retocando el texto de entrada; por ejemplo, marcar apellidos como ‘Casado’ para que no sean traducidos. Hoy en día son muy pocos los sistemas que precisan este tipo de retoques, aunque bien es cierto que un gran porcentaje de los errores que cometen los sistemas de traducción automática provienen de un texto de entrada mal escrito en cuanto a su estilo, faltas de

ortografía, mal uso de la puntuación, etc. La postedición y corrección del texto de salida una vez concluido el proceso de traducción por parte del sistema es inevitable. Sin embargo, ésta varía sustancialmente dependiendo del uso que se vaya a dar a la traducción. Así, por ejemplo, si solamente se buscaba una traducción de la terminología relevante para obtención de información y no para publicación del texto traducido, la postedición puede llegar a ser mínima. Actualmente muchos sistemas de traducción automática incluyen macros para facilitar la tarea del posteditor; por ejemplo, cambiar la palabra en femenino, ponerla en plural, eliminar el artículo definido, añadir un acento, etc. La necesidad de la postedición ha sido una de las críticas que más se han utilizado en contra de la traducción automática. Sin embargo, no se suele tener en cuenta que en gabinetes o departamentos de traducción se suelen revisar todas las traducciones, incluidas las realizadas por traductores profesionales.

En los sistemas interactivos el programa informático se detiene y consulta al traductor cada vez que detecta una ambigüedad, léxica o estructural, que él mismo no puede resolver. Se presentan las alternativas para que el traductor elija y posteriormente continúa el proceso. Estos sistemas han tenido poco éxito por razones obvias. El traductor ha de estar presente durante el proceso, esperando a que el sistema le consulte; el sistema no 'aprende' de consultas anteriores y vuelve a preguntar ante una situación similar; no se asegura que el producto final sea totalmente correcto, ya que no se puede asegurar que donde no preguntó obtuvo la traducción adecuada.

Como alternativa a este tipo de escenario, y en relación a los lenguajes controlados, están desarrollándose en la actualidad sistemas de producción multilingüe de documentos. Estos sistemas interactúan con el usuario en la confección de un documento final (Hartley y Paris, 1997).

Integración en un entorno de trabajo

Inevitablemente, a pesar de describir con cierto detalle la complejidad de los sistemas de traducción automática y los distintos enfoques y posibilidades que se pueden ofrecer, surge la pregunta: ¿cuál es el mejor sistema?

Ante esta pregunta, sólo cabe responder que el mejor sistema es el que resuelve satisfactoriamente las necesidades del usuario. Éstas conforman el entorno de traducción o de trabajo en el que se deberá integrar el sistema de traducción automática, que puede variar bastante entre un gabinete de traducción a una empresa con gran volumen de traducción interno, un traductor autónomo, una institución internacional, una pequeña empresa con intención de expandirse en el mercado internacional etc.

Asimismo, desarrollar un sistema de traducción automática conlleva mucho esfuerzo y recursos. De cualquier modo, es necesario un período de adaptación del sistema al entorno de producción en el que se pretende utilizar: compatibilidad de procesadores de texto y bases de datos, requisitos informáticos, sistemas operativos, etc.

Por tanto, antes de decidimos por un sistema debemos evaluar diversos factores, algunos de los cuales trataré de describir a continuación:

Cuestiones relevantes en la evaluación

A grandes rasgos, un sistema de traducción automática se puede evaluar desde tres puntos de vista: Evaluación lingüística por parte del usuario, evaluación por parte del diseñador del sistema y evaluación de costes y beneficios (Lehrberger y Bourbeau, 1988). La evaluación lingüística por parte del usuario se divide a su vez en tres grandes áreas: determinar el tipo de textos que se van a traducir, donde es necesario conocer el dominio al que pertenecen los textos y saber si existen glosarios o diccionarios especializados sobre la materia, o si el dominio exige una especialización muy fuerte por parte de los traductores. La segunda gran área es determinar el volumen de textos que se han de traducir en términos de palabras/año. Y en tercer lugar se deben establecer las características lingüísticas de los textos teniendo en cuenta, al menos, los siguientes factores: tamaño del vocabulario en el dominio, nivel de homonimia que presentan los textos, sintaxis y tipo de estructura más frecuentes para atender a aquellos fenómenos que suelen presentar problemas para los sistemas de traducción automática; como por ejemplo, coordinación, elipsis, listas, compuestos, extraposiciones, etc.

