

La máquina de traducir: una demo para el discurso jurídico

Luis CEREZO CEBALLOS
Georgetown University

Como citar este artículo:

CEREZO CEBALLOS, Luis (2003) «La máquina de traducir: una demo para el discurso jurídico», en MUÑOZ MARTÍN, Ricardo [ed.] *I AIETI. Actas del I Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. Granada 12-14 de Febrero de 2003*. Granada: AIETI. Vol. n.º 2, pp. 755-773. ISBN 84-933360-0-9. Versión electrónica disponible en la web de la AIETI:
<http://www.aieti.eu/pubs/actas/I/AIETI_1_LCC_Maquina.pdf>.



Las máquinas de traducir: una demo para el discurso jurídico

Luis CEREZO CEBALLOS
Georgetown University
lc98@georgetown.edu

Resumen

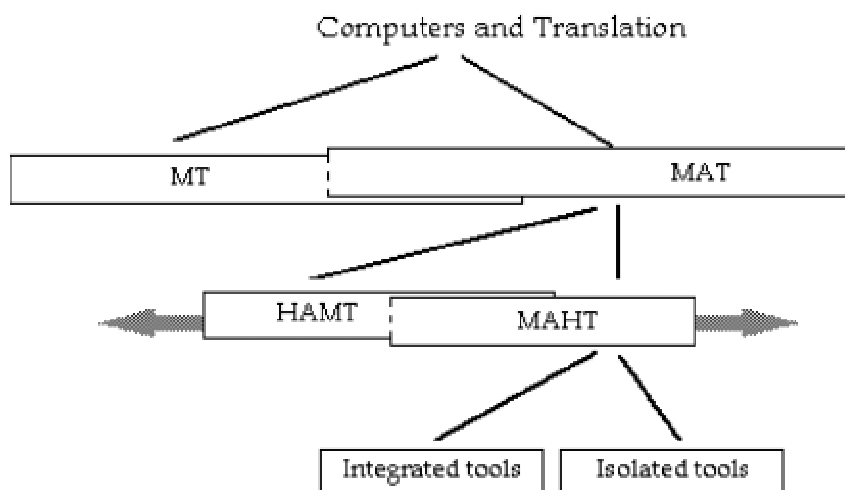
Desde sus orígenes hasta nuestros días la —para muchos, controvertida— relación entre el traductor humano y la máquina ha atravesado por varias etapas. En su fase preliminar, cuyo hito principal fue la aparición del primer sistema de traducción automática (Georgetown University - IBM 1954), los investigadores aspiraron a crear máquinas capaces de emular a los humanos. Tras la publicación del polémico informe ALPAC (1966), sin embargo, los investigadores abandonaron casi por completo la traducción automática (TA) y se concentraron en lo que hoy conocemos bajo el nombre de herramientas de traducción asistida por ordenador (TAO). Desde los primeros estudios en el seno de las Comunidades Europeas, a finales de los años setenta, la TA y la TAO se han contemplado no como realidades necesariamente incompatibles, sino más bien interrelacionables. Hoy, gracias a la aparición de numerosos sistemas comerciales de TA y TAO (muchos de ellos de uso gratuito a través de Internet), la traducción automatizada es una realidad cada vez menos ajena al público general. A lo largo de estos cincuenta años de historia ha quedado patente que los textos más aptos para la traducción a máquina son aquellos encasillables dentro de lo que se ha dado en llamar sublenguajes. Ello se debe a su carácter cerrado a nivel pragmático, comunicativo, textual, léxico, morfológico, sintáctico y semántico. Con el espíritu de aunar ambas ideas —1) el proceso de traducción puede automatizarse, y 2) sobre todo si el TO está escrito en un sublenguaje—, nace la presente comunicación, aplicada al ámbito específico del discurso jurídico. En la primera sección se ofrece una introducción histórica sobre las herramientas informáticas para la traducción. En la segunda, se describen los textos más apropiados para la traducción ‘a máquina’. En la tercera, se enfrentan las nociones de TA y TAO. En la cuarta, se presenta un caso concreto de sistema de TA, EC-Systran. En la quinta, se hace lo propio con un sistema particular de TAO, la memoria de traducción. La sección sexta resume nuestras conclusiones.

1. Las máquinas: introducción histórica

En sus cincuenta años de historia, la I+D en traducción por ordenador ha seguido caminos divergentes (cf. Hutchins 1986). Durante su primera etapa (década de 1950), marcada por la presentación del primer sistema de traducción automática (TA) de la historia (Georgetown University - IBM 1954), la mayoría de los investigadores aspi-

raba a que los ordenadores obtuvieran, por sí solos, resultados comparables a los de los traductores humanos.

En una segunda etapa (década de 1960), esta aspiración fue tachada de demasiado ambiciosa. El primero en criticarla fue Bar-Hillel (1960), considerado el primer investigador a tiempo completo en TA (Massachusetts Institute of Technology, MIT). Para éste, las barreras semánticas sólo podrían traspasarse *alimentando* a los ordenadores con conocimientos acerca del mundo real. O lo que es lo mismo, los avances en TA dependerían de los progresos en inteligencia artificial (IA). Un lustro después, un informe redactado por un comité evaluador del gobierno estadounidense, el llamado ALPAC (Pierce *et al.* 1966), auguró pocas expectativas para la TA a corto plazo y recomendó que los fondos de investigación se desviarán al desarrollo de distintas herramientas de procesamiento del lenguaje natural (PLN), entre ellas, las de traducción asistida por ordenador (TAO). Así, la I+D en traducción por ordenador quedó escindida en dos ramas: 1) la TA como tal y 2) la TAO, que muchos autores (*cf.* Hutchins & Somers 1992; Sager 1994) subdividen a su vez en 2a) TAAH (traducción automática asistida por humanos, del inglés *human-aided machine translation*, HAMT) y 2b) THAO, traducción humana asistida por ordenador (del inglés *machine-aided human translation*, MAHT). Así lo muestra la siguiente ilustración (líneas de investigación en traducción por ordenador según Hahn 1995):



La publicación del informe ALPAC supuso el inicio de una tercera etapa en la I+D en traducción por ordenador, marcada por una aplastante sequía de proyectos de TA durante casi una década en Estados Unidos (Trujillo 1999:253). Por ello, muchos investigadores se refieren a este informe como *el libro negro de la traducción automática* (*cf.* Freigang 2001) o *el informe famoso/infame* (Hutchins 1996). Durante esta década, el centro neurálgico de la investigación en traducción por ordenador se trasladó de Estados Unidos a Canadá y a Europa, en gran parte debido a las necesidades de traducción de sus estructuras gubernamentales. Así, en 1976, se marcaron dos grandes hitos a uno y otro lado del Atlántico: en Canadá, un grupo de investigación de Montreal hizo público su sistema Météo, hoy considerado el primer gran logro en la historia de sistemas de TA para textos escritos en un sublenguaje (en este caso, el de los partes meteorológicos). En Europa, la Comisión de las Comunidades Europeas adquirió una licencia de uso de Systran, que hoy se ha convertido en el sistema de TA más popular.

