

**EVALUACIÓN DE TRES TRADUCTORES AUTOMÁTICOS EN LÍNEA CUANDO
TRADUCEN UNIDADES ESPECIALIZADAS EVENTIVAS (ESPAÑOL – INGLÉS)**

MARÍA FERNANDA TENECHE SÁNCHEZ

**UNIVERSIDAD AUTÓNOMA DE MANIZALES
FACULTAD DE ESTUDIOS SOCIALES Y EMPRESARIALES
PROGRAMA DE MAESTRÍA EN TRADUCCIÓN
MANIZALES**

2014

**EVALUACIÓN DE TRES TRADUCTORES AUTOMÁTICOS EN LÍNEA CUANDO
TRADUCEN UNIDADES ESPECIALIZADAS EVENTIVAS (ESPAÑOL – INGLÉS)**

MARÍA FERNANDA TENECHE SÁNCHEZ

Trabajo de grado para optar al título de Maestro en Traducción

Directora

MARÍA MERCEDES SUÁREZ DE LA TORRE

Doctora en Lingüística Aplicada

UNIVERSIDAD AUTÓNOMA DE MANIZALES

FACULTAD DE ESTUDIOS SOCIALES Y EMPRESARIALES

PROGRAMA DE MAESTRÍA EN TRADUCCIÓN

MANIZALES

2014

DEDICATORIA

A Andrés, mi esposo y compañero, con cuyo apoyo incondicional e infinita paciencia he contado siempre. Y a Gabriela por la alegría que ha traído a nuestras vidas.

CONTENIDO

pág.

CAPÍTULO I. INTRODUCCIÓN	11
1.1 Objeto de estudio	14
1.1.1 Objetivo General.....	15
1.1.2 Objetivos Específicos	15
1.2 Metodología.....	15
1.3 Corpus.....	17
1.4 Programas de TA	18
CAPITULO II. ANTECEDENTES	19
CAPITULO III. EL PROBLEMA	28
3.1 Planteamiento del problema.....	28
CAPITULO IV. MARCO TEÓRICO.....	30
4.1 Traducción y Terminología	30
4.2 Traducción Automática.....	31
4.2.1 Arquitectura interna de los programas de TA.....	32
4.2.1.1 Primera generación de traductores automáticos: Modelo de traducción directa.....	32
4.2.1.2 Segunda generación de traductores automático: Modelo de transferencia	34
4.2.1.3 Tercera generación de traductores automáticos: Modelos basados en interlingua	34
4.2.1.4 Enfoques modernos de TA: Modelos empíricos.....	36

4.2.1.4.1 Modelo estadístico	36
4.2.1.4.2 Modelo basado en ejemplos	37
4.2.2 Evaluación de los programas de TA	38
4.2.2.1 Sistema de puntuación para evaluar la inteligibilidad.....	41
4.2.3 Terminología: de la TGT a la TCT	44
4.2.3.1 Unidades Fraseológicas Especializadas – UFE.....	48
4.2.3.2 Unidades Fraseológicas Especializadas eventivas – UFE eventivas	51
CAPITULO V. METODOLOGÍA.....	54
5.1 Selección de traductores automáticos	55
5.1.1 Criterios de selección de los programas de TA	55
5.1.2 Criterios de selección del corpus textual	62
5.1.3 Identificación y selección del corpus de análisis	63
5.2 Metodología de análisis de las UFE eventivas en los productos de los TA	68
5.3 Síntesis	72
CAPITULO VI. ANÁLISIS Y RESULTADOS.....	75
6.1 Comparación productos de los programas de TA con la TR.....	75
6.1.1 Aspectos morfo-sintácticos.....	76
6.1.2 Aspectos léxico – semánticos	86
6.1.3 Aspectos contextuales.....	93

CAPÍTULO VII. CONCLUSIONES, RECOMENDACIONES Y LIMITACIONES	101
7.1 Conclusiones	101
7.2 Limitaciones.....	106
7.3 Recomendaciones	107
CAPÍTULO VIII. REFERENCIAS BIBLIOGRÁFICAS	109

LISTA DE CUADROS

	pág.
Cuadro 1. Tipología para análisis de resultados.....	26
Cuadro 2. An example Intelligibility Scale.....	42
Cuadro 3. Selección de traductores automáticos.....	57
Cuadro 4. Ejemplo (1).....	70
Cuadro 5. Ejemplo (2).....	78
Cuadro 6. Ejemplo (3).....	80
Cuadro 7. Ejemplo (4).....	83
Cuadro 8. Ejemplo (5).....	91
Cuadro 9. Ejemplo (6).....	95
Cuadro 10. Ejemplo (7).....	97

LISTA DE FIGURAS

	pág.
Figura 1. Modelos de traducción automática	36
Figura 2. Diseño de constitución del corpus de análisis.....	54
Figura 3. Interfaz Google translator	58
Figura 4. Interfaz Promt.	61
Figura 5. Ingreso de la estructura superficial <i>generar</i> al programa ANTCONC 3.2.0w y sus apariciones en la lengua origen.	65

LISTA DE GRÁFICOS

	pág.
Gráfico 1. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google con la TR	76
Gráfico 2. Similitud en la traducción de los NV y NT con la TR – Google	77
Gráfico 3. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Systran con la TR	80
Gráfico 4. Similitud en la traducción de los NV y NT con la TR - Systran.....	80
Gráfico 5. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Promt con la TR	82
Gráfico 6. Similitud en la traducción de los NV y NT con la TR - Promt	83
Gráfico 7. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google, Systran, Promt con la TR y el TB	85
Gráfico 8. Similitud en la traducción de las UFE eventivas entre la TR y la TA	85
Gráfico 9. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google, Systran, Promt con la TR	86
Gráfico 10. Traducción NT - Google	88
Gráfico 11. Traducción NT - Systran	89
Gráfico 12. Traducción NT - Promt	90
Gráfico 13. Triangulación Traducción NT – Google, Systran, Promt	93
Gráfico 14. Similitud en el número de palabras – Google, Systran, Promt	99

LISTA DE SIGLAS

GA: Grupo A NV + NT

GB: Grupo B Ndev + NT

GC: Grupo C NT + NVpart

Ndev: Nombre deverbal

Npart: Nombre participio

NT: Núcleo Terminológico

NE: Núcleo Eventivo

TA: Traducción Automática

TR: Traducción de Referencia

TM: Texto Meta

TO: Texto Origen

UFE: Unidades Fraseológicas Especializadas

UFE eventivas: Unidades Fraseológicas Especializadas eventivas

V: Verbo

CAPÍTULO I. INTRODUCCIÓN

La traducción no ha sido ajena a los avances tecnológicos de los últimos años. El desarrollo de los computadores, el internet y el almacenamiento de información han permitido crear herramientas que apoyan el proceso de traducción humana. Entre ellas se encuentran los programas de Traducción Automática (en adelante TA). El interés de esta investigación está enfocado en la relación existente entre la terminología, la traducción y el desarrollo de nuevas tecnologías que apoyan los procesos de traducción. Este interés ha llevado a cuestionamientos acerca de la manera como los programas de TA traducen Unidades Fraseológicas Eventivas (en adelante UFE eventivas).

Hutchins & Somers (1992) plantean la siguiente definición del término Traducción Automática: “The term Machine Translation (MT) is the now traditional and standard name for computerised systems responsible for the production of translations from one natural language into another, with or without human assistance” (p. 3). Gouadec, (1990, p. 40) la define como “traduction effectuée intégralement par un ou plusieurs automates”.

Al ser una máquina la que debe procesar el lenguaje natural debe enfrentar las ambigüedades y las variaciones propias de este. De acuerdo con Gonzáles (2010) un traductor humano usa por lo menos cinco tipos de conocimiento de la lengua origen, de la lengua meta, de los equivalentes entre la lengua origen y la lengua meta, del campo de especialidad y conocimiento general de la lengua origen y la lengua meta, de los aspectos socioculturales, entre los que se encuentra el conocimiento pragmático y del mundo, o sea, de las costumbres y convenciones de

las culturas de origen y meta. La carencia de este conocimiento del mundo representa uno de los límites de esta tecnología. Sin embargo, los programas de TA son una herramienta incluida en el proceso de trabajo del traductor humano por las facilidades que brinda para procesar información en diferentes lenguas. Guerberof (2010) añade que la TA es utilizada por docentes como una estrategia para ayudar a sus estudiantes en el aprendizaje de lenguas, en la enseñanza de procesos de traducción y edición de textos.

Para que la TA cumpla con las expectativas de sus usuarios, se requieren estudios centrados, no sólo en el desarrollo de su estructura y tecnología, sino también investigaciones que analicen aspectos lingüísticos que hacen parte del proceso de traducción. Cobran especial importancia aquellos estudios de TA que describen procesos de traducción y analizan aspectos terminológicos puesto que la TA puede desempeñar un papel muy importante a la “hora de traducir textos técnicos que, por su precisión y características semánticas, sintácticas y estilísticas, son los más apropiados para obtener una traducción automática aceptable” (González, 2010, p. 191). López (2002) confirma que, con más frecuencia, los sistemas de traducción automática se basan en sublenguajes o lenguajes de especialidad (el médico, el jurídico, el científico) donde se pueden obtener los mejores resultados. Fomicheva *et al.*, (2013) reconocen la importancia del tratamiento de las unidades terminológicas para evaluar la calidad de la traducción automática ya que “el discurso especializado tiene por objetivo la transmisión de conocimiento, que en los textos especializados se vehicula a través de los términos” (p. 18). Se habla de una tecnología claramente avanzada pero que trabaja con la lengua y su complejidad. Para un traductor automático es difícil hacerse cargo de las ambigüedades del lenguaje. Por tal razón, tendrá que haber conciencia de que tales sistemas

pueden ser muy útiles para determinadas aplicaciones y áreas específicas, sin embargo, no para todas.

Aunque sigue vigente lo expresado por Hutchins (1992) 20 años atrás “Fully automatic high quality translation is not at present possible” (p. 161), cada día el campo de la TA se desarrolla y los traductores disponibles en línea mejoran sus características. Siguiendo a Gonzáles (2010) es fundamental analizar su funcionamiento, evaluar sus resultados y conocer sus aciertos y errores así aprovecharlos de la mejor manera posible.

