

ANA PANO

DESAFÍOS PARA LA TRADUCCIÓN AUTOMÁTICA EN LENGUA ESPAÑOLA: COMBINACIÓN DE MÉTODOS Y ONTOLOGÍAS

UN ANÁLISIS DEL SERVIDOR DE TRADUCCIÓN DEL INSTITUTO CERVANTES

Hoy en día el rápido aumento de la información en formato digital y el desarrollo vertiginoso de las tecnologías de la información y de la comunicación permiten el tratamiento de grandes volúmenes de datos así como la aplicación de programas informáticos para su adaptación y posterior traducción automatizada. La creación y elaboración de sistemas y programas de Traducción automática, en adelante TA, ha sido irregular y su utilización es todavía problemática si no va acompañada de la supervisión del traductor humano o de recursos lingüísticos como diccionarios o córpora paralelos, entre otros.

En este artículo analizamos las *dimensiones* que condicionan la viabilidad de cualquier sistema de TA, como son la densidad léxica de un texto, la distancia lingüística y cultural entre dos textos y la definición de los objetivos de la traducción, para abordar desde un punto de vista teórico y práctico los *métodos* adoptados hasta hoy para el diseño de sistemas de traducción. El estudio del servicio de TA propuesto por el Instituto Cervantes a través de su portal de tecnologías lingüísticas (Oficina de Español en la Sociedad de la Información-OESI¹) y la observación de algunos resultados obtenidos con este instrumento, nos permite ilustrar las cuestiones presentadas y esbozar un cuadro de los desafíos que se presentan en este campo. Desafíos que pasan por la creciente combinación de técnicas y de herramientas; la adaptación de los sistemas a los datos que se deben tratar; la elaboración de ontologías y

Dr ANA PANO – Dipartimento di Lingue e Letterature Straniere, Università di Bologna; adres do korespondencji: Via Cartoleria 5, 40129 Bologna, Italy; e-mail: ana.pano@unibo.it

¹ <http://oesi.cervantes.es/introTraduccionAutomatica.jsp>

formalismos basados no sólo el conocimiento lingüístico sino también en conocimiento del mundo; y finalmente la interacción eficaz entre el ordenador y el traductor humano.

En todo proyecto de automatización de la traducción hay que considerar los distintos tipos de conocimiento a los que recurre un traductor: conocimiento de la lengua origen (LO) y de la lengua meta (LM); conocimiento de las distintas correspondencias entre LO y LM, en el nivel más básico de los equivalentes de traducción de las distintas unidades léxicas; conocimiento del dominio sobre el que se traduce; y conocimiento del entorno cultural y de las convenciones sociales de los hablantes de LO y LM.

A menudo, el enfoque teórico adoptado para proyectar un determinado sistema considera un número limitado de cuestiones sobre la lengua, sobre todo las reglas gramaticales, cuestiones que se concentran en el nivel morfosintáctico en detrimento del nivel léxico-semántico de un texto. En otras ocasiones, los objetivos que se plantean a la hora de proyectar estos sistemas son demasiado ambiciosos, como por ejemplo, considerar un único sistema o método para traducir una amplia gama de textos. En cualquier caso, los errores conceptuales de fondo sobre lo que conlleva la traducción y sobre lo que deberían hacer los sistemas de TA explican porqué se trata de una disciplina que ha generado y suele generar escepticismo.

Existen diversas tipologías textuales que presentan diferencias sobre todo desde el punto de vista del léxico y de la sintaxis. Para abordar estas diferencias se suele hacer referencia a conceptos como género, registro, estilo o tipo (cfr. TROSBORG 1997) que permiten, entre otras cosas, diferenciar los textos por la longitud de sus oraciones, por su densidad léxica o por el uso de conectores así como por factores extralingüísticos, es decir, por la función pragmática que desempeñan (BIBER Y FINEGAN 1986). La tipología textual incide en otras dos dimensiones directamente relacionadas con la extensión de las unidades de traducción, es decir, el número de palabras que constituyen el texto, y su cohesión, o la posibilidad de interpretar cada una de estas unidades de manera conjunta o separada. En otras palabras, la unidad de traducción² y el grado de cohesión están directamente relacionadas con la gra-

² Las unidades de traducción habituales corresponden a las categorías y subcategorías morfosintácticas que permiten establecer abstracciones sobre las unidades léxicas (*ellthe* > Det; *eatlcomer* > V) y generar gramáticas de estructura sintagmática como: SN = Det N. También se tienen en cuenta los “hábitos de colocación” de cada categoría y subcategoría; las palabras compuestas y las locuciones; los giros idiomáticos, las frases hechas o los proverbios; los títulos de obras y otros elementos fijos del discurso.

nularidad y con la densidad léxica de un texto. Un programa de TA que no tiene cuenta estas dimensiones y que trata dos textos con diferentes grados de granularidad y de densidad de la misma manera, ofrece traducciones de baja calidad que requieren necesariamente una intervención a posteriori.

Generalmente, la distancia lingüística y cultural no han sido tenidas en cuenta hasta épocas recientes en la mayor parte de proyectos. Éste ha sido el caso de EUROTRA, basado en la fórmula de las arquitecturas de transferencia, que no prevé integrar un subsistema para reducir la distancia lingüística entre lenguas pertenecientes a las dos grandes familias lingüísticas europeas: la latina, con el italiano, el francés, el español y el portugués y la germánica, con el alemán, el holandés y el inglés, por ejemplo. Sistemas como INTERNOSTRUM permiten obtener resultados eficaces entre el español y el catalán o el gallego o entre el español y el francés porque se ha tenido en cuenta la menor distancia lingüística entre estas lenguas. Sin embargo, construir sistemas de traducción entre dos lenguas implica considerar no sólo los aspectos lingüísticos sino también los aspectos, que en la lengua, vehiculan convenciones culturales, y que, como veremos, conllevan en el paso de una lengua a otra, un cambio de significado en las expresiones más sencillas, lo que supone adoptar soluciones que recurran a una adecuada representación del conocimiento del mundo.

Hasta ahora, muchos sistemas han sido diseñados para servir un propósito general como es traducir cualquier tipo de texto basándose en sistemas combinados de transferencia de una lengua a otra y con la ayuda de diccionarios o córpora paralelos. Estos sistemas han proporcionado generalmente traducciones ineficaces porque el objetivo de la traducción y el ámbito de aplicación de ésta no se han tenido en consideración en el momento de desarrollar las aplicaciones. No obstante, algunos sistemas generalistas como Systran, han sido progresivamente adaptados a ámbitos concretos de aplicación como son las traducciones de textos altamente especializados en campos del saber como la economía o el derecho.

