

1. Cuestiones preliminares: definición y objetivos

1.1. Definición. ¿Qué es la TA?

- La traducción automática (TA) consiste en utilizar ordenadores para automatizar algunos o todos los procesos que intervienen en la traducción de una lengua a otra (Ward, Jurafsky y Martin 2000:799).
- La TA es una tecnología del campo de la informática que permite traducir textos de una lengua a otra mediante un programa de ordenador (Juan Alberto Alonso, *Curso de industrias de la lengua*, Soria, julio, 1998).

1.2. Ideas preconcebidas sobre la TA

Ahora bien, la tarea de la TA a veces se encuentra con algunos prejuicios, como los siguientes:

- 1) *La traducción automática es inútil porque no puede enfrentarse con textos literarios.* Sin embargo, este tipo de traducciones representan un porcentaje mínimo de las traducciones profesionales, por lo que un programa de TA pocas veces se va a tener que enfrentar con ellas y, de hacerlo, no se va a esperar una traducción igual a la de un traductor humano, que dispone de unos conocimientos culturales y contextuales de los que carece el programa.
- 2) *La traducción automática es inútil porque produce traducciones de muy baja calidad.* Sin embargo, muchas empresas y organismos la utilizan de forma productiva; además, no siempre es necesario contar con traducciones perfectas; por otra parte, con frecuencia las traducciones realizadas por personas también deben ser revisadas con posterioridad.
- 3) *La traducción automática amenaza el trabajo de los traductores.* Sin embargo, se puede considerar que, más bien, libra al traductor de la parte más repetitiva y rutinaria de su trabajo.

1.3. Factores que motivan la TA (Hutchins y Somers¹). ¿Por qué la TA?

Hay una serie de factores que han movido a los investigadores a indagar sobre las posibilidades de automatizar el proceso de la traducción:

- 1) *Factores prácticos* (volumen de información). Traducir por medios mecánicos es sin duda un desafío científico. Pero la razón por la que la traducción automática despierta tanto interés no es de índole científica, sino de mera necesidad práctica. Hubert Murray (cf. Abaitua²) cifró en 20 millones el número de palabras de información técnica que se generaban en el mundo cada día, cifra que crece exponencialmente cada día que pasa. Un lector capaz de leer 1 000 palabras por minuto necesitaría 45 días, a una media de 8 horas diarias, para digerir la producción de solo un día. Al cabo de esos 45 días, su desfase sería

¹ Hutchins, W. J. y Somers, H. L. (1995 [1992]): *Introducción a la traducción automática*, Madrid: Visor. Traducción dirigida por Joseba K. Abaitua.

² J. Abaitua: La traducción automática: presente y futuro. URL: <http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/ta97.htm> y <http://www.foreignword.com/es/Technology/art/Abaitua/>

de 5,5 años. En resumen, una comunidad lingüística necesitaría 2 000 traductores para traducir diariamente todos estos documentos.

2) *Factores sociopolíticos* (multilingüismo). La necesidad práctica de disponer de este tipo de sistemas está estrechamente relacionada con la existencia de una demanda creciente de traducciones en mercados más internacionales y en los que los diferentes tipos de documentos han de redactarse de forma cada vez más rápida y en un mayor número de lenguas. En este contexto, la TA ofrece posibilidades que de otra forma implicarían un coste elevado tanto de tiempo, como de dinero. Abaitua (*cf.* referencia anterior) proporciona los siguientes datos:

- Volumen de publicaciones. En los últimos treinta años se ha generado más información que en los cinco mil anteriores. P.e. solo en EE.UU. se editan más de 9 000 publicaciones periódicas cada año, sin contar los libros que salen a la luz diariamente en el mundo.
- Traducción e interpretación. En organismos multilingües como la Unión Europea estas tareas consumen casi el 50% de sus costes administrativos. Cada vez que un país se incorpora o que se elabora una ley nueva, la labor se incrementa considerablemente. Lo mismo ocurre en la ONU, en instituciones bilingües como el Parlamento Canadiense o, más cercanas a nosotros, las Administraciones de las Comunidades Autónomas españolas en las que existen dos lenguas oficiales y toda la información debe aparecer publicada en ambas: español-catalán, español-gallego y español-euskera. En este contexto, los avances de las nuevas tecnologías aplicadas a la TA pueden contribuir a agilizar el volumen de información implicado.

3) *Factores comerciales y tecnológicos*. La globalización en el marco político tiene repercusiones económicas tales como el incremento del comercio mundial, lo que a su vez genera la necesidad de producir más traducciones. A su vez, la evolución tecnológica genera de forma simultánea más documentación (sobre todo manuales) que debe traducirse a múltiples idiomas de forma rápida.

1.4. Tipos de textos y TA

No todos los textos ofrecen las mismas posibilidades para obtener una traducción automática aceptable. Por este motivo, en la actualidad la TA se aplica a dos tipos fundamentales de documentos:

- a) Documentos que tratan materias muy específicas y que se abordan de forma rutinaria.
- b) Documentos destinados a lectores a los que les baste con una traducción aproximada.

La traducción automática funciona relativamente bien con textos escritos en un lenguaje formal y estereotipado, hasta cierto punto predecible que, por otra parte, son mecánicos y rutinarios:

- Disposiciones legales y administrativas.
- Textos jurídicos (normativas, contratos, etc.).
- Manuales técnicos.
- Boletines informativos (partes meteorológicos, información de bolsa, teletexto, anuncios por palabras, ofertas de empleo, etc.).
- Resúmenes de publicaciones científicas, informes técnicos y textos expositivos.

Por el contrario, la traducción automática es inadecuada para procesar textos creativos (narrativa, poesía, teatro, ensayo...) o textos expresivos, en los que abunde el lenguaje coloquial, el lenguaje humorístico, los chistes, los juegos de palabras, etc. Como ya se ha señalado, los textos literarios representan un porcentaje mínimo en el mercado de la traducción.

1.5. Objetivos de la TA

De acuerdo con lo anterior y siguiendo a Hutchins y Somers, los objetivos que podemos esperar de la TA son dos:

- 1) *Traducciones en borrador.* En la actualidad, los programas de TA nos permiten obtener una primera versión de un texto en otra lengua.
 - En ocasiones, esa versión es suficiente por sí misma para los propósitos perseguidos por el usuario, que se da por satisfecho con ella.
 - En ciertas áreas, por lo generalmente restringidas temáticamente, interesa corregir o revisar la traducción (*postedición*) para disponer de versiones finales de calidad por unas tarifas económicas.
 - Incluso en algunos casos, si se aplican restricciones al texto original en cuanto al tipo de estructuras y vocabulario que se pueden utilizar (*preedición*), es posible obtener de forma automática resultados de mayor calidad que apenas requieren revisión.
- 2) *Textos sin valor literario.* La mayor parte de traducciones que se realizan tienen que ver con documentos técnicos y científicos, documentos administrativos, jurídicos y comerciales, manuales de instrucciones, libros de texto, etc. Los programas de traducción automática ayudan al traductor a agilizar la tarea y liberarle de la parte repetitiva de la misma.

Fidelidad, precisión, inteligibilidad, estilo y registro apropiados, son todos ellos criterios subjetivos que pueden aplicarse para determinar si una traducción es buena o mala. Sin embargo, en la práctica, lo que realmente importa, en lo que se refiere a la TA, es qué cantidad de correcciones deben realizarse para que una traducción alcance un *nivel de calidad aceptable* para un traductor o un lector. Con un concepto tan impreciso de lo que es la traducción, los investigadores y diseñadores de sistemas de TA tan solo pueden aspirar a:

- Realizar traducciones "útiles" en situaciones concretas (*objetivos de investigación claros*),
- Buscar aplicaciones adecuadas para las "traducciones" que dichos sistemas sean capaces de producir.

El problema para producir de forma automática una traducción de alta calidad de un texto cualquiera de una lengua a otra es la dificultad de automatizar el proceso completamente. Sin embargo, ciertas tareas más simples sí que se pueden abordar con los modelos computacionales actuales. En concreto, los sistemas de TA se centran en:

- (1) Tareas en las que basta con una traducción aproximada; p.e. la adquisición de información en la Web.

- (2) Tareas en las que puede intervenir un posteditor humano para mejorar la calidad del *output*; el borrador que proporciona el sistema puede ser útil, al acelerar el proceso, como primer paso para obtener la traducción definitiva. Se trataría de una traducción humana asistida por ordenador y no de una traducción totalmente automática. Se utiliza sobre todo en trabajos de grandes dimensiones y aquéllos que exigen una rápida respuesta. El ejemplo más conocido es el de los manuales de *software* o el de las noticias financieras.
- (3) Tareas limitadas a dominios pequeños o sublenguajes, en los que es posible obtener una TA totalmente automática de alta calidad. Las predicciones meteorológicas son un ejemplo de sublenguaje que no precisa de postedición. Se trata de dominios con un vocabulario limitado y que manejan unos pocos tipos básicos de frases. Apenas existe ambigüedad y si la hay se puede resolver fácilmente en función del contexto local, utilizando categorías gramaticales o rasgos semánticos. Otros ejemplos de sublenguajes son: manuales de mantenimiento de equipos, búsquedas relacionadas con vuelos, recomendaciones de restaurantes, etc.

1.6. Funciones de la TA

De acuerdo con lo expuesto, dos son los usos principales de la TA en nuestros días:

- 1) Hacer más rápida la traducción de los textos (*difusión de información*).
- 2) Facilitar a los interesados la comprensión de textos en idiomas que desconocen (*adquisición de información*).