Teniendo en cuenta lo anterior, se plantea un estudio bajo un esquema de análisis cualitativo y cuantitativo donde se incluye la traducción de unidades terminológicas de textos especializados y se realiza la evaluación de cómo los programas de TA traducen estas unidades.

El análisis cualitativo se elaboró a partir de la definición y observación de las categorías emergentes que surgieron del contraste de los productos de los traductores automáticos y la traducción de referencia (en adelante TR), y de la triangulación de los datos obtenidos del análisis de los productos de cada uno de los programas de TA. Mediante el contraste de la TR y los productos de los traductores automáticos se elaboró el análisis cuantitativo concebido sobre la idea de medir la similitud entre la traducción automática y la traducción humana.

Se considera pertinente analizar cómo los traductores automáticos traducen unidades especializadas porque:

- Primero, los programas de TA hacen un intento por traducir el lenguaje natural y hacen parte de las herramientas que agilizan la labor del traductor humano.
- Segundo, la traducción de unidades terminológicas y textos especializados usando los programas de TA es una realidad que debe documentarse. Esta práctica fue respaldada por la Comisión de las Comunidades Europeas en 1976 cuando firmó un contrato con Systran para desarrollar programas de TA para la traducción de sus documentos legales a las lenguas de los países miembros (Hutchins, 1986).

De esta manera, se involucra el proceso de traducción realizado por los programas de TA Google Translator, Systran, Prompt, y la terminología con las UFE eventivas que se ingresarán a los traductores automáticos y así evaluar sus productos y su desempeño.

1.1 Objeto de estudio

A partir de la evaluación lingüística de los productos de tres traductores automáticos gratuitos, disponibles en línea y con arquitecturas internas diferentes, se describe cómo estos programas de TA traducen UFE eventivas y se ilustra su desempeño en esta tarea.

Este trabajo busca evaluar los productos de los programas de TA desde la observación y descripción del producto y su utilidad en la lengua meta.

1.1.1 Objetivo General

Evaluar tres traductores automáticos cuando traducen UFE eventivas (español-inglés).

1.1.2 Objetivos Específicos

- Describir y analizar las principales dificultades con las que se enfrentan los tres programas de traducción automática, cuando traducen UFE eventivas (español-inglés).
- Determinar si el uso de una estructura equivalente¹ en las lenguas en cuestión garantiza un mejor desempeño de los programas de TA.
- Determinar cuál de los tres traductores automáticos objeto de estudio es más efectivo cuando traducen UFE eventivas (español-inglés).

1.2 Metodología

Se ha planteado una investigación descriptiva – cualitativa – cuyo objeto de análisis se centra en el ámbito de la evaluación lingüística de los productos de tres traductores automáticos realizada

¹ Una estructura equivalente es aquella que se usa con similar frecuencia en contextos y circunstancias afines en la lengua origen español y en la lengua meta inglés.

por el usuario/traductor del programa. A partir de la evaluación de los productos se observa como los programas de TA traducen UFE eventivas.

Utilizando un corpus paralelo español - inglés perteneciente al ámbito de las Energías Renovables se realizó un rastreo para identificar las UFE eventivas que tenían aparición en la estructura subyacente Núcleo Verbal (en adelante NV) + Núcleo Terminológico (en adelante NT) según la caracterización de estas unidades realizada por Bevilacqua (2004). Se identificaron un total de 20 UFE eventivas que se ingresaron a cada uno de los programas de TA seleccionados con sus respectivos contextos.

El proceso de evaluación de los traductores automáticos se desarrolló de la siguiente manera:

1. Se definieron tres aspectos en los que se centra el análisis de los productos de los traductores automáticos. Estos emergen de la observación inicial de los productos de los programas de TA² y se fundamentan en la comparación de los productos con la TR:

- Aspecto morfo-sintáctico
- Aspecto léxico - semántico
- Aspecto contextual

² Estos aspectos se basan en los criterios de análisis de las UFE planteados por (Cabré *et al.*, 1996, p. 71).

Los dos primeros aspectos están enfocados en el análisis de las UFE eventivas de manera aislada: la forma que adoptan y la precisión con que se traducen los NT; el tercer aspecto se orienta al análisis del contexto donde tienen su aparición las UFE eventivas, la similitud de este con la TR, la selección y organización de las unidades léxicas y cómo esto influye en la precisión, la transmisión del nivel de especialidad y la inteligibilidad del texto y las UFE eventivas, específicamente.

Finalmente, el análisis general del desempeño de los traductores automáticos al traducir UFE eventivas se obtiene mediante la triangulación de los datos obtenidos para cada traductor automático después de realizar las etapas descritas anteriormente.

1.3 Corpus

Se seleccionó un corpus paralelo cuya lengua origen es el español y su lengua meta es el inglés, a nivel divulgativo titulado *Renovables made in Spain energías renovables en España / Renewable energies in Spain*. Estos textos hacen parte de un proyecto impulsado por el Ministerio de Industria Español a través del Instituto para la Diversificación y Ahorro de la Energía (IDEA), pertenecientes al dominio de las energías renovables pero que a su vez son interdisciplinarias y pueden tocar ámbitos como la química, la física, el medio ambiente, entre otros. Con anticipación se verificó la presencia de *UFE eventivas* en su estructura subyacente NV+NT.

Se eligió un corpus paralelo porque al utilizar una TR como criterio de evaluación del desempeño de los programas de TA se requiere una traducción realizada por traductores humanos que sirva como referente para valorar los productos de los traductores automáticos (Amigó,

Gimenez, Verdejo, 2009). De esta manera, el texto base (en adelante TB) es en español y la TR en inglés.

1.4 Programas de TA

En cuanto a los programas de TA utilizados en esta investigación, se seleccionan tres programas de TA Google Translator, Systran y Prompt. Estos traductores automáticos cumplen los criterios de selección: son gratuitos, están disponibles en línea y tienen arquitectura interna diferente, es decir, que la manera general como procesan la organización de la información o la organización abstracta de sus módulos de procesamiento de datos difiere entre ellos (Arnold *et al*, 1994); Google Translator utiliza la traducción estadística; Systran basa su proceso de traducción en una estructura híbrida entre la traducción basada en reglas y la estadística, y Prompt utiliza una estructura multidimensional de diccionarios que almacena un número ilimitado de variantes de traducción e información gramatical.³

³ Esta información puede ser consultada en cada uno de los sitios web de los traductores automáticos.

Google Translator: <http://research.google.com/pubs/MachineTranslation.html>

Systran: <http://www.systransoft.com/systran/corporate-profile/translation-technology/systran-hybrid-technology/>

Prompt: http://www.prompt.es/company/technology/machine_translation/

CAPITULO II. ANTECEDENTES

Desde el surgimiento de los primeros sistemas que buscaban la automatización del proceso de traducción se han realizado estudios que pretenden evaluar el desempeño de los programas de TA. Los estudios se han orientado tanto al análisis y desarrollo informático de los sistemas como a examinar sus productos para comprender cómo funcionan y traducen estos programas.

A continuación se reseñan primero, estudios centrados en la observación de programas de TA que traducen textos especializados. Posteriormente, estudios centrados en la evaluación de traductores automáticos en línea utilizados para traducir Unidades Terminológicas.

Dieguez & Riedemann (1998) realizaron un análisis contrastivo de las formas [-ing] presentes en textos origen en inglés y sus respectivas traducciones al español realizadas por el traductor automático Power Translator Professional. El corpus está constituido por dos textos en inglés; uno del área de las ciencias sociales (economía) y otro de la minería. Los objetivos eran extraer conclusiones prácticas para la enseñanza de la traducción técnico-científica y evaluar la calidad de los traductores automáticos utilizados.

Para determinar los aciertos y errores presentes en los productos del traductor automático elegido se revisaron manualmente los productos y se identificó cómo habían sido traducidas las formas -ing del inglés al español teniendo como punto de partida las siguientes categorías correspondientes a las formas que puede adoptar la estructura del inglés -ing: a) Paradigma verbal, b) Paradigma nominal y c) paradigma atributivo. Como resultado, se obtuvo un análisis cualitativo

de los aciertos y errores presentes en los productos del traductor automático al traducir la estructura –ing del inglés al español.

Entre las conclusiones obtenidas en este estudio se destacan:

- En el texto económico se presenta un margen de error relativamente mayor (69,18%) en lo que respecta a la traducción de las formas –ing al español que en el texto de minería (59,34%) corroborando la hipótesis planteada por algunos autores en el sentido de que aquellos textos que pertenecen a las ciencias exactas, en comparación con aquellos pertenecientes a las ciencias sociales, pueden ser traducidos en forma más correcta por un programa de TA (Dieguez & Riedemann, 1998).
- La incidencia de las diferentes funciones gramaticales que conllevan las formas –ing no depende de las áreas temáticas ni de los tipos de texto involucrados, sino más bien responde a un patrón general del “genio”⁴ de la lengua inglesa.
- Se observa en ambos corpus que el programa de TA rinde buenos resultados cuando se enfrenta a oraciones breves y a una sintaxis simple.
- Entre los errores más frecuentes encontrados están: elipsis de partes de la oración como método de economía de lenguaje, secuencias de dos o más lexemas sin conectores o unidos

⁴ Expresión utilizada por Dieguez & Riedemann (1998) para referirse a las características morfológicas, sintácticas, semánticas y pragmáticas de la lengua inglesa.

por guión o barra, frases intercaladas entre comas o entre paréntesis que rompen la estructura sintáctica normal de la oración.

Esta investigación plantea inquietudes sobre varios aspectos:

- El uso de un traductor comercial para traducir textos especializados.
- El campo lexical en que se inscribe el texto puede determinar la exactitud con que es traducido automáticamente.
- El planteamiento de las categorías para evaluar los productos de la TA pueden variar según los requerimientos y objetivos de la evaluación.

Dieguez (2001) desarrolla un estudio donde plantea una metodología para la formación de futuros traductores. Sin embargo, de este trabajo se desprende un análisis de los aciertos y errores de los traductores automáticos utilizados.

Su corpus parcial de análisis está compuesto por dos textos que corresponden a dos géneros textuales específicos: un artículo semi especializado sobre un instituto de biotecnología y un resumen especializado de una comunicación en un seminario de medio ambiente. Estos textos fueron traducidos del inglés al español por el traductor automático en línea Systran.