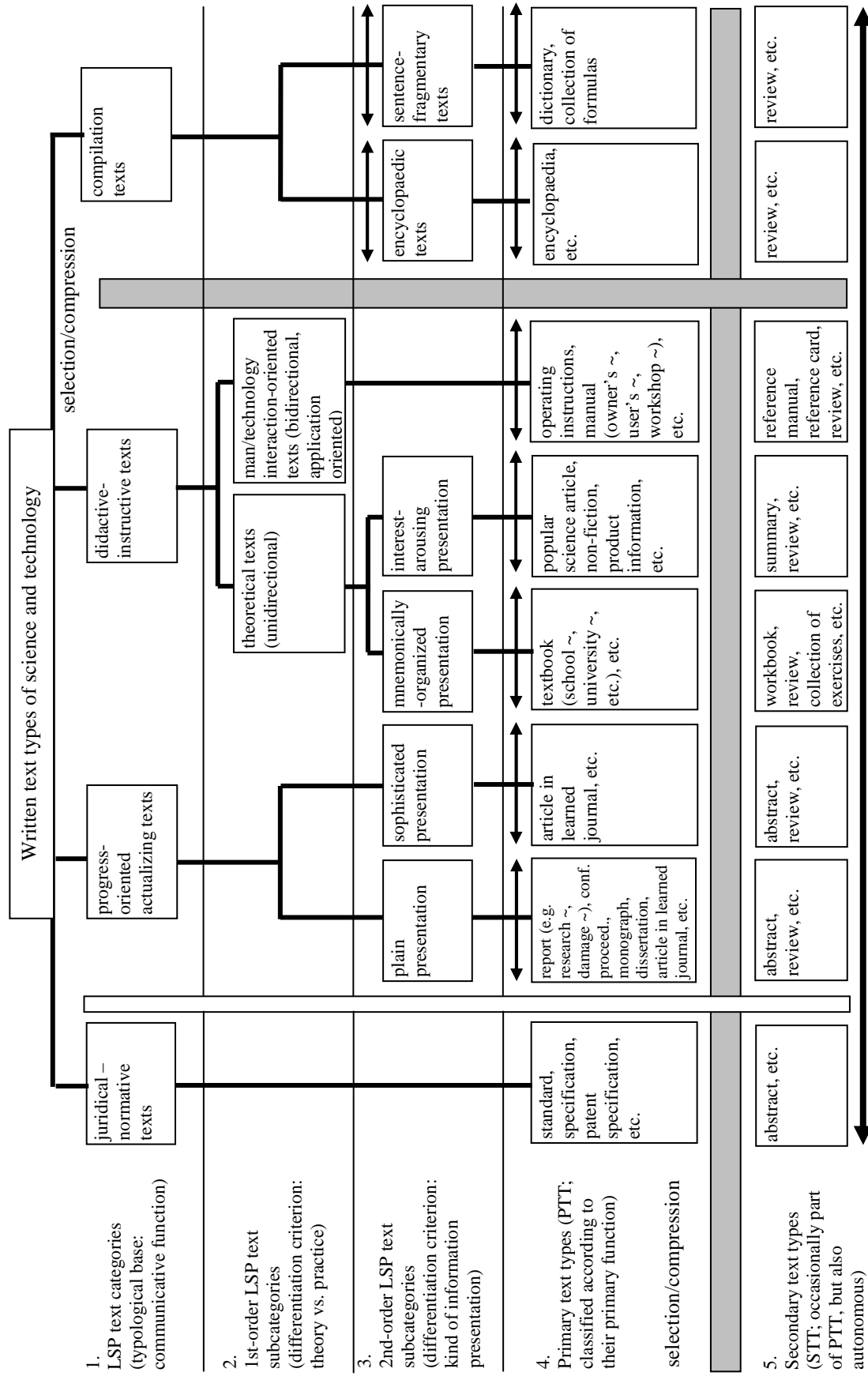
La cuarta y última etapa en la historia de la traducción por ordenador arrancó en los años ochenta, con la aparición de numerosos sistemas comerciales de TA y TAO.

La popularización de los ordenadores personales a partir de esa década, y de la red Internet desde comienzos de los noventa, han beneficiado enormemente al sector. La TA ha dejado de ser una gran desconocida para el consumidor, y han surgido nuevos productos, concebidos específicamente para la traducción de páginas web y correos electrónicos o incluso diálogos sincrónicos. Del mismo modo, el sector de la TAO ha evolucionado drásticamente, como refleja la enorme aceptación de las memorias de traducción desde su salida a la venta al público, a principios de los noventa.

2. Los textos: noción de sublenguaje

A lo largo de cincuenta años de I+D en TA, ha quedado patente que los textos más aptos para la traducción *a máquina* son aquellos encasillables dentro de lo que se ha dado en llamar *sublenguajes* o *lenguas con fines específicos* o LFE (en inglés *languages for special purposes*, LSPs). Ello se debe a su carácter cerrado a nivel *pragmático*, porque el papel del receptor con respecto al texto suele estar claramente definido; *comunicativo* porque emisor y receptor comparten una serie de conocimientos acerca del tema; *textual*, porque el tipo de texto determina la presencia o ausencia de determinados patrones; *léxico*, porque el vocabulario suele ser bastante reducido; *morfológico*, porque el abanico de inflexiones morfológicas es muy limitado; *sintáctico*, porque los patrones sintácticos pueden predecirse en función del tipo textual, y *semántico*, porque las voces suelen ser marcadamente unívocas (Kittredge & Lehrberger 1982).

Al hablar de textos con fines específicos (TFE), nos estamos refiriendo fundamentalmente a tres grandes categorías: economía y comercio, discurso jurídico, y ciencia y tecnología (cf. Webb 1998). Hasta la fecha, una de las tipologías más exhaustivas de TFE es la presentada por Göpferich (1995b, publicada originalmente en alemán, cf. 1995a). En este estudio, la autora organiza los TFE en varios niveles, según su finalidad. En el primer nivel se distinguen cuatro categorías de textos, de acuerdo con su función comunicativa. Estos son los textos jurídico-normativos, los difusores de últimos avances, los didáctico-instructivos, y los recopilatorios. En el segundo nivel, estas categorías textuales se subdividen en función de su naturaleza teórica o práctica, criterio únicamente aplicable a los textos didáctico-instructivos, que quedan así subdivididos en textos teóricos (unidireccionales) y textos orientados a la interacción hombre/tecnología (bidireccionales). En el tercer nivel, cada categoría se escinde en varias ramas según sus diversos tipos textuales, esto es, sus formatos de difusión de la información. Así, los textos sobre últimos avances se subdividen en presentaciones simples y sofisticadas; de igual manera, los textos didáctico-instructivos de naturaleza teórica se bifurcan en presentaciones organizadas mnemotécnicamente y presentaciones confeccionadas para despertar el interés del receptor; finalmente, los textos recopilatorios se dividen en enciclopédicos y suboracionales. Todos ellos, a su vez, se subdividen en los niveles 4º y 5º, dando lugar a los llamados tipos textuales primarios y secundarios, respectivamente. De esta clasificación da cuenta la siguiente ilustración (tipología de TFE escritos según Göpferich 1995b: 309).



3. El papel del humano: TA frente a TAO

La automatización del proceso de traducción a la que conducen muchos sistemas de TAO (STAO), especialmente los sistemas de memoria de traducción (SMT) que abordamos en el presente artículo, puede inducir a confundirlos con los sistemas de TA (STA). Nocionalmente, ambos sistemas están muy relacionados. Así, Abaitua (2001:3) explica el origen de los SMT como un subproducto de un enfoque particular de la TA introducido por Nagao en 1984. Nos estamos refiriendo a la llamada *example-based machine translation* (EBMT), que podríamos verter al español como *traducción automática basada en ejemplos*. Hoy día, los SMT y los STA no sólo están relacionados en el plano nocional, sino también de facto: cada vez son más los SMT con acceso a algún tipo de STA. El objetivo de esto es garantizar que el usuario siempre recibe una propuesta de traducción, aun en aquellos casos en los que el SMT no ha encontrado ningún tipo de coincidencia. Así lo expone Priamo (1999:7):

L'associazione di un motore di traduzione automatica a un ambiente di traduzione assistita consente di sfruttare appieno le caratteristiche dei due sistemi. In questo ambiente integrato, il traduttore oltre a disporre della funzionalità di consultazione delle memorie di traduzione (fondamentale per la traduzione di testi ripetitivi o simili ad altri precedentemente tradotti), può avvalersi delle proposte di traduzione automatica relative a frasi completamente nuove non presenti nelle memorie.¹

Similitudes aparte, conviene tener claro lo siguiente: el cometido de un SMT es *ayudar* al traductor humano mediante la recuperación automática de unidades de traducción (UT) almacenadas previamente, mientras que la finalidad de un STA, por el contrario, es *sustituir* al traductor humano, automatizando para ello todas las fases del proceso traslaticio (cf. Esselink 1998:134; Somers, en prensa, pág. 43).

