

Análisis del tratamiento de...  
M. Fomicheva, I. da Cunha, G. Sierra

## ANÁLISIS DEL TRATAMIENTO DE LA TERMINOLOGÍA EN LA TRADUCCIÓN AUTOMÁTICA: IMPLICACIONES PARA LA EVALUACIÓN

Marina Fomicheva\*  
Iria da Cunha\*\*  
Gerardo Sierra\*\*\*

**Resumen:** En este artículo se presenta una propuesta metodológica para el análisis comparativo de Traducciones Automáticas [TAs] y Traducciones Humanas [THs] a nivel léxico-terminológico. El objetivo general de la investigación es estudiar las diferencias lingüísticas sistemáticas entre la TA y la TH de cara a la problemática de la evaluación automática de sistemas de TA. Los objetivos específicos son: a) detectar las diferencias en la distribución de unidades terminológicas entre la TA y la TH; b) identificar las condiciones en las que se producen dichas diferencias teniendo en cuenta los Textos Originales [TOs] y las estrategias de traducción. La metodología del estudio incluye las fases siguientes: selección de dos sistemas de TA basados en estrategias diferentes, constitución de un corpus paralelo inglés-español de textos especializados del ámbito médico, análisis estilométrico de los Textos Traducidos [TTs] para caracterizar el lenguaje de la TA en oposición al de la TH y, finalmente, clasificación de las modificaciones que realizan con respecto al TO los sistemas de TA y los traductores humanos. Los resultados de la investigación indican que las diferencias relacionadas con las modificaciones opcionales realizadas por los traductores y las diferencias que se deben a la falta de modificaciones obligatorias en la TA no tienen la misma relevancia para evaluar la calidad de esta última.

**Palabras clave:** Traducción automática; *Translation shift*; Discurso especializado; Terminología bilingüe.

**Resumo:** Este trabalho apresenta uma metodologia para a análise comparativa de traduções humanas e automáticas no plano léxico-terminológico. Esta proposta é aplicada a um corpus inglês-espanhol paralelo de textos especializados na área médica. O objetivo geral da pesquisa é estudar as diferenças sistemáticas lingüísticas entre tradução automática e tradução humana no contexto do problema de avaliação automática de sistemas. Os objetivos específicos são: a) detectar diferenças na distribuição das unidades terminológicas entre a tradução humana e traduções automáticas de sistemas baseados em estratégias diferentes; b) identificar as condições em que essas diferenças ocorrem considerando os textos originais e as estratégias de tradução humana e automática. A metodologia do estudo envolve, em primeiro lugar, a utilização de técnicas estilométricas para caracterizar a linguagem da tradução automática contra o da tradução humana e, por outro lado, a classificação das alterações em relação o texto original feitas por tradutores humanos e sistemas de tradução automática. Os resultados da pesquisa indicam que as diferenças entre tradução automática e tradução humana relacionadas com modificações opcionais feitas por tradutores e as diferenças que se devem à falta de mudanças obrigatórias na tradução automática não são igualmente importantes para avaliar a qualidade desta última.

**Palavras-chave:** Tradução automática; *Translation shift*; Discurso especializado; Terminologia bilingüe.

**Abstract:** This paper presents a methodology for the comparative analysis of human and machine translation at a lexical-terminological level. This proposal is applied to an English-Spanish parallel corpus of specialized texts in the medical domain. The main aim of the study is to explore systematic linguistic differences between machine and human translation in the light of the problem of automatic system evaluation. The specific objectives are: a) detect differences in the distribution of terminological units between human and machine translation, b) identify the conditions under which such differences occur considering original texts and strategies for human and machine translation. The study methodology involves, on the one hand, the use of stylometry techniques to characterize the language of machine translation versus human translation and, on the other hand, the classification of translation shifts performed by human translators and modifications made by machine translation systems to the original text. The research results indicate that the differences between machine translation and human translation related to optional translation shifts performed by translators and the differences related to the lack of obligatory changes in machine translation are not equally important to assess the quality of the latter.

**Palabras clave:** Machine Translation; Translation Shift; Specialized Texts; Bilingual Terminology.

**Cómo citar este artículo:** FOMICHEVA, Marina; CUNHA, Iria da; SIERRA, Gerardo. Análisis del tratamiento de la terminología en la traducción automática: implicaciones para la evaluación. *Debate Terminológico*. No. 9, Nov. 2013; pp. 16-36

\* Becaria predoctoral del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra (Barcelona). Email: [marina.fomicheva@upf.edu](mailto:marina.fomicheva@upf.edu)

\*\* Investigadora Juan de la Cierva del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra (Barcelona). Email: [iria.dacunha@upf.edu](mailto:iria.dacunha@upf.edu)

\*\*\* Director del Grupo de Ingeniería Lingüística del Instituto de Ingeniería de la Universidad Nacional Autónoma de México (Ciudad de México). Email: [GSierraM@iingen.unam.mx](mailto:GSierraM@iingen.unam.mx)

## 1. Introducción

En las últimas dos décadas, el campo de la Traducción Automática [TA]<sup>1</sup> ha tenido un desarrollo vertiginoso motivado por las necesidades de la comunicación multilingüe en un mundo globalizado. Con todo, incluso los mejores sistemas de TA todavía están muy lejos de sustituir al traductor humano.

Un procedimiento crucial para el desarrollo de la TA es su evaluación. La evaluación permite detectar carencias y mejoras, así como comparar el desempeño de diferentes sistemas ilustrando las ventajas y desventajas de los métodos empleados. Medir la calidad de una traducción es un trabajo laborioso que requiere de tiempo y recursos. Por este motivo, en los últimos años se han propuesto diversos métodos para automatizar este proceso.

Las métricas de evaluación automática existentes parten del siguiente supuesto: la TA sería una tarea resuelta si fuera imposible distinguirla de la Traducción Humana [TH]. Por ello, para medir la calidad de la TA se calcula el grado de similitud entre ésta y una o varias Traducciones Humanas de Referencia [THRs]. Las métricas más utilizadas en la actualidad calculan la similitud en términos de coocurrencia de n-gramas de palabras (Papineni et al., 2001). Una limitación importante de dichas métricas es que penalizan de la misma manera cualquier tipo de diferencia TA-THR a nivel léxico, sin hacer la distinción necesaria entre la variación aceptable y las divergencias que realmente reflejan la calidad de la traducción. No todas las diferencias TA-THR resultan de errores de los sistemas. Algunas se relacionan con las modificaciones del original que realizan los traductores humanos, bajo la influencia de diversos factores, tanto lingüísticos como extralingüísticos.

Con los avances recientes en el campo de los estudios descriptivos de la traducción se ha demostrado que los Textos Traducidos [TTs] poseen características lingüísticas que los distinguen de otros tipos de textos (Puurtinen, 2003) y se ha postulado la existencia del lenguaje de traducción o *translationese* como una variante particular de la expresión lingüística. Una de las fuentes de dichos rasgos es la interferencia de la Lengua Fuente [LF], es decir, la influencia de sus características sistémicas, pragmáticas, estilísticas, etc. en el TT, que se pronuncia en la reproducción de la sintaxis y los patrones colocacionales del Texto Original [TO]. Asimismo, la actividad traductora tiene ciertos aspectos inherentes, independientes de la lengua de partida, que afectan la producción lingüística. Baker (1993) identifica las siguientes tendencias universales de la traducción, independientes de la LF: los TTs tienden a ser más explícitos y menos ambiguos que otros textos producidos en la Lengua Meta [LM]; asimismo, los traductores tienden a evitar la repetición que ocurre en los originales y a exagerar las convenciones de uso de la LM.

Estas regularidades en la selección de recursos por parte del traductor, que se manifiestan en los rasgos distintivos de los TTs, están relacionadas con las diferencias sistémicas y pragmáticas entre las lenguas, las particularidades del proceso traductor como una actividad cognitiva compleja y las condiciones de la mediación cultural, que supone una gran distancia entre el autor del original y el receptor de la traducción. Debido a los factores mencionados, la traducción implica ciertas modificaciones con respecto a la forma y el contenido del TO, denominadas *translation shifts*<sup>2</sup>. Algunas de ellas son obligatorias, es decir, se deben a las diferencias sistémicas entre las lenguas, mientras que otras son opcionales, producto de elección consciente o inconsciente del traductor.