Consideradas estas dimensiones, cabe ahora preguntarse cuál es su incidencia en los enfoques teóricos en los que se basa todo sistema de TA. Por un lado, los que ven en la LO una serie de aspectos lingüísticos estáticos y universales que tienen su equivalente en la LM. Y, por otro lado, los que ven en la lengua una serie de rasgos que evolucionan rápidamente y que presentan diferencias considerables entre la LO y la LM, sobre todo si se tienen en cuenta también los aspectos cuya formalización o representación mediante reglas es más compleja, como la semántica o las convenciones culturales.

Los primeros –sistemas por transferencia e interlingua–, se basan en la utilización de conocimiento lingüístico o del mundo y en una serie de reglas que explicitan cómo debe ser utilizado dicho conocimiento. Las traducciones llevadas a cabo con sistemas de este tipo presentan, sin embargo, limitaciones porque no tienen en cuenta o no suficientemente la distancia lingüística y cultural entre los textos y reducen el problema de la granularidad y de la densidad léxica a la equivalencia más o menos aceptable entre expresiones tomadas separadamente. Con un enfoque fundamentalmente empírico, los segundos se basan en ejemplos reales de traducción y en analogías que permiten extraer traducciones a partir de corpus alineados de textos en lenguas afines. Entre los motivos que han favorecido este cambio de perspectiva destacan la creación y puesta a disposición inmediata del usuario de herramientas informáticas y de recursos lingüísticos sofisticados, como lexicones computacionales o diccionarios electrónicos, de fácil acceso y manejo. La evolución de estos recursos ha llevado a los sistemas de “memoria” o *Memory-based Machine Translation*, que consisten en almacenar traducciones, validadas por un traductor humano, para reutilizarlas posteriormente en la traducción de textos similares.

Actualmente asistimos a un proceso de hibridación de métodos. Puesto que la traducción de lenguas naturales requiere una vasta cantidad de información lingüística y extra-lingüística se hace necesaria la extracción de datos significativos que den cuenta de las regularidades sintácticas y semánticas, la información colocacional, la polisemia, el doble sentido, las convenciones culturales, mediante la creación y utilización de extensos corpórea textuales monolingües y bilingües. Puesto que para conseguir una traducción de calidad es precisa una completa interpretación semántica del texto origen, se debe recurrir, por otro lado, a la representación de conocimiento lingüístico y del mundo mediante reglas que expliciten cómo se organiza dicho conocimiento. En este sentido, uno de los paradigmas más recientes y que perfecciona los sistemas de interlingua es el que se basa en el conocimiento (*Knowledge-Based Machine Translation*). El paradigma se apoya en la creación y utilización en el proceso de traducción de un vasto depósito de información, no sólo lingüística, gramatical, sino también relacionada con el conocimiento de campos específicos del saber. Esta información se obtiene a partir de extensas y bien estructuradas bases de datos y bases de conocimiento u ontologías³, así como de un conjunto de reglas que permiten proyectar la

³ Existen proyectos eficaces de creación de ontologías como Mikrokosmos con aplicaciones a subdominios del conocimiento como la medicina, y dentro de ella, el subdominio del cáncer (PÉREZ HERNÁNDEZ 2002).

semántica de la LO sobre representaciones semánticas de la LM (MORENO ORTIZ 2000).

El servicio de traducción propuesto por el Instituto Cervantes a través de su portal de tecnologías lingüísticas es un sistema integrado en un portal de servicios y de información relacionados directamente con la lengua y con el conocimiento lingüístico, lo que garantiza, de alguna manera, que se trata de un sistema destinado prevalentemente al traductor o a los usuarios que se ocupan de cuestiones relacionadas con la lengua española, ya sean profesores, investigadores o estudiantes. El sistema ha sido concebido por la empresa AutomaticTrans para las traducciones entre español-catalán y español-gallego. Para las traducciones entre español y francés e inglés, recurre a Systran, usado por los traductores de la UE, generalmente con buenos resultados. Como reza la presentación del servicio, este programa “facilita el desarrollo de procesos automatizados de traducción”. A través de lo que parece un clásico formulario en línea, el servicio permite traducir textos y archivos de texto, así como páginas web, del español al catalán, francés, gallego, inglés o portugués y viceversa. El sistema traduce un texto escribiendo o copiando el fragmento que se quiere traducir en el espacio “Opción TEXTO”, seleccionando el par de lenguas (LO y LM) de la traducción y pulsando el botón “Enviar”. El resultado es el texto traducido a la lengua seleccionada. La velocidad con que se obtiene el resultado es mayor o menor en función de la longitud del texto pero el proceso requiere en general pocos segundos. Como se indica en la página en que se obtiene el texto traducido, el sistema necesita una intervención posterior o en fase de post-edición sobre las palabras traducidas tomadas separadamente y sobre las estructuras sintácticas más o menos reconocidas. Así, los resultados obtenidos van siempre acompañados de la nota “El resultado precisa siempre de corrección. El Instituto Cervantes no se responsabiliza del texto que se traduce ni de la traducción resultante”.

Para ilustrar cuanto se ha dicho hasta ahora sobre las dimensiones que inciden sobre cualquier traducción y explorar los desafíos que tales dimensiones plantean a todo tentativo de traducción automatizada, se han seleccionado y sometido a traducción una serie de textos pertenecientes a diversas tipologías, cuyas temáticas y funciones comunicativas son diferentes: texto de diccionario, texto político, texto jurídico, noticia, poesía, narrativa naturalista del siglo XIX, narrativa modernista del siglo XX; y en diversas lenguas: de español hacia catalán; y de catalán, francés e inglés hacia español.

En cuanto a la selección de textos, hemos querido poner en evidencia la manera en que la tipología textual determina en gran parte: por un lado, el registro y el estilo, el grado de cohesión de las unidades morfosintácticas, el grado de especialidad en el léxico, las estructuras sintácticas más o menos complejas, la ambigüedad léxica, el objetivo de la traducción, y su dimensión cultural; y por otro lado, la necesidad de adoptar estrategias diversificadas e híbridas a la hora de diseñar un sistema de TA. En la medida de lo posible, se han intentado cubrir todas las dimensiones enunciadas más arriba de manera transversal con el objeto de señalar los problemas comunes a todas las tipologías textuales seleccionadas así como las cuestiones específicas de cada una.