Para facilitar el análisis de los errores se utilizó una tipología de errores ampliada (Wilss y Thome, 1984, citado por Dieguez, 2001) atendiendo al criterio funcionalista del impacto del error en el texto y en el eventual destinatario de la traducción. Este criterio se rige por patrones lingüísticos y pragmáticos.

Los tipos de errores son:

- Error léxico
- Error sintáctico
- Error léxico –sintáctico
- Error léxico-semántico
- Error sintáctico-semántico
- Error léxico-sintáctico-semántico

En sus conclusiones la autora describe los aciertos y errores más recurrentes observados en los productos de los traductores automáticos:

Tabla 1. Aciertos y errores más recurrentes observados en los productos de los traductores automáticos

Aciertos	Errores
<ul style="list-style-type: none"> - Reconocimiento de la función adjetiva de las palabras y frases más extensas antepuestas al sustantivo que modifican y adecuación al orden sintáctico de texto meta. - Uso de terminología correcta. 	<ul style="list-style-type: none"> - Orden sintáctico incorrecto (adjetivo-sustantivo, orden incorrecto de las palabras en la oración). - No reconoce y traduce erróneamente términos compuestos (terminología y fraseología). - Preposiciones mal traducidas (elección incorrecta, eliminación o uso inadecuados).

Fuente: Dieguez (2001).

Entre las posibles causas para estos errores la autora plantea:

- El lexema o término no figura en el lexicón.
- Presencia de fraseología en el texto origen.
- Lista de términos descontextualizados.

De esta investigación es posible destacar dos aspectos:

1. La evaluación de traductores automáticos mediante una tipología de errores.
2. La tipología de aciertos y errores emerge a partir del análisis de los productos del traductor automático.

Se debe mencionar que este estudio aborda el análisis de los textos especializados de una manera amplia y general sin detallar aspectos relevantes en ellos como la traducción de unidades terminológicas.

Las investigaciones más recientes están dirigidas al análisis lingüístico de los productos de los programas de TA en línea por ser un recurso gratuito y de libre acceso para cualquier usuario.

González, (2010) realizó un análisis comparativo de tres traductores automáticos en línea: Systran, Reverso y Google. En este estudio se ingresa a los traductores automáticos oraciones y frases breves del inglés al español que han sido elegidas de diferentes textos no especificados. En las oraciones y frases seleccionadas se incluyeron casos de ambigüedad léxica, unidades

fraseológicas (en adelante UF) – colocaciones, refranes, construcciones pasivas, entre otras unidades.

Se eligieron tres parámetros de evaluación:

- Número de lenguas que traducen Systran, Reverso y Google.
- Enfoque de traducción.
- Análisis de traducción de frases y oraciones breves del inglés al español.

En los resultados se presentan los siguientes hallazgos:

- Google es el traductor que mayor número de lenguas traduce, 52 en total.
- Respecto al enfoque de traducción, el estudio plantea que aquellos traductores que integran técnicas de diferentes enfoques tienen mejores resultados. Sin embargo, no define cuál de los enfoques de los traductores usados en el estudio genera mejores productos.
- En el análisis realizado, se encuentra que Google resuelve, de manera más apropiada, las ambigüedades léxicas, calcos léxicos y sintácticos, las colocaciones y las construcciones pasivas.

De esta investigación se destacan los siguientes aspectos:

- La consideración de la arquitectura interna de los programas de TA se utiliza como una oportunidad para evaluar qué estructura funciona mejor con el tipo de oraciones y frases ingresadas.

- El uso de UF para evaluar traductores automáticos.
- La evaluación de distintos traductores automáticos es conveniente y necesaria ya que su arquitectura incluye distintos modelos y pertenecen a diferentes generaciones de TA.

Botella (2008) resalta la importancia de la fraseología y reconoce que es un aspecto de la lengua que después de ser analizado puede encontrar soluciones pertinentes y viables para los traductores automáticos. La autora desarrolla su proyecto de investigación con el objetivo de analizar cómo reaccionan distintos traductores en línea, Systran, Promt y Apertium, ante un corpus de 300 “frases hechas” relacionadas con la mesa y la bebida con el par de lenguas español- francés. Además, se pretende evaluar si los diccionarios de los programas de TA contienen “frases hechas” y si pueden reconocerlas. La autora menciona algunas dificultades de centrarse en la traducción de las UF:

- La identificación de las “frases hechas” como tales dada su variabilidad (distintas formas, flexión verbal, complementos, etc.)
- Dificultad para encontrar una equivalencia correcta. En traducción automática esto se debe a la limitación de los lexicones.
- La dificultad de interpretar si la frase tiene sentido literal o figurado.

La elección de los traductores automáticos Systran⁵, Promt y Apertium se debe a que están basados en reglas y no en corpus.

⁵ En el 2008 Systran era un programa de TA basado en reglas. Actualmente se puede constatar en su sitio web que es un traductor automático híbrido.

El criterio de evaluación de los fragmentos se realizó en función de si estaban bien, o mal traducidos y el por qué fueron etiquetados de esa manera.

Para el análisis de los resultados se establece la siguiente tipología.

Cuadro 1. Tipología para análisis de resultados

MWE bien
MWE pero mal
Traducción aceptable
Traducción literal
Palabra por palabra
Agramatical
No reconoce palabras
Traducción inaceptable

Fuente: Botella, (2008, p. 9)

Algunos ejemplos de esta tipología presentados en el estudio son:

Tabla 2. Ejemplos de tipología

“Frasas hechas”	Traducciones
1. “bebe a hurtadillas”	« il boit à la dérobée » « Il boit à hurtadillas » las palabras – Traducción inaceptable « il boit à la dérobée »
2. “come como un puerco”	(Systran) « il mange comme un puerco » (Apertium) « Il mange comme un puerco » (Promt) « il mange comme un cochon »

Fuente: Botella, (2008, p. 9)

Systran y Apertium no reconocen las palabras, la traducción es inaceptable.

Prompt traduce puerco por « cochon » aunque la traducción acertada es « porc ».

Este estudio concluye que, para el tema seleccionado, la fraseología no se destaca en los programas de TA seleccionados ya que pocas frases fueron traducidas correctamente. Sin embargo, hubo muchas traducciones que sorprendieron a la investigadora.

Del estudio realizado por Botella (2008) se destacan los siguientes aspectos:

- Las “frases hechas” utilizadas para este estudio corresponden a un campo restringido como lo es la mesa y la bebida. El uso de estas frases fue un punto de partida para definir las unidades que se ingresaron a los traductores automáticos del presente estudio seleccionando las UFE eventivas en un corpus paralelo del campo de la energía solar.
- El desempeño “pobre” de los programas de TA basados en reglas, utilizados en este estudio, plantea que este enfoque no es el apropiado para traducir textos u oraciones que puedan pertenecer a un campo especializado.

La revisión de estos estudios orienta el planteamiento del presente trabajo y muestra la necesidad profundizar en el estudio de los productos de los programas de TA.

CAPITULO III. EL PROBLEMA

3.1 Planteamiento del problema

La TA está mejorando. En la última década, ha habido muchos avances e indiscutiblemente estos sistemas de traducción permiten a muchas personas el acceso a una gran cantidad de información que no aparece en su lengua. Como señalan Porta & Hampshire (2010), hoy por hoy, en internet, los programas de TA son una aplicación real que está siendo utilizada. Un sinnúmero de usuarios de internet utiliza estos programas, así como muchas organizaciones internacionales los usan en la traducción de determinados textos, como por ejemplo la Unión Europea, que utiliza el sistema Systran, o la Organización Panamericana de la Salud (PAHO, por sus siglas en inglés [*Pan American Health Organization*]), que tiene su propio sistema de TA, Spanam & Engspan.

Para los programas de TA, la traducción de términos, unidades fraseológicas, entre otros elementos característicos de los textos especializados, representan un problema ya que sus corpus y diccionarios tienen limitaciones respecto a la cantidad de información almacenada y su procesamiento. La falta de trabajos descriptivos que permitan conocer cómo funcionan los algoritmos aplicados en cada uno de los métodos de traducción de los programas de TA limita las posibilidades de uso de esta herramienta.

Como lo menciona Botella (2008) la fraseología no es un aspecto destacado de los traductores automáticos, Systran, Promt, Aprentium, que se emplearon en su investigación con relación al tema seleccionado. Sin embargo, en sus conclusiones, la autora, reconoce que los

traductores automáticos utilizados en su estudio realizaron traducciones de “frases hechas” con gran precisión.

Al hacer una revisión bibliográfica puede concluirse que la fraseología ha sido ampliamente analizada por la lexicografía y la terminología (Cowie, 1998; Anderson, 2006; Granger & Meunier, 2008). No obstante, el número de estudios que analizan el funcionamiento de los programas de TA, sus errores y aciertos, no se encuentran en igual número.

Después de revisar documentos de autores relevantes de la TA (Hutchins, 1992, 2003; Arnold *et al.*, 1994), y teniendo en cuenta lo mencionado respecto a la fraseología, se generan diversas inquietudes con relación a cómo son traducidas las Unidades Fraseológicas Especializadas y, específicamente, las UFE eventivas mediante el uso de programas de TA. Como lo indica Cabré (1999) “per se estas unidades léxicas no son ni término ni palabra, sino que adquieren el carácter de término en función de su uso” (p. 102-103) En otras palabras, estas unidades provienen del léxico general pero adquieren su grado de especialidad al estar inmersas en un ámbito especializado. Esta característica hace pensar que los programas de TA pueden tener herramientas para traducirlas acertadamente.

Partiendo de estas inquietudes se formularon las siguientes preguntas problematizadoras:

¿Cómo traducen los programas de TA en línea las UFE eventivas? ¿Qué similitudes y diferencias pueden evidenciarse al comparar una traducción realizada por un traductor humano con el mismo texto traducido por un programa de TA? ¿Existen diferencias entre los productos de traductores automáticos que puedan evidenciar la eficacia de su estructura interna?

CAPITULO IV. MARCO TEÓRICO

En el presente apartado se hará un abordaje teórico de conceptos fundamentales como la evaluación de la TA y las UFE eventivas. Estos conceptos se revisarán desde su base, la traducción y la terminología, hasta llegar a su descripción particular.