Dado que la TA es una tarea de Inteligencia Artificial, podríamos considerar que uno de sus objetivos principales es modelar el comportamiento de los traductores humanos. Así, un análisis lingüístico comparativo TA-TH que tome en consideración el trabajo descriptivo realizado en el marco de la traductología sería de gran utilidad para la evaluación y el desarrollo de sistemas de TA, ya que ayudaría a responder a una pregunta importante: ¿qué aspectos de la TH queremos y podemos modelar por medio de técnicas computacionales?

El presente estudio se enmarca en una investigación más amplia (Fomicheva, 2012) que tiene por objetivo desarrollar una propuesta metodológica para el análisis comparativo TA-TH a tres niveles de la

<sup>1</sup> El Anexo 1 contiene la lista de abreviaciones utilizadas en el presente artículo.

<sup>2</sup> Para una descripción detallada de la evolución de las concepciones teóricas desarrolladas al respecto, véase Cyrus (2009).

lengua: léxico-terminológico, morfosintáctico y discursivo. Dicha propuesta fue aplicada a un corpus paralelo inglés-español de textos especializados del ámbito médico, que incluía TOs, THs y TAs producidas por un sistema estadístico (*Google Translate*<sup>3</sup>) y un sistema basado en reglas (*Lucy LT*<sup>4</sup>). Con base en este corpus se compararon las distribuciones de expresiones lingüísticas (unidades terminológicas [UTs], n-gramas de etiquetas morfosintácticas y relaciones discursivas) en las TAs y las THs, y se identificaron las condiciones en las que se producían las diferencias detectadas teniendo en cuenta los TOs y las estrategias de traducción empleadas por los traductores humanos y por los sistemas de TA.

En este artículo presentamos el análisis realizado a nivel léxico-terminológico. Los sistemas de TA se usan con frecuencia para trabajar con textos especializados. El discurso especializado tiene por objetivo la transmisión del conocimiento, que en este tipo de textos se vehicula a través de los términos. Por tanto, el tratamiento de UTs es crucial para la evaluación de la calidad global de la traducción. Los términos son unidades del léxico que activan un significado preciso, es decir, su valor terminológico, cuando se usan en el contexto de la comunicación especializada (Cabré, 2003). Por un lado, en el caso de los términos, la complejidad del procesamiento automático se ve reducida, ya que presentan menor grado de ambigüedad y de variación que otras unidades del léxico. Por otro lado, el procesamiento de UTs supone dificultades adicionales para la TA, puesto que, además del conocimiento lingüístico, se requiere tomar en consideración el contexto del ámbito de especialidad. En palabras de Ghoniem et al. (2011: 115), "the translation of terminology presents a challenge to machine translation as it requires domain specific knowledge in addition to a linguistic mastery of both the source and the target languages".

El artículo está organizado de la siguiente manera. En la sección 2 se describe la metodología del estudio. En la sección 3 se ilustra el análisis realizado y se presentan los resultados. Finalmente, en la sección 4 se exponen las principales conclusiones, se discuten las limitaciones y las posibles líneas de trabajo futuro.

## 2. Metodología

Uno de los objetivos principales de la evaluación es explicar en qué y por qué la TA difiere de la TH. En la presente investigación, identificamos las características distintivas de la TA frente a la TH a partir de la observación de la distribución de UTs en tres grupos de textos: TH, TA de *Google Translate* y TA de *Lucy LT*. A continuación, relacionamos las características identificadas con el contexto dado por el TO y las estrategias empleadas por los traductores humanos y los sistemas de TA, es decir, intentamos identificar las fuentes de las diferencias detectadas. El estudio incluye las fases siguientes: selección de sistemas de TA, construcción del corpus de estudio, análisis cuantitativo de las diferencias en el tratamiento de la terminología, y clasificación de las diferencias detectadas en términos de *translation shifts* en la TH y errores de la TA.

### 2.1. Selección de sistemas de TA

En el campo del Procesamiento del Lenguaje Natural [PLN] existen dos enfoques generales: el simbólico y el empírico. El enfoque simbólico supone la codificación explícita del conocimiento lingüístico. En el caso de la TA se trata de sistemas que realizan la traducción con base en lexicones y gramáticas computacionales y se denominan sistemas basados en reglas (*Rule-Based Machine Translation*). En el marco del enfoque empírico las inferencias sobre el uso del lenguaje se realizan a partir de grandes cantidades de datos por medio de técnicas estadísticas. Los sistemas de TA que se basan en este enfoque se denominan sistemas estadísticos (*Statistical Machine Translation*). Una de las maneras más eficientes de procesar el lenguaje humano es a partir de una combinación de estrategias estadísticas y lingüísticas, tal como indica Dale (2000: 1): "[...] it is now widely recognized that the key to automatically processing human languages lies in the appropriate combination of symbolic and nonsymbolic techniques". Para saber de qué manera podemos realizar tal combinación es imprescindible

---

<sup>3</sup> <http://translate.google.es/>

<sup>4</sup> <http://www.lucysoftware.com/english/machine-translation/>

identificar las ventajas y las limitaciones de diferentes tipos de sistemas. Precisamente, para los fines del presente estudio seleccionamos un sistema basado en reglas y un sistema estadístico. De esta manera, es posible averiguar qué tipo de modificaciones realizadas por el traductor humano son capaces de reproducir los sistemas de TA basados en estrategias diferentes.

*Lucy LT* es un sistema de TA comercial creado por la empresa Lucy Software. El sistema está basado en el modelo de transferencia y realiza la traducción en tres etapas: análisis del TO, transferencia léxica y sintáctica, y generación del TT. Para el procesamiento del léxico *Lucy LT* cuenta con vocabularios monolingües en la LF y la LM, que se usan en las fases de análisis y de generación, y un vocabulario bilingüe para la fase de transferencia. Los vocabularios están organizados de acuerdo con el criterio temático: vocabulario general, vocabulario social general (que a su vez se divide en subtemas: arte, humanidades, etc.), vocabulario técnico general (que se divide en informática, medicina, etc.). Para traducir los textos de nuestro corpus, escogimos la opción vocabulario técnico general. El sistema resuelve la ambigüedad léxica por medio de pruebas de contexto en las que se toman en consideración las categorías gramaticales y las clases semánticas de las palabras adyacentes. Si no es capaz de resolver la ambigüedad, ofrece al usuario todas las traducciones posibles de la palabra ambigua.<sup>5</sup>

*Google Translate* es un sistema de TA desarrollado por el grupo de investigación de la empresa Google. *Google Translate* está basado en técnicas estadísticas y genera traducciones a partir de los patrones identificados en un gran número de textos disponibles en Internet. El sistema utiliza los modelos de TA estadística basados en frases. La calidad de la traducción es dependiente de lengua, ya que varía en función del tamaño del corpus de entrenamiento disponible para un par de lenguas determinado.

## 2.2. Construcción del corpus

Nuestro corpus de análisis está compuesto de textos que forman parte del Corpus Técnico del IULA (Institut Universitari de Lingüística Aplicada) (Cabré et al., 2006). Para los fines de esta investigación seleccionamos artículos de divulgación científica del ámbito médico publicados en la revista *Scientific American* y su versión en español *Investigación y Ciencia* entre los años 1994 y 2000. La Tabla 1 incluye las estadísticas del corpus.

	TO	TH	<i>Google Translate</i>	<i>Lucy LT</i>
nº de textos	10	10	10	10
nº de oraciones	1,134	1,156	1,134	1,134
nº de palabras	24,053	25,305	27,547	26,206

Tabla 1. Estadísticas del corpus

En la segunda columna se indica el número de textos, oraciones y palabras correspondientes a los artículos originales publicados en inglés en la revista *Scientific American*. En la tercera columna, se presentan los datos correspondientes a las THs de estos textos al español publicadas en *Investigación y Ciencia*. En la cuarta y en la quinta columnas, se ofrecen los datos correspondientes a las TAs inglés-español producidas por los sistemas *Google Translate* y *Lucy LT*, respectivamente.

<sup>5</sup> Para poder realizar el análisis comparativo necesitamos seleccionar una única opción entre las variantes que ofrece *Lucy LT*. Decidimos hacerlo con base en la comparación de las opciones que propone el sistema basado en reglas con la TA de *Google Translate*. Así, escogemos la opción más adecuada, si *Google Translate* también realiza una selección apropiada. De lo contrario, se selecciona la opción equivocada. En el siguiente ejemplo *Lucy LT* ofrece dos variantes para la traducción del término *treatments* (una expresión ambigua desde el punto de vista del sistema). Escogemos la opción adecuada - "tratamientos", ya que *Google Translate* ofrece esta misma opción.