En cuanto a la selección de las lenguas nuestro objetivo era verificar si la menor o mayor distancia lingüística y cultural entre los pares de lenguas facilita o hace más compleja la traducción y en qué medida esto es válido para todos los textos. Hay que considerar que para los pares español-inglés y español-francés, el servicio utiliza Systran, empleado desde hace años en el ámbito de las traducciones oficiales de la UE. Aunque es un sistema muy desarrollado parece útil integrar ejemplos y equivalencias basadas en ontologías de subdominios jurídicos y culturales relacionados con la Unión como institución política. Los ejemplos no pretenden ser exhaustivos y los análisis se han limitado, por cuestiones de espacio, a ofrecer algunas consideraciones sobre los límites de la TA basada en modelos tradicionales y enumerar algunos de los desafíos y posibles respuestas a los que se enfrentan los actuales sistemas.

El primer texto en lengua española corresponde a las primeras definiciones de *traducción* del Diccionario de la Real Academia de la Lengua Española (22ª edición). Este texto ha sido traducido al catalán, lengua cuya proximidad lingüística con el español permite obtener traducciones generalmente correctas. No hay que olvidar que entre estas lenguas la definición de reglas de equivalencia está actualmente muy desarrollada gracias a numerosos trabajos de investigación centrados en las lenguas románicas. La descripción detallada de reglas de representación de las categorías morfosintácticas y de sus relaciones a nivel gramatical entre lenguas afines así como el establecimiento de objetivos de traducción específicos para un determinado tipo de texto (en este caso, un diccionario), permiten obtener resultados válidos.

El diccionario se caracteriza por el uso de estructuras oracionales cortas y unidades léxicas cuyo grado de cohesión es medio. Las definiciones se componen generalmente de sintagmas nominales cuyo núcleo suele llevar actualizadores y modificadores (*Obra del traductor; traducción libre*), y subordinadas sustantivas introducidas por artículo seguido de relativo (*La que se*

hace de un idioma extranjero al idioma del traductor). El texto que se obtiene (*Obra del traductor; traducció lliure; La que es fa d'un idioma estranger a l'idioma del traductor*) presenta las mismas estructuras y una serie de unidades léxicas (sustantivos, adjetivos y verbos) equivalentes en la LM. Debido al grado de especialización del texto, cuyo léxico, propio de los diccionarios, no suele tener ambigüedad y cuyas estructuras sintácticas son generalmente fijas, el sistema recurre sin problemas a equivalencias basadas en reglas formales lo cual facilita la traducción automatizada, al menos entre las lenguas español-catalán-gallego, cuya afinidad desde un punto de vista lingüístico es mayor.

La traducción del francés hacia el español de las definiciones de *traduction automatique* y de *traduction instantanée ou simultanée* de la última versión del *Trésor de la Langue française informatisé*, presenta algunas diferencias (léxico, sintaxis y ortografía) respecto a la traducción precedente. Si, por un lado, la distancia lingüística se amplía, por otro lado, el estilo del texto en la LO es sensiblemente diferente respecto al del ejemplo anterior. A esto cabe añadir que el cambio de sistema, en este caso Systran, conlleva también un cambio importante en cuanto a las reglas de transferencia en las que se basa el sistema. Aquí las oraciones y las estructuras sintácticas son más complejas (*Opération visant à assurer la traduction de textes par des moyens informatiques*) probablemente porque el término *traduction informatique* requiere una definición un poco más descriptiva y rica que la de *traducción*. El texto tiene en efecto un grado de cohesión mayor puesto que aparecen en él grupos sintagmáticos lexicalizados como *traduction assistée par ordinateur* o *moyens informatiques* y estructuras verbales compuestas más o menos estables que, como *visant à assurer* (verbo + preposición + verbo en infinitivo), necesitan un modelo de representación lingüística que considere las categorías morfosintácticas de manera conjunta. Además, el *Trésor* ofrece ejemplos de uso del término extraídos de otros textos, en este caso jurídicos y del ámbito de la política, que hacen más complejo el proceso de traducción.

Se han traducido sin problemas cada una de las unidades léxicas tomadas en consideración ofreciendo el equivalente correcto en la LM (*traducción asistida por ordenador; medios informáticos; destinada a garantizar*). Como en el caso anterior, el léxico especializado, poco ambiguo, es compatible con los actuales métodos de representación lingüística integrados en estos sistemas. Los hábitos de colocación (sustantivo + adjetivo) son generalmente fijos y similares en la LM lo cual permite, también en este caso, una más fácil formalización de las equivalencias entre las dos lenguas. Aunque no parece

necesario recurrir a c rpora con ejemplos, s  puede ser  til recurrir a ellos o recurrir a modelos de equivalencias de abreviaturas, acr nimos o siglas, para afinar la traducci n, por ejemplo, de abreviaturas como *d'apr.* para *d'apr s* o *Sc.* para *Sciences*, que han sido trasladadas a la LM manteniendo la misma forma, o que simplemente se han traducido de manera incorrecta: *col.* aparece como *cuello* y no como *colecci n*.

En la traducci n de las definiciones de *translation* y de *machine translation* del *Oxford English Dictionary* hacia el espa ol, el sistema realiza una traducci n aceptable que no requiere un gran trabajo de post-edici n. El texto original no presenta dificultades puesto que est  constituido por unidades l xicas, que como en los casos anteriores, no presentan ambigüedad, dada la tipolog a textual. Se trata de sintagmas nominales con modificadores y construcciones de relativo, *text or word that is translated*. En este caso, el resultado ha sido *texto o una palabra se traducen que*, en el que hay un problema en cuanto al orden de las unidades cuya soluci n pasa a trav s de formalismos sofisticados y focalizados en la sintaxis para establecer equivalencias entre las estructuras de relativo de ambas lenguas. Aunque la distancia lingüística y cultural es mayor entre el espa ol y el ingl s, la baja cohesi n entre las unidades l xicas tomadas en consideraci n y la relativa uniformidad de las estructuras sint cticas as  como la precisi n del significado de las palabras permiten al sistema traducir palabra por palabra sin grandes problemas.

Los textos que siguen son un poco m s extensos que las definiciones anteriores y el registro var a sensiblemente. El primero corresponde a la *D cision 508/2000/CE* del Parlamento europeo y del Consejo del 14 de febrero de 2000, que establece el programa 'Cultura 2000' de la UE. El segundo texto, en ingl s, corresponde a la *Council Joint Action 2005/190/CFSP on the European Union Integrated Rule of Law Mission for Iraq*, del 7 de marzo de 2005, publicada en el EUJUST LEX *Official Journal L 062, 09/03/2005*. Este texto es similar al anterior aunque se trata de un texto jur dico vinculante que presenta una mayor precisi n l xica.