Cuando se realiza la evaluación del componente lingüístico se deben incluir los siguientes pasos: construir oraciones de prueba y traducirlas, seleccionar un texto típico y traducirlo, clasificar e interpretar los

resultados, formular hipótesis sobre las limitaciones y, por último, confirmar con otros textos. Además habrá que tener en cuenta los siguientes aspectos: número de entradas que tiene el diccionario, posibilidad de modificarlas, caracteres disponibles (ñ, ó, ...), longitud máxima de una palabra en el diccionario, número de textos que se pueden procesar de una vez, requisitos de memoria y configuración del *hardware*, integración con herramientas disponibles en el entorno habitual, lenguaje de programación utilizado y estrategia de análisis que se sigue (directo/transferencia/interlingua).

En la evaluación por parte del diseñador, una vez identificados los posibles errores y limitaciones del sistema, él es el único que puede solucionarlos satisfactoriamente. Él es quien tiene acceso completo al sistema y quien puede clasificar los errores por su causa para determinar el esfuerzo requerido para solucionarlos. Habitualmente se acometerán primero aquellos que tendrán una mayor incidencia en la calidad de la traducción debido a su frecuencia relativa o absoluta en los textos propuestos por el usuario. En la evaluación de costes y beneficios, el usuario habrá de calcular los costes directos e indirectos derivados y comparar los beneficios que le aportará el sistema con relación a la traducción encargada a un profesional en las condiciones habituales que lo viniera haciendo. Entre otros, habrán de evaluarse los siguientes aspectos: coste de revisión de la primera traducción, costes de personal, tiempo que se invierte en todo el proceso de traducción, coste de mantenimiento de los diccionarios, incremento de productividad, decremento de costes, período de amortización, apoyo de los diseñadores y posibilidad de contar con futuras versiones, calidad de la traducción en bruto; es decir oraciones que no necesitan revisión, oraciones que necesitan poca revisión y oraciones que han de ser totalmente revisadas.

Conclusión

En este artículo hemos descrito de forma muy sucinta algunas cuestiones relativas a los sistemas de traducción automática tomados como sistemas completos y a su integración en un entorno de trabajo. Como conclusión se puede apuntar que actualmente hay una gran variedad de herramientas que pueden automatizar en mayor o menor medida el proceso de traducción. Sin embargo, serán las necesidades últimas del usuario las que determinen qué herramienta es la más apropiada teniendo en cuenta no solamente cuestiones de calidad, sino de complejidad técnica y evaluación de los costes y beneficios asociados.

Bibliografía

- AECMA: *AECMA Simplified English: A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language*, Bruselas, 1995.
- Grimaila, A.; Chandiooux, J.: «Made to measure solutions». En: John Newton, ed.: *Computers in Translation: A Practical Appraisal*, Londres, Routledge, 1992: 33-45.
- Hartley, A.F.; Paris, C.L.: «Multi-lingual document production: from support for translating to support for authoring», *Machine Translation* (Special Issue on new tools for human translators) 1997; 12 (12): 109-129.
- Ide, I; Véronis, J.: «Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art», *Computational Linguistics* 1998; 24 (1): 1-40.
- Lehrberger, L.; Bourbeau, L.: *Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation*, Amsterdam/Filadelfia, John Benjamins, 1988.

José Gabriel Amores Carredano

Profesor titular en el Departamento de Lengua Inglesa de la Universidad de Sevilla. En 1990 obtuvo el grado de Masters en Machine Translation por UMIST, Reino Unido, y posteriormente trabajó durante dos años en el Departamento de Traducción Automática de la Organización Panamericana de la Salud, en Washington DC, en el desarrollo de los sistemas ENGSPAN y SPANAM. Actualmente es investigador principal del grupo de investigación en Procesamiento del Lenguaje Natural *Julietta*, que cuenta con contratos y proyectos de investigación nacionales y europeos. Ha publicado numerosos artículos sobre traducción automática y es coautor (junto con el Dr. José Francisco Quesada) de la obra *Diseño e implementación de sistemas de traducción automática*, publicada por el Servicio de Publicaciones de la Universidad de Sevilla en el año 2000.

jgabriel@cica.es