TO: Most of these approaches have yet to pass even the most preliminary clinical tests demonstrating their overall safety and efficacy, but these ideas may lead to better cancer **treatments** in the future.

*Google Translate*: La mayoría de estos enfoques tienen aún que pasar aún las pruebas clínicas más preliminares que demuestran su seguridad y eficacia general, pero estas ideas pueden dar lugar o mejores **tratamientos** contra el cáncer en el futuro.

*Lucy LT*: La mayor parte de estos enfoques tiene que pasar todavía incluso las pruebas clínicas más preliminares que demuestran su seguridad global y eficacia, pero estas ideas pueden conducir a <A[**tratamientos**][tratos]>-de cáncer mejores en el futuro.

Cabe mencionar que el género de divulgación es interesante desde el punto de vista de la investigación en TA, puesto que en él se combinan los rasgos prototípicos de los textos de alto nivel de especialización, que tienden a ser sistemáticos y evitar la ambigüedad al máximo posible, y de los textos periodísticos, que presentan más diversidad en la selección del léxico.

Como sabemos, un texto especializado es aquella producción lingüística que sirve para expresar y transmitir conocimiento especializado, que tiene una serie de rasgos lingüísticos que le dan una especificidad en el conjunto de textos producidos en la lengua, y que, además, presenta una serie de características pragmáticas determinadas por los elementos específicos del proceso de comunicación (el tema, los interlocutores y la situación comunicativa) (Cabré, 2003).

Los artículos de divulgación pertenecen al discurso especializado, pero presentan ciertos rasgos lingüísticos particulares relacionados con el objetivo general de este tipo de textos: familiarizar al público no experto con los avances científicos. Esto supone una reformulación del mensaje original para una nueva audiencia. La particularidad del destinatario y de los objetivos de la divulgación científica se refleja en sus características lingüísticas en comparación con los artículos científicos. Concretamente, a nivel léxico, se produce una reducción del vocabulario de especialidad y su sustitución por términos de uso más extendido, así como la utilización de diversas estrategias para aclarar su significado.

*Scientific American* es una prestigiosa publicación mensual editada en Estados Unidos. Todos los artículos publicados en la revista están escritos por especialistas en el tema específico del artículo. *Investigación y Ciencia* es la versión española de *Scientific American*. La traducción de los artículos de *Scientific American* es encargada a un investigador experto en la materia. Fernández Polo (1999) realiza una encuesta entre los traductores que han colaborado regularmente con la revista, en la cual se les pide que valoren la importancia de una serie de factores relacionados con su actividad para descubrir sus prioridades a la hora de llevarla a cabo. De acuerdo con dicha encuesta, los traductores conciben las finalidades de su trabajo, en orden descendente de importancia, como muestra Fernández Polo (1999: 100):

- 1 Conservar toda la información del original
- 2 Velar por la legibilidad de la versión traducida
- 3 Emplear un léxico y sintaxis correctos
- 4 Emplear una terminología apropiada
- 5 Hacer que el texto suene como un original en castellano
- 6 Procurar que la traducción sea lo más amena posible

La importancia concedida por los traductores al componente informativo y la poca importancia que conceden al componente de entretenimiento se debe a que la versión en español prevé unos lectores distintos de los correspondientes a la versión inglesa. En la misma encuesta se sondeó la opinión de los traductores acerca del tipo de lectores que preveían para la revista, con el fin de comprobar si coincidían con el perfil de los lectores de la versión original. En orden descendente de importancia, los lectores prototípicos de *Investigación y Ciencia* serían, en opinión de los traductores encuestados (Fernández Polo, 1999: 101):

- 1 Investigadores o profesionales expertos (en el tema del artículo)
- 2 Estudiantes universitarios expertos
- 3 Investigadores o profesionales no expertos
- 4 Profesores de enseñanza secundaria
- 5 Estudiantes universitarios no expertos
- 6 Otros.

El mayor grado de especialización previsto para los receptores de los TTs, según esta escala, podría explicar algunos de los cambios que encontramos en las traducciones: en concreto, el cambio de nivel de

especialización que se manifiesta en la selección de UTs con un nivel de pertenencia al dominio más alto que el de los términos originales correspondientes.

### 2.3. Procedimientos del análisis

A continuación presentamos los procedimientos del análisis del tratamiento de la terminología en las TAs y la TH. En primer lugar, realizamos la extracción automática de las UTs del corpus. Para ello, empleamos la herramienta propuesta por Vivaldi y Rodríguez (2011), que obtiene términos de un dominio de especialidad utilizando las estructuras de páginas y categorías de Wikipedia. El extractor proporciona para cada Candidato a Término [CT]<sup>6</sup> un coeficiente de dominio [CD] que indica su grado de pertenencia al ámbito de especialidad. Dada la temática de los textos que conforman nuestro corpus, definimos "Medicina" y "Biología" como categorías representantes del dominio. Al realizar la extracción terminológica obtenemos tres listas de CTs: los de la TH, los de la TA de *Google Translate* y los de la TA de *Lucy LT*. Para cada CT el extractor proporciona la información siguiente: lema, patrón morfosintáctico y CD. Sometemos las listas de CTs a un proceso de filtrado manual en el cual: a) se eliminan todos los CTs con un CD menor de 0.5<sup>7</sup>; b) se eliminan las unidades cuyo significado identificado por el sistema no coincide con el significado que activan en el contexto (por ejemplo, "echar **mano** de", "tener en **mente**"). No realizamos ningún otro tipo de validación o corrección manual con respecto a estas listas para lograr la máxima sistematicidad en el análisis cuantitativo.

En segundo lugar, identificamos las características cuantitativas del uso de la terminología en los TTs. Una de las estrategias ampliamente utilizadas en los estudios de traducción basados en corpus para medir la variedad del vocabulario y comprobar si la traducción conlleva simplificación léxica es el *type-token ratio*, que se calcula de acuerdo con la fórmula siguiente:

$$\text{Type-token ratio} = \frac{n^{\circ} \text{ de types} \times 100}{n^{\circ} \text{ de tokens}} \quad (1)$$

El número de *types* es el número de palabras diferentes en el corpus. El número de *tokens* es el número total de palabras en el corpus. Esta medida se aplica, normalmente, a todas las palabras de los textos que se quiere caracterizar. En el presente estudio calculamos el *type-token ratio* de las listas de CTs extraídos automáticamente del corpus con el fin de proporcionar una caracterización cuantitativa de la variedad del vocabulario especializado en los TTs. Asimismo, calculamos la proporción de los CTs con los CDs altos (CD = 1) en cada grupo de textos para comprobar el grado de pertenencia al dominio de los términos usados por los traductores humanos y por los sistemas de TA.

En tercer lugar, calculamos el número total de diferencias TA-TH de la siguiente manera. Desarrollamos un programa, por medio del cual realizamos la extracción de todas las ocurrencias de los CTs junto con los números de las oraciones en las que aparecen. Después, detectamos las ocurrencias de los CTs que aparecen en la TH pero no se encuentran en las TAs, y las ocurrencias de los CTs que se presentan en las TAs pero no aparecen en la TH, calculando las proporciones correspondientes.

En cuarto y último lugar, seleccionamos una muestra aleatoria de 50 oraciones, en las que realizamos una alineación manual de los CTs identificados en los TTs y las unidades correspondientes de los TOs. En los casos en los que no era posible establecer la equivalencia a nivel del CT detectado por el extractor, la unidad de análisis se amplía. Por ejemplo, en la frase "agente etiológico del SIDA" el extractor asigna el CD a la unidad "agente etiológico". La unidad correspondiente en el original es *AIDS-causing agent*; por tanto, en este caso tomamos como unidad de análisis la frase "agente etiológico del SIDA". Al organizar de esta manera los CTs, clasificamos las diferencias TA-TH en términos de su origen especificando el tipo de modificaciones que realizan los traductores humanos y los sistemas de TA con

<sup>6</sup> Debido a que no existe un sistema de extracción automática de términos libre de errores, se considera que la salida de un extractor está conformada por CTs en lugar de UTs.