La mayor apertura de los textos a la descripci n de un programa marco y a la toma de una decisi n sobre la posici n de la UE en Irak; su contextualizaci n en un determinado periodo hist rico y la definici n de una serie de objetivos pol ticos y jur dicos conlleva el uso de un l xico m s rico que en los casos anteriores y la utilizaci n de oraciones m s complejas con un mayor n mero de conjunciones (coordinadas), de relativos (subordinadas) y de preposiciones. Algunas caracter sticas del texto parecen facilitar la automatizaci n de la traducci n: el uso regular de sintagmas nominales generalmente formados

por núcleo sustantivo + modificador adjetivo; la regularidad en la construcción de las oraciones; la utilización de fórmulas, nombres oficiales y títulos.

El léxico, en los textos francés e inglés, presenta mayor ambigüedad respecto a los diccionarios. Aunque tratándose de una decisión que afecta las políticas de la Unión en materia de cultura y de un texto jurídico en el que se adopta una posición oficial en materia de relaciones exteriores, la voluntad de precisión del redactor contrarresta la posible ambigüedad semántica de algunas palabras. Ésta se resuelve sin dificultad por medio de equivalencias entre ambas lenguas recurriendo a ejemplos extraídos de otros textos publicados en las instituciones europeas. En efecto, el sistema establece la equivalencia directa para cada categoría y subcategoría morfosintáctica y traduce aparentemente sin dificultad. Algunos grupos de palabras lexicalizados presentan mayor cohesión y parecen ser menos flexibles, desde un punto de vista sintáctico. En el primer ejemplo, los problemas aparecen al traducir las formas verbales compuestas como *mettre en œuvre* (*poner en marcha*) que el sistema ha traducido a mitad: *poner en?*. En el caso del verbo *apparaître*, en *Un large consensus est apparu*, el sistema no ofrece la palabra correcta debido probablemente a la imposibilidad de representar formalmente los diversos significados que un verbo como *apparaître-apparaître* puede tener en diversos contextos oracionales. Así en español, *aparecer* se muestra impropio, aunque no incorrecto, cuando va unido al concepto de *consenso* en *Un amplio consenso apareció*. Similar a éste es el caso de *consultation approfondie* que traduce por *consulta profunda*. Se produce aquí una no correspondencia entre las palabras *approfondie* y *profunda* ya que *approfondie* corresponde normalmente a *profundizada* y *profunda* equivale a *profonde* en francés. Esta no correspondencia puede explicarse por la ausencia de la palabra *profundizada* en el repositorio léxico en español al que recurre el programa, lo que conlleva la búsqueda por parte del sistema de la palabra cuyo lexema y significado se acercan más a *profonde*. En este sentido, es importante recordar que en muchos casos, la constitución de repositorios y de bases de datos léxicas que estos sistemas interrogan se basa en la lematización y no en la realización completa de las palabras (e.g. formas verbales compuestas, participios con valor de adjetivo, estructuras fuertemente lexicalizadas) lo cual puede limitar una traducción correcta y completa. La solución a estos problemas pasa pues, en parte, por la combinación de métodos basados en reglas que tengan en cuenta la mayor cohesión de algunas unidades léxicas (verbos, en particular), y memorias de traducción con ejemplos extraídos de tipos de textos similares.

Las normas y leyes se componen generalmente de fórmulas jurídicas y

administrativas, por este motivo, el segundo texto presenta un alto grado de densidad. Las estructuras son poco complejas desde un punto de vista sintáctico aunque se encuentran en ellas combinaciones de palabras difícilmente separables como *criminal justice system*, fórmulas fijas como *Joint Fact Finding Mission* y estructuras recurrentes en los textos de la UE como *European Council*. La traducción basada en formalismos de representación de equivalencias entre unidades léxicas de menor cohesión como las categorías morfosintácticas sustantivo o adjetivo tomadas separadamente y la no consideración del fuerte carácter formulario de este tipo de textos conlleva incorrecciones desde un punto de vista sintáctico: *sistema de la justicia criminal*, semántico: *hecho común que encontraba la misión; Consejo de Europa* (cuando se trata del *Consejo europeo*). Para paliar este problema pueden establecerse representaciones comunes o intermedias entre las dos lenguas a través de modelos basados en gramáticas de caso y valencias, por un lado, y representaciones conceptuales (e.g. ontologías para el subdominio jurídico de la UE), por otro, que permitan la transferencia y la desambiguación léxica de las unidades tomadas en consideración.

Para este estudio también se ha procedido a la traducción automatizada de un texto periodístico. El texto, en francés, corresponde al primer párrafo de una noticia extraída de *Le Monde* del 28 enero de 2006 sobre los violentos enfrentamientos entre palestinos en Gaza. Las dificultades se concentran en el léxico. Por ejemplo, *confrontaciones* y *rabia* son efectivamente traducciones correctas de los términos *affrontements* y *rage* respectivamente pero no parecen ser adecuados en el contexto oracional en el que aparecen: *las confrontaciones hacían aún rabia* no es correcto en español mientras sí lo es *Les affrontements faisaient encore rage*, en francés. Lo mismo ocurre con *más fatales* que no corresponde a *les plus meurtriers* (ambigüedad de tipo gramatical) y *apoderarse de un partidario* por *s'emparer d'un partisan* (ambigüedad léxica). En estos casos, la solución puede ser recurrir a ejemplos y estadísticas de frecuencia de aparición de estas unidades en determinados contextos y sobre todo establecer modelos de representación lingüística que recojan de manera explícita las diferentes valencias semánticas de la unidad en cuestión para una lengua y otra. Otro orden de dificultades se presenta en el uso de las preposiciones y apócope (preposición + artículo), de artículos que no se acuerdan correctamente con el sustantivo al que actualizan o modifican y de artículos que delante de los nombres de los movimientos palestinos enfrentados han sido erróneamente mantenidos, por ejemplo en español, Fatah y Hamás no van precedidos de artículo determinado como en francés. Esto puede deberse a una

incorrecta definición de las reglas de representación intermedia entre las dos lenguas en lo que respecta al uso de preposiciones y de artículos.

Hasta aquí hemos considerado una serie de problemas relacionados con la traducción automatizada de textos cuya función descriptiva y comunicativa en los contextos lingüístico, político-jurídico y periodístico respectivamente, determina un estilo, en la mayor parte de los ejemplos, expositivo, claro; una mayor densidad y cohesión de las unidades morfosintácticas del texto, y una ambigüedad sintáctica y léxica que aumenta a medida que la distancia lingüística y cultural entre las lenguas se amplía. Para cubrir otra serie de aspectos relacionados con la traducción automatizada, se han traducido algunos textos literarios, cuya voluntad expresiva está determinada por la búsqueda de un estilo propio, aun en el seno de una corriente literaria, lo cual conlleva una complejidad mayor, sobre todo en lo que concierne el léxico.