4. Un ejemplo de TA: EC-Systran

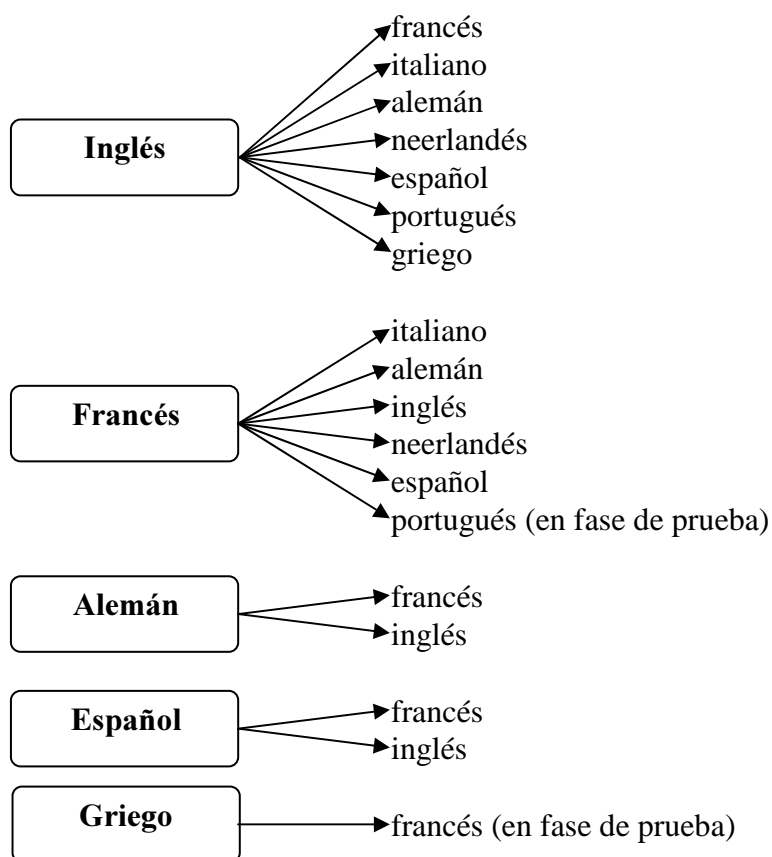
A lo largo de su historia, la TA se ha abordado desde perspectivas muy diferentes, lo que ha dado lugar a un amplio abanico de sistemas. En líneas generales, los sistemas de TA pueden catalogarse en función de los siguientes cuatro parámetros básicos (cf. Hutchins & Somers 1992: 69-80):

- 1) El número de pares de lenguas abordadas: hablamos así de sistemas bilingües frente a sistemas multilingües.
- 2) El enfoque metodológico utilizado: que puede ser tradicional —en cuyo caso hablamos de arquitectura directa (*direct systems*), arquitectura de transferencia (*transfer systems*) o arquitectura basada en una interlingua (*interlingua-based systems*)— o moderno —que enfrenta, de una parte, a los sistemas de enfoque empírico, de los que existen dos tipos, el estadístico (*statistics-based systems*) y el basado en ejemplos (*example-based systems*), y de otra parte, a los sistemas basados en el conocimiento (*knowledge based systems*).

- 3) El grado de interacción con el humano: hablamos así de traducción no intervencionista frente a traducción interactiva.
- 4) El tipo de datos almacenados en el lexicón.

Hoy, tras cincuenta años de investigación en TA, podemos afirmar que el ideal original de reemplazar por completo al traductor humano dista mucho de ser alcanzado, si es que acaso es posible. Sin embargo, como hemos adelantado en las dos primeras secciones del presente trabajo, se ha avanzado considerablemente en la traducción de TFE. A continuación, a modo de ejemplo, presentamos un breve estudio de EC-Systran, sin duda uno de los sistemas de TA más exitosos de cuantos existen en el panorama internacional.

Como recuerda Petrits (2001), Systran, acrónimo de *system translation*, nació en 1956 a manos de Peter Toma, investigador formado en el grupo de Georgetown University que dio a luz al primer STA de la historia. Systran fue producto de la guerra fría, y no en vano se utilizó por primera vez en las fuerzas aéreas estadounidenses (1969, par ruso-inglés) y la NASA (1973, par inglés-ruso). En 1976, el Servicio de Traducción (SdT) de la Comisión de las Comunidades Europeas adquirió los derechos de Systran para el par de lenguas inglés-francés, piedra fundamental de lo que hoy conocemos como EC-Systran. En 1977, el SdT invirtió dicho par de lenguas, y al año siguiente introdujo el par inglés-italiano. Durante la década de los ochenta se fueron añadiendo progresivamente nuevos pares, hasta llegar a los 18 actuales, mostrados en la siguiente ilustración (pares de lenguas disponibles actualmente en EC-Systran según Petrits 2001:3):



Con respecto a su enfoque metodológico, EC-Systran nació como un claro ejemplo de arquitectura directa, si bien con el tiempo se ha ido aproximando cada vez más a una arquitectura de transferencia. En lo que se refiere a su modalidad de traducción, se trata de un sistema no intervencionista. Por último, en lo tocante a la información contenida en el lexicón, Systran funciona en torno a dos tipos de diccionarios, que interactúan constantemente durante el proceso de traducción: un diccionario de raíces (STEM), para palabras aisladas, y un diccionario de expresiones idiomáticas e información semántica (IDLS), para palabras en contexto.

El acceso a Systran está restringido a los traductores del SdT de la Comisión, ya sean en plantilla o *free-lance*. Existen dos modalidades de acceso: a través de la Web o por correo electrónico. En la primera modalidad el usuario puede interactuar directamente con la interfaz del sistema; en la segunda, en cambio, debe mandar un correo electrónico a una dirección especial, especificando las lenguas origen y meta, y adjuntando, en fichero aparte, el TO que desea traducir. De manera opcional, el usuario puede especificar el campo temático en el que se inserta el texto.

El uso de Systran por parte de los traductores de la sección española del SdT ha ido creciendo exponencialmente en los últimos años, como demuestra la siguiente tabla (Martínez 1998:2):

	1994	1995	1996	1997
<i>Total de páginas traducidas</i>	16 000	16 000	15 000	16 000
<i>Total de páginas traducidas con ayuda de SYSTRAN</i>	900	2 700	7 000	8 500

De acuerdo con Martínez (1998:2), si el uso de Systran no se hizo popular antes entre los traductores al español no fue porque éstos se mostraran reticentes, sino simplemente porque todavía no se había facilitado el acceso a través de la Web o por correo electrónico.

Como apuntábamos en la sección segunda del presente trabajo, los textos jurídicos son especialmente adecuados para la TA. La siguiente tabla ofrece un extracto de un artículo de Beaven (1998) con múltiples ejemplos, en los que se invita al lector a juzgar por sí mismo. Obsérvese que los fragmentos están organizados en tres filas: la primera contiene el TO; la segunda, la traducción de EC-Systran en 1997; la tercera, la traducción del mismo fragmento diez años antes.