<sup>7</sup> Dado que en este trabajo no hemos realizado experimentos con diferentes umbrales, decidimos determinar el valor intermedio entre 0 y 1, que son los valores mínimo y máximo, respectivamente, que ofrece el extractor.

respecto a las unidades originales correspondientes (los CTs alineados y la clasificación de las diferencias TA-TH se ofrece en el Anexo 2).<sup>8</sup>

Dicha clasificación nos sirve para identificar las regularidades en el comportamiento de los traductores y de los sistemas de TA que contribuyen a la conformación de los rasgos distintivos de la TA detectados mediante el análisis cuantitativo. Para llevar a cabo esta tarea, nos basamos en la propuesta de clasificación de *translation shifts* desarrollada por van Leuven-Zwart (1989). En relación con la comparación TA-TH introducimos las modificaciones siguientes a la propuesta original de esta autora:

1. La propuesta de van Leuven-Zwart es demasiado detallada para implementarla en el presente estudio, ya que la cantidad de oraciones que analizamos de manera manual no es grande, y tener una clasificación detallada no nos permitiría llegar a generalizaciones a partir de los datos. Por esta razón, tomamos en cuenta únicamente las categorías generales de la clasificación de van Leuven-Zwart: modulación (generalización vs. especificación), modificación y mutación (omisión vs. adición).
2. Para las categorías "modulación/generalización" y "modulación/especificación" de van Leuven-Zwart usamos los términos explicitación e implícitación, respectivamente, y los empleamos en un sentido amplio, siguiendo a Becher (2011: 18), quien define estas categorías de la manera siguiente:

Implicitness is the non-verbalization of information that the addressee might be able to infer.  
Explicitness is the verbalization of information that the addressee might be able to infer if it were not verbalized.  
Explicitation is observed where a given target text is more explicit than the corresponding source text.  
Implication is observed where a given target text is less explicit (more implicit) than the corresponding source text.

3. A diferencia de van Leuven-Zwart (1989), tomamos en cuenta no solamente los *translation shifts* opcionales, sino también aquellos que son obligatorios debido a las diferencias sistémicas entre la LF y la LM. Consideramos que esta distinción es relevante, puesto que la agramaticalidad de las oraciones generadas por los sistemas de TA se relaciona frecuentemente con la falta de modificaciones obligatorias.
4. La unidad de análisis en el estudio de van Leuven-Zwart (1989) es el transema<sup>9</sup>. En la presente investigación, realizamos la clasificación únicamente con respecto a las UTs; es decir, hablamos de explicitación, implícitación o modificación del significado léxico codificado por medio de las UTs. Siguiendo a van Leuven-Zwart (1988), distinguimos entre explicitación/implícitación como uso de expresiones más o menos informativas y adición/omisión de las unidades de análisis.
5. Introducimos dos categorías adicionales relacionadas con la naturaleza de los textos analizados y con los propósitos de la investigación. Por un lado, utilizamos la categoría "alteración" para los casos en los que el significado de la unidad de análisis del original es alterado a causa de los errores de la TA. Por otro lado, anotamos como "variación" aquellos casos en los que el humano y la máquina ofrecen variantes, que, de acuerdo con los recursos de referencia que usamos, son posibles equivalentes del término.<sup>10</sup> Así, la clasificación se lleva a cabo a partir de la taxonomía siguiente: explicitación, implícitación, modificación, variación, omisión, adición y alteración.

<sup>8</sup> Para realizar el análisis comparativo y comprobar el estatus terminológico de las unidades de análisis utilizamos diversos recursos terminológicos: diccionarios de lengua general, diccionarios de especialidad, ontologías y bases de datos terminológicas, entre ellos:

Oxford Advanced Learner's Dictionary [OALD] (Hornby, 2011) - <http://oald8.oxfordlearnersdictionaries.com>

The Dictionary of cell biology (Lackie y Dow, 1995)

Diccionario de la lengua española [DRAE] (Real Academia Española, 2001) - <http://lema.rae.es/drae>

Diccionario de bioquímica y biología molecular (Widmer y Beffa, 2000)

Base de datos terminológica InterActive Terminology for Europe [IATE] (European Commission, Translation Centre, 1999) - <http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load>

<sup>9</sup> Van Leuven-Zwart delimita las unidades de análisis, que denomina transemas, a partir de los criterios derivados de la Gramática Funcional de Dik (1978). De acuerdo con estos criterios, los transemas pueden ser de dos tipos: a) *state of affair transeme* (predicado y sus argumentos); b) *satellite transeme* (especificación o amplificación adverbial de los transemas del primer tipo).

<sup>10</sup> En la terminología tradicional se niega la existencia de sinonimia en textos especializados y se postula la biunivocidad de los términos, lo cual implica que a un concepto, a modo de etiqueta, le corresponde una única denominación. En cambio, los planteamientos más recientes en el ámbito de la terminología (Cabré, 1999, 2003) consideran la variación como un hecho real en la comunicación especializada. Freixa (2002: 54) define la variación denominativa como "el fenómeno por el cual a una misma noción le corresponden diversas denominaciones". La noción de variación denominativa es relevante para el análisis de la traducción especializada y para la evaluación de las traducciones, ya que permite, dada una UT del TO, delimitar el grado de variación

### 3. Análisis y resultados

A continuación presentamos los resultados del análisis cuantitativo y cualitativo, y ofrecemos algunos ejemplos ilustrativos. La Tabla 2 incluye las características cuantitativas del tratamiento de la terminología en la TH y las TAs.

	CTs			<i>Type-token ratio</i>	Diferencias TA-TH
	<i>Types</i>	<i>Tokens</i>	CD = 1		
TH	200	1,528	27%	13%	
<i>Google Translate</i>	183	1,635	22%	11%	17%
<i>Lucy LT</i>	136	1,612	24%	8%	26%

Tabla 2. Características cuantitativas del tratamiento de la terminología en la TH y las TAs de *Google Translate* y *Lucy LT*

En la segunda y la tercera columnas de la Tabla 2 se indica el número total de tipos (*types*) y de ocurrencias (*tokens*) de CTs. En la cuarta columna se presenta el porcentaje de tipos de CTs con el grado de pertenencia al dominio alto (CD = 1). En la quinta columna se indica el *type-token ratio*. En la sexta columna se muestra el número total de diferencias TH-TA de *Google Translate* y TH-TA de *Lucy LT*.

El número de tipos de CTs identificados por el extractor terminológico en la TH es mayor que en las TAs. Al procesar los términos originales, los sistemas de TA ofrecen variantes que no pertenecen al dominio de especialidad a causa de problemas de ambigüedad o falta de cobertura del vocabulario. Además, el extractor de Vivaldi y Rodríguez (2011) sólo detecta las unidades con función referencial y no asigna ningún coeficiente a los verbos, de manera que el número elevado de nominalizaciones en la TH tiene un impacto en los resultados. Por último, en la TH se observa una tendencia a cambiar el nivel de especialización, al usarse UTs incluso en aquellos casos en los que en los TOs se utilizan términos de uso más extendido (de ahí que el número de CTs con el CD más alto sea mayor en la TH, tal como lo indica la cuarta columna de la Tabla 2).

El número de CTs extraídos de la TA de *Google Translate* es mayor que el número de CTs identificados en la TA de *Lucy LT*. Al mismo tiempo, el número de diferencias TA-TH es mayor en el caso del sistema basado en reglas. Ello se debe a que, dada la naturaleza de los sistemas estadísticos, las traducciones de *Google Translate* se acercan más a la selección léxica de la TH. El número total de las diferencias confirma la hipótesis ya comprobada en otros trabajos (Coughlin, 2003) de que los sistemas de TA estadística superan los sistemas basados en reglas en cuanto a la selección del léxico.

El *type-token ratio* es más alto en la TH debido al uso de sinónimos, hipónimos e hiperónimos para la transmisión del mismo significado original. La Tabla 3 ilustra las diferencias en el grado de la variedad del vocabulario entre las TAs y la TH.