4.1 Traducción y Terminología

La traducción y la terminología establecen una relación necesaria como disciplinas. Suarez (2004) explica que la terminología y la traducción han dejado de ser vistas como, en los planteamientos prescriptivos, meras actividades donde intervienen las lenguas o donde una denominación tiene un solo significado para ser reconocidas como, en los planteamientos descriptivos, actos comunicativos y textuales que se desarrollan y deben analizarse en un contexto social y en su uso real. Por su parte Cabré (2000) explica que la traducción como práctica es un proceso de transferencia de información entre las lenguas distintas en el que la terminología juega un papel relevante. La importancia de la terminología y su interrelación con la traducción se visualiza claramente en la comunicación especializada donde el conocimiento especializado se condensa especialmente en las unidades terminológicas.

Los avances teóricos actuales en ambas disciplinas y los desarrollos tecnológicos informáticos que abren nuevas posibilidades y retos para el traductor humano exigen investigaciones enmarcadas en propuestas de tipo comunicativo.

4.2 Traducción Automática

Para lograr el desarrollo de los programas de TA se han realizado estudios que permitan evaluar y mejorar su usabilidad y resultados. En las últimas décadas, estos estudios se han orientado al análisis de cómo estos programas traducen y a observar los aciertos y errores encontrados en sus productos. Además, se evidencia la necesidad de evaluar la utilidad de la TA con relación a los lenguajes de especialidad. Los programas de TA se han convertido en una herramienta que requiere una evaluación detallada y constante para así mejorar la calidad de sus productos y ofrecer nuevas posibilidades que los traductores humanos puedan aprovechar convenientemente. Como lo advierte Hutchins (2003) es necesario realizar investigaciones y análisis de diversos tipos para, en un futuro, poder hacer comparaciones que permitan establecer el avance en la calidad de estos programas.

Arnold *et al.* (1994) definen la traducción automática como el intento de automatizar la totalidad o parte del proceso de traducir un lenguaje humano a otro. Por su parte, Berner (2003), la define como “the use of computer software to translate text or speech from one natural language into another”.

Estas definiciones podrían hacer llegar a pensar que el proceso de TA es simple pero como lo explican Christianson *et al.*, (2011) la TA enfrenta grandes retos, ya que, aunque las capacidades de una máquina superen en muchos aspectos las humanas, actualmente, ningún computador o programa ha podido superar las habilidades de lenguaje que un humano adquiere naturalmente.

El avance tecnológico de estos sistemas requiere de investigaciones que analicen sus productos y, a partir de allí, se pueda modificar su arquitectura. Los procesos de investigación y análisis han llevado a la construcción de diversas arquitecturas que se clasifican en tres generaciones de traductores automáticos y los enfoques modernos o empíricos. Seguidamente, explicaremos cada uno de ellos.

4.2.1 Arquitectura interna de los programas de TA

Hoy en día las tres generaciones de traductores automáticos siguen vigentes incluso se han generado sistemas que responden a combinaciones de las tres reconocidos como los enfoques modernos o empíricos de la TA.

Hutchins (1992) plantea una clasificación clásica de los traductores en tres generaciones según el esquema de organización del procesamiento de datos que utilicen. Aunque Moreno (2000) plantea que hoy en día esta clasificación no es, estrictamente, pertinente puesto que los sistemas modernos incorporan técnicas de varios enfoques que no están contemplados en ella, esta clasificación permite mostrar las técnicas más importantes utilizadas en TA.

4.2.1.1 Primera generación de traductores automáticos: Modelo de traducción directa

Hutchins (1992) define este modelo como una metodología de TA que no tiene ninguna etapa intermedia en el proceso de traducción. Podría decirse que la traducción es casi palabra por palabra. Arnold *et al.*, (1994) establecen que este modelo está diseñado para un par de lenguas, y

constituido por una colección de reglas gramaticales. La traducción es considerada como una búsqueda de palabras de una lengua origen en un diccionario bilingüe y por esto, las traducciones que realizan, generalmente, son entre pares de lenguas muy cercanas que tengan estructuras sintácticas similares ya que estos sistemas contienen poco o ningún desarrollo en el área semántica. Moreno (2000) resume la idea principal de esta arquitectura como la transformación de frases de la lengua origen en frases de la lengua meta de la manera más simple posible, reemplazando palabras de la lengua origen con sus correspondientes de la lengua meta siguiendo determinado diccionario bilingüe, y después reordenándolas según las reglas especificadas en la lengua meta en cuestión. Por su parte, Hernández (2002) explica que el modelo de traducción directa es válido cuando se halla ante un texto con un vocabulario y estilo bien definido y limitado. La calidad de sus productos está en relación directa con la calidad de la información que contiene el diccionario.

La implementación de este modelo es simple, rápido, económico y procesa grandes volúmenes de texto pero la calidad de sus resultados es muy baja.

Entre los programas de TA más conocidos que utilizaron este sistema en sus inicios está Systran, Moreno (2000).

4.2.1.2 Segunda generación de traductores automático: Modelo de transferencia

Esta generación de traductores automáticos está basada en la implementación de reglas lingüísticas⁶ tanto de la lengua origen como la lengua meta y, la representación profunda de las estructuras de la lengua origen para poder ser traducida a otras.

Como lo describe Tertoolen (2010), este modelo utiliza conocimiento contrastivo de las dos lenguas implicadas en el proceso de traducción y genera un producto de la siguiente manera: el sistema analiza el Texto Base (en adelante TB) y genera un texto abstracto en la lengua origen después con esta representación genera una del Texto Meta (en adelante TM) para finalmente, obtener una traducción en la lengua meta.

Las fortalezas de este modelo radican en ser un modelo simple multilingüe y con un amplio conocimiento sintáctico-semántico de las lenguas tratadas, sin embargo, presenta dificultades en el proceso de análisis y en la desambiguación. Ejemplos de este modelo son METAL, Ariane, BCI y Kaplan, Moreno (2000)⁷.

4.2.1.3 Tercera generación de traductores automáticos: Modelos basados en interlingua

El proceso de traducción en estos sistemas discurre en dos etapas: Se lleva a cabo un análisis profundo de la estructura del TB para obtener una representación en un lenguaje universal del

⁶ Estas reglas son las que definen la correspondencia entre una lengua respecto de otra.

⁷ Metal: Mechanical Translation and Analysis of Language - Hoy, Siemens-Nixdorf -

significado del texto. Partiendo de esta estructura, se puede generar el texto correspondiente de cualquier lengua meta sin importar cuál fue la lengua origen. (López, 2002)

Esta separación entre el conocimiento de la lengua origen y la lengua meta es la principal motivación de la TA de interlingua afirma Moreno (2000).

Este modelo tiene dificultades para crear las representaciones universales de cada una de las lenguas de trabajo (Arnol *et al.*, 1994).

Ejemplos de este tipo de TA son: ULTRA (New Mexico State University) y PANGLOSS (CMU, NMSU y University of Southern California), entre otros (Benotti & Roger, 2004).

La figura 1, adaptación del triángulo de Vauquois, ilustra el proceso de traducción automática según los modelos descritos anteriormente.

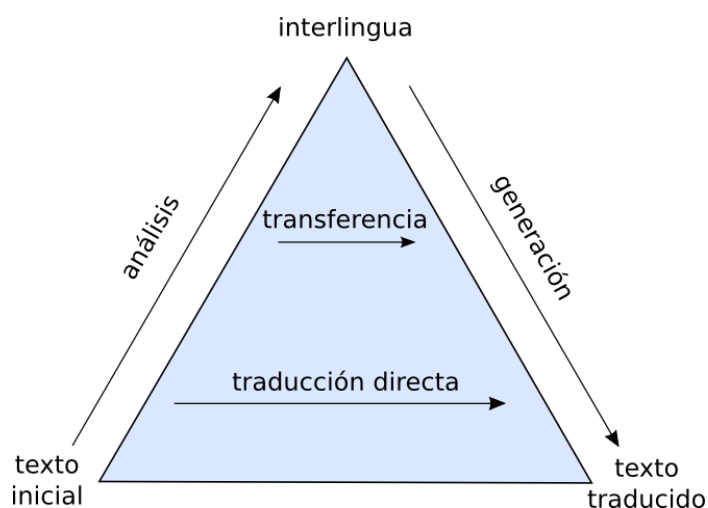


Figura 1. Modelos de traducción automática

4.2.1.4 Enfoques modernos de TA: Modelos empíricos

Moreno (2000) explica que los enfoques empíricos suponen una ruptura radical con los modelos de interlingua y de transferencia. Estos modelos aprovechan la disponibilidad de grandes cantidades de texto en formato electrónico para aplicar técnicas estadísticas o de búsqueda de patrones tanto para traducir como para extraer información para ser utilizada posteriormente en el proceso de traducción. Dentro de estas nuevas metodologías cabe distinguir dos modelos distintos: la TA basada en estadística y la TA basada en el ejemplo.

4.2.1.4.1 Modelo estadístico

El modelo estadístico utiliza corpus paralelos bilingües (textos traducidos por un humano) que contienen oraciones con el mismo significado en dos o más lenguas. El sistema de TA es entrenado para seleccionar la secuencia más probable entre las secuencias posibles y, bajo este criterio, seleccionar el equivalente pertinente para el contexto del TM.

Blackwood (2010) reconoce algunas ventajas a este modelo:

- Un alto nivel de conocimiento lingüístico especializado no es requerido.
- Los procesos del modelo son, ampliamente, independientes de la lengua, es decir, que excluyen la necesidad de conocimiento lingüístico o de cualquier otro tipo, y por lo tanto no contemplan, en principio, la utilización de un repositorio de información léxica.

- Producción de traducciones naturales y fluidas, es decir, traducciones que pueden ser leídas como un texto de la cultura meta por la habilidad que tiene este método de aprender traducciones en contexto.

Un ejemplo de este modelo es Google Translator.

4.2.1.4.2 Modelo basado en ejemplos

Este modelo tiene componentes matemáticos y funciona de manera similar al modelo estadístico, sin embargo, se va a presentar separadamente para hacer una descripción más detallada y hacer mayor claridad al lector.

