

## APLICACIONES DE TRADUCCIÓN BASADAS EN MEMORIAS DE DATOS. DESARROLLO Y PERSPECTIVAS DE FUTURO

*José Manuel Muñoz Muñoz*

*Mercedes Vella Ramírez*

Universidad de Córdoba

### ABSTRACT

In dealing with memory-based translation tools a distinction is frequently drawn between machine translation applications, which some translators consider a useless endeavour, and computer assisted translation tools, which are considered essential for some translation tasks. In the present paper we look at memory based translation tools and underline the fact that the technologies they use have their origin in machine translation research. Finally we examine new trends in the application of memory-based (machine and assisted) applications as a consequence of globalization and universal access to the Internet.

**KEYWORDS:** Memory-based translation tools, Machine translation applications, Computer assisted translations.

### RESUMEN

Al hablar de tecnología aplicada a la traducción con frecuencia se establece una distinción entre los conceptos de traducción automática, que se considera un empeño quimérico, y la traducción asistida, a cuyas herramientas de ayuda al traductor se concede gran importancia. En el presente artículo analizamos las tecnologías basadas en la reutilización de segmentos de traducción que tienen su origen en los modelos estadísticos de traducción automática y que surgieron como consecuencia del aparente fracaso de los modelos basados en reglas. Por último analizamos las tendencias actuales de aplicación de esas herramientas de traducción en el entorno actual de globalización y de generalización del acceso a la red.

**PALABRAS CLAVE:** Herramientas de traducción basadas en memorias de datos, traducción automática (TA), traducción asistida por ordenador (TAO).

## 1. INTRODUCCIÓN

Es frecuente que en la literatura relacionada con la aplicación de las nuevas tecnologías al ámbito de la traducción se distinga entre los sistemas de traducción automática, a los que se concede escasa, cuando no nula, validez y los sistemas de traducción asistida por ordenador tales como los gestores de memorias de traducción, de terminologías, glosarios, etc., a los que se les reconoce una gran utilidad en determinados ámbitos de la traducción. Una observación más cercana de estos sistemas y de su historia nos llevará, sin embargo, a concluir que los segundos nacieron como consecuencia de la aplicación práctica de principios y procedimientos que estaban ya presentes en los primeros y que fueron aplicados a tareas menos ambiciosas, más cercanas a las necesidades reales del traductor. En el presente artículo se analizan las circunstancias en las que se produjo esa transición, desde los objetivos iniciales de conseguir una traducción completamente automática hasta la utilización de elementos que se habían desarrollado en el seno de esas investigaciones, para objetivos más asequibles y más acordes con los instrumentos, tanto teóricos como prácticos, de los que se dispone.

Las primeras propuestas de sistemas de traducción asistida basados en la recuperación de traducciones previas se producen a principios de los años ochenta como alternativa a los sistemas de traducción automática basados en el conocimiento lingüístico. En efecto esos sistemas, que habían nacido con el objetivo de lograr una traducción de alta calidad sin intervención humana, no habían logrado hasta entonces producir resultados mínimamente aceptables en muestras de lenguaje no restringido, e incluso en lenguajes altamente normalizados la calidad de la traducción que ofrecían hacía necesaria su post-edición exhaustiva. Martin Kay (1980) se refiere con demoledora ironía a la traducción automática mediante estos programas y los describe como situaciones en las que:

*Passages of innocent prose (...) were subjected to a process of vivisection at the hands of an uncomprehending electronic monster that transformed them into stammering streams of verbal wreckage.*

La enorme cantidad de actividad que se generó en los años cincuenta en torno a la consecución del gran objetivo de lograr una traducción automática de calidad sin intervención humana (FAHQT) sufre un brusco descenso en la década de los sesenta como consecuencia de dos factores fundamentales: por un lado, la precariedad de los resultados que se iban obteniendo hasta ese momento y que no respondían en absoluto a las expectativas creadas, ni guardaban relación alguna con los recursos invertidos hizo que muchas grandes compañías redujesen significativamente sus inversiones en este área. En esa situación de recorte de

recursos algunas compañías supieron replantearse a tiempo sus objetivos y estrategias y obtuvieron una importante ventaja estratégica frente a sus competidores. Tal es el caso de ALPS, como veremos más adelante (cf. Lonsdale, 2007).