---

aceptable en el TT. Si las UTs que ofrecen el sistema de TA y el traductor humano son variantes denominativas, la diferencia no debe penalizarse en la evaluación.



nº	TO	TH (CD)	Google Translate (CD)	Lucy LT (CD)
909	stroke	accidente cerebrovascular (0.65)	carrera (0.00)	golpe (0.00)
911	stroke	accidente cerebro vascular (0.65)	derrame cerebral (0.62)	golpe (0.00)
915	stroke	ictus (0.62)	accidente cerebrovascular (0.65)	golpe (0.00)
852	seizure	convulsión (1.00)	convulsión (1.00)	toma (-1.00)
911	seizure	convulsión (1.00)	embargo (0.00)	toma (-1.00)
915	seizure	convulsión (1.00)	ataque (0.85)	toma (-1.00)

Tabla 3. CTs extraídos de los TTs y las unidades correspondientes de los TOs

En la primera columna registramos el número de la oración. En la columna "TH" encontramos los CTs extraídos de la TH. En las columnas "TO", "Google Translate" y "Lucy LT" se registran las unidades correspondientes de los TOs y las TAs. Para cada unidad objeto de análisis se registra la información sobre su CD. Estos datos nos muestran cómo *Lucy LT*, aun realizando una selección poco apropiada, al ofrecer "golpe" como equivalente de *stroke*, y "toma" como equivalente de *seizure*, lo hace siempre de la misma manera, de acuerdo con el diccionario bilingüe que utiliza. En cambio, *Google Translate* realiza la traducción en función del contexto lingüístico inmediato y ofrece varias opciones para el mismo término ("derrame cerebral", "accidente cerebrovascular" y "carrera" para *stroke*, y "convulsión", "embargo" y "ataque" para *seizure*), algunas de las cuales son adecuadas, mientras que otras son completamente erróneas, con lo cual el TT pierde coherencia a nivel del léxico (véase Rogers, 2008, para una discusión sobre la importancia de este aspecto en la traducción de textos especializados).

Ahora bien, con respecto a la naturaleza de las diferencias detectadas, a partir del análisis de una muestra aleatoria de 50 oraciones de nuestro corpus, obtenemos los resultados incluidos en la Tabla 4.

Translation shift	TH	Google Translate	Lucy LT
Adición	3 (6%)	0 (0%)	0 (0%)
Omisión	2 (4%)	0 (0%)	0 (0%)
Explicitación	8 (15%)	2 (8%)	0 (0%)
Implicitación	8 (15%)	0 (0%)	0 (0%)
Modificación	17 (32%)	5 (21%)	0 (0%)
Variación	15 (28%)	10 (42%)	11 (31%)
Alteración	0 (0%)	7 (29%)	25 (69%)
Total	53 (100%)	24 (100%)	36 (100%)

Tabla 4. Clasificación de las diferencias TH-TA de *Google Translate* y TH-TA de *Lucy LT*

En primer lugar, la Tabla 4 indica que el número de diferencias que se originan en la TH es mayor que el número de diferencias procedentes de las TAs. Ello se debe, por un lado, a que a nivel de la UT la calidad de la TA no es baja y, por otro lado, a que en la TH se producen numerosos *translation shifts* con respecto a las UTs de los originales.

En segundo lugar, tanto el número total de diferencias como el número de alteraciones es mayor en el caso de *Lucy LT*. Ello se explica por el hecho de que, como se ha mencionado anteriormente, los sistemas estadísticos superan a los sistemas basados en reglas en cuanto a la selección del léxico.

En tercer lugar, al registrar la naturaleza obligatoria u opcional de los *translation shifts* (véase el Anexo 2) observamos que el número de *shifts* opcionales es mayor que el número de *shifts* obligatorios,

tanto en la TH como en la TA. Suponemos que esto se debe a que, en la estructura conceptual del ámbito médico, las divergencias léxicas entre el inglés y el español que impliquen una necesidad de modificación obligatoria se presentan con poca frecuencia.

Veamos algunos ejemplos. El *translation shift* más frecuente en la TH es la **modificación**. Al analizar casos concretos (véase ejemplo 1) observamos que este procedimiento está relacionado con los factores de uso.

1. TO: The tiny cells in our **bodies** harbor amazing internal communication networks.  
TH: Las células de nuestro **organismo** contienen unas redes de comunicación interna sorprendentes.  
*Google Translate*: Las pequeñas células en nuestros **cuerpos** albergan increíbles redes de comunicación interna.  
*Lucy LT*: Las células diminutas en nuestros **cuerpos** albergan cadenas de comunicación internas asombrosas.

Suponemos que, en los textos especializados, la UT "célula" tiene mayor fuerza asociativa al relacionarse con el término "organismo" que con el término "cuerpo" (que en este caso actúan como sinónimos contextuales). Por ello, el traductor se permite la licencia de ofrecer la UT "organismo" como equivalente de *body*.

Asimismo, la **modificación** se produce en el contexto de las diferencias sintácticas entre el español y el inglés. Ilustramos esta observación con el ejemplo 2.

2. TO: On a simplistic level, neurobiologists associate the brainstem with the most basic functions: **breathing, eating, balance, motor coordination** and so forth.  
TH: Solemos asociar el tronco cerebral con funciones básicas, desde la **respiración** hasta la **deglución**, pasando por el **equilibrio**, la **coordinación motora** y otros.  
*Google Translate*: En un nivel simple, los neurobiólogos asocian el tronco cerebral con las funciones más básicas: **respirar, comer, el equilibrio, la coordinación motora** y así sucesivamente.  
*Lucy LT*: En un nivel simplista, neurobiologists asocian el brainstem con las funciones más básicas: **respirando, comiendo, equilibrio, coordinación de motor** y así sucesivamente.

Las unidades léxicas *breathing, eating, balance* y *motor coordination* son UTs, ya que hacen referencia a las "funciones básicas" del organismo. De la misma manera que en el ejemplo anterior, en el original se utilizan expresiones de uso más generalizado, mientras que en la TH se observa un incremento en el grado de especialización (*eating* > "deglución"). No obstante, puede ser que en este caso la decisión del traductor no esté relacionada con una tendencia hacia el cambio de nivel de especialización, sino con el aspecto estilístico: los términos de una enumeración tienen que ser similares en cuanto a su función sintáctica y su significado.

En cuanto a la TA, *Lucy LT* produce una oración agramatical al no realizar la modificación obligatoria (al gerundio nominal del inglés le corresponde en español el infinitivo o el nombre deverbal) debido a la ambigüedad de las formas en *-ing* del inglés. Por su parte, *Google Translate* selecciona una opción estilísticamente imperfecta: "respirar, comer, el equilibrio, la coordinación motora".

Asimismo, este ejemplo ilustra uno de los problemas de la TA en el ámbito de la terminología, la traducción de los sintagmas nominales complejos del inglés al español (Maxwell, 1992). Esta tarea resulta problemática para el sistema basado en reglas, debido a que en inglés no hay marcas explícitas que permitan determinar la categoría gramatical del modificador en la frase *motor coordination* (ambigüedad categorial). *Lucy LT* asigna a esta frase el patrón "Nombre + Nombre" y, de acuerdo con las reglas sintácticas del módulo de transferencia, la traduce por defecto como "Nombre + Frase Preposicional", cuando en realidad en este caso *motor* es un adjetivo que significa, de acuerdo con *Oxford Advanced Learner's Dictionary* [OALD] (Hornby, 2011), "connected with movement of the body that is produced by muscles; connected with the nerves that control movement".

En el ejemplo 2 se presenta otra alteración típica del sistema de TA basado en reglas, que se debe a la falta de cobertura de su vocabulario. Los términos *neurobiologists* o *brainstem* no se encuentran en el diccionario del sistema y, por ello, *Lucy LT* no es capaz de encontrar sus equivalentes en la LM. Hay que mencionar que en otros ejemplos de nuestro corpus en los que el término *brain stem* está escrito por separado, *Lucy LT* propone una traducción adecuada, "tallo de cerebro". Así, una de las limitaciones de la TA basada en reglas consiste en que no es robusta y presenta errores de traducción si en el original hay erratas o variantes ortográficas que no aparecen en el diccionario del sistema.

La siguiente categoría más frecuente en la TH es la **variación**. Las diferencias TA-TH que pueden caracterizarse como variación denominativa con frecuencia están relacionadas con el proceso de adaptación al dominio en la TH, la cual se manifiesta en una preferencia por el uso de UTs con mayor grado de especialización. Por ejemplo:

3. TO: **Blood-forming stem cells**, for example, give rise to every other type of blood cell (red cells, white cells of the immune system, and so on) and reconstitute the blood as needed; they also make more copies of themselves.

TH: Las **células hematopoyéticas**, por ejemplo, originan todos los tipos celulares de la sangre (eritrocitos, leucocitos del sistema inmunitario, etc.) y reconstituyen la sangre cuando es necesario; también producen copias de sí mismas.

*Google Translate*: **Que forman la sangre células madre**, por ejemplo, dar lugar a cualquier otro tipo de células sanguíneas (glóbulos rojos, glóbulos blancos del sistema inmunológico, etc.) y la reconstitución de la sangre como sea necesario, sino que también hacen más copias de sí mismos.

*Lucy LT*: **Formando de sangre las células indiferenciadas**, por ejemplo, dan lugar a cada dos tipos de célula sanguínea (células rojas, células blancas del sistema inmunitario, y así sucesivamente) y rehidratan la sangre como necesitada; también hacen más copias de ellos mismos.