1.7. Dimensiones de la TA

Teniendo en cuenta todo lo anterior, J. Abaitua³ enumera una serie de dimensiones o factores que hay que considerar a la hora de utilizar la traducción automática:

- Tipo
 - Densidad
 - Distancia
 - Desarrollo
 - Medio
 - Fines
- a) *Tipología de textos*. Cada texto presenta peculiaridades en función de parámetros como:
- *Género*: novela, poesía, teatro, etc. Se trata de diferencias extralingüísticas entre textos.
 - *Tipo*. Permite distinguir textos dentro de un mismo género. P.e. dentro del género novela, podemos encontrar: novelas históricas, novelas eróticas, novelas breves, etc.
 - *Registro*: culto, coloquial, etc. Suele variar según la situación y el contexto social o profesional. En traducción automática destaca el concepto de *sublenguajes* o *lenguajes de especialidad*, en contraposición con la *lengua común*, como señala

³ Introducción a la traducción automática (en 10 horas). URL:
http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/mt10h_es/index.html

Melby. Cada sublenguaje se define por el empleo de términos y estructuras sintácticas específicas; frente a la lengua común, cuyas expresiones y vocabulario son de uso genérico.

- *Estilo* (formal, informal). Introduce diferencias dentro de un mismo registro, ya que un mismo texto puede ser más o menos formal, etc. Tiene en cuenta parámetros lingüísticos de los textos: longitud de oraciones, utilización de perífrasis verbales, densidad léxica, uso de conectores, etc.

En general, en TA las consideraciones se reducen a lo que M. Kay denominó:

- *Lenguaje singular* ("remarkable language"): textos creativos, no aptos para la automatización.
- *Lenguaje no singular* ("unremarkable language"): textos no creativos, aptos para la automatización, también conocidos como *sublenguajes*.

Esta distinción es muy similar a la que se establece entre *lengua común* y *lenguajes de especialidad* (o lenguaje dinámico frente a estático, según otros autores).

<i>Lenguaje singular o creativo</i>	<i>Lenguaje no singular o no creativo</i>
Literatura Textos espontáneos Libros	Lenguaje repetitivo Lenguaje controlado Lenguaje normalizado P.e. disposiciones administrativas, legislación, informes y manuales técnicos, boletines divulgativos, teletipos, etc.

- b) *Densidad léxica y granularidad*. Estos dos conceptos están relacionados con la distinción anterior entre lenguajes de especialidad y lengua común.
- La *granularidad* (Melby) se refiere al tamaño de las unidades léxicas de un texto. Una granularidad más gruesa implica la presencia en el texto de frases hechas, locuciones, proverbios o refranes, nombres propios, léxias complejas (combinaciones de palabras que se lexicalizan, como *comida rápida*), colocaciones (combinaciones de palabras que con frecuencia aparecen juntas pero que todavía no se han lexicalizado), etc. Antes de proceder a la traducción, el sistema debe identificar estas unidades mayores que la palabra, uno de los principales problemas de la teoría de la traducción.
 - La *densidad* es inversamente proporcional a la noción anterior. Una mayor densidad léxica (léxias simples) implica una menor granularidad. Es lo que suele ocurrir en los lenguajes de especialidad que, por su mayor densidad, son más adecuados para la TA, frente a la lengua común, que tiene menos densidad.
- c) *Distancia lingüística y cultural* entre las lenguas objeto de traducción, dimensión sólo tenida en cuenta recientemente y que explica el fracaso de proyectos que no la consideraron, como EUROTRA, que tenía como objetivo traducir entre las lenguas de la Unión Europea.
- *Proximidad lingüística*. Cuanto más próximas son las lenguas, menores son los problemas relacionados con el orden de palabras o la ambigüedad estructural y, por lo tanto, más fácil y rápida resultará la traducción. Es lo que sucede en los sistemas de TA del español al catalán y al gallego, ya que todas son lenguas de la misma familia lingüística. Cuando la distancia aumenta, como ocurre entre el español y el

euskera, la traducción es más costosa, debido a las diferencias estructurales entre las lenguas.

- *Proximidad cultural.* Dos lenguas próximas estructuralmente, pueden diferir considerablemente en aspectos culturales que tienen su reflejo en la lengua, lo que dificultará el proceso de la traducción. En este sentido, es paradigmático el caso del japonés, lengua que ofrece obstáculos de este tipo para las traducciones, sobre todo a lenguas europeas.
- d) *Desarrollo: estandarización, normalización y recursos lingüísticos.* El diseño de un programa de TA exige la existencia previa de recursos lingüísticos para las lenguas implicadas, tales como diccionarios, gramáticas, analizadores, bases de conocimiento y corpus. Estos recursos implican una estandarización de la escritura y una normalización, así como la existencia de grupos de trabajo en ingeniería lingüística encargados de desarrollar los recursos enumerados anteriormente. Por lo tanto, la TA será más fácil cuando las lenguas objeto de la traducción sean equiparables en este aspecto.
- e) *Medio y modo.* Esta dimensión considera las diferencias que existen entre traducción de:
- *Medio oral:* trata de traducir intervenciones producidas en la modalidad hablada de la lengua. Se denomina *interpretación*. Existe un interés comercial grande por desarrollar programas de traducción del habla (C-STAR, JANUS, VERMOBIL, EUTRANS, etc.).
 - *Medio escrito:* se ocupa de traducir textos pertenecientes a la modalidad escrita de la lengua. Recibe el nombre de *traducción* propiamente dicha. Los esfuerzos de la TA en este medio se centran en mejorar los resultados.
 - *Medio electrónico:* en este caso, los textos objeto de traducción están publicados en formato electrónico. Se habla de *localización* para referirse a este tipo de traducciones. Muchas veces, los textos se conciben de manera controlada, teniendo en cuenta que pueden ser objeto de una TA. Al ser multimedia, pueden integrar tanto texto escrito como habla. En la actualidad la localización centra gran parte de las preocupaciones comerciales.

<i>Medio oral (Interpretación)</i>	<i>Medio escrito (Traducción)</i>
No tiempo para pensar	Tiempo para pensar
No preparación del texto	Preparación del texto
No consulta a diccionarios	Consulta a diccionarios y otras fuentes
No corrección o revisión	Corrección y revisión
Simultánea	Diferida
Interacción social	Aislamiento
Condicionado por el contexto, protocolos y convenciones	Condicionada por la fecha de entrega
▪ Características específicas de la modalidad oral de la lengua (espontaneidad, contextualización, discontinuidad, posibles agramaticalidades...)	Características específicas de la modalidad escrita de la lengua (planificación, no contextualización, continuidad, menor presencia de errores gramaticales...)

f) *Fines*. Los resultados de la TA puede que no estén a la altura de las expectativas, pues son muchos los factores que intervienen. Por eso, es importante tener en cuenta el uso (recuperar información, informar, publicar, etc.) que se quiere dar al programa de TA. Se distinguen los sistemas de TA de:

- *Propósito general*: traducir cualquier tipo de texto, habitual en épocas anteriores, pero las traducciones obtenidas no eran aceptables.
- *Propósito específico*: traducir textos de ámbitos muy concretos, lo que redundaba en mejores resultados, al poder elaborarse módulos sintácticos y léxicos *ad hoc*. Con frecuencia, sistemas de propósito general, se han adaptado a una determinada aplicación. Es el caso del sistema SYSTRAN en la Unión Europea, empleado para realizar traducciones urgentes, ojear los contenidos de documentos, obtener versiones preliminares, etc.

2. Breve historia de la TA: orígenes y desarrollo

Mucho antes de que existieran los ordenadores ya se especulaba seriamente con la idea de traducir por medios mecánicos. Existen trabajos anteriores a la aparición del ordenador desarrollados en la antigua URSS que dieron lugar a patentes, pero que en Occidente no se conocieron hasta finales de los años 50. Por lo tanto, había un gran interés por mecanizar la traducción. Con la invención del ordenador en la década de 1940, la TA pasó a convertirse inmediatamente en una de las aplicaciones más importantes de la recién nacida Lingüística Computacional. Siguiendo el esquema de J. Slocum⁴, que aborda la historia de la TA⁵ por décadas, y añadiendo dos más, podemos considerar estas etapas:

1ª) Años 40

A raíz del desarrollo de ENIAC en 1946, investigadores de diversa procedencia pasaron a utilizar esta herramienta como banco de pruebas para sus teorías. Entre los investigadores pioneros que llevaron a cabo los primeros experimentos en TA hay que citar a Warren Weaver, quien en un memorándum publicado en 1949 dio a conocer públicamente la disciplina proponiendo posibles métodos científicos para abordarla:

- El uso de técnicas criptográficas como las empleadas durante el período de guerra para descifrar mensajes en clave.
- La aplicación de la teoría de la información de C. Shannon.
- La utilidad de la estadística.
- La posibilidad de aprovechar los aspectos universales del lenguaje.

El problema era que tanto los ordenadores como las técnicas de programación de aquellos años eran muy rudimentarias.

⁴ Slocum, J. (1988): *Machine Translation Systems*, Cambridge: Cambridge University Press.