El primer texto, en catalán, corresponde a los primeros versos del *Amb mans de paper fi* del poeta Miquel Martí i Pol, recogidos en el libro de poemas del autor *La pell del violí* (1972-73). Se trata de una traducción hacia el español, lengua próxima desde un punto de vista gramatical. Aquí el problema aparece cuando se traducen algunas palabras como el verbo *esquinço* que va acompañado de *rocs*; la construcción a la *post*; los verbos *resta* y *corseca*; y el sustantivo *gep*. Cabe suponer que el repositorio al que recurre el servidor AutomaticTrans no ha encontrado la equivalencia en español de estas unidades léxicas, que han sido tomadas separadamente. Siendo el texto poético, las palabras se alejan del uso común. Éstos son términos poco habituales que contribuyen a elaborar imágenes apoyadas en figuras como la metáfora (*romper rocas con las manos*), dificultando la automatización de la traducción.

Aunque este tipo de traducciones son difícilmente automatizables porque se alejan del lenguaje específico o de un dominio específico del saber, es posible en estos casos recurrir a memorias de traducciones y a modelos sofisticados basados en redes semánticas que permitan la transferencia léxica y la desambiguación léxica para poder traducir las unidades polisémicas del texto. Sin olvidar además que la posibilidad de obtener traducciones de calidad será mayor cuanto mayor sea la proximidad lingüística y cultural entre las dos lenguas.

El texto narrativo tiene una dificultad añadida que concierne generalmente las estructuras sintácticas, más complejas. Hemos seleccionado un fragmento extraído de la novela de Guy de Maupassant *Une vie* (1883), que se encuadra generalmente en la narrativa de corte naturalista del siglo XIX. El fragmento corresponde al primer capítulo de la novela en la edición digitalizada disponible en el catálogo electrónico de la Bibliothèque Nationale de

France. El texto está compuesto fundamentalmente por oraciones complejas, coordinadas copulativas y subordinadas, cuyos núcleos verbales van acompañados de diversos complementos. En el fragmento traducido, las oraciones son, dejando de lado el estilo, correctas sintácticamente: *L'averse, toute la nuit, avait sonné contre les carreaux et les toits*, que traduce por *El chubasco, toda la noche, había sonado contra los cuadrados [las baldosas] y los techos* y *Le ronflement des ruisseaux débordés emplissait les rues désertes où les maisons, comme des éponges, buvaient l'humidité*, donde se obtiene en español *El ronquido [rugir] de los arroyos desbordados llenaba las calles desérticas donde las casas, como esponjas, bebían la humedad*.

El texto está cargado de términos polisémicos en una búsqueda por parte del autor de imágenes evocativas del tiempo, de la noche, del aspecto de la ciudad tras la tormenta. Abundan pues las metáforas (*casas como esponjas y ronquido del arroyo*) que el sistema reconoce gracias a una serie de reglas de transferencia léxica basadas en equivalencias entre el significado de la unidad léxica en la LO y el primer significado que esa misma unidad tiene en el lenguaje común. En otros casos, no consigue trasladar el significado correctamente porque el sistema limita la equivalencia a pocos significados: uno común y otro especializado. Recuérdese que el sistema utilizado aquí para el par francés-español es Systran, aplicado a las traducciones de textos en la UE. Así, *chargé* se traduce como *encargado*, que es un término correcto en un informe económico o jurídico, pero incorrecto en este fragmento donde *chargé* significa *cargado*, respecto al cielo lluvioso.

Otro texto que plantea dificultades similares es *Métaphoriquement*, uno de los *exercices de style* que Raymond Queneau publicó por primera vez en 1947 y que hemos extraído de la edición de 1995. En este caso, el sistema no parece haber reconocido ciertas unidades léxicas como *tas*, que traduce por *montó* o *voyageuses*, *déplumé* y *oisillon*, que simplemente mantiene en francés, probablemente porque no encuentra equivalente en la LM o porque el analizador semántico no ha identificado correctamente la unidad en la LO. Los analizadores sintácticos, como se ha visto en los ejemplos precedentes, no suelen tener problemas en trasladar las estructuras sintácticas más o menos complejas de la LO a la LM, a excepción de algunas combinaciones de palabras lexicalizadas⁴, así como algunas construcciones apocopadas (preposición + artículo).

⁴ En un sistema de TA, lo ideal es que el sistema reconozca la unidades de mayor cohesión antes de tratar las de menor cohesión, es decir, reconocer segmentos completos antes de tratar cada palabra separadamente.

Si ampliamos la distancia lingüística, las dificultades aumentan. *Ulysses* de James Joyce (1922) está considerada una novela “masiva” que contiene 267.000 palabras de un total de 30.000 palabras extraídas del vocabulario. Como Queneau en *Exercices de style*, el autor de *Ulysses* recurre constantemente a los dobles sentidos, los juegos de palabras y la parodia. La traducción de este fragmento, que corresponde al tercer párrafo del primer capítulo de la novela (edición digital de la Penn State University), presenta resultados igualmente interesantes. La densidad léxica es mayor y la presencia de unidades lingüísticas complejas gramaticalmente como *faced about*, *mounted the round gunrest* y *catching sight*, complican la tarea del sistema, que tiende a traducir palabra por palabra: *hizo frente alrededor*, *montó el redondo gunrest* y *vista catching*; y a no considerar las locuciones de una y otra lengua basándose, por ejemplo, en analogías y ejemplos que presenten equivalencias sintácticas. El problema de las correspondencias entre LO y LM, en el nivel más básico de los equivalentes de traducción de las distintas unidades morfosintácticas se acentúa cuando se traduce entre el inglés y el español, cuya distancia lingüística es mayor. Algunas dificultades persisten cuando aparecen verbos con preposición como *he bent towards*, que se traduce por *se dobló hacia* y no *se inclinó hacia*; palabras o unidades sintagmáticas con diversos sentidos: *the surrounding country* por el *país circundante*; unidades para las cuales el sistema no traduce la palabra porque no encuentra un equivalente en español: *gunrest*, *untonsured* y *hued*; y unidades como *shaking* y *sleepy* que, aunque sí se haya encontrado equivalente, no han sido correctamente traducidas, presentando errores de carácter morfológico: *sacudariendo*, *sacudaría* y *soñoliento*.