OALD define *stem cell* como "a basic type of cell which can divide and develop into cells with particular functions". De acuerdo con la base de datos terminológica InterActive Terminology for Europe [IATE] (European Commission, Translation Centre, 1999), el equivalente de este término en español es "célula madre" (cf. la definición del término "célula madre" que ofrece el Diccionario de la lengua española de la Real Academia Española [DRAE] (Real Academia Española, 2001): "célula indiferenciada que puede dar lugar a distintos tipos de tejidos, como los constituidos por células hepáticas, nerviosas, epiteliales o a las diversas estirpes de células sanguíneas"). En la terminología médica en inglés existe el término *hematopoietic stem cell*, que significa "cell that can develop into any type of specialized blood cell" (OALD). Dado el carácter divulgativo de los textos, en el original se usa una variante más transparente, más accesible para el lector. En cambio, en la TH se emplea el término "células hematopoyéticas", lo cual conlleva, por un lado, la implicación de la información que concierne a las características de las células madre, y, por otro, el aumento en el grado de pertenencia al dominio de especialidad.

Ahora bien, antes de relacionar este caso con la adaptación al dominio como una tendencia universal de la traducción, habría que pensar en las diferencias lingüísticas entre el inglés y el español que, probablemente, condicionan el *translation shift* realizado por el traductor humano. El empleo del participio presente con función adjetival es usual en inglés, pero el gerundio en español no posee esta función; por tanto, para traducir el modificador *blood-forming*, el traductor se vería obligado a recurrir a otros recursos de modificación, por ejemplo, introducir una cláusula subordinada de relativo (una opción menos concisa). Precisamente, *Google Translate* realiza esta modificación obligatoria pero no logra cambiar el orden de palabras. Por su parte, *Lucy LT* ofrece una traducción literal de la estructura sintáctica del inglés, que resulta en una oración agramatical.

El cambio de nivel de especialización en la TH se refleja con más claridad en el tratamiento de los términos *red cells* y *white cells*. En este caso no hay ninguna explicación tipológica que dé cuenta de la selección del traductor. En inglés existen las UTs *erythrocyte* y *leukocyte*, pero, al igual que en el caso anterior, en el original se opta por un término más accesible para el lector. En cambio, el traductor prefiere usar las UTs con un grado de pertenencia al dominio mayor, debido, posiblemente, al tipo de

lectores a quienes va dirigida la versión española de la revista *Scientific American*. Los sistemas de TA proponen variantes diferentes. *Google Translate* selecciona una opción adecuada, más frecuente en el contexto de los textos especializados ("glóbulos rojos" y "glóbulos blancos"). *Lucy LT* ofrece una traducción literal ("células rojas" y "células blancas") que en este caso tampoco es inaceptable.

La **implicitación** (y otros tipos de *translation shifts*) en la TH se produce con frecuencia cuando una traducción literal conlleva la repetición léxica. En el ejemplo 4, a la segunda ocurrencia de la UT *cancer* le corresponde en la TH el término "patología". Cáncer es un tipo de patología, de manera que en este caso el traductor usa el hiperónimo para evitar la repetición léxica presente en el original.

4. TO: They are also finding that microsatellites change in length early in the development of some **cancers**, making them useful markers for early **cancer** detection.

TH: Se está comprobando que los microsatélites cambian de longitud en fases precoces de ciertos **cánceres**, lo que les convierte en valiosos marcadores para el diagnóstico precoz de tales **patologías**.  
*Google Translate*: Ellos también están encontrando que el cambio microsatélites de longitud temprana en el desarrollo de algunos tipos de **cáncer**, lo que los marcadores útiles para la detección temprana del **cáncer**.

*Lucy LT*: Están encontrando también que los microsatélites cambian en longitud temprano en el desarrollo de algunos **cánceres**, haciéndoles marcadores útiles para la primera detección de **cáncer**.

Asimismo, la **implicitación** ocurre en aquellos casos en los que ciertos rasgos del significado de la UT original pueden inferirse con facilidad a partir del contexto lingüístico inmediato. En el ejemplo 5 observamos que el traductor ofrece una expresión más concisa para la UT *phagocytic cells*.

5. TO: At other times, it is strategically more advantageous for the bacterium not to interact with host cells - particularly **phagocytic cells**, which engulf and destroy bacteria.

TH: En otras ocasiones le resulta a ésta más ventajoso no interactuar con la célula huésped; en particular si se trata de **fagocitos**, que destruyen las bacterias tras atraparlas en su interior.

*Google Translate*: En otras ocasiones, es estratégicamente más ventajoso para la bacteria que no interactúan con las células huésped, particularmente las **células fagocíticas**, que se tragan y destruyen las bacterias.

*Lucy LT*: En otros tiempos, es estratégicamente más ventajoso que la bacteria no interactúe con células de anfitrión - **células especialmente fagocitarias**, que inundan y destruyen bacterias.

En cuanto a la **omisión**, uno de los contextos en los que este *translation shift* se produce con regularidad es la traducción de las UTs originales que designan a los investigadores. En numerosas ocasiones los términos que hacen referencia a los participantes de la investigación quedan implícitos en la TH. Se omiten dejando lugar al uso del verbo en primera persona del plural, lo cual podría explicarse por el carácter divulgativo de los textos, en los que es importante destacar el protagonismo de los investigadores. Por ejemplo:

6. TO: From this population of mutagenized bacteria or yeast, the **geneticists** can identify individuals not capable of replicating their DNA.

TH: A partir de esta población de bacterias o levaduras que han experimentado mutagénesis, **identificaremos** individuos incapaces de replicar su ADN.

*Google Translate*: De esta población de bacterias o levaduras mutadas, los **genetistas** pueden identificar a los individuos que no son capaces de replicar su ADN.

*Lucy LT*: De esta población de bacterias de mutagenized o levadura, los **genetistas** pueden identificar individuos no competentes de reproducir su ADN.

El *translation shift* de **explicitación** se relaciona con la preferencia de los traductores por los términos con mayor grado de pertenencia al dominio, más específicos. Así, en el ejemplo 7 a la frase *AIDS-causing agent* le corresponde en la TH la unidad "agente etiológico del sida":

7. TO: Early clinical experiments using this strategy are now under way in cancer patients, as well as in those infected with HIV, the **AIDS-causing agent**, and other pathogens.

TH: Y se han puesto en marcha los primeros ensayos clínicos que emplean este método en el tratamiento no sólo del cáncer, sino también en infectados por el VIH, el **agente etiológico del SIDA**, y por otros patógenos.

*Google Translate*: Los primeros experimentos clínicos utilizando esta estrategia están en marcha en pacientes con cáncer, así como en las personas infectadas con el VIH, el **agente causante del SIDA**, y otros agentes patógenos.

*Lucy LT*: Los primeros experimentos clínicos que utilizan esta estrategia están ahora en curso en pacientes de cáncer, así como en esos infectados con HIV, el **agente que causa el SIDA**, y otros pathogens.

De acuerdo con el DRAE, "etiológico" significa "perteneciente o relativo a la etiología", y "etiología" se refiere a "1. f. Fil. Estudio sobre las causas de las cosas. 2. f. Med. Estudio de las causas de las enfermedades. 3. f. Med. Estas causas", siendo esta última la acepción con la que se emplea el término en la TH. Así, el traductor utiliza una unidad léxica más informativa y en cierto sentido redundante (etiológico <causante de enfermedades> del SIDA).

La **adición** está condicionada por los factores pragmáticos. En el ejemplo 8, el traductor, además de dar el equivalente del término en español, ofrece, entre paréntesis, la expresión correspondiente en inglés con el fin de proporcionar al lector la información adicional sobre el término original.

8. TO: Walter Schaffner and Steven Lanier McKnight, among others, had additionally identified an unusual set of regulatory elements called **enhancers**, which facilitate transcription.

TH: Walter Schaffner, Steven Lanier McKnight y otros identificaron, además, unos elementos reguladores nuevos, los **intensificadores (enhancers)**, que estimulan la transcripción.

*Google Translate*: Walter Schaffner y Steven McKnight Lanier, entre otros, han identificado, además, un inusual conjunto de elementos reguladores llamados **potenciadores**, que facilitan la transcripción.

*Lucy LT*: Walter Schaffner y Steven Lanier McKnight, entre otros, habían identificado adicionalmente un conjunto inusual de elementos legales llamados **potenciadores**, que facilitan transcripción.