TO (IN)	<i>The Commission's difficulties in checking legality, regularity and sound financial management aspects of the Union's finances undermine its ability to play its central role in managing the budget.</i>
1997	Las dificultades de la Comisión en la comprobación de la legalidad, de la regularidad y de aspectos sanos de gestión financiera de las finanzas de la Unión socavan su capacidad de desempeñar su papel central en gestionar el presupuesto.
1987	Las dificultades de la comisión en la comprobación de los aspectos financieros de administración de legalidad, de regularidad y de sonido de las finanzas de la unión minan su capacidad al juego su papel central en conseguir el presupuesto.

TO
(IN) *No formal list of supporting arguments was compiled but a number of points were common to the papers and discussions, including the following:*

1997 No se compiló ninguna lista formal de argumentos favorables sino que varios puntos eran comunes a los documentos y a las discusiones, incluida la siguiente:

1987 *Ninguna lista formal de mantener las discusiones fue compilada pero varios puntos eran comunes a los papeles y a las discusiones, con inclusión del siguiente:*

TO
(FR) *Sur la base d'un rapport publié le 5 juillet, la Commission a organisé à Louvain, du 23 au 25 septembre, un séminaire réunissant des représentants des autorités douanières et des opérateurs économiques, sur les perspectives de réforme du régime de perfectionnement actif*

1997 Sobre la Base de un informe publicado el 5 de julio, la Comisión organizó en Lovaina, del 23 al 25 de septiembre, un seminario reuniendo representantes de las autoridades aduaneras y de los agentes económicos, sobre las perspectivas de reforma del régimen de perfeccionamiento activo.

1987 *Sobre la base de un informe publicado los 5 julio, commission ha organizado con Lovaina, de los 23 a los 25 septiembre, un seminario que arma repräsentants de las autoridades douanizeres y de los operadores económicos, en las perspectivas de la reforma de la regla perfectionnement activa.*

TO
(FR) *La Commission a d'autre part clôturé la procédure relative aux exportations de fibres de polyester de la Communauté vers la Turquie.*

1997 La Comisión por otra parte dio por finalizado el procedimiento relativo a las exportaciones de fibras de poliéster de la Comunidad hacia Turquía.

1987 *Commission de otra parte ha cercado el procedimiento relativo a las exportaciones de las fibras del poliéster de la comunidad hacia el pavo.*

TO
(FR) *Dans le cadre de la mise en oeuvre du règlement CE n° 3381/94 instituant un régime communautaire de contrôle de ces biens, la Commission a par ailleurs organisé une audition des représentants des exportateurs concernés par ce régime.*

1997 En el marco de la aplicación del Reglamento CE n° 3381/94 que instituía un régimen comunitario de control de estos bienes, la Comisión por otra parte organizó una audiencia de los representantes de los exportadores afectados por este régimen.

1987 *Dans.le.cadre.de la aplicación del reglamento la CE n° 3381/94 que instituye una regla comunitaria del control estos bienes, commission par.ailleurs ha organizado una audiencia repräsentants de los exportadores concern3es por esta regla.*

5. Un ejemplo de TAO: las memorias de traducción

Según Melby (1995:225ss), la idea original de los sistemas de memoria de traducción (SMT) surgió de su propio grupo de investigación en Brigham Young University durante los años setenta, y se implementó por primera vez en 1981 en ALPS, un sistema comercial de TA. La mayoría de los investigadores, sin embargo, hinca las raíces de la noción de SMT en dos influyentes artículos publicados por Kay (1980) y Arthern (1981) (cf. Somers, en prensa, pág. 33). Comercialmente hablando, los primeros SMT no se pusieron a la venta hasta la década de los noventa. Hoy, apenas transcurrida una década, tales sistemas constituyen, como afirma Somers (en prensa, pág. 41), las herramientas de traducción más populares de cuantas han surgido de la investigación y desarrollo en TA y TAO.

Como tantas otras herramientas de TAO, los SMT beben de tres nociones fundamentales de la lingüística computacional (LC), a saber, *reutilización, integración e intercambio de información* (Cámara 2001:1). Por la primera se apela al reciclaje continuo de los recursos disponibles para el estudio de las distintas disciplinas lingüísticas; por la segunda, al desarrollo de software que se adapte perfectamente a los entornos informáticos existentes; por la tercera, finalmente, a la creación de herramientas de procesamiento de lenguaje natural (PLN) que garanticen una fácil gestión de la información.

La noción misma de SMT parte de la premisa de que como afirma Isabelle (1993a:8) *existing translations contain more solutions to more translation problems than any other available resource* (cf. también Macklovitch 1992; Kay & Röscheisen 1993:122). Según Isabelle (1993b), los diccionarios y las bases de datos terminológicas, en su concepción actual, sólo proporcionan *equivalencias virtuales*, esto es, traducciones potenciales de una palabra o frase en varios contextos; por ende, las únicas fuentes de *traducciones reales* son, según el autor, los corpus de traducciones previas.

El grupo creado por EAGLES (1996) para la evaluación de herramientas de PLN, propone la siguiente definición de *memoria de traducción*:

a translation memory is a multilingual text archive containing (segmented, aligned, parsed and classified) multilingual texts, allowing storage and retrieval of aligned multilingual text segments against various search conditions.²

En otras palabras, una memoria de traducción es un *sistema* concebido para comparar una determinada secuencia del TO con otras secuencias almacenadas en uno o varios *bancos de datos* de traducciones previas, con el fin de reutilizar cada segmento original y su correspondiente traducción. Como se desprende de los términos anteriores en cursiva, surge aquí cierta confusión terminológica, ya que como apunta Somers (en prensa, pág. 31), a menudo se emplea el término *memoria de traducción* para referirse tanto al software genérico como a cada uno de los bancos de segmentos alineados que se han almacenado mediante su uso. Muchos autores en lengua inglesa han optado por deslindar ambas nociones empleando *translation memory (system)* para referirse al software y *memory database (MDB)* para referirse a los bancos de segmentos, noción que a su vez

enfrentan a *terminology database* (TDB), reservada a los bancos de términos exclusivamente. En español, sin embargo, no se ha impuesto ninguna precisión terminológica, al menos de manera generalmente aceptada. A ello puede haber contribuido el hecho de que muchos paquetes de SMT no se han localizado a nuestra comunidad, y aquellos que sí se han localizado no acaban de resolver el problema. Así, por ejemplo, el manual de instrucciones de Déjà Vu, sistema de SMT español de amplia difusión internacional, habla de *memorias* para referirse a *memory databases* y de *bases de datos terminológicas* para referirse a *terminology databases*, y en ningún momento se refiere al sistema como una *memoria de traducción*, sino como un *sistema integrado de traducción para traductores*. Ello se debe, a su vez, a que la mayoría de los distribuidores de SMT suelen vender tales sistemas integrados en arquitecturas cada vez más complejas —las llamadas *translation workstations*—, en donde la SMT puede utilizarse conjuntamente con un programa de edición de textos, una herramienta de alineación de segmentos, un gestor de terminología y una aplicación de conversión de ficheros. Pero volviendo a la cuestión que nos ocupa, esto es, la confusión conceptual a la que induce el término *memoria de traducción* en español, proponemos aquí la siguiente distinción: para referirnos al software, hablaremos explícitamente de *sistema de memoria de traducción*, abreviado SMT; de esta manera evitaremos también utilizar las siglas MT por sí solas para referirnos a tales sistemas, ya que éstas se emplean profusamente en la literatura científica para hablar sobre *machine translation* y pueden inducir a confusión. Por lo que respecta a las bases de datos que contienen segmentos alineados —laxamente llamadas *memorias*—, utilizaremos la etiqueta *bancos de segmentos*, y sobre todo, su abreviación BS. Del mismo modo, a las bases de datos que contienen términos los llamaremos *bancos de términos*, abreviados BT. La siguiente tabla resume nuestra propuesta de equivalencias:

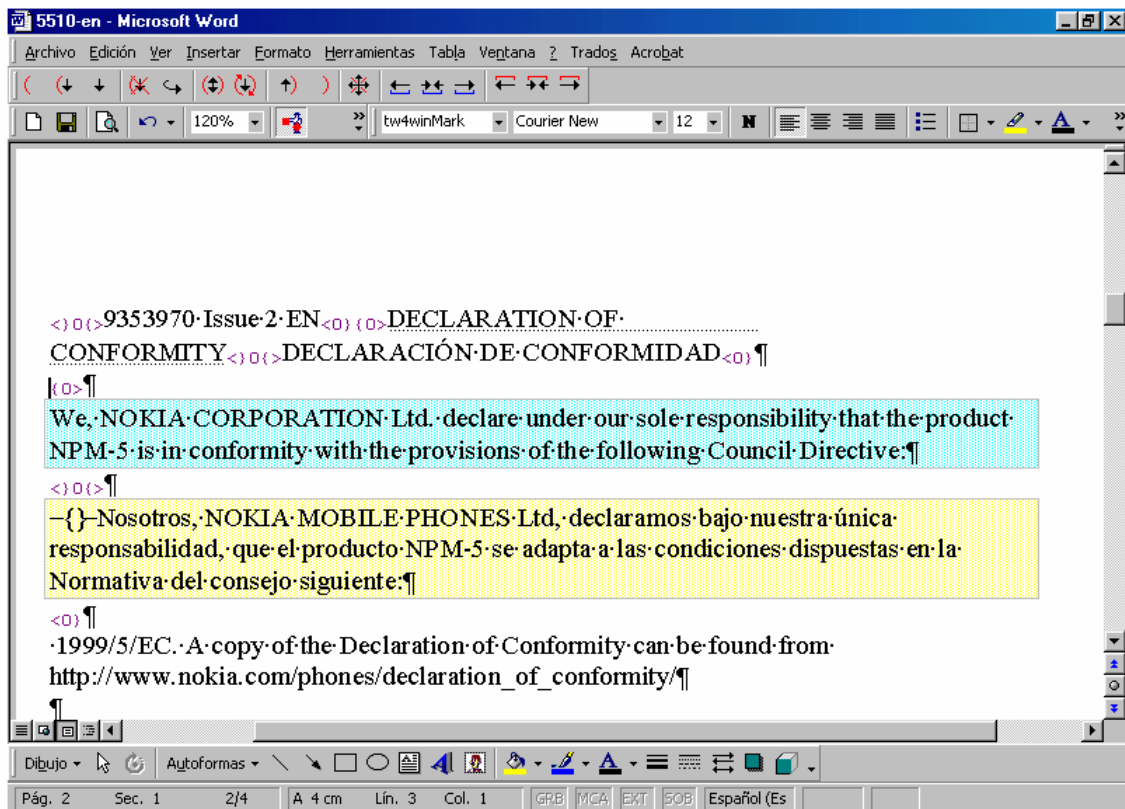
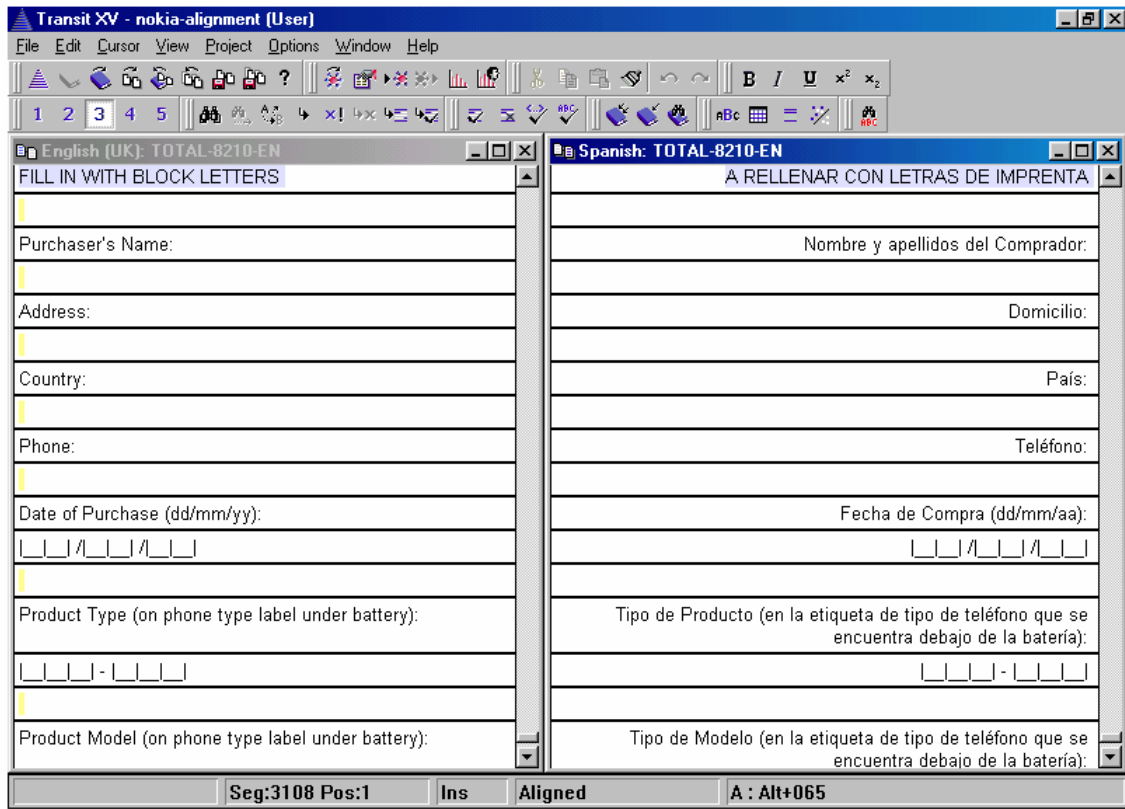
INGLÉS		ESPAÑOL	
Translation Memory (System)	TM(S)	Sistema de Memoria de Traducción	SMT
Memory database	MDB	Banco de segmentos	BS
Terminology database	TDB	Banco de términos	BT

A continuación profundizamos en la noción de *memoria de traducción*, apoyados en la distinción que acabamos de realizar (BS y BT frente a SMT).

Básicamente, un **BS** puede contemplarse como un *corpus paralelo* de segmentos del TO alineados de manera explícita con sus correspondientes equivalencias del TT. Tal estructura se conoce a veces con el nombre de *bitexto*. En palabras de Harris (1998:41, *apud* Groß 1998: 6):

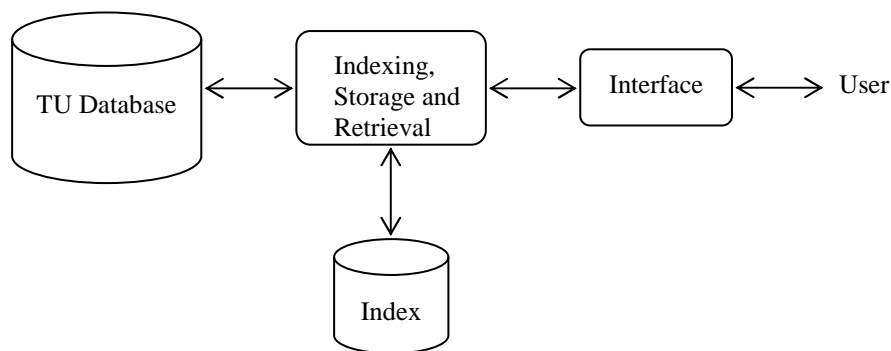
*Bi-text is [...] the juxtaposition of a translation's source (ST) and target (TT) texts on the same page or screen: either parallel – in which case the two texts appear side by side on a split page or screen – or interlinearly – whereby each ST line is intercalated between a corresponding line of TT.*³

La siguiente ilustración muestra capturas de pantalla de un bitexto paralelo generado con Transit XV (arriba) y de un bitexto interlineal generado con Trados 5.5 (abajo).