Por otro lado, el informe del Comité Asesor sobre el Procesamiento Automático del Lenguaje (ALPAC) de la National Science Foundation, que concluye sus trabajos en 1966, confirma las dudas que ya había manifestado Bar Hillel (1960) unos años antes en el sentido de considerar la traducción automática de alta calidad (FAHQT) como poco menos que imposible dado el nivel de desarrollo de la lingüística en ese momento. Los trabajos que se habían llevado a cabo hasta entonces habían tropezado con lo que Bar Hillel denominaba ‘barrera semántica’, que él ejemplificaba, de una manera que ha llegado a ser un clásico en la literatura especializada, con la necesidad de recurrir al conocimiento extralingüístico para resolver el diferente significado de la palabra “pen” en las frases ‘the pen is in the box’, ‘the box is in the pen’. Para Bar-Hillel ningún programa de ordenador podría incorporar el conocimiento semántico suficiente como para poder asignar el significado correcto a estas dos ocurrencias del término “pen”.

El informe ALPAC ha pasado a la historia como el factor determinante en el desprestigio de la investigación en traducción automática y con frecuencia se hace alusión exclusivamente a los aspectos del informe que resultan negativos para el desarrollo de la disciplina. Se suele soslayar el hecho de que el informe contiene también elogios y recomendaciones de apoyo para actividades relacionadas con la elaboración de herramientas de apoyo a la traducción tales como los bancos de datos terminológicos, los glosarios específicos y otras herramientas que se consideran prometedoras aportaciones para la consecución de traducciones de mayor calidad. El informe recomendaba incentivar los trabajos incipientes que se estaban llevando a cabo en Alemania para la elaboración de glosarios por parte de la Agencia de Traducción de las Fuerzas Armadas y de bancos de datos terminológicos como DICAUTOM, que estaba elaborando la Comunidad Europea del Carbón y del Acero y que sería el germen del futuro EURODICAUTOM.

## 2. LA MEMORIA COMO NUEVO PARADIGMA

Como consecuencia de estos dos factores, y también de factores de desarrollo tecnológico, como la disponibilidad de memorias de almacenamiento de datos, capaces de gestionar un volumen ingente de información de manera más eficaz y económica, se produce un cambio de tendencia cuya característica fundamental es la orientación hacia sistemas basados en la memoria (‘memory-based systems’), también llamados sistemas basados en los datos (‘data-driven systems’). Como afirma Hutchins (1998:3) este cambio de orientación no es

consecuencia del trabajo de una sola persona o grupo, sino del desarrollo de una misma idea en diferentes lugares y desde diferentes puntos de vista. En efecto, los fundamentos para la reutilización de segmentos previamente traducidos estaban ya presentes de modo seminal en los trabajos desarrollados a lo largo de los años sesenta y setenta con el establecimiento de los grandes bancos de datos terminológicos tales como EURODICAUTOM en Europa o *Termium* en Canadá. La misma tecnología que se aplicaba al almacenamiento, búsqueda y recuperación de información terminológica podía aplicarse a la búsqueda y recuperación de segmentos textuales ya traducidos.

A principios de los ochenta surgen las primeras propuestas metodológicas que se desarrollarán años después. Entre ellas destacaremos las de Arthern, Kay, Melby y Nagao.

En 1979, Peter Arthern, con ocasión del debate que se originó en la entonces Comunidad Europea acerca del uso en la Comisión Europea de texto traducido por *Systran* sin revisión humana, observa que la gran mayoría de los textos que se traducen en la comisión son muy repetitivos, y contienen gran cantidad de citas literales de otros textos que a su vez ya han sido traducidos previamente, lo que hace que hayan de traducirse repetidamente con el consiguiente gasto y esfuerzo de traducción. Arthern sugiere que, en estas circunstancias, una estación de trabajo específicamente diseñada para los traductores contribuiría a reducir los altos costes de traducción. Se apunta de manera clara a lo que hoy conocemos como memorias de traducción:

It must in fact be possible to produce a programme (sic) which would enable the word processor to remember whether any part of a new text typed into it had already been translated and to fetch this part, together with the translation which had already been translated... One advantage over machine translation proper would be that all the passages so retrieved would be grammatically correct<sup>1</sup>.

Un año más tarde, Martin Kay, en su artículo ya clásico sobre el lugar que corresponde al hombre y a la máquina en el proceso de traducción, se muestra muy crítico con los procesos de traducción automática en los que la labor del traductor consiste en la post-edición de la producción del ordenador. Para Kay:

When the computer is improperly used, its effects are, of course, quite different. This happens when the attempt is made to mechanize the non-mechanical or something whose mechanistic

---

<sup>1</sup> Arthern P. J. (1979): "Machine translation and computerized terminology systems: a translator's viewpoint" en Snell, B. M. (ed): *Translating and the computer. Proceedings of a seminar, London, 14<sup>th</sup> November 1978*, Amsterdam, North-Holland, pág. 318).