Tal como observamos en la Tabla 4 en las traducciones de *Google Translate* (pero no en las de *Lucy LT*) también se realizan *translation shifts* que en ocasiones coinciden con la TH. En el ejemplo 9, tanto en la TH como en la TA de *Google Translate* se usa el término "cicatrización". Esto supone una explicitación con respecto al significado léxico de la UT *healing*, ya que, de acuerdo con el OALD, el verbo *heal* significa "to become healthy again; to make something healthy again", y el término "cicatrización" en español significa "completar la curación de las llagas o heridas, hasta que queden bien cerradas" (DRAE). En la TA de *Lucy LT* el significado de la UT *healing* es alterado, debido a la ambigüedad de las formas en *-ing* del inglés, que ya hemos discutido anteriormente.

9. TO: These reactions ultimately stimulate proteins in the nucleus to activate genes that cause the cells to divide, an action that promotes wound **healing**.

TH: Ras pone en marcha una serie de procesos enzimáticos que terminan por estimular proteínas del núcleo que activan genes promotores de la división celular, lo que redundante en la **cicatrización** de la herida.

*Google Translate*: Estas reacciones en última instancia, estimular las proteínas en el núcleo para activar los genes que causan que las células se dividan, una acción que promueve la **cicatrización** de heridas.

*Lucy LT*: Estas reacciones en el fondo estimulan proteínas en el núcleo activar genes que causan las células para dividirse, una acción que promueve herida **que se cura**.

Como hemos visto en los ejemplos anteriores, la alteración en la TA de *Lucy LT* se produce en el contexto de ambigüedad léxica o estructural, o a causa de falta de cobertura del vocabulario. En cuanto a

la TA de *Google Translate*, los errores de los sistemas estadísticos son difíciles de explicar a partir de un análisis lingüístico. Lo que observamos es que se manifiestan con frecuencia en una ordenación inadecuada de los elementos. En el ejemplo 10 (véase también el ejemplo 3), *Google Translate*, al procesar el sintagma terminológico *blood type*, no cambia el orden de las palabras, con lo cual se altera el significado del término original.

10. TO: Although many other genes appear in several forms - for example, the genes that encode eye color or **blood type** - highly conserved genes are not commonly found in multiple versions (also known as polymorphic alleles, or allelic variants).

TH: Asimismo, muchos otros genes aparecen en formas diversas; por ejemplo, los que determinan el color de los ojos o el **grupo sanguíneo**. No ocurre así con los genes muy conservados, de los que no suele haber alelos polimórficos, o variantes alélicas.

*Google Translate*: Aunque muchos otros genes aparecen en varias formas, por ejemplo, los genes que codifican el color de ojos o la **sangre de tipo** altamente conservadas genes no se encuentran comúnmente en varias versiones (también conocido como alelos polimórficos, o variantes alélicas).

*Lucy LT*: Aunque muchos otros genes aparecen en varias formas - por ejemplo los genes que codifican color de ojo o **tipo de sangre** - los genes altamente conservados no se encuentran comúnmente en versiones múltiples (también conocido como alelos polimorfos, o alelomorfos).

#### 4. Conclusiones

En este artículo hemos presentado una metodología para el análisis comparativo de TAs y THs a nivel léxico-terminológico. El objetivo general del trabajo era estudiar las diferencias lingüísticas sistemáticas entre la TA y la TH en el contexto de la evaluación automática de sistemas.

Entre los resultados relevantes de este estudio, encontramos que las diferencias asociadas a la TH son más frecuentes que las diferencias relacionadas con la TA, debido a que a nivel de la UT la traducción que ofrecen los sistemas es cercana al original, mientras que en la TH se producen modificaciones opcionales que están condicionadas por factores estilísticos o pragmáticos y no por las divergencias léxicas entre los sistemas de LF y LM. Los traductores humanos adaptan los TOs a las expectativas del receptor y a las convenciones de uso de la LM, lo que en el caso de nuestro corpus resulta en un cambio de nivel de especialización.

La TA de *Google Translate* se asemeja más a la TH que la TA de *Lucy LT*, puesto que el sistema estadístico es capaz de reproducir aquellas modificaciones que los traductores realizan con regularidad en un contexto lingüístico determinado. En cambio, la selección léxica en la TA de *Lucy LT* está limitada por la información presente en los diccionarios del sistema.

Con respecto a las implicaciones del análisis realizado para la evaluación de los sistemas de TA, observamos que las diferencias TA-TH relacionadas con las modificaciones opcionales en la TH y las diferencias que se deben a la falta de modificaciones obligatorias en la TA no tienen la misma relevancia para evaluar la calidad de esta última.

Somos conscientes de que el presente trabajo tiene varias limitaciones. La representación de datos, las herramientas utilizadas y el corpus de estudio limitan las posibilidades de observación. Así, los resultados del análisis cuantitativo con la medida *type-token ratio* se ven afectados por los errores del extractor terminológico y el tamaño reducido de las listas de CTs.

Como trabajo futuro, primero, realizaremos experimentos con sistemas de TA entrenados para la traducción de textos especializados del ámbito médico. Segundo, compararemos las THs con textos originalmente escritos en la LM, lo que aportará datos interesantes sobre las características distintivas de los TTs. Tercero, llevaremos a cabo experimentos con otros umbrales para el CD del extractor empleado, así como con otros sistemas de extracción terminológica. Cuarto, aplicaremos la metodología a otro tipo de textos (por ejemplo, textos de alto nivel de especialización) y la automatizaremos el máximo posible. Finalmente, usaremos la metodología para otros pares de lenguas, ya que consideramos que ésta es

independiente de lengua; el único recurso de PLN necesario para ello es una herramienta de extracción terminológica para la LM.

## 5. Referencias bibliográficas

Baker, Mona. Corpus Linguistics and Translation Studies: Implications and Applications. En M. Baker, G. Francis y E. Tognini-Bonelli (eds.), *Text and Technology. In Honour of John Sinclair*. Amsterdam: Benjamins, 1993, 233–250.

Becher, Viktor. *Explicitation and implicitation in translation*. Hamburgo: Universidad de Hamburgo [Tesis Doctoral], 2011.

Cabré, M. Teresa. *La Terminología. Representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 1999.

Cabré, M. Teresa. Theories of terminology. Their description, prescription and explanation. *Terminology*, 2003, 9(2): 163–200.

Cabré, M. Teresa; Bach, Carme; Vivaldi, Jorge. *10 anys del Corpus de l'IULA*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 2006.

Coughlin, Deborah. Correlating Automated and Human Assessments of Machine Translation Quality. En *Proceedings of Machine Translation Summit IX*, 2003, 23–27.

Cyrus, Lea. Old concepts, new ideas: approaches to translation shifts. *MonTi. Monografías de Traducción e Interpretación*, 2009, 1: 87–106.

Dale, Robert. Symbolic Approaches to Natural Language Processing. En D. Robert, M. Hermann y Somers, H. L. (eds), *Handbook of Natural Language Processing*. New York: M. Dekker, 2000, 1–9.

Dik, Simon. *Functional Grammar*. Amsterdam: North-Holland, 1978.

European Commission, Translation Centre. *IATE — Services for the Development of an Interactive Terminology Database System*. Open Call for Tenders DGIII/99/050-IDA-101.02/01/IATE1, Luxemburg/Brussels, 1999.

Fernández Polo, Francisco. *Traducción y retórica contrastiva: A propósito de la traducción de textos de divulgación científica del inglés al español*. Santiago de Compostela: Servicio de Publicaciones da Universidade de Santiago de Compostela, 1999.

Fomicheva, Marina. *Análisis lingüístico de la traducción automática para su evaluación*. México, D.F.: Universidad Nacional Autónoma de México [Tesis de Máster], 2012.

Freixa, Judit. *La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra [Tesis Doctoral], 2002.

Ghoniem, Mohammad; Tokal, Khaled; Tawfik, Ahmed. An Analysis and Evaluation of English Arabic Statistical Machine Translation of Terminology-Rich Text. *Short papers of the 9th International Conference on Terminology and Artificial Intelligence*, 2011, 115–118.

Hornby, Albert. *Oxford Advanced Learner's Dictionary*. 2011. Disponible en: <http://oald8.oxfordlearnersdictionaries.com>.