Para este tipo de textos, la elaboración de modelos de análisis semántico es más compleja y la necesidad de recurrir a representaciones conceptuales válidas para diversas lenguas plantea mayores desafíos que pasan por la creación de lexicones especializados en determinados campos del conocimiento y su integración con corpórea constituidos por ejemplos de traducciones precedentes. El sistema podría integrar redes semánticas y ontologías que, sin embargo, como demuestra la práctica, son solamente eficaces si se quiere describir una porción del conocimiento o subdominio del conocimiento. En la elaboración de una ontología de este tipo hay que tener en cuenta que el número de conceptos, sus propiedades y atributos y las relaciones que establecen entre sí es enorme y que siempre existe el riesgo de no cubrir todos los aspectos semánticos ligados a un concepto.

En el caso de los textos literarios, en los que la búsqueda de imágenes a través de figuras retóricas como la metáfora favorecen la ambigüedad léxica, es posible establecer lexicones específicos para cada lengua pero es muy complejo elaborar ontologías de conceptos independientes de las dos lenguas que sean eficaces y capaces de generar representaciones semánticas interlingua para expresar, por ejemplo, una metáfora. Sí pueden, sin embargo, establecerse o usarse bases de datos en diversas lenguas y representaciones conceptuales ligadas a la obra de un determinado autor o a una determinada corriente literaria o movimiento y crear subdominios del saber en torno a ese autor o ese movimiento. Sería posible relacionar determinados conceptos a las imágenes recurrentes en las obras de ese autor y establecer representaciones semánticas interlingua en torno a esos conceptos facilitando la automatización de la traducción.

El tipo de TA más utilizado hoy en día es la que pasa a través de servicios como el del Instituto Cervantes, que se presentan en línea y que ofrecen traducciones rápidas a través del fácil manejo de los sistemas. El tratamiento de unidades léxicas y terminológicas es además muy consistente y rápido y funciona mejor entre lenguas afines; entre textos especializados en los que aparecen estructuras gramaticales más o menos fijas y unidades léxicas con poca polisemia. Las dificultades surgen, en la mayor parte de los casos, cuando una palabra tiene diferentes significados y las restricciones de significado que tiene una lengua no tienen porqué ser evidentes en otra. E incluso cuando no se tienen en cuenta, en la elaboración de modelos y reglas de equivalencias, la densidad y la granularidad del texto. Los sistemas actuales han superado el enfoque reductivo de traducción palabra por palabra mientras tienden a una combinación de métodos que permite al traductor automático integrar de forma modular no sólo una serie de reglas formales y ontologías que representen equivalencias a nivel lingüístico (fonético, morfológico, sintáctico y semántico) y extra-lingüístico (convenciones culturales) sino también traducciones previamente realizadas recogidas en corpórea paralelos en los que es posible visualizar una determinada unidad léxica o sintagmática en ambas lenguas y en contexto.

En esta combinación de métodos sería útil tener en consideración los procesos de redacción de un texto, si su léxico pertenece a un determinado sublenguaje o tipología textual, para ofrecer modelos similares de textos previamente tratados con equivalencias. Esto facilitaría la definición de conceptos para un determinado dominio (ontologías) y la proyección, a través de reglas bien definidas, de estructuras sintácticas sobre estructuras semánticas y viceversa (MORENO ORTIZ 2000).

Por otro lado, en cualquier proyecto es fundamental contar con la interacción entre el ordenador, el sistema, y el traductor humano. En su célebre artículo *The Proper Place of Men and Machines in Language Translation*, Martin Kay pone una serie de condiciones aún válidas para una TA realista. Según este autor, para desarrollar una TA eficaz es necesario equilibrar la participación entre la máquina y el traductor humano, favoreciendo la *interactividad* entre ambos. De esta manera la máquina ayuda al traductor con todas las herramientas puestas a disposición por la informática y el traductor “ayuda” al ordenador (LAVID 2005: 258-259), en particular, en los casos en que se presenten problemas de ambigüedad léxica y allí donde la traducción requiera un conocimiento extra-lingüístico, cuya representación mediante reglas y ontologías es todavía compleja.

BIBLIOGRAFÍA

- BIBER, D. / FINEGAN, E. (1986). An initial typology of text-types. In: J. AARTS, W. MEIJIS (eds.), *Corpus linguistics II*, Amsterdam: Rodopi, p. 19-46.
- HUTCHINS, W. J. (2002) Machine translation today and tomorrow. In: G. WILLÉE, B. SCHRÖDER, H.-C. SCHMITZ H-C. (hrsg): *Computerlinguistik: was geht, was kommt? Festschrift für Winfried Lenders*, Sankt Augustin: Gardez! Verlag, p. 159-162.
- HUTCHINS, W. J. / SOMERS, H.L. (1992). *An Introduction to Machine Translation*, New York: Academic Press.
- KAY, M. (1997). The Proper Place of Men and Machines in Language Translation, *Machine Translation* 12:1-2, p. 3-23.
- LAVID, J. (2005). *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*, Madrid: Cátedra.
- MELBY, A.K. / WARNER, T. (1995). *The Possibility of Language. A discussion of the nature of language with implications for human and machine translation*, Amsterdam-Philadelphia: John Benjamins Publishing.
- MORENO ORTIZ C. (2000). Diseño e implementación de un lexicón computacional para lexicografía y traducción automática, *EliEs – Estudios de Lingüística del Español* 9, <http://elies.rediris.es/elies9/> consultado el 10/12/2006
- NORD, C. (1993). Scopus, Loyalty, and Translational Conventions, *Target*, p. 95-107.
- PÉREZ HERNÁNDEZ, C. (2002.) Explotación de los corpóra textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento, *EliEs – Estudios de Lingüística española* 18, <http://elies.rediris.es/elies18/> consultado el 19/09/2006
- TROSBORG, A. (ed.) (1997). *Text typology and translation*, Amsterdam-Philadelphia: John Benjamins Publishing.

EJEMPLOS⁵**Texto original** (DRAE 22ª edición):

traducción.

(Del lat. traductiō, -ōnis).

1. f. Acción y efecto de traducir.

2. f. Obra del traductor.

[...]

~ directa. 1. f. La que se hace de un idioma extranjero al idioma del traductor.

Texto traducido (español-catalán):

traducció.

(Del lat.[llat.] traductio, -onis).

1. f. Acció i efecte de traduir.

2. f. Obra del traductor.

[...]

~ directa. 1. f. La que es fa d'un idioma estranger a l'idioma del traductor.