Como lo detallan Kit *et al.*, (2002), la base para la existencia de este modelo es un gran volumen de textos traducidos por profesionales quienes, además de la proficiencia en las lenguas, son especialistas en el campo del conocimiento al que pertenece el texto que traduce. El modelo basado en ejemplos consiste en codificar el conocimiento de una manera sencilla de tal forma que permita hacer coincidir un trozo de texto en una lengua con un trozo de texto en otra lengua.

Entre las fortalezas de este modelo se encuentran su flexibilidad, la posibilidad de combinar diferentes tipos de traducción y de incluir el componente semántico. Por el contrario, presenta dificultades con la polisemia, la metáfora y con algunas formas compuestas de la lengua.

Moreno (2000) asegura que hasta la fecha no ha habido muchas implementaciones de este tipo de motores de traducción, siendo los más sobresalientes MBT y PanEBMT.

La clasificación de los programas de TA descrita anteriormente facilita el conocimiento de la arquitectura que utilizan para traducir y cómo sus características inciden en la calidad de los

productos. Aunque en esta investigación no se evalúan las características y el funcionamiento interno de los traductores automáticos, su consideración enriquece el análisis de los productos de los programas de TA y permite concluir las ventajas y desventajas que cada una tiene al traducir las UFE eventivas.

A continuación se comentarán algunas estrategias para la evaluación de los programas de TA.

4.2.2 Evaluación de los programas de TA

Hovy *et al.*, (2002) hacen un recorrido por los diferentes enfoques para evaluar los programas de TA. Ente ellos:

Norma ISO/IEC 9126, 14598: establece normas para evaluar la calidad del software basado en su funcionalidad, fiabilidad, usabilidad, portabilidad y mantenibilidad. Estos criterios están enfocados en el desempeño de la arquitectura tecnológica con la que está construido el programa más que en la calidad lingüística de la traducción lograda.

Las métricas: las métricas de evaluación automática comparan la salida del sistema con traducciones de referencia realizadas por traductores humanos. Esta evaluación se considera iterativa, es decir, repetitiva sin que se requieran nuevos juicios humanos (Amigó *et al.*,2009). El objetivo de las métricas es medir la similitud entre el producto del traductor automático y la traducción referencia y así, establecer aciertos y errores en ellos. Peinado (2004) asegura que

cuanto más se aproxime una traducción automática a la de un traductor humano profesional, mejor traducción será.

Para su estudio Amigó *et al.*, (2009) eligieron métricas a tres niveles de procesamiento:

- Nivel léxico: métricas basadas en la distancia de edición, la medida de frecuencia (GTM y METEOR), la cobertura léxica (ROUGE) y la precisión léxica (NIST y BLEU).
- Nivel sintáctico: métricas basadas en procesamiento sintáctico superficial y análisis completo: relaciones de dependencia y árboles constituyentes.
- Nivel semántico: métricas que emplean los roles semánticos y estructuras de discurso.

Entre los proyectos más conocidos y actuales relacionados con la evaluación de programas de TA basado en métricas se encuentran: FEMTI (por sus siglas en inglés: a Framework for the Evaluation of Machine Translation in ISLE) y BLEU (BiLingual Evaluation Understudy).

Estos sistemas permiten ahorrar tiempo y costos, además pueden procesar grandes cantidades de texto. Sin embargo, presentan dificultades cuando deben comparar sistemas con arquitecturas diferentes; realizan mediciones de estructuras no mayor a la oración o frase y, al tener como base traducciones ideales, no contempla factores contextuales y pragmáticos que son indispensables en la comprensión de un texto.

Reconociendo las limitaciones descritas anteriormente se retoma de las métricas el uso de una traducción humana como referencia con el objetivo de validar la precisión con que los

programas de TA traducen unidades terminológicas, específicamente las UFE eventivas analizando aspectos morfológicos, sintácticos y semántico.

El uso de la traducción de referencia no responde a una visión idealizada de la traducción sino que pretende ser un criterio para el análisis de la traducción de las UFE eventivas y sus contextos.

Algunos autores han evidenciado la dificultad que representa evaluar los programas de TA a través de sus arquitecturas internas y han reconocido la posibilidad de evaluarlos a través de sus productos. Al respecto Christianson *et al.*, (2011) consideran que tradicionalmente han existido dos paradigmas para la evaluación de los programas de TA: “Glass-box evaluation, which measures the quality of a system based upon internal system properties, and Black-box evaluation, which measures the quality of a system based solely upon its output, without respect to the internal mechanisms of the translation system” (p. 745).

Respecto a este tipo de evaluación Hutchins (1992) considera: “Here there are important differences between glass-box evaluations and black-box evaluations, between assessment by those who have access to all the workings of the system and assessment by those who can work only with inputs and output” (p. 163). El autor afirma que la evaluación de caja negra genera datos más objetivos. (p. 163)

En consecuencia, el objetivo central es evaluar los programas de TA a través del análisis lingüístico de sus productos y no de las características de sus arquitecturas internas, esto indica, una “black-box evaluation”.

4.2.2.1 Sistema de puntuación para evaluar la inteligibilidad

Para Arnold *et al.*, (1994) este sistema evalúa el nivel de comprensión asignando puntos a las oraciones traducidas por los programas de TA. La escala asigna los puntajes más altos a las oraciones cuyas características (sintácticas, léxicas, morfológicas, semánticas) les permiten ser leídas como un texto de la lengua meta y, por lo tanto, el mensaje del TB es transmitido en el TM. En consecuencia, se asignan puntajes bajos a aquellas oraciones que no cumplen con estas características. Entre estos extremos se asigna un mayor o menor puntaje a las oraciones según el grado de facilidad o dificultad con que se pueda comprender la traducción evaluada.

El ejemplo más conocido donde se utiliza este sistema es el reporte del ALPAC⁸. En este se describe una escala de 9 puntos para evaluar no solo la inteligibilidad de los productos de la TA sino que se usó para evaluar traducciones humanas.

El sistema de evaluación por puntos es un sistema flexible que permite evaluar la comprensión de un producto de la TA con los recursos y la visión del traductor/usuario.

⁸ ALPAC por sus siglas en inglés Automatic Language Processing Advisory Committee fue un comité de siete científicos liderados por John R. Pierce establecido en 1964 por el gobierno de los Estados Unidos para evaluar el progreso de la lingüística computacional y en particular el progreso de la TA.

Un ejemplo del sistema de puntos para evaluar la inteligibilidad de los productos de los programas de TA es presentado por Arnold *et al.*, (1994) en el siguiente cuadro:

Cuadro 2. An example Intelligibility Scale

An Example Intelligibility Scale	
1	The sentence is perfectly clear and intelligible. It is grammatical and reads like ordinary text.
2	The sentence is generally clear and intelligible. Despite some inaccuracies or infelicities of the sentence, one can understand (almost) immediately what it means.
3	The general idea of the sentence is intelligible only after considerable study. The sentence contains grammatical errors and/or poor word choices.
4	The sentence is unintelligible. Studying the meaning of the sentence is hopeless; even allowing for context, one feels that guessing would be too unreliable.

Fuente: Arnold *et al.*, (1994).

El cuadro muestra que en el sistema de puntos para evaluar la comprensión de un producto de un traductor automático se establece una escala numérica (de 1 a 4) con su correspondiente descripción cualitativa. Esta descripción expresa las cualidades que deben tener la traducción y el nivel en que se manifiestan para que se le asigne una determinada valoración.

La evaluación manual es una de las más utilizadas pero poco documentada. Generalmente, se utiliza para describir y evaluar los productos del traductor automático con relación a aspectos lingüísticos.

Quien lleva a cabo este tipo de evaluación, en su mayor parte traductores en ejercicio, debe determinar el aspecto de la lengua que desean analizar en los productos del traductor automático

(sintáctico, semántico, etc.), y establecer un método que le permita verificar la calidad de la traducción y las dificultades encontradas sin la intervención de la subjetividad humana.

La importancia de la evaluación manual radica en la proximidad de esta con la realidad de los traductores profesionales, la individualización y caracterización del contexto que quien la realiza puede hacer. Sin embargo, este tipo de evaluación es criticada por los altos costos que genera, los riesgos de subjetividad al evaluar los productos y la dificultad de generalizar los resultados obtenidos.

Hutchins (1992) aprueba que los productos de los traductores automáticos puedan ser evaluados por traductores humanos o comparándolos con el mismo texto traducido por un humano. Usualmente, se establecen criterios, escalas, que permiten concluir si la traducción es apropiada o no.

Considerando las ventajas de este tipo de evaluación, para esta investigación se llevó a cabo una revisión manual de los productos de los traductores automáticos a partir de los aspectos establecidos a analizar.

En síntesis, después de revisar los métodos, técnicas y estrategias utilizadas para la evaluación de productos de los traductores automáticos y establecer sus ventajas y limitaciones, consideramos pertinente evaluar los programas de TA seleccionados a través del análisis manual de sus productos, un análisis lingüístico a partir de los principios de la “black- box evaluation” y utilizando como criterio de evaluación la similitud entre el producto de la TA y la TR.

Después de describir las estrategias de evaluación de los programas de TA pasaremos a explicar las unidades terminológicas que se ingresaron a los traductores automáticos.

4.2.3 Terminología: de la TGT a la TCT

Un recorrido por la historia de la terminología expone como esta disciplina tuvo sus inicios en enfoques prescriptivistas y poco a poco avanzó hacia estudios descriptivos.

La Teoría General de la Terminología (TGT) ha sido el marco desde donde se han expuesto los planteamientos de los enfoques tradicionalistas siendo E. Wüster su máximo exponente. Estos enfoques se desarrollan en los inicios del siglo XX, época en que “la terminología deviene un instrumento de trabajo para la desambiguación de la comunicación científica y técnica” (Suarez, 2004, p. 18), tomando como punto de partida el carácter sistemático de los términos.

Santamaría (2006), resume las características de la TGT:

- Se considera que la comunicación especializada debe ser fluida y sin ambigüedades, por tanto, es necesario controlar y armonizar los vocabularios terminológicos.
- Se busca la normalización internacional de la terminología por medio del Comité Técnico de la Terminología.
- El objeto de estudio es el término. Este es considerado como unidad específica de un ámbito de especialidad y su uso queda restringido a él; por tanto, la terminología solo tiene interés por la comunicación entre especialistas.

- Se parte del concepto considerado independiente de la denominación lo que no da cabida a la polisemia y obliga a apoyar la monosemia esto indica que a cada denominación le corresponde un solo significado y una sola referencia.