- Lackie, John; Dow, Julian (eds.). *The Dictionary of cell biology*. London: Academic, 1995.
- van Leuven-Zwart, K. M. Translation and original: Similarities and dissimilarities. *Target*, 1989, 1(2): 151–181.
- Maxwell, Kerry. Automatic Translation of English Compounds: Problems and Prospects. En *Studies in MT and NLP*, 1992, 8: 37–56.
- Papineni, Kishore; Roukos, Salim; Ward, Todd; Zhu, Wei-Jing. *Bleu: a method for automatic evaluation of machine translation*. RC22176 (Technical Report), IBM T.J. Watson Research Center, 2001.
- Puurtinen, Tiina. Genre-specific Features of Translationese? Linguistic Differences between Translated and Non-translated Finnish Children's Literature. *Literary and Linguistic Computing*, 2003, 18(4): 389–406.
- Real Academia Española. *Diccionario de la lengua española, 22ª edición*. 2001. Disponible en: <http://lema.rae.es/drae/>
- Rogers, Margaret. Terminological equivalence: probability and consistency in technical translation. En H. Gerzymisch-Arbogast, G. Budin y G. Hofer (eds.), *LSP Translation Scenarios. Selected Contributions to the EU Marie Curie Conference Vienna; MuTra Journal* 2008, 2: 101–108.
- Vivaldi, Jorge; Rodriguez, Horacio. Extracting terminology from Wikipedia. *Procesamiento del lenguaje natural*, 2011, 47: 65–73.
- Widmer, François; Beffa, Roland. *Diccionario de bioquímica y biología molecular* [Traducido por M. J. Arrizubieta Balerdi]. Zaragoza: Acribia, 2000.



**ANEXO 1. Lista de abreviaciones**

CD	Coeficiente de Dominio
CT	Candidato a Término
DRAE	Diccionario de la Real Academia Española
IATE	<i>InterActive Terminology for Europe</i>
LF	Lengua Fuente
LM	Lengua Meta
OALD	<i>Oxford Advanced Learner's Dictionary</i>
PLN	Procesamiento del Lenguaje Natural
TA	Traducción Automática
TH	Traducción Humana
THR	Traducción Humana de Referencia
TO	Texto Original
TT	Texto Traducido
UT	Unidad Terminológica

**ANEXO 2. Clasificación de las diferencias entre la TH y las TAs de *Google Translate* y de *Lucy LT***

Oración	TO	TH	Shift TO-TH	<i>GoogleTranslate</i>	Shift TO- <i>Google Translate</i>	<i>Lucy LT</i>	Shift TO- <i>Lucy LT</i>
383	0	patogénesis	adición (obligatoria)	0	alteración	0	alteración
837	AIDS-causing agent	agente etiológico de el sida	explicitación (opcional)	agente causante de el sida	-	agente que causar el sida	-
159	animal	ratón mutante	explicitación (opcional)	animal	-	animal	-
898	animal	roedor	explicitación (opcional)	animal	-	animal	-
441	animal	individuo manipulado	modificación (opcional)	animal	-	animal	-
394	biomedical researcher	biomedicina	modificación (opcional)	investigador biomédico	-	investigador biomédico	-
1060	blood type	grupo sanguíneo	variación	sangre de tipo	alteración	tipo de sangre	variación
267	body	cuerpo humano	explicitación (opcional)	cuerpo	-	cuerpo	-
2	body	organismo	modificación (opcional)	cuerpo	-	cuerpo	-
1005	brain stem	tronco cerebral	variación	tronco cerebral	-	tallo de cerebro	variación
307	cancer	patología	implicitación (opcional)	cáncer	-	cáncer	-
1005	cell body	soma	implicitación (opcional)	cuerpo celular	-	cuerpo celular	-
361	defense	ataque	modificación (opcional)	defensa	-	defensa	-
303	disease-causing bacterium	bacteria patógeno	implicitación (opcional)	bacteria causante de enfermedad	-	bacteria que causar enfermedad	-

1013	diagnose	diagnóstico	modificación (opcional)	diagnosticar	-	diagnosticar	-
307	early detection	diagnóstico precoz	explicitación (opcional)	detección temprano	-	primero detección	alteración
1023	eat	deglución	modificación (opcional)	comer	-	comiendo	alteración
662	enhancer	intensificador (enhancer)	adición (opcional)	potenciador	-	potenciador	-
662	enhancer	intensificador (enhancer)	-	potenciador	alteración	potenciador	alteración
316	evolution	proceso evolutivo	explicitación (opcional)	evolución	-	evolución	-
111	gene targeting	sustitución dirigida de genes ( <u>gene targeting</u> )	adición	orientación de genes	-	gen que apunta	-
111	gene targeting	<u>sustitución dirigida de genes</u> (gene targeting)	modificación (obligatoria)	orientación de genes	alteración	gen que apunta	alteración
128	geneticist	0	omisión (opcional)	genetista	-	genetista	-
89	healing	cicatrización	explicitación (opcional)	cicatrización	explicitación (opcional)	que curar	alteración
622	heart disease	cardiopatía	variación	enfermedad cardiaco	variación	enfermedad de corazón	variación
340	host	huésped	variación	huésped	-	anfitrión	variación
756	human	genoma humano	explicitación (opcional)	humano	-	humano	-
808	infectious disease	enfermedad infeccioso	variación	enfermedad infeccioso	-	enfermedad contagioso	variación
227	inherited ill	enfermedad hereditario	-	mal heredado	alteración	ills heredado	alteración

782	interleukin	interleuquina	variación	interleucina	variación	interleukin	alteración
847	life	biología	modificación (opcional)	vida	-	vida	-
763	marker genes	marcador genético	modificación (opcional)	gen marcador	-	gen de marcador	alteración
303	microbe	microorganismo	implicitación (opcional)	microbio	-	microbio	-
443	microinjecting	microinyección	modificación (opcional)	microinyección	modificación (opcional)	microinyectar	-
106	mouse	múrido	implicitación (opcional)	ratón	-	ratón	-
153	mouse	animal	implicitación (opcional)	ratón	-	ratón	-
157	mutant animal	animal mutante	-	animal mutante	-	animal de mutante	alteración
2	neurobiologist	0	omisión (opcional)	neurobiólogo	-	neurobiologist	alteración
1014	neurobiology	neurobiología	-	neurobiología	-	neurobiology	alteración
306	neurological condition	enfermedad neurológico	modificación (obligatoria)	enfermedad neurológico	modificación (obligatoria)	condición neurológico	alteración
940	neurosurgery	neurocirugía	-	neurocirugía	-	neurosurgery	alteración
837	pathogen	patógeno	-	agente patógeno	explicitación (opcional)	pathogen	alteración
831	patient	organismo	modificación (opcional)	paciente	-	paciente	-
341	phagocytic cell	fagocito	implicitación (opcional)	célula fagocítico	-	célula fagocítico	-
808	preventing	prevención	modificación (opcional)	prevención	modificación (opcional)	prevenir	-
263	red cell	eritrocito	variación	glóbulo rojo	variación	célula blanca	variación

322	respiratory tract	tracto respiratorio	variación	tracto respiratorio	-	aparato respiratorio	variación
363	respiratory tract	tracto respiratorio	variación	vías respiratorias	variación	aparato respiratorio	variación
852	seizure	convulsión	-	convulsión	-	toma	alteración
911	seizure	convulsión	-	embargo	alteración	golpe	alteración
915	seizure	convulsión	variación	ataque	variación	golpe	alteración
316	sickle cell anemia	anemia falciforme	implicitación (opcional)	anemia de célula falciforme	-	anemia de célula de hoz	alteración
911	stroke	accidente cerebrovascular	-	accidente cerebrovascular	-	toma	alteración
926	stroke	accidente cerebrovascular	-	accidente cerebrovascular	-	golpe	alteración
909	stroke	accidente cerebrovascular	-	carrera	alteración	golpe	alteración
915	stroke	ictus	variación	derrame cerebral	variación	toma	alteración
782	T cell	linfocito	variación	célula T	variación	célula T	variación
350	tissue fluid	líquido tisular	variación	fluido tisular	variación	fluido de tejido	alteración
908	transplant	trasplante	modificación (opcional)	trasplantar	-	trasplantar	-
926	trauma	traumatismo	modificación (opcional)	traumatismo	modificación (opcional)	trauma	-
745	treating	tratamiento	modificación (opcional)	tratar	-	que invitar	alteración
729	treating	tratamiento	modificación (opcional)	tratamiento	modificación (opcional)	tratar	-
263	white cell	leucocito	variación	glóbulo blanco	variación	célula roja	variación
746	white cell	leucocito	variación	célula blanca	variación	célula blanca	variación