⁵ Resulta interesante, para conocer la historia de la TA, Hutchins, W. J. (2001): "Machine translation over fifty years". *Histoire, Epistemologie, Langage*, XXII, 1, 7-31. URL: <http://ourworld.compuserve.com/homepages/WJHutchins/HEL.pdf>. Este investigador inglés ha dedicado un gran número de trabajos a esta cuestión. Se pueden consultar en su página web: <http://ourworld.compuserve.com/homepages/WJHutchins/>

2ª) Años 50

Con la llegada de la nueva década, nos encontramos con el primer investigador contratado para trabajar exclusivamente en TA: Y. Bar-Hillel en el MIT (Instituto de Tecnología de Massachusetts). En 1954 se efectuó la primera demostración pública de un sistema de TA en la Universidad de Georgetown, apoyado por IBM, que consistió en la traducción del ruso al inglés de 49 oraciones previamente seleccionadas. El éxito fue tal que en E.UU. se invirtieron grandes cantidades, sobre todo a cargo del Ministerio de Defensa para traducir del ruso, francés y alemán⁶. La ingenuidad del acercamiento provocó una euforia inicial que llevó a los investigadores a pensar en el objetivo de lograr una *traducción totalmente automática de alta calidad* (FAHQT, *fully automatic high-quality translation*). Entre los sistemas pioneros de aquella década hay que destacar los de las universidades de Georgetown y Texas, donde se sentaron las bases para el posterior desarrollo de SYSTRAN y METAL respectivamente, dos sistemas de TA que todavía están en uso en la actualidad, aunque mejorados.

Esta primera generación de sistemas de TA se caracteriza por:

- La utilización del denominado método directo: la traducción se lleva a cabo básicamente mediante el uso de diccionarios bilingües, análisis morfológico, resolución de homógrafos y un análisis sintáctico mínimo.
- Pares de lenguas más trabajados: inglés-ruso.
- Tipo de textos: documentos científicos y técnicos.
- Necesidad de postedición y/o preedición.

3ª) Años 60

Sin embargo, ante la ausencia de resultados, pronto surgieron voces críticas, como la del propio Bar-Hillel. Los sistemas se topaban con la llamada "barrera semántica": si no se incorporaban conocimientos semánticos y pragmáticos sería imposible lograr traducciones de calidad. Estos problemas motivaron que en 1964 el *National Research Council* de EE.UU. constituyera un comité, ALPAC (*Automatic Language Processing Advisory Committee*), para evaluar la situación de la TA y sus expectativas de cara al futuro. Las conclusiones que se publicaron dos años más tarde (1966) tuvieron efectos considerables:

- No eficiencia: la TA era más lenta, menos precisa y dos veces más cara que la traducción humana.
- No resultados: "...no se ha obtenido traducción automática para textos científicos genéricos, y tampoco parece que se vaya a obtener a corto plazo".
- No utilidad: "no existe expectativa inmediata o previsible de que la traducción automática resulte útil".
- Asimismo, desaconsejaba realizar nuevas inversiones en la investigación sobre TA.
- En su lugar, se recomendaba que se desarrollasen herramientas informáticas para los traductores, como diccionarios automáticos, y que se mantuviera el apoyo a la investigación fundamental en lingüística computacional.

⁶ No hay que olvidar el contexto de guerra fría en que se inscriben estas primeras investigaciones.

El resultado de estas conclusiones fue un drástico recorte financiero que casi terminó con la investigación en los EE.UU. durante más de una década. Pese a ello, se produjeron avances teóricos tanto en Lingüística, en especial los del lingüista N. Chomsky y su gramática generativa, como en el campo de la Informática.

Los trabajos descritos hasta el momento constituyen la denominada "primera generación" de TA:

- La TA se concebía de una forma muy simplista, como un proceso que consistía en introducir texto en una lengua y generarlo en otra, de tal forma que el significado permaneciera inalterado (método directo). No obstante, en algunos proyectos ya se ensayaron versiones iniciales de los métodos de transferencia (en el CLRU, *Cambridge Language Research Unit*) y de interlingua (en el MIT).
- Por otra parte, hay que señalar que gran parte de la investigación llevada a cabo durante este período iba a ejercer una influencia notable, no sólo en el campo de la TA sino también en la lingüística computacional y la inteligencia artificial, especialmente en lo relativo al desarrollo de diccionarios automatizados y de técnicas de análisis sintáctico.
- Sin embargo, no se logró alcanzar el objetivo básico de desarrollar sistemas capaces de realizar traducciones de calidad. En un ambiente de optimismo desbordado, se habían hecho numerosas predicciones sobre inminentes y espectaculares avances pero, en la medida en que la complejidad de los problemas lingüísticos se volvía cada vez más evidente (la llamada "barrera semántica"), un sentimiento de desilusión fue ganando terreno (Bar-Hillel). Esta fue la situación hasta 1966, fecha del Informe ALPAC, a partir del cual se redujo drásticamente el presupuesto dedicado a TA. A partir de entonces, el objetivo de los sistemas será más realista: se reducirá a tratar de ayudar al traductor en sus tareas.

4ª) Años 70-80

El informe ALPAC afectó dramáticamente a la TA en EE.UU. Los americanos se habían centrado en traducir al inglés materiales técnicos y científicos escritos en ruso. Sin embargo, en Canadá, Europa y Japón, lugares a los que en la década de los 70 se desplazará el interés por la TA, las necesidades eran muy distintas:

- La política bicultural canadiense creó una demanda de traducción inglés-francés (y en menor medida francés-inglés) superior a la capacidad del mercado. En Montreal se constituyó un grupo de investigación (TAUM, *Traduction Automatique de l'Université de Montréal*) que no obtuvo un éxito pleno en la creación de un ambicioso programa de traducción inglés-francés para manuales de aeronáutica, pero que es en la actualidad célebre por la creación, en 1976, del sistema arquetípico MÉTÉO, que traduce textos limitados a un sublenguaje, el de los partes meteorológicos de emisión diaria. Es un sistema que ha hecho historia, por la idoneidad de la aplicación y el diseño.
- Por su parte, la entonces CEE planteaba ya en aquel momento una gran demanda de traducción de documentación legal, administrativa, técnica y científica que comprendía todas las lenguas de la Comunidad. En 1976 la CE decidió recurrir a la TA para hacer frente a la desbordante demanda de traducciones internas en sus diversas sedes administrativas. La Comisión compró a las fuerzas aéreas de EE.UU. las licencias para desarrollar SYSTRAN (usado desde 1970 para la traducción ruso-inglés) y adaptarlo a sus necesidades. Poco después, con la idea de impulsar la investigación en Europa y

elevar la calidad de las traducciones, la propia Comisión financia el ambicioso proyecto multilingüe EUROTRA, basado en los más recientes avances en TA y LC, que pretendía abordar todas las lenguas de la Comunidad.

Los años 80 se caracterizan por un fuerte desarrollo de los métodos simbólicos y una gran vitalidad de la investigación en sintaxis (gramáticas basadas en la unificación de rasgos) y en semántica (formalismos basados en la lógica de predicados). Sin embargo, los descubrimientos en el plano teórico no acababan de trasladarse al terreno de los resultados. Con los avances en teorías lingüísticas y en lenguajes de programación, vuelven a resurgir con fuerza nuevos proyectos a gran escala (Logos, PAHO...). Además, se prueban nuevas estrategias, como la transferencia (sistemas de los 70) y la interlingua (sistemas de los 80), ya utilizada en los años 60. Son los "Sistemas de segunda generación", entre los que destacan:

- Grupo de investigadores franceses en Grenoble: sistema Ariane, basado en la transferencia y centrado en el par ruso-francés.
- Grupo de Saarbrücken, Alemania: programa SUSY, multilingüe, también basado en la transferencia.
- LRC (Linguistic Research Center) en Austin, Texas: METAL (multilingüe, transferencia).
- Universidad de Kyoto, Japón: sistema Mu, basado también en la transferencia, para el par japonés-inglés.
- Universidad de Carnegie Mellon, Pittsburgh, EE.UU.: sistemas de interlingua, traducción a través de representaciones intermedias basadas en elementos universales de índole extralingüística (conocimiento del mundo), de acuerdo con las premisas de la IA.
- Utrecht: sistema DLT, que adopta el esperanto como interlingua.
- Philips en Eindhoven: sistema Rosetta, que utiliza como interlingua la semántica de Montague.

Esta segunda generación de sistemas de TA se caracteriza por:

- Aplicación de la TA a sublenguajes, dominios específicos del lenguaje.
- Ampliación del número de pares de lenguas objeto de estudio.
- Métodos indirectos: transferencia e interlingua.
- Mayor papel de las investigaciones teóricas tanto en Lingüística (sintaxis, semántica) como en Inteligencia Artificial (organización del conocimiento del mundo).
- Además de los grandes sistemas para corporaciones o instituciones, llegan los primeros productos comerciales para usuarios particulares.

5ª) Años 90

Pese a los extraordinarios esfuerzos e inversiones realizados durante las pasadas décadas, los resultados no acababan de materializarse. De hecho, a principios de los 90 (1992-1995) cabría hablar incluso de retroceso, por lo menos en lo que respecta a la trayectoria de los métodos de investigación.

- Existía la sensación de que los métodos preponderantes desde los 70 y 80 habían llegado al límite y que se volvía a la metodología de hace 30 años, más simple conceptualmente. Este sentimiento de decepción ya había sido vaticinado en 1960 por Y. Bar-Hillel.
- A raíz de las conclusiones del *informe Danzin*, encargado en 1991 por la CE, en Europa la CE decidió cancelar definitivamente la financiación de EUROTRA; y, en el mundo empresarial, PHILIPS inesperadamente da por terminado uno de los proyectos de más prestigio entre los especialistas, ROSETTA.
- Paralelamente en Japón, se aplica una política de moderación presupuestaria.

Por lo tanto, a mediados de los 90, cambio de estrategias y actitudes: Si la década de los sesenta, por influencia de la gramática generativa, y la de los setenta por la de la inteligencia artificial, marcaron un giro en el enfoque de los estudios hacia la consideración de la sintaxis y de la semántica, en los noventa presenciamos un retorno a postulados más simples, más rudimentarios, pero que son también más realistas y prácticos (memorias de traducción, traducción por ejemplos, etc.).