substructure science has not yet been revealed. In other words, it happens when we attempt to use computers to do something we do not really understand. History provides no better example of the improper use of computers than machine translation.<sup>2</sup>

Y propone una introducción gradual de la máquina en tareas concretas a modo de “translator’s amanuensis”:

I want to advocate a view of the problem in which machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process. First they will take over functions not essentially related to translation. Then, little by little, they will approach translation itself. The keynote will be modesty. At each stage, we will do only what we know we can do reliably. Little steps for little feet!<sup>3</sup>

Paralelamente, Alan Melby (1981) propone la fragmentación del texto en segmentos que pueden utilizarse como unidades para establecer concordancias bilingües que puedan ayudar al traductor en la toma de decisiones, basada en traducciones previas de segmentos coincidentes. Esta propuesta de Melby sería más adelante reformulada por Harris (cf. Harris, 1988) con la introducción del concepto unidad de traducción (“translation unit”) con la que el texto y su traducción no son vistos ya como entidades independientes sino como una misma entidad con dos dimensiones.

Más ambiciosa resulta la propuesta de Nagao, formulada en 1981 y publicada tres años más tarde, que propugna un cambio de enfoque en la traducción automática y que explora la posibilidad de establecer un modelo estadístico de traducción, basado en ejemplos o semejanzas, para el establecimiento de patrones de coincidencias formales entre segmentos de textos paralelos anotados (EBMT- Example-based Machine Translation). Si bien la propuesta de Nagao utiliza prioritariamente criterios estadísticos, no se descarta en ella la incorporación de determinados niveles de conocimiento lingüístico junto al puramente estadístico. Aunque con frecuencia el concepto de EBMT se identifica con el de memoria de traducción, se trata de propuestas significativamente distintas en cuanto que, como apunta Sommers (2003), la EBMT tiene como objetivo la utilización de modelos estadísticos para la construcción de un sistema autónomo de traducción automática, y presenta, por tanto, problemas mucho más complejos que los que presenta un gestor de memorias de traducción que está concebido no como sistema autónomo sino como simple ayuda al traductor humano.

---

<sup>2</sup> Kay, M. (1980): *The Proper Place of Men and Machines in Language Translation*. CSL-80-11, Xerox Corporation.

<sup>3</sup> *Op. cit.*

A pesar de estas propuestas, hasta finales de los ochenta no comienzan a aparecer formulaciones prácticas de sistemas estadísticos de traducción automática. En 1988 Peter Brown (cf. Brown *et al.*, 1988) propone lo que en el momento aparecía como un planteamiento radicalmente nuevo del problema de la traducción automática, basado en métodos puramente empíricos y no en el racionalismo lingüístico que había informado todos los intentos anteriores. El sistema *CANDIDE*, desarrollado por IBM, plantea el objetivo de utilizar para la traducción información exclusivamente estadística, mediante un corpus paralelo inglés-francés de unos tres millones de palabras procedentes del *Hansard*, medio oficial del Parlamento canadiense. A pesar de la enorme repercusión que tuvieron las primeras demostraciones del sistema, ésta se mantuvo, al menos en los términos precisos de la propuesta de Brown, en los ámbitos puramente académicos y no llegó a desarrollarse como alternativa operacional a los sistemas comerciales tradicionales.

### 3. APLICACIONES COMERCIALES

En el ámbito de la traducción asistida, la primera aplicación comercial no vio la luz hasta que ALPS (Automatic Language Processing Systems) lanzó su TSS (Translation Support System), un procesador de textos multilingüe que contenía un innovador sistema de procesamiento de repeticiones (*repetitions processing*), que, aunque estaba limitado a coincidencias exactas de cadenas alfanuméricas, contenía todos los elementos de los actuales gestores de memorias de traducción (nombre que no se utilizaría hasta mucho más adelante).

ALPS venía trabajando desde principios de los ochenta en un sistema de traducción automática por lotes que presentaba la peculiaridad, frente a otros sistemas, de que la edición de la traducción podía llevarse a cabo de manera interactiva. En 1983 el departamento de traducción del gobierno canadiense reunió a un nutrido grupo de psicólogos, lingüistas, traductores y expertos en documentación para llevar a cabo un estudio del rendimiento del sistema, que Lonsdale (2007) califica como la evaluación más exhaustiva jamás llevada a cabo sobre un sistema de traducción. El informe concluyó que el sistema resultaba decepcionante, económicamente poco rentable e incomprensible para los traductores profesionales<sup>4</sup>. Se hacía aconsejable un cambio de orientación si la compañía quería sobrevivir en un mercado altamente competitivo. Y el cambio vino con el TSS que aprovechaba muchas de las tecnologías y la experiencia acumulada en el sistema anterior para crear un procesador de textos orientado hacia la traducción. A mediados de los ochenta se produce paralelamente a este cambio de orientación en la compañía un avance muy significativo en la difusión