En un BS los textos no se almacenan como tales, sino fraccionados en *unidades de traducción* (UT) conocidas frecuentemente como *segmentos*. Como explica Abaitua (1997 y 2001), la noción de *segmento* ha redefinido sustancialmente la noción tradicional de UT. Por ejemplo, desde la adopción de formatos estándar como el TMX, un *segmento* se define como cualquier secuencia de caracteres contenida entre las etiquetas <TU> y </TU> (Gómez 2001). En la práctica, esto se traduce en que la longitud de un segmento puede oscilar entre unidades muy pequeñas (p. ej., las entradas de cada una de las celdas de una tabla, cada elemento de una lista o incluso palabras aisladas) y unidades considerablemente mayores (por ej., oraciones, párrafos o incluso un texto entero). En otras palabras, y siguiendo la tipología de UTs de Bennett (1994), los segmentos de un BS abarcan desde los llamados *átomos de traducción* (los segmentos más pequeños de un enunciado cuyas partes constituyentes no pueden traducirse por separado debido a su fuerte cohesión; Vinay & Darbelnet 1958:16, 36ss) hasta las llamadas *macrounidades de traducción* (esto es, el texto, definido como la mayor unidad lingüística que debe contemplar el traductor —Bennett, *op. cit.*), pasando por todo tipo de unidades léxicas. Cabe señalar, no obstante, que por fines puramente funcionales, la gran mayoría de segmentos suelen ser oraciones, razón por la cual los SMT se conocen a veces con el sobrenombre de *memorias de oraciones* (del inglés *sentence memories*; O'Brien 1998:116).

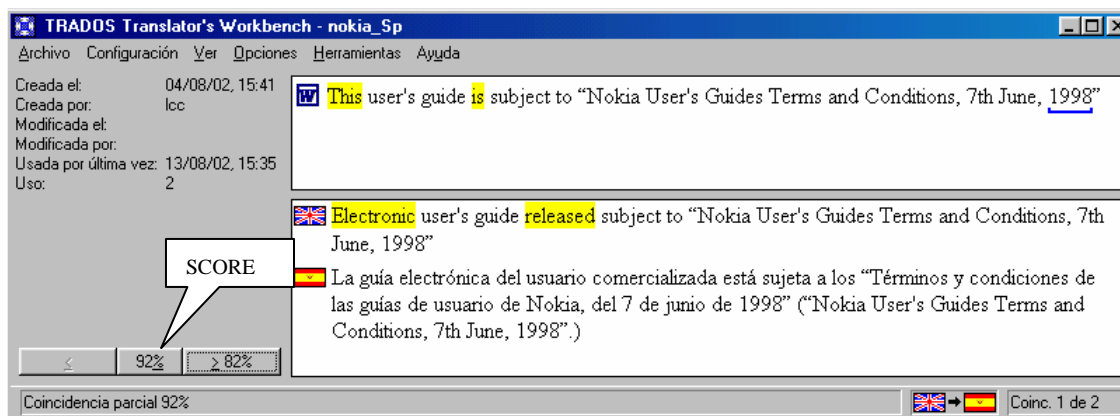
Hasta ahora hemos esbozado los rasgos fundamentales de un BS. A continuación nos centraremos en la otra cara de la dualidad, esto es, el sistema. En líneas generales, un SMT consta, básicamente, de un motor que permite interactuar con el BS para almacenar, buscar y recuperar UT. Así lo ilustra Trujillo (1999:69):



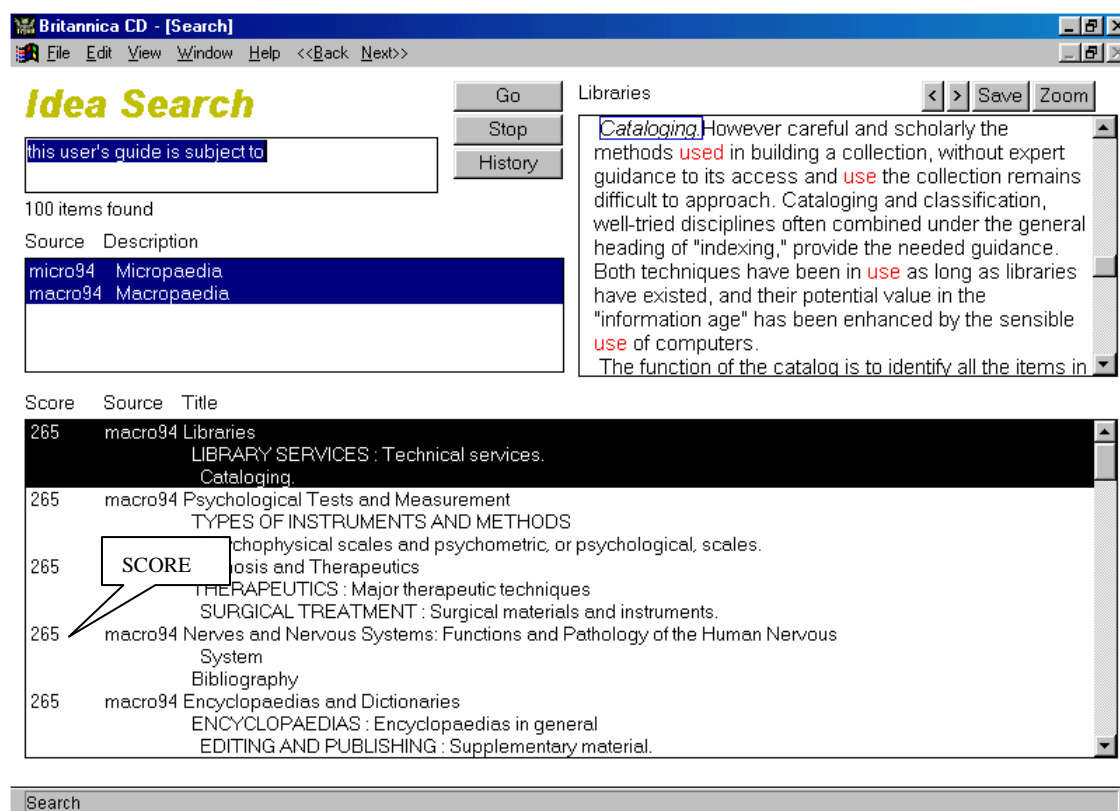
Funcionalmente, un SMT es muy similar a las bases de datos (BD) objeto de estudio de la RI, o Recuperación de la Información (del inglés *Information Retrieval, IR*), subárea del PLN. Nos referimos, entre otros ejemplos, a los motores de búsqueda de Internet, los catálogos bibliográficos informatizados y los diccionarios electrónicos sensibles al contexto. Así lo expresa Trujillo (1999: 61):

some of the techniques used in IR can be applied directly to TM by viewing the query as the input sentence, and the document collection as the TU[']s DB.⁴

La siguiente captura de pantalla ilustra un ejemplo de recuperación de datos mediante un SMT (Trados 5.5).



Por su parte, la siguiente captura de pantalla es un ejemplo de enciclopedia electrónica sensible al contexto (Encyclopaedia Britannica 1.01).