Características de la TA en los 90:

- IBM publica los resultados de su sistema *Candide*, en el que había empleado métodos estadísticos.
- En Japón se ensaya con métodos de TA basados en corpus de ejemplos de traducciones, lo que hoy en día se conoce como *traducción basada en ejemplos*, en la que no se establecen reglas sintácticas o semánticas, sino que se emplean textos ya traducidos recopilados en un corpus.
- También se inician las investigaciones en la traducción oral, que integran las tecnologías del habla (reconocimiento y la síntesis) con la traducción por reglas y basada en corpus. Destacan los proyectos ATR en Japón, JANUS (ATR, Univ. de Carnegie-Mellon en EE.UU. y Univ. de Karlsruhe) y Verbmobil en Alemania.
- Otro aspecto destacable de esta etapa es el desarrollo de Internet, así como el cambio de enfoque de la traducción hacia la *localización*.
- Por último, se pasa de la investigación pura a aplicaciones prácticas: estaciones de trabajo para traductores profesionales, lenguajes controlados, sistemas de dominio restringido, integración de componentes de TA en sistemas de información multilingües, etc. Destacan las memorias de traducción como Translation Manager de IBM, Translator's Workbench de la empresa alemana TRADOS, TRANSIT, de la también alemana STAR, y DÉJÀ-VU, de la española ATRIL.

6ª) Actualidad

Desde finales de los noventa se observa:

- Una continuación de las tendencias anteriores.
- Un aumento del uso de la TA y de las ayudas a la traducción en grandes corporaciones.
- La localización de *software*: necesidad de adaptar productos y servicios.
- Un aumento de la venta de productos para ordenadores personales.
- Disponibilidad de servicios de traducción en línea.
- Nuevos sistemas que conviven con versiones mejoradas o reducidas de sistemas previos.
- Un crecimiento todavía mayor de la TA para aplicaciones directas en Internet en las que importa la rapidez y no la calidad, como los servicios para traducir páginas web.

3. Métodos

Las estrategias para llevar a cabo la TA se engloban en dos categorías generales:

- a) *Traducción basada en reglas (Rule-based Machine Translation, RBMT)*. Se base teórica se la aportan la Lingüística Generativa y la Inteligencia Artificial, sobre todo a partir de los 70. Se considera la traducción como una tarea en la que el objetivo es lograr equivalencias semánticas. Si se logra determinar la estructura lógica que comparten todas las lenguas y formalizarla, la TA podría llevarse a cabo sin problemas. Es lo que persiguen los métodos basados en la transferencia o en la interlingua que, sin embargo, no han producido sistemas útiles para los traductores.
- b) *Traducción basada en analogías (Analogy-based Machine Translation, ABMT)*. De aparición más reciente, en la década de los 90, utilizan técnicas estadísticas para determinar la mejor traducción posible tomando como referencia muestras de textos previamente traducidos. Surgen como consecuencia de las críticas vertidas al enfoque anterior: no universalidad conceptual entre las lenguas, importancia de otros niveles de equivalencia diferentes del semántico (equivalencia estilística, equivalencia pragmática e, incluso, equivalencia semiótica).

3.1. La traducción basada en reglas

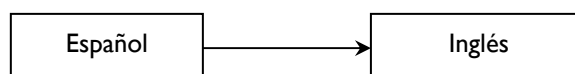
Destacan dos métodos, llamados indirectos, porque entre el texto de la lengua de origen y su traducción a la lengua de llegada, emplean algún tipo de representación intermedia que trata de dar cuenta de esa pretendida universalidad estructural y conceptual de las lenguas. Así pues, se puede establecer una distinción preliminar entre:

- *Métodos directos*
- *Métodos indirectos*
 - *Interlingua*
 - *Transferencia*

a) Método directo

Es la estrategia más antigua y la más simple. Fue la primera desde un punto de vista histórico, adoptada por la mayor parte de lo que se ha denominado primera generación de programas de TA. Se realiza un análisis mínimo del texto de entrada, sin construir ningún tipo de representación sintáctica. La traducción se puede decir que se efectúa palabra a palabra, igual que si se utilizara un diccionario bilingüe como única herramienta para llevar a cabo la tarea.

El enfoque directo representa una estrategia que carece de cualquier tipo de fase intermedia: el procesamiento de los textos de la lengua de origen conduce "directamente" a los textos de la lengua meta deseada. De forma gráfica:



En determinadas circunstancias, este enfoque continúa siendo válido hoy en algunos programas bilingües unidireccionales; incluso en programas de enfoque indirecto (como Météo) pueden encontrarse características propias del enfoque directo. No obstante, el grado de ingenuidad lingüística es menor que en el pasado; sin embargo, se mantiene la idea general de 'volcar' (reestructurando) el texto de origen sobre el texto traducido en la lengua meta: los programas de este tipo explotan las similitudes tanto de la estructura como del vocabulario entre las lenguas de origen y meta con el fin de traducir cuanto sea posible conforme al enfoque "directo"; de este modo, los diseñadores pueden centrar sus esfuerzos en aquellas áreas de la gramática y la sintaxis donde las lenguas difieren en mayor medida.

A la hora de evaluar el funcionamiento de los programas de TA de primera generación, hay que tener presente que los ordenadores disponibles a finales de los años 50 eran muy rudimentarios, incluso si los comparamos con las más modestas calculadoras electrónicas de la actualidad. Además, no existían lenguajes de programación de alto nivel.

A grandes rasgos, la primera generación de programas de TA directos abordaba lo que podríamos denominar fase de análisis morfológico, en la que se llevaba a cabo la identificación de las terminaciones de las palabras y la reducción de las formas flexivas a sus formas básicas sin flexión, que se consultaban posteriormente en un diccionario bilingüe de grandes dimensiones. Por lo tanto, no se efectuaba ningún tipo de análisis de la estructura sintáctica o de las relaciones semánticas. En otras palabras, la identificación léxica dependía exclusivamente del análisis morfológico y llevaba directamente a una búsqueda en el diccionario bilingüe de los términos equivalentes en la lengua meta. A continuación, se aplicaban algunas reglas de reordenación local con el fin de que lo producido en la lengua meta resultase más aceptable, ya fuera mediante el desplazamiento de algunos adjetivos, o de algunas formas verbales.

El enfoque directo se resume en la figura 4.1. (Hutchins y Somers):

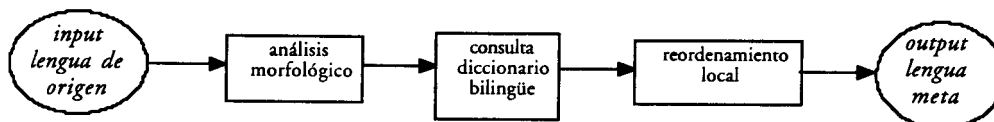


Figura 4.1. Sistema de TA basado en el método directo

Críticas

Las severas limitaciones impuestas por este enfoque resultan obvias:

- El producto final que se obtiene puede caracterizarse como una traducción "palabra por palabra", a la que se aplican algunos ajustes de carácter local que afectan al orden de palabras.
- La calidad de las traducciones obtenidas mediante estos programas era la que cabría esperar de una persona que contase únicamente con un diccionario bilingüe muy barato y con un conocimiento muy rudimentario de la gramática de la lengua meta: frecuentes errores de índole léxica en la traducción y estructuras sintácticas inapropiadas que reflejaban, en gran medida, las estructuras propias de la lengua de origen.

Ejemplos de las traducciones llevadas a cabo por un sistema ruso-inglés de esta clase (la traducción correcta al inglés se ofrece en segundo lugar):

(1) *My trebuem mira.*

[trad. palabra por palabra] *We require world.*

lit. 'Nosotros requerimos al mundo.'

[trad. correcta al inglés] *We want peace.*

'(Nosotros) queremos la paz.'

(2) *Nam nužno mnogo uglja, železa, elektroenergii.*

[trad. palabra por palabra] *To us much coal is necessary, gland, electric power.*

lit. 'Para nosotros mucho carbón es necesario, glándula y energía eléctrica.'

[trad. correcta al inglés] *We need a lot of coal, iron and electricity.*

'Necesitamos mucho carbón, hierro y electricidad.'

(3) *On dopisal stranitsu i otlžil rucku v storonu.*

[trad. palabra por palabra] *It wrote a page and put off a knob to the side.*

lit. '(Ello) escribió una página y apartó un pomo de la puerta a un lado.'

[trad. correcta al inglés] *He finished writing the page and laid his pen aside.*

'Acabó de escribir la página y dejó la pluma a un lado.'

(4) *Včera my tselyi čas katalis' na lodke.*

[trad. palabra por palabra] *Yesterday we the entire hour rolled themselves on a boat.*

lit. 'Ayer, nosotros la totalidad de la hora se balancearon en un bote.'

[trad. correcta al inglés] *Yesterday we went out boating for a whole hour.*

'Ayer salimos a dar una vuelta en bote durante toda una hora.'

(5) *Ona navarila ščeie na neskol' ko dnei.*

[trad. palabra por palabra] *It welded on cabbage soups on severas days.*

lit. 'Se fundió en sopas de berza durante varios días.'

[trad. correcta al inglés] *She cooked enough cabbage soup for several days.*

'Hizo sopa de berza para varios días.'