---

<sup>4</sup> Las palabras exactas según Lonsdale fueron: "disappointing", "uneconomical", "mesmerizing".

de la informática con la aparición de los IBM PC/AT, IBM VM/CMS, que constituyen las plataformas sobre las que se apoyará el éxito del TSS (cf. Lonsdale, 2007). Es significativo, sin embargo, el hecho de que el nacimiento del TSS se produce no como fruto de una investigación centrada en la consecución de sistemas de apoyo a la traducción humana, sino más bien como un efecto secundario de los trabajos que se venían llevando a cabo para la consecución de un sistema autónomo de traducción automática.

El TSS presentaba el original y la traducción en pantallas paralelas e incorporaba la posibilidad de recuperar traducciones anteriores mediante un mecanismo conocido como 'repstraction' que procesaba en tiempo real las repeticiones de cadenas alfanuméricas exactas que se produjesen durante la traducción permitiendo al traductor incorporarlas o desecharlas en cada caso. El sistema también incluía un innovador sistema de tratamiento de las coincidencias terminológicas, *AutoTerm*, que se presentaban en una ventana en la parte inferior de la pantalla conforme iban apareciendo en el texto.

Sin embargo, el éxito del TSS fue efímero y ALPS lo retiró del mercado a principios de los noventa. Quizá había llegado demasiado pronto y ni la extensión de la informática en los ámbitos profesionales ni los hábitos de trabajo de los traductores eran los adecuados para una herramienta tan sofisticada (cf. García, 2005).

A principios de los noventa se produce la gran expansión de la informática en el ámbito profesional y los traductores abandonan la máquina de escribir en favor del ordenador personal, lo que hace que las grandes compañías de software se vean obligadas a adaptar las aplicaciones que habían desarrollado para ser ejecutadas en grandes servidores. En 1990 la competencia en el ámbito de los sistemas operativos para los ordenadores personales era feroz y Hummel y Knyphausen, fundadores de *Trados*, tuvieron la intuición de apostar por *Windows*, adaptando *Multiterm* (en 1992) y más adelante, en 1994, el *Translator's Workbench II* al nuevo sistema operativo, cuya versión 3.0 había aparecido en 1990. Esta intuición les proporcionó una enorme ventaja competitiva incluso sobre rivales de tamaño muy superior como *Eurolang*.

En la segunda mitad de los noventa, los sistemas de apoyo al traductor basados en la recuperación de datos de memorias conocieron una enorme difusión entre los traductores profesionales que trabajaban para organismos oficiales, en agencias de traducción o autónomamente. Esta difusión contribuyó también a una rápida mejora de los programas, que iban incorporando elementos cada vez más sofisticados para la gestión de las memorias de traducción, glosarios, bases de datos terminológicas, filtros para la conversión entre formatos, etc. Entre los más extendidos cabe destacar, además de *Trados*<sup>5</sup>, *Déjà-Vu*, *TRANSIT*, *SDLX*, *Eurolang*

---

<sup>5</sup> Trados fue adquirido en 2005 por SDL, pasando su producto estrella a denominarse SDL-Trados.

*Optimizer* (LANT), etc., productos que tienen una utilidad especial en entornos de traducción de texto restringido, formulaico o altamente repetitivo, donde representan un ahorro significativo del rendimiento de los traductores profesionales. La TM Survey (Lagoudaki, 2006) hace una evaluación de la extensión que el uso de estas memorias de traducción tiene en el Reino Unido.

#### 4. MODELOS HÍBRIDOS PARA LA TRADUCCIÓN AUTOMÁTICA

En el ámbito de la traducción automática, a lo largo de la década de los noventa se ha ido imponiendo una metodología empírica de desarrollo que representa, en cierto modo, un nuevo paradigma (cf. Jones 1992), propiciado tanto por los prometedores resultados de los primeros intentos como por el más que aparente estancamiento de los modelos lingüísticos tradicionales. Con el paso del tiempo se ha ido imponiendo una visión más realista y más cooperativa del enfrentamiento, que se produjo en un principio, entre los modelos tradicionales y los modelos estadísticos: hoy se puede afirmar que muy pocos grupos centran su investigación en un enfoque estadístico puro, al estilo de las propuestas de Brown *et al.* (1988). La mayoría de los sistemas, aunque mantienen los presupuestos básicos de este enfoque inductivo, aceptan la necesidad de incorporar conocimiento lingüístico ya sea en la fase de etiquetado del corpus paralelo, en el que se incorporan identificadores de unidades sintagmáticas de acuerdo con estructuras dependenciales (cf. Sadler, 1991) o simplemente como ayuda a la identificación estadística (cf. Chen y Chen, 1995).