Hasta la fecha se han desarrollado tres tipos de motores en los SMT (Cámara 2001:6), a saber: 1) el *estático*, que sólo recupera *coincidencias totales* (*exact matches*); 2) el *interactivo*, que recupera *coincidencias parciales* (*fuzzy matches*) según un umbral de semejanza predefinido por el usuario, y 3) una combinación de ambos, que recupera tanto las coincidencias totales como las parciales. Además, los últimos avances en I+D están suscitando la aparición de motores de búsqueda que recuperan nuevos tipos de coincidencias, tales como la *completa* (*full match*) y la *subsegmental* (*sub-segment match*).

Debido a la fuerte competencia del mercado, los distribuidores de SMT tienen por norma no revelar los detalles de los llamados *algoritmos detectores de coincidencias* (del inglés *matching algorithms*). Según Somers (en prensa, pág. 38),

la mayoría de ellos se basa en la similitud de las secuencias de caracteres; sin embargo —añade— muchos sistemas están comenzando a ofrecer a los usuarios la posibilidad de ponderar la búsqueda utilizando otros tipos de coeficientes, tales como el grado de fiabilidad del segmento en memoria (fuente, autoría, etc.), las diferencias de formato, e incluso la importancia de determinadas palabras.

Como se desprende de lo anterior, los SMT están concebidos para la traducción de TFE, entre los que ocupan un lugar destacado los textos jurídicos. Así lo demuestra la creciente implementación de SMT en los distintos entornos de la traducción jurídica, hecho que suscita cada vez mayor interés por parte de la comunidad académica internacional (*cf.*, por ej., los proyectos de investigación de Paulsen Christensen, Syddansk Universitet, y Cerezo, Universidad de Málaga/Georgetown University).

Para ilustrar la aplicación de SMT en los entornos profesionales de la traducción jurídica recurriremos de nuevo al ejemplo paradigmático del SdT de la Comisión Europea. Como ya vimos en el primer apartado del presente trabajo, el afán del SdT por agilizar el proceso de traducción mediante el desarrollo de herramientas informáticas nace ya en la década de los setenta, primero con la adquisición de Systran (*cf.* apartado 4), y más tarde (finales de los ochenta, comienzos de los noventa), con la adquisición de diversas herramientas TAO. Hoy, el flujo de trabajo del SdT se encuentra altamente informatizado en todas sus fases, como demuestra la tabla del final del apartado adaptada de McCluskey (2002):

La traducción de textos jurídicos en el seno del SdT se aborda, pues, desde un entorno altamente informatizado, que se caracteriza por los siguientes puntos:

- 1) La presencia de SMT por partida doble (*cf.* Strandvik 2001):
 - a. Un SMT central con más de 50 millones de UT en las 11 lenguas de la Unión, al puede accederse independientemente, mediante su propia interfaz, o a través de la red del apartado *b*.
 - b. Una red de SMT locales (TWB de Trados) y sus correspondientes gestores terminológicos (MultiTerm de Trados).
- 2) La existencia de una serie de recursos informáticos que pueden utilizarse en tándem con los SMT:
 - c. *Celex*, un banco de datos jurídicos que engrosa día a día el BS del SMT central.
 - d. *Eur-lex*, un servicio de la oficina de publicaciones de la Comisión Europea que proporciona libre acceso al boletín oficial de las Comunidades, así como a todos los documentos legislativos de la Unión, entre otros textos jurídicos.
 - e. *Eurodicautom*, que nutre el BT de los SMT locales, a través de MultiTerm.

FASE DEL PROCESO TRASLATIO	SOFTWARE UTILIZADO	FUNCIÓN
El SdT recibe un encargo de traducción desde instancias superiores.	Poetry	Transmisión electrónica de encargos de traducción de los diversos clientes al SdT.
Se acepta el encargo y se asigna a la unidad correspondiente.	Winsuivi	Gestión electrónica de encargos de traducción dentro del propio SdT.
Se evalúa si el documento es traducible mediante TWB y se prepara la traducción.	DossierManagement	Gestión de traducciones.
El jefe de la unidad de traducción correspondiente recibe el encargo.	Winsuivi	
	Capacity Planning	Distribución electrónica de los encargos de traducción a las distintas unidades.
Se crea un proyecto de traducción.	Dossier Management	
Un profesional traduce el documento; otro lo revisa.	Word	Procesamiento de texto.
	TWB	SMT local para la gestión de datos procedentes de la memoria central.
	Excel	Hoja de cálculo.
	Celex	Banco de textos jurídicos de la UE.
	MultiTerm	Gestor terminológico local.
	Eurodicautom	BD terminológica de la Comisión Europea.
	SdTVista	Repertorio de documentos traducidos con opción de búsqueda en todo el texto.
	EC SYSTRAN	STA.
Se da via libre al documento.	DossierManagement	
Se archiva el documento.	SdTVista	
Se envía la traducción al cliente.	Winsuivi	

6. Conclusiones

Tras cincuenta años de I+D en TA, el ideal de reemplazar al traductor humano dista mucho de ser alcanzado, si acaso es posible. Los mayores logros de la traducción por ordenador se han dado en el ámbito de los textos para fines específicos, entre los que ocupa un lugar destacado el discurso jurídico. La TA es a veces percibida con recelo por parte de los traductores, ante el temor de que ésta pueda llegar a reemplazarlos. Sin embargo, esto no es posible, ya que los ordenadores carecen del conocimiento y la experiencia de los humanos acerca del mundo real.

Por otro lado, la TAO recibe una mayor acogida por parte de los traductores, que utilizan de manera cada vez más generalizada sistemas de memoria de traducción. No obstante, estos sistemas también pueden repercutir negativamente sobre el traductor, ya que plantean nuevos temas como la reducción de tarifas por coincidencias parciales y los derechos del cliente sobre la propiedad de la memoria. En nuestra opinión, los avances en traducción por ordenador no deben contemplarse como una amenaza, sino como una ayuda al humano, ya que, de una parte, le liberan del tedio de traducir secuencias recurrentes, y de otra, han abierto nuevas salidas laborales, como la de revisor o post-editor.

Notas

1. «La asociación de un motor de traducción automática a un entorno de traducción asistida permite explotar al máximo las características de ambos sistemas. En este entorno integrado, el traductor, además de disponer del potencial de consulta propio de las memorias de traducción (fundamental para la traducción de textos repetitivos o textos similares a otros traducidos con anterioridad), puede servirse de propuestas de traducción automática para frases completamente nuevas no presentes en las memorias.»
2. «una memoria de traducción es un archivo de textos multilingües (segmentados, alineados, analizados sintácticamente y clasificados), que permite almacenar segmentos de textos multilingües alineados y recuperarlos mediante diferentes estrategias de búsqueda.»
3. «Por bi-texto se entiende [...] la yuxtaposición de los textos origen (TO) y meta (TM) en una misma página o pantalla, bien en paralelo –cada texto aparece a un lado de la página o de la pantalla– o de manera interlineal –cada línea del TO va seguida de su línea correspondiente en el TM.»
4. «algunas de las técnicas utilizadas en el ámbito de la RI pueden aplicarse directamente a la TA; la secuencia de búsqueda puede equipararse a la oración por traducir, y la colección de documentos, al BD de UT.»

El presente trabajo ha sido realizado en el seno del proyecto PB98-1399 "Diseño de un tipologizador textual para la traducción automática de textos jurídicos (español <-> inglés/alemán/italiano/árabe)" (DGICYT, 1999-2002).