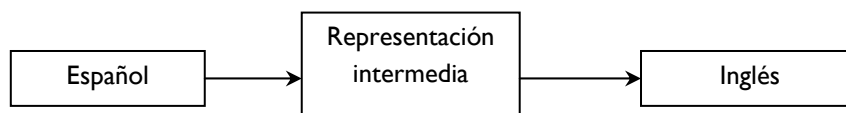
(i) *Some developing countries have less than 4 US dollars to spend per person on health care over an entire year - less than the small change many people in developed countries keep in their pockets or purses.*

Algunos países de desarrollo tienen menos de 4 dólares americanos para gastar por la persona en cuidado de salud durante un año entero - menos del menudo mchuas personas en países desarrollados contienen sus bolsillos o colectas.

La ingenuidad lingüística y computacional de este enfoque quedó patente de inmediato. Desde el punto de vista lingüístico, lo que se echaba de menos era el análisis de la estructura interna del texto de origen, en particular, de las relaciones gramaticales entre las principales partes de las oraciones. La falta de sofisticación computacional fue, en gran medida, un reflejo del estado primitivo en el que se encontraba la informática por aquel entonces, si bien fue determinante, asimismo, la total carencia de base lingüística que caracterizó a los proyectos de TA de finales de los años cincuenta.

b) Métodos indirectos

El fracaso de los sistemas de primera generación (método directo) condujo al desarrollo de modelos lingüísticos de traducción más sofisticados. En particular, fue ganando terreno la idea de proyectar el análisis de los textos de origen en un tipo de representación intermedia -que de algún modo representa el "significado"- a partir de la cual se genera el texto meta. Es lo que se conoce como *métodos indirectos*, en ocasiones también conocidos como programas de segunda generación, en los que podemos distinguir dos variantes principales, a saber: el método basado en la interlingua y el método basado en la transferencia. De forma gráfica:



Técnica de interlingua

El método **interlingüe** o de **interlingua** -que precedió al de transferencia- concibe la traducción en dos fases, pasando por una única representación intermedia de carácter universal, común a las lenguas entre las que se va a traducir:

- a) Fase de análisis: texto de origen-interlingua. El análisis del texto en la lengua fuente da como resultado una representación semántica abstracta o interlingua.
- b) Fase de generación: interlingua-texto de salida. A partir de la interlingua se genera la traducción a la lengua meta.

La fundamentación teórica de este método es la de los universales lingüísticos: los significados o conceptos son los mismos en todas las lenguas, lo que varía es la forma de expresarlos. Se inspira en investigaciones realizadas en Inteligencia Artificial sobre cómo representar y organizar el conocimiento del mundo.

El método es interlingüe en el sentido de que la representación entre dos o más lenguas es neutral, es decir, la misma representación sirve para diferentes lenguas. En el pasado se pretendía o esperaba poder obtener una representación interlingüe que fuera verdaderamente "universal" (independiente de la lengua) y que, de este modo, pudiese actuar como intermediaria entre cualesquiera lenguas naturales. En la actualidad, los objetivos de los programas basados en la interlingua son menos ambiciosos.

Fue muy popular en Japón durante los setenta y ochenta, en proyectos como: ATLAS (Fujitsu), PIVOT (NEC), etc. Y también en otras partes del mundo: en EE.UU. destacan los proyectos KBMT (Carnegie Mellon University), ULTRA (New Mexico State University), PANGLOSS (CMU, NMSU y University of Southern California), UNITRAN (University of Maryland), aunque se trata en todos los casos de proyectos de investigación realizados en centros universitarios, ninguno de los cuales ha llegado a la fase de comercialización. En Europa se pueden reseñar dos proyectos de inspiración interlingüe: DLT, de la empresa holandesa BSO, que utiliza el esperanto como interlingua, y ROSETTA (PHILIPS), basada en la gramática de Montague.

El enfoque basado en la interlingua es, sin ningún género de dudas, el más atractivo en lo que toca a los sistemas multilingües. Cada módulo de análisis puede ser independiente, tanto del resto de los módulos de análisis como de los de generación (figura 4.2. de Hutchins y Somers). Las lenguas meta no tienen efecto alguno sobre los procesos de análisis; el objetivo del análisis es la derivación de una representación "interlingüe".

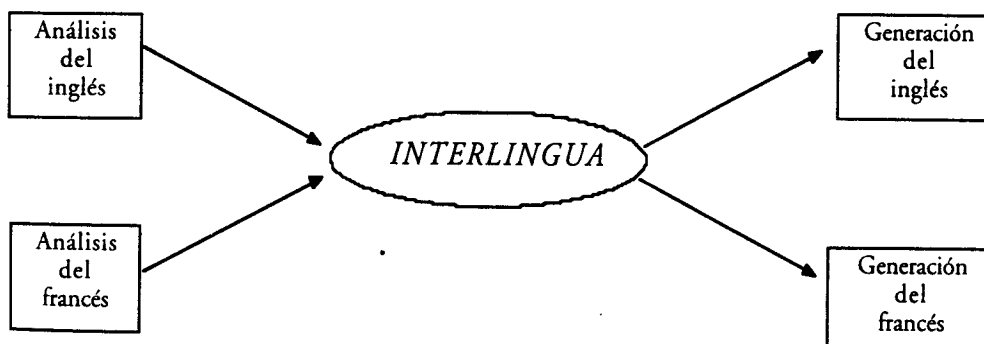


Figura 4.2. Modelo basado en la interlingua con dos pares de lenguas

La ventaja de este enfoque reside en que la incorporación de una nueva lengua al sistema conlleva la creación de tan sólo dos nuevos módulos, a saber: una gramática destinada al análisis y una gramática destinada a la generación. Mediante la incorporación de un módulo de análisis en la figura 4.2., por ejemplo, una gramática destinada al análisis del alemán, el número de sentidos en los que se puede llevar a cabo la traducción se incrementa de dos (del inglés al francés, y del francés al inglés) a cuatro (mediante la incorporación de las traducciones del alemán al francés, y del alemán al inglés). La introducción de otro módulo de generación, una gramática destinada a la generación del alemán, añade dos nuevos pares (del inglés al alemán y del francés al alemán). Compárese la figura 4.3. con la Figura 4.2. (Hutchins y Somers):

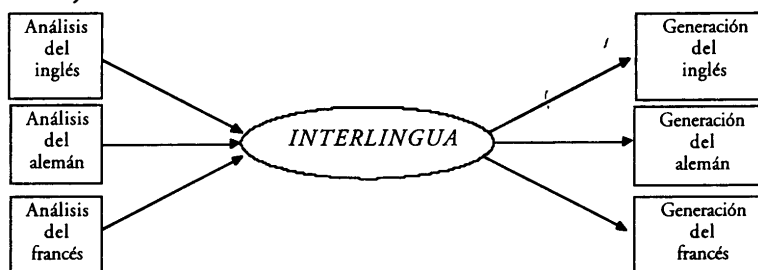


Figura 4.3. Modelo basado en la interlingua con seis pares de lenguas

La adición de dos nuevos módulos destinados al análisis, digamos, del español, incrementaría el número de pares de lenguas en seis (del inglés, francés y alemán al español, y del español al inglés, francés y alemán), y así exponencialmente.

Cabe señalar asimismo que una configuración como ésta permite la 'traducción' desde y a la misma lengua: por ejemplo, la conversión de un texto de origen en inglés a una representación interlingüe y, nuevamente, a un texto meta en inglés. Esta capacidad de 'retraducción', que es aparentemente innecesaria, puede resultar, de hecho, extremadamente valiosa durante el proceso de desarrollo del programa con el fin de verificar los módulos de análisis y generación. Adviértase que el texto meta regenerado no tiene por qué ser idéntico al texto de origen del que parte el análisis, si bien cabe esperar que ambos textos sean pragmáticamente equivalentes.

Aunque la incorporación de nuevas lenguas en un programa interlingüe puede parecer sencilla, existe, sin embargo, un serio obstáculo, el mayor inconveniente que se ha señalado para este método: la propia dificultad de definir la interlingua, incluso para lenguas estrechamente emparentadas (por ejemplo, las lenguas romances: francés, italiano, español o portugués). La creación de una interlingua independiente de las lenguas y verdaderamente "universal" ha supuesto un desafío a los mejores esfuerzos de los lingüistas y filósofos desde el siglo XVII en adelante.

Técnica de transferencia

La segunda variante del enfoque indirecto se conoce como método basado en la transferencia. En sentido estricto, todos los sistemas de traducción incluyen un proceso de "transferencia" de alguna clase, esto es, la conversión de un texto o representación de origen en un texto o representación meta. El aspecto distintivo del método de transferencia es que propone dos representaciones intermedias (módulo bilingüe), una por cada lengua del par de traducción. Estas representaciones, a diferencia de aquéllas propias de los programas interlingües, son dependientes de la lengua que caracterizan. La función de los módulos bilingües es convertir las representaciones intermedias de la lengua de origen en representaciones intermedias de la lengua meta, como se muestra en la figura 4.4. (Hutchins y Somers). Es decir, para traducir entre el inglés y el francés se deberían diseñar 6 módulos:

- Módulo de análisis del inglés
- Módulo de transferencia del inglés al francés
- Módulo de generación del inglés
- Módulo de análisis del francés
- Módulo de transferencia del francés al inglés
- Módulo de generación del francés

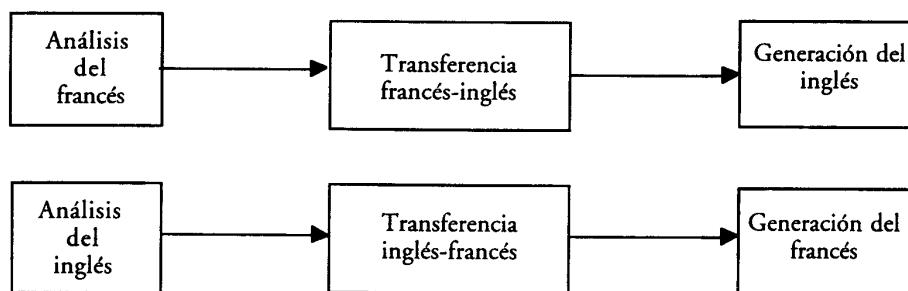


Figura 4.4. Modelo basado en la transferencia con dos pares de lenguas

Sobre estas representaciones se organizan el análisis del texto de origen y la generación del texto meta, de modo que la traducción se realiza en tres fases:

- a) *Análisis*: del texto de la lengua de origen se obtiene una representación léxica, semántica y/o sintáctica.
- b) *Transferencia*: se establecen equivalencias entre las representaciones del par de lenguas objeto de la traducción.
- c) *Generación*: a partir del módulo de transferencia se produce el texto en la lengua meta.