Los modelos estadísticos presentan frente a los tradicionales ciertas ventajas indudables entre las que pueden destacarse las siguientes:

- Los ejemplos que constituyen el elemento central del sistema son datos de lenguaje real y, por tanto, su uso reduce la sobre-generación de ejemplos acordes con la regla pero ajenos al lenguaje real.
- El conocimiento lingüístico del sistema puede enriquecerse fácilmente simplemente añadiendo más ejemplos.
- Al tratarse de sistemas basados en los datos más que en una teoría, se reduce el problema de la congruencia de la teoría y de los posibles conflictos entre reglas. (La contrapartida es que al añadir más y más ejemplos pueden surgir conflictos entre ellos)
- Es posible desarrollar muy rápidamente sistemas para traducir entre lenguas a partir de sólo un corpus paralelo de ejemplos, lo que es especialmente conveniente en el caso de lenguas para las que aún no se han desarrollado herramientas tales como ‘parsers’, diccionarios, etc.



Si bien estas ventajas son claras, al menos desde el punto de vista teórico, debe reconocerse también que estos sistemas no están libres de errores y que su rendimiento y sus ámbitos de aplicación en la práctica distan mucho de ser los que auguraban sus creadores hace un cuarto de siglo: las palabras de Somers representan lo que puede considerarse hoy una corriente de opinión generalizada:

EBMT is certainly here to stay, not as a rival to rule-based methods but as an alternative, available to enhance and sometimes replace it. (...) The problem of scaling up remains, as do a large number of interesting translation problems, especially as new uses for MT emerge. The new paradigm is now approaching its teenage years: the dust has settled and the road ahead is all clear.<sup>67</sup>

En los próximos años, cabe esperar un avance estable en la integración de los sistemas basados en ejemplos y los sistemas de traducción tradicionales, de manera que cada uno de ellos contribuya a mejorar el rendimiento global, ayudando a resolver aquellas partes del proceso para las que está mejor equipado. En esta dirección se encaminan algunas propuestas, como la de Carl y Hansen (1999), que tienen como objetivo encontrar la manera en la que dos motores de traducción, basados en estrategias contrapuestas, puedan participar en el proceso de traducción de modo que el rendimiento del sistema integrado supere el de los dos motores actuando de modo independiente.

## 5. RESPUESTAS A NUEVAS NECESIDADES SOCIALES

En la década de los noventa se producen una serie de acontecimientos que hacen que se replanteen las ideas que se habían venido fraguando respecto al papel de la tecnología en la traducción. Este cambio de mentalidad viene determinado por la generalización del uso de Internet y por el fenómeno social de la globalización, que plantean nuevas necesidades, nuevas actitudes y nuevos enfoques ante la traducción.

La generalización del uso de la red hace que se incremente de manera exponencial la demanda de traducciones en tiempo real, lo que lleva a que se desarrollen sistemas específicamente diseñados para responder a ese tipo de demanda, donde el resultado final no es evaluado tanto en función de la calidad «publicable» de la traducción, sino de parámetros tales como la inmediatez, la simplicidad, la facilidad de acceso, o el bajo coste. Se trata en ocasiones de sistemas

---

<sup>67</sup> Sommers, H. (2003): "An Overview of EBMT" en Carl, M y Way, A. (eds.): *Recent advances in example-based machine translation*. Dordrecht: Kluwer Academic Publishers, pág. 150.

que deben ser operativos incluso en situaciones en las que el texto original está plagado de expresiones idiosincrásicas cuando no de errores, intencionados o no, como puede ocurrir en la comunicación via IRC-chat o SMS.

El desarrollo de las nuevas tecnologías de la comunicación también ha hecho que aparezcan nuevas necesidades de traducción, especialmente en el ámbito de la recuperación de información. La red reúne a millones de personas en todo el mundo en busca de información, o de comunicación, en procesos que se caracterizan por el rápido intercambio de información con frecuencia efímera. La traducción en bruto aparece aquí como un elemento más, integrado en el sistema de intercambio de información, de modo que la lengua en la que ésta se codifica no constituya un obstáculo para su recuperación eficaz en tiempo real. Como respuesta a esta nueva necesidad surge un campo muy amplio de aplicación centrada en el desarrollo de sistemas multilingües (cross-lingual information retrieval systems).