Suárez (2004, p. 20) afirma que

Los planteamientos de la TGT dejan entrever que la concepción inicial de la terminología era prescriptiva [...], se centraba en la idea de establecer un consenso de los términos científico-técnicos con el fin de lograr la transferencia de conocimientos entre los especialistas a nivel internacional.

Se busca entonces una unidad terminológica que “denomina un concepto que es a su vez uniforme, ahistórico, atemporal, asocial e ideológicamente neutro. Desde este punto de vista meramente denominativo, los aspectos sintácticos y gramaticales del término, como la fraseología, son irrelevantes” (García, 2001).

El paso del estructuralismo, base de la TGT, a la aparición del funcionalismo lingüístico y, posteriormente, la aplicación al lenguaje de las teorías de base cognitiva propuestas por la filosofía y la psicología han permitido acuñar una propuesta adecuada de descripción de los términos que considere sus aspectos semánticos y pragmáticos (Cabré, 2005).

Como lo explica Cabré (2005)

La diversificación de los focos de interés dentro de la lingüística ha abierto la mirada no solo a las oraciones, sino también al texto, permitiendo abordar así tanto las similitudes como las diferencias de los textos de especialidad con relación a los generales, como la observación de las unidades terminológicas en su contexto lingüístico.

Es esta apertura de la lingüística y las insuficiencias de los postulados de la TGT lo que posibilita el surgimiento de la Teoría Comunicativa de la Terminología (TCT) siendo María Teresa Cabré su exponente más representativa.

Seoane (2008) resume las características de la TCT:

- El objetivo de la TCT es describir formal, semántica y funcionalmente las unidades que pueden adquirir valor terminológico, explicar la activación, el valor y explicar las relaciones que se establecen con otros signos del mismo sistema de conocimiento u otro sistema distinto.
- Su objeto de estudio son las unidades terminológicas propiamente dichas.
- Los términos son unidades léxicas, activadas singularmente por sus condiciones pragmáticas de adecuación a un tipo de comunicación determinada; se constituye de forma (constante) y significado (se adecua a la situación comunicacional).
- El término es considerado una unidad multifacetada, de carácter poliédrico ya que puede ejercer distintas funciones según el contexto y la situación presente.

- Cada unidad poliédrica es al mismo tiempo una unidad lingüística, cognitiva y sociocultural.
- Los términos son unidades lingüísticas de conocimiento específico y de comunicación especializada que tienen una doble funcionalidad: Representar el conocimiento especializado y constituir el vehículo para la comunicación de dicho conocimiento.
- Los términos deben ser estudiados en función de su uso utilizando una metodología descriptiva.

Bajo los postulados de la TCT se han orientado diversos estudios sobre UF que tienen como objetivo la descripción de estas unidades en diversos ámbitos de especialidad. Son representativos de esta línea de investigación los trabajos de Estopà (1999), de Bevilacqua (1999) y de Orente (2001).

De las dos corrientes teóricas expuestas anteriormente, este trabajo retoma los planteamientos de la TCT por considerar los términos unidades que se activan según el contexto donde se realizan y reconocer la fraseología como unidades terminológicas, como un fenómeno real y relevante del lenguaje. Así mismo, por considerar los términos como unidades poliédricas. Lo anterior permite visualizar las unidades fraseológicas, específicamente las UFE eventivas, desde tres perspectivas:

- Desde una perspectiva lingüística describiendo su traducción en los productos de los programas de TA partiendo de la traducción humana. En este caso, el traductor humano ha construido el valor especializado de los términos según el uso real que tienen las unidades en el área temática.

- Desde una perspectiva cognitiva se establece si los programas de TA conservan el nivel de especialidad y transmiten el nodo conceptual de la UFE eventiva.
- Desde una perspectiva de la comunicación se evalúa la usabilidad del producto de los traductores automáticos por medio del análisis de aspectos que emergen de una observación inicial de los productos y, a partir de estos, se determina el uso que pueden tener estas traducciones.

Respondiendo a estos postulados, las UFE eventivas se ingresaron con su contexto inmediato a los programas de TA igualmente, el análisis parte del estudio de los contextos traducidos y cómo ellos influyen en la precisión, comprensión del texto y transmisión del nivel de especialidad y nodo conceptual de las UFE eventivas.

4.2.3.1 Unidades Fraseológicas Especializadas – UFE

La fraseología se estudia desde el campo de la terminología y es allí donde se establece su relación. Puede decirse que el nacimiento de ambas disciplinas tiene un origen común en la necesidad de resolver los problemas léxico-sintácticos que plantean los términos al aparecer combinados con otras palabras (Aguado, 2007).

La definición, caracterización y delimitación de la fraseología, y concretamente de las UFE, es una labor en la que siguen trabajando los investigadores. Como lo señala Orente (2001), por ser un área de estudio reciente, entre las distintas dificultades que presenta su abordaje, se encuentra

el “uso de una terminología muy variada y ambigua, donde no ha habido lugar a consensos respecto de denominaciones y contenidos”. Igualmente, no es una tarea fácil identificar unidades en el ámbito profesional que tienen un uso frecuente en el lenguaje general. Al consultar la definición de *salto*, por ejemplo, en un diccionario⁹ del lenguaje general se observa que existen once entradas relacionadas con el significado de esta unidad. La primera entrada define *salto* como “*acción y efecto de saltar*” y es solo en la quinta entrada donde se hace referencia a una definición cercana al sentido que toma la unidad en un contexto especializado “*caída de un caudal importante de agua, especialmente en una instalación industrial*”. La UFE eventiva que contiene esta unidad se forma con el NV *aprovechar*, el artículo *los* y el NT *saltos*. Al consultar la IATE¹⁰ la unidad *salto* es reconocida como término en el área de la energía y está relacionada con las centrales hidroeléctricas y las caídas de agua para la producción de energía eléctrica. Por tanto, una unidad de uso frecuente en el lenguaje general adquiere su carácter especializado cuando se encuentra en un discurso de este tipo.

Guadec (1997, p. 167) define la unidad fraseológica como: “Un phraséologisme est une chaîne de caractères, autre qu’une désignation, formant bloc spécifique, soit en vertu de sa spécialisation, soit en vertu de son itération/repetition”.

Por su parte Aguado (2007), define las UF como “cadenas léxico sintácticas transmisoras del conocimiento”.

⁹ Diccionario de la Real Academia Española.

¹⁰ IATE: InterActive Terminology for Europe.

Estas definiciones concuerdan en que las UFE son series de unidades léxicas transmisoras de conocimiento y que adquieren su carácter especializado en contexto.

Aguado (2007) propone las siguientes perspectivas de análisis para las UFE:

- Perspectiva léxico-sintáctica: para la formación de las UFE se identifican los mismos tipos categoriales presentes en los patrones colocativos. Los cinco más comunes son: N+N, V+N, N+Adj, V+Adv y Adj+Adv. El análisis de la manera como se relacionan los elementos que constituyen cada patrón colocativo permite describir las características de cada elemento, las variaciones que admite, la frecuencia con que se manifiestan, etc.
- Perspectiva léxico-semántica: en esta perspectiva existen dos corrientes, una de carácter verbal que se ocupa del contenido especializado de los verbos y otra que analiza las relaciones semánticas que se establecen entre los verbos y las unidades así, *producir* – verbo ; *Energía* – término relacionado semánticamente con el verbo.
- Perspectiva conceptual: desde aquí se observan las relaciones conceptuales que establecen los términos compuestos por varias unidades léxicas. Este análisis permite establecer relaciones entre los campos temáticos, delimitar definiciones, incluir nuevos términos, etc.

Lo antes citado permite considerar que las UFE eventivas pueden ser analizadas a partir de estas perspectivas de análisis.

Desde la perspectiva léxico-sintáctica, puede considerarse que las UFE eventivas son apropiadas para ingresar a los traductores automáticos y, analizadas bajo esta perspectiva, dado que se manifiestan dentro de los tipos categoriales presentes en los patrones colocativos descritos anteriormente como V+N. Adicionalmente, esta estructura es equivalente en la lengua origen y en la lengua meta.

Desde la perspectiva léxica – semántica, se considera relevante analizar si, al ser traducidas automáticamente, las UFE eventivas mantienen su estructura origen - V+N- y si el contenido especializado de los verbos y los nombres se transmite.

4.2.3.2 Unidades Fraseológicas Especializadas eventivas – UFE eventivas

En su trabajo Bevilacqua (2004) describe las UFE eventivas como unidades sintagmáticas formadas a partir de un NT y un Núcleo Eventivo (en adelante NE). El NT representa el nodo de conocimiento en la estructura o mapa conceptual del ámbito especializado, es de categoría nominal, tiene valor referencial y posee carácter denominativo. Por su parte el NE procede de un verbo y se manifiesta como verbo, nombre deverbal o participio.

Estas unidades se caracterizan por estar compuestas por palabras de uso general que toman una connotación especializada al inscribirse en un área específica del conocimiento. Así, la expresión *consumir energía* está compuesta por dos palabras de uso general pero al estar en el contexto de la energía solar cobran un significado especializado.

Esta cualidad convierte a estas unidades en las apropiadas para ingresar a los traductores automáticos pues se considera que estos traductores han sido creados para lenguajes generales pero que han arrojado buenos resultados en investigaciones con lenguajes especializados o campos del conocimiento restringidos (Fomicheva *et al.*, 2013; Botella, 2008; Dieguez, 2001; Gonzales, 2010).

De acuerdo con Bevilacqua (2004) estas unidades son consideradas eventivas porque se derivan o forman por un verbo que denota procesos y acciones contrario a los que denotan situación.

Estas unidades se definen por las siguientes propiedades:

- Son unidades sintagmáticas formadas por uno o más términos, que constituyen un núcleo terminológico (NT), y un núcleo eventivo (NE), de carácter terminológico o no, procedente de verbo y que se manifiesta como verbo – *consumir energía*-, nombre deverbal –*consumo de energía*- o participio –*energía consumida*-;
- El NT representa un nodo de conocimiento en la estructura o mapa conceptual del ámbito especializado, es de categoría nominal, tiene valor referencial y posee carácter denominativo. En los casos *consumir energía* - *consumo de energía* - *energía consumida*, el NT corresponde a *energía*;
- El NE es de categoría verbal o derivada de verbo (nombre deverbal o participio), es relacional y denota los procesos y acciones propios de determinada área de conocimiento o temática;

- Entre estos dos núcleos se establecen relaciones de tipo sintáctico, pero sobre todo de carácter semántico determinadas por su utilización en un ámbito temático, lo que le confiere a la unidad cierta estabilidad, es decir, un carácter semifijo;
- Son, por tanto, unidades que se conforman por y en el discurso o texto de un ámbito determinado y que pasan a tener valor especializado por la temática especializada del texto en que son utilizadas.