Por lo tanto, en el enfoque de transferencia no se encuentran representaciones independientes de las lenguas: la representación intermedia (o módulo de transferencia) depende de la lengua de origen y de la lengua meta. No es lo mismo este módulo para la traducción, p.e. del español al inglés que para la traducción del inglés al español.

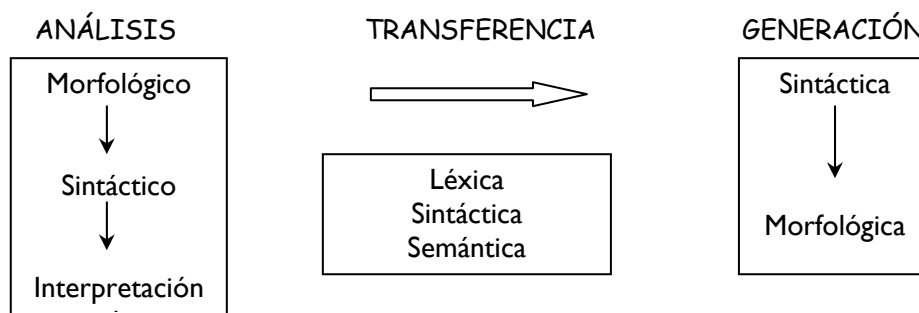
Esta división modular pretende separar el conocimiento perteneciente a cada lengua:

- El análisis depende exclusivamente de la lengua de origen, sin tener en cuenta la lengua de llegada.
- La generación depende únicamente de la lengua de llegada.

Por otra parte, el nivel al que se realiza la transferencia puede variar de un sistema a otro, ya que la transferencia se puede realizar tanto a nivel léxico, como sintáctico, o semántico.

- *Transferencia léxica*: la búsqueda del término equivalente en la lengua meta se realiza a partir de la información contenida en el diccionario.
- *Transferencia sintáctica*: el árbol de análisis de la oración de origen se transforma en un árbol de generación equivalente para la oración meta.
- *Transferencia semántica*: se transforman representaciones profundas, como patrones de casos, redes semánticas, o estructuras lógicas. Este tipo de transferencia se aproxima bastante a la interlingua.

Una posible representación gráfica de este método puede ser esta:



Los sistemas más conocidos diseñados con este método han sido: METAL, MÉTÉO, SUSY, EUROTRA, LOGOS y GETA (Universidad de Grenoble).

Críticas

El método de transferencia presenta claras desventajas si lo comparamos con el de interlingua en lo que respecta a un programa multilingüe. La incorporación de una nueva lengua conlleva no sólo los dos módulos correspondientes al análisis y a la generación, sino también la adición de nuevos módulos de transferencia, cuyo número puede variar de acuerdo con el número de lenguas con que cuente el programa en cuestión: en el caso de un programa diseñado para dos lenguas, una tercera lengua exigiría la inclusión de cuatro nuevos módulos de transferencia. Compárese la figura 4.5. (Hutchins y Somers) con la figura 4.4. y con el gráfico correspondiente al programa de interlingua (figura 4.3.).

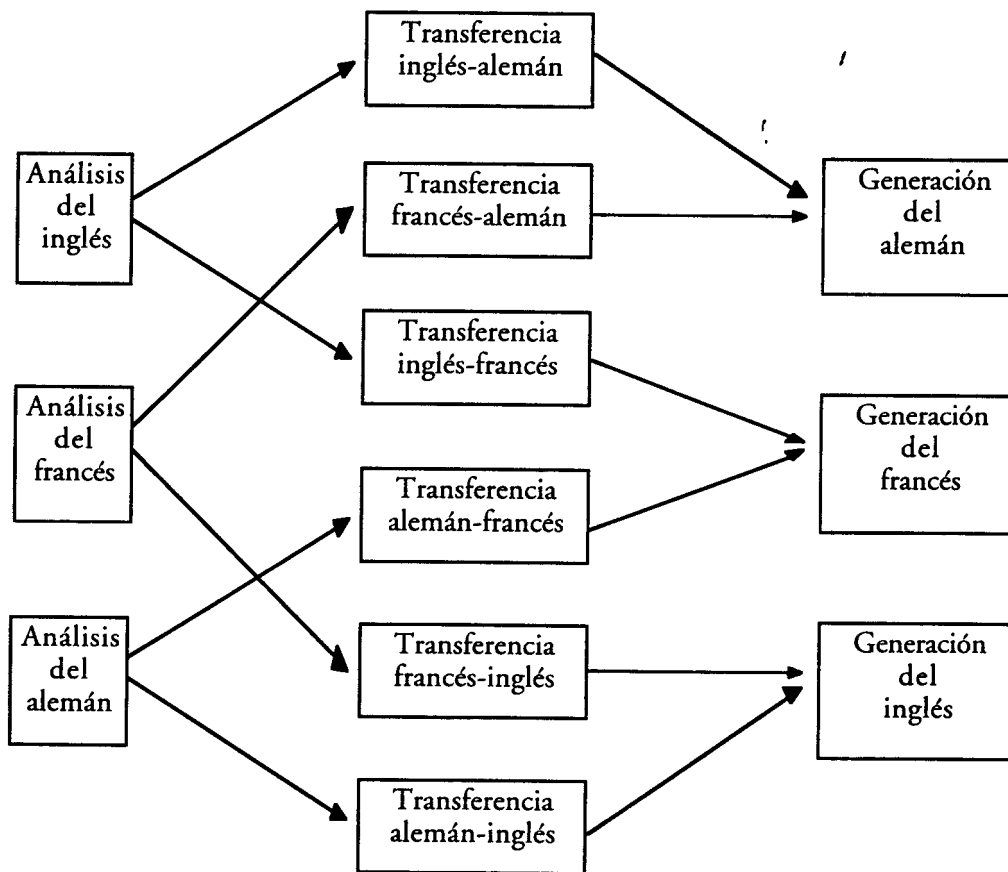


Figura 4.5. Modelo basado en la transferencia con seis pares de lenguas

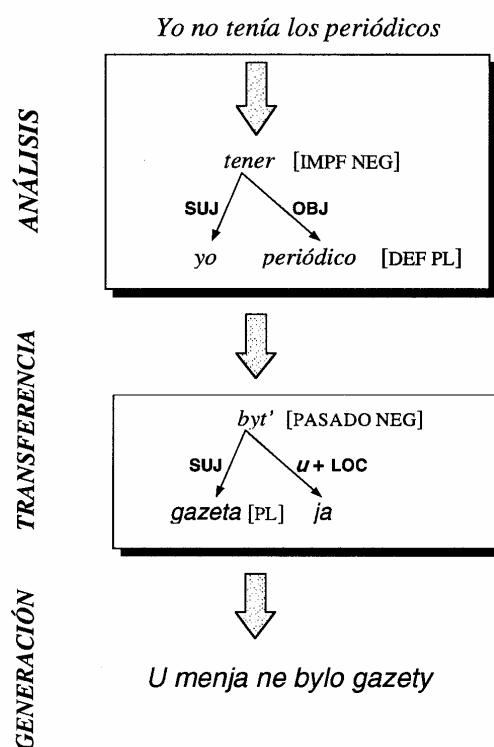
La incorporación de una cuarta lengua supondría el desarrollo de seis nuevos módulos de transferencia, y así sucesivamente, como queda ilustrado en la tabla 4.1. (Hutchins y Somers):

<i>Número de lenguas</i>	<i>Módulos de análisis</i>	<i>Módulos de generación</i>	<i>Módulos de transferencia</i>	<i>Módulos totales</i>
2	2	2	2	6
3	3	3	6	12
4	4	4	12	20
5	5	5	20	30
9	9	9	72	90
n	n	n	$n(n-1)$	$n(n+1)$

Tabla 4.1. Módulos necesarios en un programa de transferencia multilingüe que contemple todos los pares de lenguas

La pregunta, es, entonces ¿por qué suele optarse por el método de transferencia frente al método interlingüe?

- La primera razón ya se ha señalado: la dificultad que entraña el diseño de representaciones independientes de las lenguas como las de la técnica de la interlingua.
- La segunda razón es la complejidad de las gramáticas de análisis y generación cuando las representaciones se distancian, de forma inevitable, de los rasgos característicos de los textos de origen y meta. En términos comparativos, la relativa complejidad de los módulos de análisis y generación en un programa de transferencia se ve reducida en gran medida, debido a que las representaciones intermedias que integran el programa continúan siendo abstracciones dependientes de la lengua. Al mismo tiempo, si el diseño resulta adecuado, la tarea llevada a cabo por los módulos de transferencia puede verse enormemente simplificada y, asimismo, la creación de nuevos módulos puede resultar menos onerosa de lo que cabría imaginar.