La globalización y los procesos económicos y sociales que la caracterizan determinan el crecimiento de la demanda de traducciones comerciales, científicas, técnicas, jurídicas, etc., que pertenecen a registros especiales en los que el elenco de recursos lingüísticos utilizados es muy limitado. En otros casos se trata de documentos de trabajo que no son definitivos por estar sujetos a continua revisión y cuya traducción por medios humanos sería económicamente inviable.

Estos cambios tecnológicos y sociales han hecho que los sistemas de traducción automática tradicionales, basados en reglas, con un enfoque típicamente lingüístico hayan encontrado un importante campo de aplicación en el ámbito de la navegación en la red, donde cada día es mayor la necesidad de una traducción, o un borrador de traducción, disponible en tiempo real, con un coste mínimo y que, aunque esté plagada de errores o tenga una calidad muy lejana a la exigible, puede ser útil para transmitir al usuario una idea aproximada del contenido del documento que se somete a traducción. Para dar respuesta a esta nueva necesidad, en 1997 *Systran* llega a un acuerdo con *Altavista* para incluir en su página un servicio de traducción de páginas web bajo el nombre *Babelfish*, que se ofrecería a los usuarios de manera gratuita. Desde entonces se han puesto a disposición del navegante en la red un buen número de traductores, en muchos casos asociados a los propios buscadores de información. Esta utilización ha venido a proporcionar una rentabilidad añadida a los sistemas comerciales tradicionales cuyas expectativas de desarrollo se veían limitadas por la ya aludida falta de calidad de la traducción.

Todo parece indicar que en el futuro estos sistemas, más que por la perfección de sus traducciones serán juzgados por su utilidad para producir textos que, sin pretender ser definitivos, pueden comunicar con más o menos eficacia lo esencial de un mensaje de un modo pragmáticamente aceptable y, por otro lado, por su disponibilidad y accesibilidad en aquellos entornos específicos en los que sean necesarios.

Sin embargo este renacimiento de los viejos sistemas no ha contribuido significativamente a la mejora de su calidad. Puede afirmarse que se han hecho grandes esfuerzos para resolver problemas técnicos relacionados con la disponibilidad en la red, la rapidez o la facilidad de uso, pero no se ha invertido en la mejora del sistema, en lo que respecta a la calidad de la producción final.

Esta utilización de los traductores clásicos en la web ha hecho que en el ámbito de la evaluación del rendimiento de sistemas se haya sustituido en la práctica el viejo ideal de la FAHQQT por uno nuevo, más asequible y más útil, la FAUT (fully automatic understandable translation). En términos comerciales podría afirmarse que la traducción en bruto ha encontrado un nuevo nicho de mercado y es previsible que se siga desarrollando de manera estable en el futuro, generalizándose a nuevos dispositivos (teléfonos móviles, e-books, videoconsolas, etc.).

Se ha producido también una diversificación de los sistemas tradicionales al menos en dos sentidos: por un lado en el número de lenguas disponibles y por otro en la cantidad de situaciones comunicativas, estilos y registros a los que pueden ser aplicados (correo electrónico, chat, estilo informal, lenguajes de especialidad, etc.) y el tipo de dispositivos en los que se oferta (teléfonos móviles, agendas electrónicas, e-books, etc.).

Junto a estos desarrollos, es también justo admitir que la falta de sistemas de control de la calidad de los productos que se ofertan ha facilitado la presencia en la red de subproductos que se autodenominan sistemas de traducción y que no son más que diccionarios bilingües de ínfima calidad ofertados por compañías poco escrupulosas para captar clientes incautos. Esta presencia contribuye a aumentar en los usuarios la falta de confianza en los sistemas de traducción automática en general, dado que no se ha desarrollado aún entre ellos una conciencia de la complejidad del problema y, por tanto una capacidad crítica para discernir la calidad del servicio que, en cada momento, están utilizando.

## 6. LA SÍNTESES

La generalización del uso de Internet ha hecho que surjan nuevos modos de aplicar la tecnología a la resolución de los problemas que plantea la traducción.

En primer lugar se han superado las separaciones que existían entre los modelos basados en el conocimiento y los modelos estadísticos. Hoy día, es frecuente encontrar propuestas que combinan ambos modelos para conseguir el mayor rendimiento posible. Un buen ejemplo de esta tendencia es *Google Translate*, un ambicioso proyecto disponible hasta el momento en cincuenta y seis lenguas diferentes. De acuerdo con la escasa información que se facilita acerca del proyecto, *Google Translate* utiliza como memoria de traducción una base de datos documental indexada de documentos en la red junto con sus equivalentes en otras