Texto original (Trésor de la Langue française informatisé):

Traduction automatique (T.A.), traduction assistée par ordinateur (T.A.O.). „Opération visant à assurer la traduction de textes par des moyens informatiques” (GDEL). Un premier congrès international sur la traduction automatique eut lieu dès 1952 (Hist. gén. sc., t. 3, vol. 2, 1964, p. 116).

Traduction instantanée ou simultanée. (Dans une conférence internationale) Traduction immédiate (et orale) du discours ou de la communication prononcé(e) par l'orateur (d'apr. BAUDHUIN 1968). La traduction simultanée des conférences est assurée ainsi que l'intervention de traducteurs dans les petits groupes (Lettres des départements sc. du CNRS, Sc. de l'homme et de la société, oct.-nov. 1989, no 20, p. 33, col. 2).

Texto traducido (francés-español):

Traducción automática (T.A.), traducción asistida por ordenador (T.A.O.)., Operación destinada a garantizar la traducción de textos por medios informáticos [informáticos]” (GDEL). Un primer congreso internacional sobre la traducción automática tuvo lugar a partir de 1952 (Hist. gén. Sc., t. 3, vol. 2.,1964, p. 116).

Traducción instantánea o simultánea. (En una conferencia internacional) Traducción inmediata (y oral) del discurso o la comunicación pronunciados por el orador (d'apr. [de] BAUDHUIN 1968). La traducción simultánea de las conferencias está garantizado [garantizada] así como la intervención de traductores en los pequeños grupos (Cartas de los departamentos Sc [C.] del CNRS, Sc de hombre y sociedad [C. del hombre y de la sociedad], noviembre de 1989, n 20, p. 33, cuello [colección] 2).

⁵ Las unidades entre corchetes son sólo propuestas de traducción que pretenden ejemplificar el error o la anomalía encontrados en el texto y sugerir alternativas válidas. No pretenden en ningún caso ser soluciones únicas.

Texto original (Oxford English Dictionary):

translation

- noun 1 the action or process of translating. 2 a text or word that is translated.

machine translation

- noun translation carried out by a computer.

Texto traducido (inglés-español):

traducción

¿? sustantivo 1 la acción o el proceso de traducir. 2 un texto o una palabra se traducen que [que se traducen].

traducción automática

¿? traducción del sustantivo realizada por una computadora/un ordenador.

Texto original (Décision 508/2000/CE):

Pour l'Union européenne, ce premier programme-cadre en faveur de la culture est l'occasion de mettre en œuvre une nouvelle approche pour l'action culturelle de la Communauté lui permettant de répondre aux défis contemporains et aux aspirations des citoyens européens, comme à celle de ses acteurs culturels. La consultation approfondie engagée par la Commission pour la préparation du programme-cadre a permis de préciser le rôle et la place de la culture dans les grands défis auxquels l'Union européenne est confrontée. Un large consensus est apparu entre les États membres, les parlementaires, la Commission, les organisations internationales (Conseil de l'Europe, Unesco) et les organisations culturelles pour considérer que la notion de culture s'étend aujourd'hui à la culture populaire, la culture industrielle de masse et la culture au quotidien.

Texto traducido (francés-español):

Para la Unión Europea, este primer programa marco en favor de la cultura es la ocasión de poner en? [en marcha] un nuevo enfoque para la acción cultural de la Comunidad que le permite [permite] responder a los retos contemporáneos y a las aspiraciones de los ciudadanos europeos, como a la de sus protagonistas culturales. La consulta profunda [amplia/profundizada] iniciada por la Comisión para la preparación del programa marco permitió [ha permitido] precisar el papel y el lugar de la cultura en los grandes retos a los cuales se enfrenta [a] la Unión Europea. Un amplio consenso apareció [se ha producido] entre los Estados miembros, los parlamentarios, la Comisión, las organizaciones internacionales (Consejo de Europa, UNESCO) y las organizaciones culturales para considerar que el concepto de cultura se extiende hoy a la cultura popular, a la cultura industrial de masa [masas] y a la cultura al diario [cotidiana].

Texto original (Council Joint Action 2005/190/CFSP):

Having regard to the Treaty on European Union, and in particular Article 14, Article 25, third paragraph, Article 26 and Article 28 (3) thereof,

Whereas:

(2) The European Council on 5 November 2004 welcomed the Joint Fact Finding Mission for a possible integrated police, rule of law and civilian administration mission for Iraq and considered its report. The European Council recognised the importance of strengthening the criminal justice system, consistent with the respect for the rule of law, human rights and fundamental freedoms. It noted the wish of the Iraqi authorities for the EU to become more actively invol-

ved in Iraq and that strengthening the criminal justice sector would respond to Iraqi needs and priorities.

Texto traducido (inglés-español):

Teniendo respeto [Considerando el] al tratado en [de] la unión [Unión] europea, y [en] particular el artículo 14, el artículo 25, el tercer párrafo, el artículo 26 y el artículo 28 (3) de eso [éste],

Mientras que [Considerando que]:

(2) el Consejo de Europa [Consejo europeo] el 5 de noviembre [de] 2004 dio la bienvenida [acogió el] al hecho común que encontraba la misión [la Joint Fact Finding Mission] para [para establecer una posible] policía integrado [integrada] posible, regla de la ley [la ley] y [una] misión [para] de la administración civil para Iraq y consideraba [consideró] su informe. El Consejo de Europa [Consejo europeo] reconoció la importancia de consolidar el sistema criminal de la justicia [sistema de justicia], constante [de acuerdo] con el respecto por regla de la ley [la ley], los derechos humanos y los freedoms [las libertades] fundamentales. Observó [Constató] el deseo de las autoridades iraquíes para que el EU [la UE] se implicó [implique] más activamente en Iraq y que consolida [la consolidación de] el sector criminal de la justicia [el sector de la justicia] respondería a las necesidades y a las prioridades iraquíes.

Texto original (Le Monde 28 enero 2006):

Les affrontements faisaient encore rage, dimanche 28 janvier, entre le Hamas et le Fatah dans la bande de Gaza, où 24 Palestiniens sont morts en moins de trois jours, dans les heurts les plus meurtriers depuis l'élection des islamistes en janvier 2006. Les affrontements se sont poursuivis en dépit des appels au calme venus de toutes part. Dans la nuit de samedi à dimanche, quatre Palestiniens ont été tués. Des militants armés du Hamas ont pris d'assaut un immeuble résidentiel à Beit Lahya, dans le nord de la bande de Gaza, pour s'emparer d'un partisan du Fatah.