3.2. Traducción basada en analogías

Durante la década de los 90, los cambios experimentados en Lingüística Computacional y los avances tecnológicos que permitieron abaratar el precio de los ordenadores, facilitaron la recopilación de grandes corpus textuales en formato electrónico. La manipulación de estas ingentes cantidades de datos también exigía nuevas técnicas, por lo que se empezaron a probar métodos probabilísticos y conexionistas ya empleados con éxito desde hacía tiempo en el ámbito de las tecnologías del habla. El resultado son nuevos acercamientos metodológicos al problema de la TA:

- traducción estadística
- traducción basada en ejemplos

La traducción basada en analogías depende crucialmente de los corpus bilingües, lo que ha hecho que las técnicas de alineación se hayan convertido en una de las áreas de investigación de mayor interés en traducción automática. Alinear es hacer explícitas las relaciones de correspondencia entre segmentos del corpus bilingüe. Se puede hablar de tres enfoques principales:

- *Enfoque estadístico*: método de alineación que aprovecha la similitud de algunos rasgos cuantitativos identificados en el corpus, como la longitud de oraciones, el número de palabras o de caracteres, etc.
- *Enfoque lingüístico*: la alineación se basa en el emparejamiento previo de unidades sintagmáticas o de estructuras dependenciales.
- *Enfoque mixto*: método que aprovecha la identificación de categorías gramaticales como apoyo para la alineación estadística.

Por otra parte, hay que señalar que la traducción basada en analogías ha experimentado su mayor desarrollo en una tecnología conocida como "memoria de traducción" (MBMT, *Memory-based Machine Translation*). La técnica consiste en almacenar traducciones, realizadas manualmente y validadas por un traductor humano, para reutilizarlas posteriormente en la traducción de textos similares. Esta tecnología ha sido llevada al mercado con un considerable éxito en paquetes de *software* que incluyen los módulos de gestión de las memorias, además de programas para crear y mantener bases de datos terminológicas, alineadores automáticos, etc.

Métodos estadísticos

Los modelos anteriores son una respuesta a la cuestión de qué representaciones utilizar y qué pasos seguir para llevar a cabo la traducción. El acercamiento estadístico a la traducción, en vez de centrarse en el proceso, se centra en el resultado.

Desde este acercamiento se parte del hecho de que no es posible lograr una traducción de una lengua a otra en sentido estricto (fiel a la lengua origen y natural en la lengua meta), ya que continuamente surgen conceptos específicos a culturas, metáforas, una construcción, una palabra o un tiempo que no tienen correspondencia exacta en la otra lengua: P.e. el hebreo *adonai roi* ("the Lord is my shepherd", "el Señor es mi pastor") no se puede traducir a la lengua de una cultura que no tenga ovejas; lo mismo sucede con la traducción del japonés *fukaku hansei shite orimasu*, ya que presenta una serie de matices difíciles de reflejar en una lengua europea.

- Puede ocurrir que la traducción en la lengua meta sea clara, pero que no sea fiel al original.
P.e. *the Lord will look after me* para el texto hebreo.
P.e. *we apologize* para el texto japonés.
- O que la traducción sea fiel al original pero resulte oscura en la lengua meta.
P.e. *the Lord is for me like somebody who looks after animals with cotton-like hair* para el texto hebreo.
P.e. *We are deeply reflecting on our past behavior, and what we did wrong, and how to avoid the problem next time* para el texto japonés.

Por lo tanto, toda traducción es producto de un compromiso entre ambos criterios: *fidelidad al original* y *naturalidad en la lengua meta*. El acercamiento estadístico se basa en cuantificar esos dos criterios (fidelidad y naturalidad) para obtener la mejor traducción posible en cada caso.

Por lo general, las estadísticas o probabilidades de traducción se extraen de corpus paralelos y/o alineados: corpus que contienen el texto original y su traducción a una o más lenguas. Si, además, están alineados, las unidades de traducción se disponen unas al lado de otra para una mejor identificación. Después, se calculan las probabilidades de traducción de una unidad de una lengua a otra unidad en otra lengua.

Aunque ya W. Weaver había planteado la posibilidad de abordar la traducción en términos estadístico en 1949, no fue hasta principios de los 90 cuando investigadores procedentes del reconocimiento del habla lo desarrollaron plenamente: sistema CANDIDE, diseñado por el grupo del *Thomas J. Watson Center* de IBM en Nueva York. El sistema no se llegó a comercializar, pero supuso un hito histórico en la reorientación de las investigaciones. Entre las aportaciones más recientes cabe destacar la del grupo ISI de la *University of Southern California*, que desarrolla EGYPT, un paquete de software para construir sistemas basados en la estadística a partir de corpus bilingües.

Traducción basada en ejemplos

Un enfoque similar a la traducción estadística también basado en la disponibilidad de corpus paralelos o alineados es la traducción basada en *ejemplos*. Fue inicialmente propuesto en 1981 por el investigador japonés M. Nagao, pero la técnica no fue probada hasta finales de la década de manera simultánea por el propio Nagao en la Universidad de Kyoto y por el grupo del proyecto DLT en Holanda. La hipótesis es que los textos traducidos pueden servir de modelo a las nuevas traducciones. El método funciona mediante la extracción y selección de oraciones o sintagmas en un corpus de textos bilingües, previamente alineado. La traducción por ejemplos está en la base del desarrollo de las denominadas memorias de traducción.

4. Problemas en la traducción

A la hora de traducir de una lengua a otra, los principales problemas para un sistema de TA suelen venir derivados de las diferencias léxicas y estructurales entre las lenguas, tal y como señalan Hutchins y Somers:

- *Diferencias léxicas*
 - *Ambigüedades de traducción*
 - *Ambigüedades estilísticas*
 - *Ambigüedades gramaticales*
 - *Ambigüedades conceptuales*
- *Diferencias estructurales*

Diferencias léxicas

La mayor parte de las diferencias entre los sistemas léxicos de las distintas lenguas tienen su origen en divergencias conceptuales, en tanto que otras son producto de diferencias estilísticas y gramaticales. Es lo que globalmente se llama *ambigüedades de traducción*. Se producen cuando una palabra de la lengua de origen puede traducirse por varias palabras o expresiones de la lengua meta, no ya porque la palabra sea ambigua en sí misma, sino porque lo es sobre todo desde la perspectiva de la otra lengua.

Las *ambigüedades estilísticas* de la traducción se producen cuando la elección de equivalentes léxicos en la lengua meta depende de diferencias de registro o de tipo de texto. Se suele producir entre lenguas que no poseen acervos culturales semejantes. En japonés, por ejemplo, las palabras que expresan las relaciones de parentesco varían según se refieran al hablante o al oyente: es el caso de *kanai* "mi esposa" y *okusan* "tu esposa".

Las *ambigüedades gramaticales* son mucho menos frecuentes y se producen cuando existe una elección alternativa, condicionada por el contexto gramatical, en la lengua meta. P.e. la traducción del verbo inglés *know* al francés o español: en los dos casos, la elección entre *connaître* y *savoir* o *saber* y *conocer* depende, en términos generales, de si el objeto directo es un sintagma nominal o una oración subordinada de relativo o un infinitivo.

- | | |
|--|--|
| <p>(1a) <i>I know the right answer.</i>
 <i>Je connais la bonne réponse.</i>
 "Yo conozco la respuesta correcta"</p> | <p>(2a) <i>I know what the right answer is.</i>
 <i>Je sais quelle est la bonne réponse.</i>
 "Yo sé cuál es la respuesta correcta".</p> |
| <p>(1b) <i>I know the author of that book.</i>
 <i>Je connais l'auteur de ce livre</i>
 "Conozco al autor de ese libro".</p> | <p>(2b) <i>I know who the author of that book is.</i>
 <i>Je sais qui est l'auteur de ce livre.</i>
 "Yo sé quién es el autor de ese libro".</p> |

En algunos casos, esto es, cuando ambos tipos de estructura son posibles en el texto meta, puede optarse por uno u otro verbo indistintamente:

- (3) *I know the quickest way to get from Norwich to Manchester.*
Je connais la route la plus rapide de Norwich à Manchester.
Je sais comment aller le plus vite possible de Norwich à Manchester.
 "Conozco el camino más rápido para ir de Norwich a Manchester".
 "Sé cómo ir más rápido desde Norwich hasta Manchester".

Las *ambigüedades conceptuales de la traducción* son la causa de los mayores problemas en traducción y constituyen el tema principal de los debates sobre metodologías y diseños de sistemas de TA. Dichas ambigüedades surgen cuando un único "concepto" representado por una palabra en una lengua corresponde a una serie de conceptos y, por ende, a una serie de palabras en otra lengua. Este no es en absoluto un fenómeno aislado, ni siquiera entre lenguas estrechamente relacionadas; al contrario, podemos encontrar multitud de ejemplos de este tipo, algunos de los cuales son:

(4a) Inglés <i>wall</i> (pared, muro)	Alemán	<i>Wand</i> (dentro de un edificio) <i>Mauer</i> (fuera)
(4b) Inglés <i>river</i> (río)	Francés	<i>Rivière</i> (concepto general) <i>Fleuve</i> (río principal que desemboca en el mar)
(4c) Inglés <i>leg</i> (pierna)	Español	<i>Pierna</i> (humana) <i>Pata</i> (animal, mesa) <i>Pie</i> (silla) <i>Etapa</i> (de jornada)
	Francés	<i>Jambe</i> (humano) <i>Patte</i> (animal, insecto) <i>Pied</i> (mesa, silla) <i>Étapa</i> (jornada)
(4d) Inglés <i>blue</i> (azul)	Ruso	<i>Goluboi</i> (azul claro) <i>Sinii</i> (azul oscuro)
(4e) Francés <i>louer</i> (alquilar)	Inglés	<i>Hire</i> o <i>rent</i>
(4f) Francés <i>colombe</i> (paloma)	Inglés	<i>Pidgeon</i> o <i>dove</i>
	Alemán	<i>taube</i>
(4g) Alemán <i>leihen</i> (prestar o pedir prestado)	Inglés	<i>Borrow</i> o <i>lend</i>
(4h) Inglés <i>rice</i> (arroz)	Malayo	<i>Padi</i> (grano no cosechado) <i>Beras</i> (sin cocinar) <i>Nasi</i> (cocinado) <i>Emping</i> (aplastado) <i>Pulut</i> (glutenoso) <i>Bubor</i> (cocinado como gachas)
(4i) Inglés <i>wear</i> (llevar puesto)	Japonés	<i>Kiru</i> (en general) <i>Haoru</i> (abrigo o chaqueta) <i>Haku</i> (zapatos o pantalones) <i>Kaburu</i> (sombbrero) <i>Hameru</i> (anillo o guantes) <i>Shimeru</i> (cinturón, corbata o bufanda) <i>Tsukeru</i> (broche o clip) <i>Kakeru</i> (gafas o collar)

Diferencias estructurales

Muchas de las diferencias sintácticas de relativa poca importancia entre las lenguas son ya bastante conocidas: en español, como en francés, por ejemplo, la mayoría de los adjetivos suceden a los sustantivos, en inglés, por el contrario, los adjetivos normalmente preceden a los sustantivos a los que califican.

Otras diferencias también conocidas son las que se registran entre lenguas como el japonés y el latín, en que el verbo principal (conjugado) de una oración se sitúa en último lugar, y en lenguas como el inglés y el alemán, en las que el verbo conjugado aparece después del primer sustantivo (sintagma nominal). Sin embargo, equiparar las estructuras de la lengua de origen y de la lengua meta no siempre resulta tan sencillo.

Un buen ejemplo es la construcción en voz pasiva. En inglés la construcción pasiva está formada mediante el verbo auxiliar *to be* y el participio del verbo principal. Pues bien, incluso cuando se traduce a una lengua que posee una construcción semejante, no siempre resulta apropiado -ni siquiera posible- emplearla. Así, la traducción de (5a) por una estructura de pasiva es aceptable, mientras que son preferibles otras alternativas para (5b, c).

- (5a) *Le bâtiment fut construit en 1923.*
The building was constructed in 1923.
 "El edificio fue construido en 1923".
- (5b) *Ces livres se lisent facilement.*
 Lit. *These books read themselves easily.*
 Lit. "Estos libros (se) leen ellos mismos fácilmente."
These books are easily read.
 "Estos libros se leen fácilmente".
- (5c) *Ici on parle anglais.*
 Lit. *One speaks English here.*
 Lit. "Uno habla inglés aquí".
English is spoken here.
 "Se habla inglés aquí".

5. Sistemas más relevantes

La bibliografía mencionada a lo largo del tema ofrece una amplia descripción de los principales sistemas de TA.

6. TA e Internet

En los últimos años Internet se ha convertido en una herramienta de trabajo omnipresente y hoy en día imprescindible. Prácticamente todo se puede buscar, encontrar, comprar o vender mediante Internet, y la TA no es una excepción. Algunos de los tipos de sistemas de TA que podemos encontrar en Internet son los siguientes:

- *Traductores integrados en otras aplicaciones.* Así, es posible encontrar sistemas de TA integrados en *buscadores* (p.e. *Babel Fish*, el sistema de TA basado en Systran utilizado por el buscador AltaVista; o la herramienta de TA integrada en Google), medios de comunicación (p.e. el traductor *Reverso* de la firma Softissimo integrado en el periódico *El Mundo* o el servicio de traducción de *El País*) o instituciones, como el servicio de TA del Instituto Cervantes, desarrollado por la empresa AutomaticTrans.
- *Servicios gratuitos.* Otros sistemas son directamente accesibles por Internet y permiten traducir palabras, frases, textos y páginas web (OneLook, interNOSTRUM, TranslateNow!, Universia, OpenTrad apertium español-gallego de la Universidad de Vigo).
- *Traducción en línea.* Otra posibilidad es la de lugares que ofertan servicios de traducción (humana y automática) a través de Internet a un módico precio (en el caso de que requiera que intervenga un traductor humano). P.e. WorldLingo ofrece

traducciones automáticas gratuitas o humanas, realizadas por un traductor profesional según presupuesto.

- *Productos y demos.* También están muy extendidos los lugares en los que se pueden comprar programas de TA para PC, y que ofrecen demos o versiones de prueba a los usuarios. P.e. Babylon, empresa especializada en productos de TA.

Sistemas de TA en Internet:

- 1) Babel Fish, servicio de TA integrado en el buscador Altavista. URL: <http://babelfish.altavista.com/>

- 2) Herramientas del idioma del buscador Google. URL: http://www.google.es/language_tools?hl=es

Traduzca

3) Servicio de traducción de El País. URL: <http://www.elpais.es/traductor/index.html>

ELPAIS.es > Servicios > Traductor

traductor

i El servicio de traducción de **ELPAIS.es** le permite seleccionar entre un amplio abanico de idiomas. Elija el servicio de traducción que desea utilizar:

▼ TRADUCIR UNA PALABRA O FRASE > TRADUCIR UNA DIRECCIÓN (URL)

1. Escriba la palabra o frase que quiera traducir (150 palabras como máximo)
2. Elija los idiomas de traducción

de inglés a español ▼ **TRADUCIR**

Esta herramienta tiene como fin ayudar a la mejor comprensión de textos en otros idiomas, aunque sus traducciones pueden no ser perfectas.

4) Traductor del periódico El Mundo. URL: <http://www.elmundo.es/traductor/>

Mi Traductor

Texto Pag.web Coment. Ayuda

1 Escriba su texto Entrez votre texte
Fügen Sie den Text ein Enter your text

2 Aquí está la traducción Ici la traduction
Die Übersetzung ist da The translation is here **3**

SP-EN
 EN-SP
 SP-FR
 FR-SP
 EN-FR
 FR-EN
 EN-DE
 DE-EN

Caracteres especiales Caractères spéciaux
Sonderzeichen Special characters

à é î ô ù Otros

Borrar - Clear - Löschen - Effacer

E-MAIL

[Condiciones de utilización](#)

5) Servicio de TA del Instituto Cervantes. URL: <http://oesi.cervantes.es/traduccionAutomatica.html>



Para realizar una traducción gratuita* de una página web o de un texto, siga los siguientes pasos:

Traducir de

Opción WEB:

Seleccione una página de la lista:

Traducir

O escribala en el siguiente cuadro:

Traducir

Opción TEXTO:

Escriba el texto en el siguiente cuadro:

Traducir

Borrar datos

6) interNOSTRUM, servicio de TA castellano-catalán desarrollado por el Dpto. de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. URL: <http://internostrum.com/indice.phpWorldLingo>



Traducción de texto - Versión estándar

➔ Pruebe la versión actual de interNOSTRUM

Introduzca o pegue un texto ANSI (hasta 16384 caracteres) en el cuadro de texto, seleccione el sentido de la traducción, el tipo de salida que quiera generar y pulse "Traducir".

Sentido

Novetat!


Da prioridad a las formas valencianas

Tipo de salida

6) OneLook. URL: <http://www.onelook.com/index.html>



7) TranslateNow! de Foreignword. URL: <http://www.foreignword.com/es/Tools/transnow.htm>



8) Universia. URL: <http://copacabana.dlsi.ua.es/es/index.php>



9) OpenTrad apertium español-galego. URL: <http://sli.uvigo.es/tradutor/>

tradutor OpenTrad apertium español <-> galego

10)

WorldLingo.

URLs:

http://www.worldlingo.com/es/products_services/computer_translation.html

y

http://www.worldlingo.com/es/products_services/human_translation.html

WorldLingo
Traducción Localización Globalización

Traducción Automática | Traducción Profesional

Traducción electrónica de sitios Web
Para traducir, escriba o pegue el URL (dirección en la Web) a continuación >

Traducción electrónica de texto
Para traducir, escriba o pegue el texto a continuación >

Caracteres especiales

Tema: General

Idioma de origen: Seleccione un idioma...

Idioma de destino: Seleccione un idioma...

Al presionar el botón Traducir acepto los [términos y las condiciones.](#) Traducción Electrónica

Traducción a cargo de un profesional

Envíe su documento ahora y obtendrá un presupuesto instantáneo para una traducción realizada por un profesional. >

Envío de archivo: Examinar...

Tema: General

Idioma de origen: Seleccione un idioma...

Idioma de destino: Seleccione un idioma...

Formato de archivo: Texto (.txt)

Presupuesto

11) Babylon, empresa especializada en productos de TA. URL:
<http://www.babylon.com/spa/index.php>

Babylon-Pro 5.0



Acceda a traducciones e información relevante al instante con un solo clic de ratón. Al hacer clic en cualquier elemento de su aplicación de escritorio, obtendrá resultados inmediatos, fruto de la amplísima colección de diccionarios lingüísticos, glosarios y herramientas de conversión de Babylon. Con las herramientas de ayuda a la redacción de Babylon, puede encontrar fácilmente la palabra que precise, cerciorarse de su uso correcto y pegarla directamente a sus documentos.

↓ Descargar

Ver Demo

🛒 Comprar

7. La situación de la TA en España

Véase el artículo de J. Abaitua: "Quince años de traducción automática en España", *Perspectives: Studies in Translatology*, Vol. 7:2, 1999. URL:
<http://sirio.deusto.es/abaitua/konzeptu/ta/ta15.htm>