lenguas. La búsqueda se realiza a través de un algoritmo de escala que intenta encontrar las unidades mayores ya traducidas. Solo cuando no encuentra una alternativa de traducción fiable en esa inmensa memoria de traducción se recurre a un sistema de traducción automática tradicional basado en reglas, similar a *Babelfish*. Este doble mecanismo con frecuencia produce resultados sorprendentes que, sin embargo, resultan poco fiables por su falta de consistencia. El sistema es interactivo e incorpora aspectos de ‘crowdsourcing’ en el sentido de que si un determinado fragmento de la traducción no resulta satisfactorio, el usuario puede sugerir una alternativa que se incorpora a la memoria de traducción para ser utilizada en sucesivas ocasiones. Del mismo modo el usuario puede a través del *Google Translator Toolkit* incorporar sus propias memorias de traducción en formato TMX o documentos traducidos en diversos formatos. Esos documentos traducidos o memorias de traducción pueden ponerse a disposición del público o solo a disposición de usuarios autorizados. *Google* incorpora *Google Translate* como un elemento más de su nuevo navegador *Google Chrome*, lo que, sin duda contribuirá a aumentar enormemente el tamaño de las memorias existentes en la actualidad.

## 7. NUEVAS HERRAMIENTAS DE TRADUCCIÓN

Otra de las tendencias recientes en el ámbito de la tecnología para la traducción es consecuencia del movimiento que se ha dado en llamar ‘cloud computing’. La generalización del uso de la red, su disponibilidad y la mejora experimentada en la velocidad de acceso permiten utilizar el ordenador como un terminal con acceso a todos los servicios, programas y aplicaciones disponibles en la red. Esta posibilidad cambia radicalmente el concepto de la traducción como tarea de un traductor aislado y permite el trabajo colaborativo de un grupo de traductores en un mismo proyecto en tiempo real. Estos traductores comparten los recursos disponibles en la “nube”. Las aplicaciones que el traductor utiliza se conciben como un servicio y no como un producto. Un buen ejemplo de este tipo de aplicación desarrollada en el ámbito del software libre es *Pootle* que se desarrolló para la localización de las interfases gráficas de usuario (GUI) de proyectos de software libre pero puede adaptarse a otros usos. *Pootle* es un sistema de traducción ‘on-line’ pensado para que un grupo de traductores colabore en proyectos de localización y traducción a través de un servidor web e incorpora el *Translate Toolkit*, que es un conjunto de herramientas para la traducción y localización.

*Pootle* ha hecho posible la localización de programas como la distribución de *Linux Ubuntu* o el paquete de ofimática, *Open Office*, en numerosas lenguas, tanto mayoritarias como muy minoritarias, en un esfuerzo de colaboración sin precedentes en el ámbito de la traducción.

En el aspecto comercial algunas multinacionales han comenzado a ser conscientes del potencial que esta tecnología encierra y están comenzando a ofrecer programas específicos para implementar estos nuevos modos de traducción, así *Lionbridge*, por citar solo un ejemplo, ofrece un sistema de gestión de archivos de traducción a través de la web basado en tecnologías de ‘cloud computing’ denominado *Translation Workspace*, y un sistema de control de proyectos denominado *Freeway*<sup>7</sup>.

La creación de tecnologías de “cloud computing” y su aplicación a procesos de “crowdsourcing” surge primariamente en el contexto del *software* libre con el objetivo de ponerse al servicio de proyectos solidarios de colaboración, y poner así, a disposición de usuarios sin recursos o pertenecientes a comunidades lingüísticas muy minoritarias, herramientas informáticas que estaban fuera de su alcance. Esas mismas tecnologías en manos de grandes compañías cuyos objetivos son exclusivamente comerciales pueden representar un serio peligro para los traductores y para la profesión en general. Así, es inquietante observar cómo estas nuevas tendencias se presentan como grandes avances en el ámbito de la tecnología de la información y las comunicaciones en foros tales como el Common Sense Advisory, que la denomina CT3 (community translation + collaborative technology + crowdsourcing translation), o la TAUS (Translation Automation User Society), bajo el nombre “community translation” (cf. TAUS, 2009).

## 8. CONCLUSIONES

Como ya apuntara Kay, en 1980, refiriéndose al ideal de la traducción automática de alta calidad:

Even if it is never achieved, it provides an incomparable matrix in which to study the workings of human language. Whether it is achieved or not, other useful, if more modest, inventions may well emerge as by-products of the attack on FAHQT provided only that the work is conducted in a healthy intellectual environment. The trouble is that no such environment exists today.<sup>8</sup>

El análisis detallado del desarrollo de las tecnologías que hoy día se consideran esenciales para la traducción se produce de manera sistemática

---

<sup>7</sup> <http://www.lionbridge.com/lionbridge/es-ES/services/localization-translation/language-asset-management.htm>

<sup>8</sup> Kay, M. (1980): *The Proper Place of Men and Machines in Language Translation*. CSL-80-11, Xerox Corporation.

como consecuencia de la persecución del ideal inalcanzable de la traducción automática.