Lo descrito anteriormente confirma que las UFE eventivas son apropiadas para ingresar a los programas de TA dado que al ser estas unidades de uso general que adquieren su carácter especializado en contexto pueden suponer un mejor desempeño de los traductores automáticos que han sido creados, inicialmente, para traducir el lenguaje general.

En consecuencia, se analiza el corpus textual para identificar las UFE eventivas que allí tienen ocurrencia y se conforma el corpus de análisis. En el siguiente apartado se describen los aspectos metodológicos que se han tenido en cuenta para desarrollar los objetivos planteados.

CAPITULO V. METODOLOGÍA

Se describe a continuación la estrategia metodológica para analizar la traducción automática de las UFE eventivas.

En primer lugar, se describe cómo se establecieron los criterios para seleccionar los programas de TA y como se constituyó el grupo de traductores automáticos que se evaluó a partir de estos criterios.

En segundo lugar, se presentan los criterios que se han tenido en cuenta para la selección del corpus de UFE eventivas y una breve descripción de los textos seleccionados. Además, se describe el método utilizado para la extracción de las UFE eventivas asumiendo las propiedades descritas por Bevilacqua (2004) como fundamentales en este proceso.

Finalmente se describe el método de análisis utilizado para analizar los productos de los programas de TA y poder concluir la manera como los programas de traducción automática traducen las UFE eventivas.

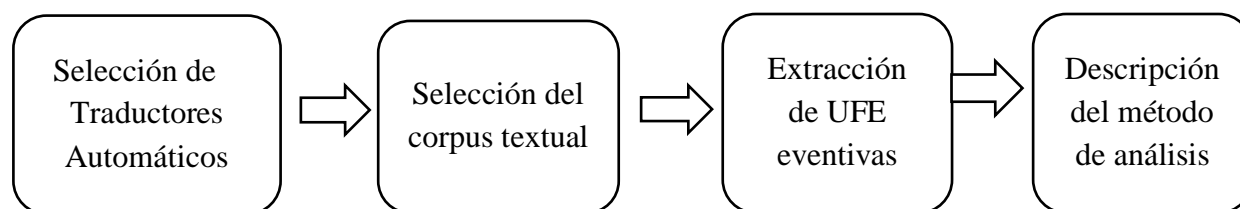


Figura 2. Diseño Metodológico

5.1 Selección de traductores automáticos

En este apartado se explica detalladamente como se seleccionaron los traductores automáticos usados en este estudio. Primero, se explica la forma como se establecieron los criterios de selección. Después, se explica cómo se constituyó el grupo de traductores automáticos analizados a partir de los criterios de selección establecidos.

5.1.1 Criterios de selección de los programas de TA

En consonancia con Botella (2008), uno de los criterios tenidos en cuenta está relacionado con el reconocimiento y permanencia en el espacio virtual de los traductores automáticos seleccionados. La autora además tiene en cuenta el criterio de método de traducción porque considera que por la manera como está concebido este método produce traducciones de “frases hechas” más adecuadas.

Por su parte, Gonzáles (2010) consideró como el criterio más relevante para la elección de sus traductores automáticos, Systran, Reverso y Google, su presencia gratuita en internet.

Para Tertoolen (2010) el método interno de traducción fue el criterio para elegir el traductor automático de su estudio. El autor considera que el método estadístico, con el que funciona Google Translator, es el más innovador y prometedor dado que está basado en alineación de corpus bilingües paralelos y la probabilidad de ocurrencia de las unidades a traducir haciendo que no

requiera que todo el vocabulario y las reglas gramaticales de una lengua estén definidos como en el método basado en reglas.

Una vez revisados los criterios de selección y los programas de TA empleados en los estudios mencionados anteriormente, se definen los siguientes criterios de selección para el presente estudio:

- Reconocimiento y permanencia en el espacio virtual, garantizando la vigencia y utilidad del análisis realizado sobre los programas de TA, es decir, que los resultados y conclusiones aquí obtenidas puedan ser relevantes para usuarios de dichos programas.
- Disponibilidad de una versión online gratuita que permita un acceso libre e ilimitado durante el desarrollo de la investigación con el fin de realizar pruebas frecuentemente y garantizar la fiabilidad de los resultados.
- Evidencia de la incorporación de diccionarios y/o corpus especializados, dado que ingresamos documentos sobre temáticas especializadas y es necesario que los traductores automáticos tengan herramientas para traducir este lenguaje, de lo contrario, es predecible que los productos presenten muchas inconsistencias.
- Participación en investigaciones similares y anteriores a la presente con el fin de tener antecedentes que corroboren que estos programas de TA tienen herramientas en sus estructuras internas y corpus para traducir unidades terminológicas especializadas.

Teniendo como base estos criterios de selección se evaluó cada uno de los traductores automáticos utilizados en los estudios mencionados anteriormente, a saber, Google Translator, Systran, Reverso, Apertium, Promt y Babelfish utilizando los símbolos A (aplica), N/A (no aplica), N/S (no se tiene información), y comentarios complementarios.

Cuadro 3. Selección de traductores automáticos

Criterios de selección	Traductores Automáticos					
	Reverso	Google	Systran	Apertium	Babelfish	Prompt
Tener reconocimiento y permanencia en el espacio virtual	A	A	A	N/A	N/A sus enlaces conducen a Bing - Microsoft	
Gratuitos y online	A	A	A	A	A	A
Deben haber sido utilizados en investigaciones similares	A	A	A	A	N/A	A
Estructura interna	N/S	Híbrido - Corpus	Híbrido - diccionarios	Basado en reglas	N/S	Basado en reglas
Diccionarios/Corpus especializados	N/S	Corpus publicados en la red	Diccionarios bilingües robustos en diferentes áreas	N/S	N/A	Se puede seleccionar entre varios campos

Tabla 3. Criterios de selección de los traductores automáticos

Criterios de selección	Traductores Automáticos		
	Google	Systran	Prompt
Reconocimiento y permanencia en el espacio virtual	Antes de Octubre, 2007	1997	1998
Gratuitos – online	http://translate.google.com/	http://www.systranoft.com/	http://www.online-translator.com/
Uso en investigaciones similares	González, M. (2010) Tertoolen, R. (2010)	(2010). Botella, MV. (2008)	Botella, MV. (2008)
Arquitectura interna	Traducción Estadística Híbrido-Corpus	Traducción basada en reglas y estadística Híbrido – Diccionario	Basado en reglas
Diccionarios / corpus especializados	Tiene incorporados corpus publicados en la red.	Posee diccionarios bilingües-robustos en diferentes áreas.	Se puede elegir entre un diccionario según la temática del texto a traducir.

Se seleccionaron, entonces, los programas de TA que cumplieran el mayor número de criterios, a saber, Google Translator, Systran y Promt. A continuación se describen las características de estos programas.

5.1.1.1 Google Translator

Este es un sistema gratuito, en línea, proporcionado por Google Inc. El traductor de Google tiene una presentación amigable para el usuario. El sitio web presenta dos rectángulos, en el primero, se ingresa el texto a traducir y, en el segundo, se genera el producto traducido. Este traductor da la posibilidad al usuario de traducir entre 71 lenguas sin ningún tipo de restricciones. Además, cuando se traducen palabras, el diccionario de Google presenta traducciones alternativas que el usuario puede elegir según su criterio.

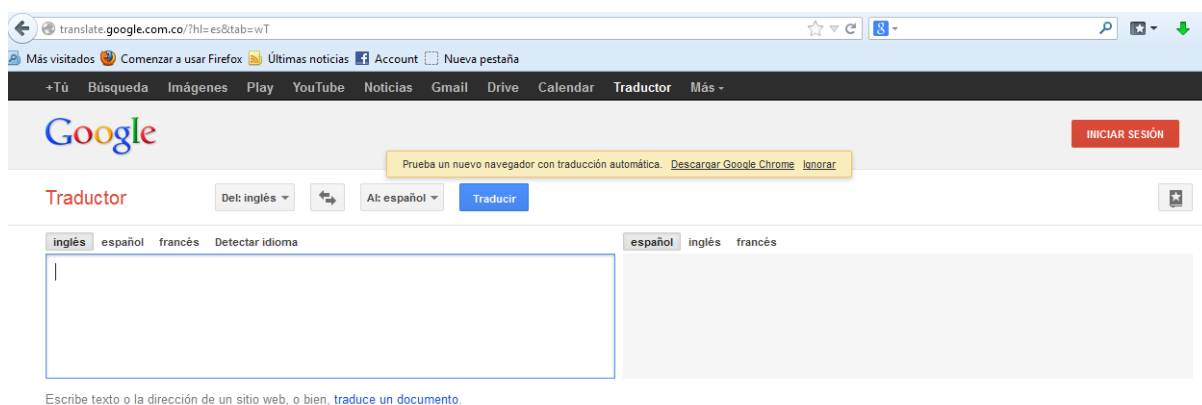


Figura 1. Interfaz Google translator

En su arquitectura de traducción no se aplican reglas gramaticales; sus algoritmos se basan en el análisis estadístico más que en el análisis por reglas. Este traductor recolecta millones de corpus paralelos bilingües de diferentes temáticas, en su mayoría producidos por traductores humanos. Primero se alinean las oraciones de los textos de la lengua origen y la lengua meta y después se calculan las probabilidades de traducción o de equivalencia, es decir, las probabilidades de que se correspondan con otras de traducciones realizadas por profesionales en ambas lenguas (González, 2010, p. 196).

Cuando Google Translate genera una traducción, este traductor busca patrones en millones de documentos para ayudar a elegir la mejor traducción según el contexto. Al detectar patrones en documentos que ya han sido traducidos por traductores humanos, Google Translate puede hacer conjeturas de lo que una traducción adecuada podría ser.