Texto traducido (francés-español):

Las confrontaciones [Los enfrentamientos] hacían aún rabia [estragos], el domingo 28 de enero, entre el Hamas [Hamás] y el Fatah [Fatah] en la Franja de Gaza, donde 24 Palestinos (se) han muerto en menos de tres días, en los choques más fatales [violentos] desde la elección del [de los] islamistas en enero de 2006. Las confrontaciones [Los enfrentamientos] (se) continuaron a pesar de las llamadas a la calma venidas de todas van [partes]. En la noche de [del] sábado a [al] domingo, se mató a cuatro Palestinos. Militantes armados del [de] Hamas tomaron de asalto un edificio residencial a [en] Beit Lahya, en el norte de la Franja de Gaza, para apoderarse de [arrestar a] un partidario del [#]Fatah.

Texto original (Amb mans de paper fi, en La pell del violí 1974):

Esquinço rocs amb mans de paper fi
i em defenso dels llamps amb un sonet
que vaig escriure fa molts anys, quan era
jove i agosarat. Ara no tinc
ni pa a la post ni borra a les butxaques.
Tot el que em resta ho duc al gep: els mots,
els desenganys i aquell desig vivíssim
mil voltes reprimít i que em corseca.

Texto traducido (catalán-español):

Esquinço rocs/Rompo rocas con manos de papel fino
 y me defiendo de los rayos con un soneto
 que escribí hace muchos años, cuando era
 joven y atrevido. Ahora no tengo
 ni pan a la post [sobre la tabla] ni borra a [en] los bolsillos.
 Todo cuanto me resto [queda] lo llevo al gep [en la chepa]: los mots [las palabras],
 los desengaños y aquel deseo vivísimo
 mil vueltas reprimido y que me corseca [carcome].

Texto original (*Une vie* 1883, ed. digital BNF):

Jeanne, ayant fini ses malles, s'approche de la fenêtre, mais la pluie ne cessait pas. L'averse, toute la nuit, avait sonné contre les carreaux et les toits. Le ciel bas et chargé d'eau semblait crevé, se vidant sur la terre, la délayant en bouillie, la fondant comme du sucre. Des rafales passaient pleines d'une chaleur lourde. Le ronflement des ruisseaux débordés emplissait les rues désertes où les maisons, comme des éponges, buvaient l'humidité qui pénétrait au-dedans et faisait suer les murs ...

Texto traducido (francés-español):

Jeanne, terminando sus portaequipajes [mallas], se acercó [a] la ventana, pero la lluvia no cesaba. El chubasco, toda la noche, había sonado contra los cuadrados [las baldosas] y los techos. El cielo bajo y encargado [cargado] de agua parecía estallado [moribundo], vaciándose sobre la tierra, regándolo en pulpa [diluyéndola como pasta], fundiéndoselo [fundióndola] como azúcar. Ráfagas pasaban llenas de un calor pesado. El ronquido [rugir] de los arroyos desbordados llenaba las calles desérticas donde las casas, como esponjas, bebían la humedad que penetraba a interior [en el interior] y hacía sudar las paredes...

Texto original (*Métaphoriquement, en Exercices de style* [1945] 1995 Gallimard):

Au centre du jour, jeté dans le tas des sardines voyageuses d'un coléoptère à l'abdomen blanchâtre, un poulet au grand cou déplumé harangua soudain l'une, paisible, d'entre elles et son langage se déploya dans les airs, humide d'une protestation. Puis attiré par un vide, l'oisillon s'y précipita. Dans un morne désert urbain, je le revis le jour même se faisant moucher l'arrogance pour un quelconque bouton.

Texto traducido (francés-español):

En el centro del día, lanzado [tirado] de las el montóces [en el montón de] sardinas voyageuses [viajeras] de un coleóptero al/de abdomen blancuzco, un pollo al [con un] gran cuello déplumé [desplumado] arengó repentinamente la [a] una, pacífica, de entre ellas y su lengua [lenguaje] se desplegó en [por] los aires, húmedo de una protesta. Luego atraído por un vacío, el oisillon [polluelo] se precipitó. En un mate desierto urbano/desierto urgano apagado, lo revivo [vi de nuevo] el mismo día haciéndose mou [limpiar] la arrogancia para [por un] botón [un botón cualquiera].

Texto original (*Ulysses* 1929, ed. digital Penn State University):

Solemnly he came forward and mounted the round gunrest. He faced about and blessed gravely thrice the tower, the surrounding country and the awaking mountains. Then, catching sight of Stephen Dedalus, he bent towards him and made rapid crosses in the air, gurgling in his throat and shaking his head. Stephen Dedalus, displeased and sleepy, leaned his arms on the top of the staircase and looked coldly at the shaking gurgling face that blessed him.

Texto traducido (inglés-español):

Él vino solemnemente adelante [avanzó] y montó el redondo man's gunrest [accedió a la redonda plaza de tiro]. Él [Ø] hizo frente alrededor [giró alrededor] y bendijo grave [con seriedad] tres veces la torre, el país [campo] circundante y las montañas que se despertaban. Entonces, vista catching de [viendo a] Stephen Dedalus, él se dobló [se inclinó] hacia él e hizo [y trazó] cruces rápidas en el aire, expressando con gorjeos en su garganta [gorjeando con la garganta] y sacudariendo su [sacudiendo la] cabeza. Stephen Dedalus, descontentado y soñoliento [contrariado y somnoliento], inclinado sus brazos en la tapa [apoyó los brazos en el inicio] de la escalera y mirado [miró] frío [con frialdad] la cara que expressaba con gorjeos que sacudaría que lo bendijo [la cara gorjeante que le bendecía]...

CHALLENGING AUTOMATIC TRANSLATION IN SPANISH:
COMBINATION OF METHODS AND ONTOLOGIES

S u m m a r y

The creation and improvement of automatic translation tools and technologies has been irregular. The application and purpose of some of these tools is still problematic, in particular, while implemented without linguistic resources like dictionaries or parallel corpora.

In this paper, I explore some fundamental dimensions within texts that should be taken into account for the design and implementation of any automatic translator. I critically consider the theoretical approaches and methods applied for the design of automatic translation software by analysing different texts translated into Spanish by the Instituto Cervantes online translation service. The purpose of this succinct study is to illustrate the challenges put by translation and suggest some answers related to the combination of methods and tools; the design and implementation of ontologies and formal rules based not only on linguistic information but also on knowledge; and the effective interaction of the machine and the human translator.

Summarised by Ana Pano

Słowa kluczowe: tłumaczenie automatyczne, reguły, korpusy, ontologie, ambiwalencja, polisemia.

Palabras claves: traducción automática, reglas, corpora, ontologías, ambigüedad, polisemia.

Key words: automatic translation, rules, corpora, ontologies, ambiguity, polysemy.