En efecto, todos los principios que actualmente se aplican en herramientas tales como las memorias de traducción, fueron concebidos y desarrollados primariamente para ser aplicados a sistemas autónomos modelos de traducción mecánica.

Puede afirmarse que la tradicional dicotomía entre sistemas de traducción automática (tanto basados en reglas como basados en analogías) y herramientas de traducción asistida se ha ido difuminando en favor de la aplicación de criterios pragmáticos que requieren la integración de ambas herramientas en el proceso de traducción. En particular, los sistemas basados en criterios estadísticos están cada vez más presentes, dadas las enormes posibilidades que ofrece la indexación de enormes cantidades de documentos de la web y la incorporación de sus traducciones a sistemas tales como *Google Translate*.

Por último, la aparición de nuevas herramientas para la traducción en la “nube” permite que la traducción y especialmente la localización en la web puedan llevarse a cabo de modo cooperativo con criterios de “*crowdsourcing*”, lo que puede plantear problemas laborales y sociales importantes para el futuro de la profesión del traductor. En este contexto se cobran renovada validez los planteamientos de Kay (1980) acerca del lugar que debería corresponder a hombres y máquinas en el proceso de traducción.

## BIBLIOGRAFÍA

- ALPAC (1966): *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, N.A.S., N.R.C. Washington, D.C., National Academy of Sciences, National Research Council, 1966.
- ARTHERN, P. J. (1979): “Machine translation and computerized terminology systems: a translator's viewpoint” en Snell, B. M. (ed): *Translating and the computer. Proceedings of a seminar, London, 14<sup>th</sup> November 1978*. Amsterdam, North-Holland.
- BAR HILLEL, Y. (1960): The present status of automatic translation of languages, *Advances in Computers* 1, 91-163.
- (1964): *Language and Information*. Reading, MA, Addison Wesley.
- BROWN, P. F., et al. (1988): “A statistical approach to language translation”, *Computational Linguistics* 16, 79-85.
- , LAI, J. C. y MERCER, R. L. (1991): “Aligning sentences in parallel corpora”, en *29th Annual Meeting of the Association for Computational Linguistics: proceedings of the conference, 1991 Jun 18, Berkeley, California, [ACL]*.

- CARL, M. y HANSEN, S. (1999): "Linking Translation Memories with Example-Based Machine Translation". *Machine Translation Summit VII*, Singapur, 617–624.
- CHEN, K. y CHEN H. (1995): "Machine Translation: An Integrated Approach", en *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation TMI 95*.
- GARCIA, I. (2005): "Long term memories: Trados and TM turn 20". *The Journal of Specialised Translation* 4, 18-31.
- HARRIS, B. (1988): "Bi-text, a new concept in translation theory", *Language Monthly* 54, 8-10.
- KAY, M. (1980): *The Proper Place of Men and Machines in Language Translation*. CSL-80-11, Xerox Corporation.
- LAGOUDAKI, E. (2006): *TM Survey: Translation Memory Systems, enlightening users' perspective*. Imperial College London. Disponible en <http://www.atril.com/docs/tmsurvey.pdf> [Consulta: 10-1-2010].
- LONSDALE D. (2007): "From ALPS to Alpnet (and beyond)". Disponible en <http://www.mtarchive.info/Lonsdale-2007.pdf> [Consulta: 12.10.10].
- MELBY, A. K. (1981): "Translators and machines – can they cooperate?" *Meta* 26(1), 23-34.
- NAGAO, M. (1984): "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle", en Elithorn, A. y Banerji, R. (eds.): *Artificial and Human Intelligence*. North-Holland, Amsterdam.
- SNELL, B. M. (ed) (1978): *Translating and the computer. Proceedings of a seminar, London, 14<sup>th</sup> November 1978*. North-Holland, Amsterdam.
- TAUS (2009): *Edinburgh, March 25-27, 2009*. Disponible en: <http://www.translationautomation.com/meetings/edinburgh-march-25-27-2009.html>. [Consulta 10.05.2009].
- SADLER, V. (1991): "The Textual Knowledge Bank: Design, Construction, Applications", en *International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP)*, Kyoto, Japon.
- SOMMERS, H. (2003): "An Overview of EBMT" en Carl, M y Way, A. (eds.): *Recent advances in example-based machine translation*. Dordrecht, Kluwer Academic Publishers.