Referencias

- ABAITUA, Joseba. 1997. Segmentos y unidades de traducción. Disponible en: <http://www.serv-inf.deusto.es/abaitua/konzeptu/12uutts.htm>.
- . 2001. Memorias de traducción en TMX compartidas por Internet. *Revista Tradumàtica: Traducció i tecnologies de la informació*. Disponible en: <http://www.fti.uab.es/tradumatica/revista/num0/articles/jabaitua/art.htm>.
- ARTHERN, Peter John. 1981. Aids Unlimited: The Scope for Machine Aids in a Large Organization. *Aslib Proceedings* 33.
- BAR-HILLEL, Yehoshua. 1960. The present status of automatic translation in languages. *Advances in Computers* 1.
- BEAVEN, John. 1998. MT: 10 Years of Development. Disponible en: http://europa.eu.int/comm/translation/reading/articles/pdf/1998_01_tt_beaven1.pdf.
- BENNETT, Paul. 1994. Translation Units in Human and Machine. *Babel*, 14.
- CÁMARA, Lidia. 2001. El papel de las herramientas TAO en la documentación técnica multilingüe. *Revista Tradumàtica: Traducció i tecnologies de la informació*. Disponible en: <http://www.fti.uab.es/tradumatica/revista/num0/articles/lcamara/art.htm>.
- CEREZO CEBALLOS, Luis. En curso. *Las memorias de traducción al servicio de los textos jurídicos*. Málaga: Universidad (Proyecto de tesis doctoral adscrito a beca FPU del MECD, ref. AP99 44582773).
- EAGLES. 1996. Design and function of translation memory. En *Evaluation of Natural Language Processing Systems: Final Report. Eagles Document EAG-EWG-PR.2*. Dinamarca: Center for Sprogteknologi. (También disponible en: <http://issco-www.unige.ch/projects/ewg96/node152.html>.)
- ESSELINK, Bert. 1998 [2000, ed. rev.]. *A Practical Guide to Software Localization*. Amsterdam: John Benjamins.
- FREIGANG, Karl-Heinz. 2001. Automation of Translation: Past, Presence, and Future. *Revista Tradumàtica: Traducció i tecnologies de la informació*. Disponible en: <http://www.fti.uab.es/tradumatica/revista/num0/articles/khfreigang/art.htm>.
- GÓMEZ, Josu. 2001. Una Guía al TMX. *Revista Tradumàtica: Traducció i tecnologies de la informació*. Disponible en: <http://www.fti.uab.es/tradumatica/revista/num0/articles/jgomez/art.htm>.
- GÖPFERICH, Susanne. 1995a. *Textsorten in Naturwissenschaften und Technik. Pragmatische Typologie - Kontrastierung - Translation*, Tübingen: Gunter Narr (Forum für Fachsprachen-Forschung 27).
- . 1995b. A Pragmatic Classification of LSP Texts in Science and Technology, *Target: International Journal of Translation Studies*, 7, 2.
- GROß, Bettina. 1998. *Vergleichende Untersuchung von Alignment-Tools*. Saarbrücken: Universität des Saarlandes (Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen).
- HAHN, Walther von. 1995. Machine Translation. Disponible en: <http://www.racai.ro/awd/awd16/hahn.html>.
- HARRIS, Brian. 1998. Are you bi-textual? *Language Technology*, 7.
- HUTCHINS, William John y Harold Leopold SOMERS. 1992. *An Introduction to Machine Translation*. Londres: Academic Press.
- HUTCHINS, William John. 1986. *Machine Translation: Past, Present, Future*, Chichester/Nueva York: Ellis Horwood/Wiley.

- . 1996. ALPAC: the (in)famous report. *MT News International* 14, 9-12. (También disponible en:
<http://ourworld.compuserve.com/homepages/WJHutchins/Alpac.htm>.)
- ISABELLE, Pierre. 1993a. Bi-textual Aids for Translators. Technical Report. Quebec: Centre for Information Technologies Innovation.
- . 1993b. Translation Analysis and Translation Support Tools. En *Thirteenth International Conference on Artificial Intelligence, Expert Systems and Natural Language*. Avignon.
- KAY, Martin. 1980. The Proper Place of Men and Machines in Language Translation. Research Report CSL-80-11. Palo Alto: Xerox Research Centre.
- KAY, Martin y RÖSCHEISEN, Martin. 1993. Text Translation Alignment. *Computational Linguistics*, 19 (3).
- KITTREDGE, Richard y John LEHRBERGER, eds. 1982. *Sublanguage: Studies of language in restricted semantic domains*. Berlin: De Gruyter.
- MACKLOVITCH, E. 1992. Corpus-Based Tools for Translators. En *Proceedings of the 33rd Annual Conference of the American Translators Association*. San Diego.
- MARTÍNEZ, Jesús. 1998. La buena salud de la traducción automática al español. Disponible en:
http://europa.eu.int/comm/translation/reading/articles/pdf/1998_01_tt_martinez.pdf.
- MCCLUSKEY, Brian, coord. 2002. *Tools and workflow at the Translation Service of the European Commission*. Bruselas: SdT de la Comisión Europea. Disponible en:
http://europa.eu.int/comm/dgs/translation/bookshelf/2002_tools_and_workflow_en.pdf.
- MELBY, Alan K. 1995. *The Possibility of Language: A Discussion of the Nature of Language*. Amsterdam: John Benjamins.
- NAGAO, Makoto. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. En *Artificial and Human Intelligence: edited review papers at the International NATO Symposium on Artificial and Human Intelligence sponsored by the Special Programme Panel held in Lyon, France, October, 1981*. Amsterdam: Elsevier Science Publishers.
- O'BRIEN, Sharon. 1998. Practical Experience of Computer-Aided Translation Tools in the Software Localization Industry. En: Lynne Bowker, Michael Cronin, Dorothy Kenny and Jennifer Pearson, eds. *Unity in Diversity?: Current Trends in Translation Studies*. Manchester: St. Jerome.
- PAULSEN CHRISTENSEN, Tina. En curso. *Kritisk vurdering af anvendeligheden af en translation memory som værktøj ved oversættelsen af juridiske dokumenter inden for selskabsretten*. Kolding: Syddansk Universitet. Disponible en:
<http://www.humaniora.sdu.dk/ifki/forskningsplan/Forskningsplan2002.pdf>.
- PETRITS, Angeliki. 2001. EC-Systran: The commission's machine translation system. Disponible en:
http://europa.eu.int/comm/translation/reading/articles/pdf/2001_mt_mtfullen.pdf
- PIERCE, John R., John B. CARROLL, Eric P. HAMP, David G. HAYS, Charles F. HOCKETT, Anthony G. OETTINGER, y Alan PERLIS. 1966. *Languages and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council*. Washington, D.C.: National Academy of Sciences, National Research Council (Publication 1416). Disponible en:
<http://www.nap.edu/books/ARC000005/html/index.html>.

- PRIAMO, Angelo. 1999. Progettazione, sviluppo e integrazione di software di Traduzione automatica e Traduzione assistita. En: AITI, ed. *Atti del seminario Terminologia e Traduzione Assistita*. Trento: AITI.
- SAGER, Juan Carlos. 1994. *Language Engineering and Translation: Consequences of automation*. Amsterdam: John Benjamins.
- SOMERS, Harold Leopold. (en prensa). Translation Memory Systems. En H.L. Somers, ed. *Computers and Translation: A handbook for translators*. Amsterdam: John Benjamins (Translation Studies).
- STRANDVIK, Ingemar. 2001. Las memorias de traducción en la Comisión Europea. *Revista Tradumàtica: Traducció i tecnologies de la informació*. Disponible en: <http://www.fti.uab.es/tradumatica/revista/num0/articles/istrandvik/imprimir.pdf>.
- TRUJILLO, Arturo. 1999. *Translation Engines: Techniques for Machine Translation*, Berlin; Heidelberg; New York: Springer.
- VINAY, Jean-Paul y Jean DALBERNET. 1958. *Stylistique comparée du français et l'anglais*. Paris: Didier.
- WEBB, Lynn Eckel. 1998. *Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis*. Monterey: Monterey Institute of International Studies. Disponible en: <http://www.webbsnet.com/translation/thesis.html>.