La exactitud de la traducción puede variar de acuerdo con el número de textos traducidos que contenga el programa revisados y alineados a nivel de frase. Respecto a esto, en 2007 Google introdujo a su traductor un gran número de textos de las Naciones Unidas ayudando a mejorar su precisión.

5.1.1.2 Systran

La interfaz de este traductor automático es simple y de fácil uso. Presenta la posibilidad de traducir textos, sitios web y, además, tiene un diccionario disponible en diferentes lenguas.

El texto a traducir se ingresa en uno de los rectángulos y, en el otro, el traductor automático arroja su producto. Systran permite traducir desde 15 lenguas origen pero tiene limitaciones para generar productos en el mismo número de lenguas. Así, de estas 15 lenguas origen solo 13 pueden ser traducidas al inglés y 6 al francés como lengua meta (González, 2010).

Hutchins & Somers (1992) explican que Systran está compuesto por una base de datos léxica dividida en un diccionario bilingüe principal de palabras y varios diccionarios multi-palabras contextualizadas.

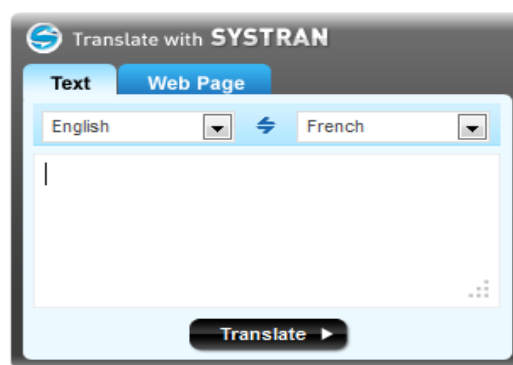


Figura 3. Interfaz Systran

Systran funciona con un sistema directo que combina la predictibilidad y consistencia de la traducción basada en reglas, y la fluidez y flexibilidad de la traducción estadística. Desde 1976, Systran diseñó una versión específica, para traducir textos legales de la Unión Europea.¹¹

¹¹ Systran cuenta también con una versión comercial no es la intencionalidad de este trabajo no se tiene la intención de realizar una comparación entre la versión gratuita y la comercial.

5.1.1.3 Promt

Como lo referencia su sitio web, el traductor automático de Promt está basado en los principios del método de transferencia.

Este sistema utiliza una arquitectura multidimensional de diccionario. Los diccionarios utilizan un método formal de descripción morfológica para todos los documentos ingresados. Lo que requiere que puedan almacenar un número ilimitado de variantes de traducciones, además de grandes volúmenes de información gramatical similar a la de los diccionarios electrónicos tradicionales. Como resultado, en Promt el ajuste temático puede efectuarse sencillamente mediante la selección de un diccionario. Entre las temáticas para elegir se encuentran la comunicación, asuntos económicos, computadoras, automóviles, medicina, viajes, deportes entre otras.



Figura 2. Interfaz Promt.

La traducción automática de PROMT cubre siete idiomas: inglés, alemán, francés, español, portugués, italiano y ruso, con un total de 24 direcciones de traducción en más de cien campos especializados. Los servicios de Internet globales de Promt, incluido el PROMT Translator, procesan aproximadamente 75 millones de peticiones de traducción en línea al mes, Alcina (2011).

5.1.2 Criterios de selección del corpus textual

Para llegar a los objetivos propuestos para esta tesis se seleccionaron los textos que forman el corpus textual a partir de los que hemos recogido las UFE eventivas.

Como se explicó anteriormente, el corpus textual está agrupado bajo el título Renovables made in Spain energías renovables en España / Renewable energies in Spain y hace parte de un proyecto impulsado por el Ministerio de Industria Español a través del Instituto para la Diversificación y Ahorro de la Energía (IDEA).

Entre los criterios¹² tenidos en cuenta para la constitución de nuestro corpus destacamos los siguientes:

- a. Paralelismo: un corpus paralelo se compone de un TB en la lengua A y sus respectivas traducciones en lengua B. (Baker, M., 1995, p. 230)

En el contexto propio de esta investigación, el TB está en español, la TR en inglés y el Texto Meta corresponde a los productos de los traductores automáticos.

¹² Para determinar los criterios de selección del corpus se tuvieron en cuenta los parámetros de Suárez (2004, p.5).

b. Nivel de especialidad: se han seleccionado artículos escritos por especialistas pero que puedan ser comprendidos por un público lego, es decir, personas no especialistas en la temática.

c. Unidades de análisis: se ha seleccionado un corpus donde, previamente, se verificó la presencia de UFE eventivas que cumplieran con las características de las unidades descritas por Bevilacqua (2004).

Así, el corpus textual paralelo elegido está conformado por 10 textos cuya lengua origen es el español con sus respectivas traducciones al inglés (68.000 palabras aproximadamente).

5.1.3 Identificación y selección del corpus de análisis

Como se explicó, las unidades que se ingresaron a los programas de TA son las UFE eventivas definidas por Bevilacqua (2004), como unidades sintagmáticas formadas por uno o más términos que constituyen su NT, y un NE. Según la autora entre estos dos núcleos se establece relaciones de tipo sintáctico y semántico determinadas por su uso en determinado campo temático.

De la investigación de Bevilacqua (2004) retomamos para nuestro estudio la aplicación y constatación de algunos de los principios de la TCT que le permiten demostrar que las UFE eventivas son unidades que hacen parte del lenguaje general del hablante pero cumplen una función discursiva en un determinado ámbito temático o de especialidad, es decir, cumplen el papel de unidades representativas y transmisoras de conocimiento especializado. Así, la expresión *consumir*

energía está compuesta por dos palabras de uso general pero al estar en el contexto de la energía solar cobran un significado especializado.

Esta cualidad convierte a estas unidades en las apropiadas para ingresar a los traductores automáticos pues se considera que estos traductores han sido creados para lenguajes generales pero que han arrojado buenos resultados en investigaciones con lenguajes especializados o campos del conocimiento restringidos (Fomicheva *et al.*, 2013; Ghoniem, Tokal, Tawfik, 2011; Toledo, 2009).

Bevilacqua (2004) confirma su hipótesis sobre la realización de las UFE eventivas en tres estructuras, verbo (*almacenar calor*), nombre deverbal (*almacenamiento de calor*) y participio (*calor almacenado*), derivadas de una misma estructura subyacente formada por un *núcleo eventivo* NE (*almacenar*) de carácter relacional y por un NT (*calor*) que corresponde a una entidad. En este trabajo recogimos y analizamos solamente las UFE eventivas formadas por la estructura Núcleo Verbal (NV) + Núcleo Terminológico (NT) por ser esta una estructura equivalente en ambas lenguas, es decir, que es una estructura que se usa con similar frecuencia en contextos y circunstancias afines en la lengua origen español y en la lengua meta inglés; por ejemplo: *Producir biodiesel / to produce biodiesel*. La similitud en el uso de esta estructura en ambas lenguas podría favorecer los resultados obtenidos por los programas de TA al traducir las UFE eventivas constituidas por esta combinación léxica.

Por lo expuesto anteriormente, se determinó identificar las unidades de análisis a partir de la estructura subyacente NV+NT. Para establecer los NV que tenían ocurrencia en el corpus textual se tomaron como base los 26 NE descritos por Bevilacqua (2004):

absorber, acumular, almacenar, aprovechar, aumentar, captar, concentrar, conservar, consumir, convertir, crear, distribuir, emplear, gastar, generar, hallar, obtener, producir, proporcionar, recibir, suministrar, tomar, transformar, transmitir, usar, utilizar.

Cabe recordar que para establecer su corpus de análisis Bevilacqua (2004) recogió solamente las unidades formadas por el término *energía* y los NV que lo acompañaban, así, surgen los 26 NE ya mencionados.

Estos NE se ingresaron a la herramienta Antconc 3.2.0w para extraer todas las concordancias presentes en el corpus textual con sus respectivos contextos.

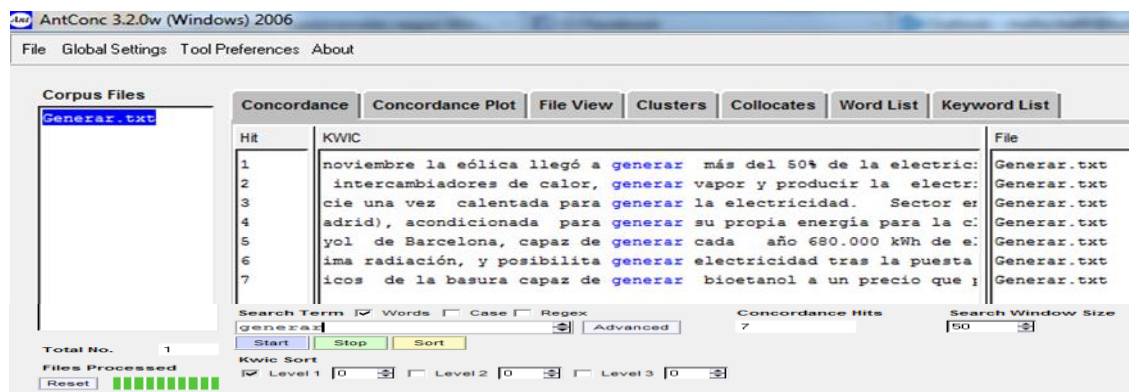


Figura 3. Ingreso de la estructura superficial *generar* al programa ANTCONC 3.2.0w y sus apariciones en la lengua origen.

De los contextos identificados a través de esta herramienta se seleccionaron aquellos donde el NV estaba acompañado por un NT que debía cumplir con las características descritas por Bevilacqua, (2004):

- El NT debería ser una unidad nominal que correspondiera a un nodo cognitivo en el ámbito tratado. En el contexto “...permiten *aprovechar saltos* de un metro gracias a turbinas hidráulicas de elevada velocidad específica...”, por ejemplo, el NT *saltos* puede tener varias acepciones en el lenguaje general que se limitan al estar inmersos en el área temática de las energías renovables; mientras el diccionario de la RAE¹³ presenta 18 acepciones para *salto* la IATE¹⁴ tiene una sola definición específica para el área de la energía.

- Debería ser, sintácticamente, su complemento o sujeto según el caso.

Siguiendo el mismo ejemplo, *saltos* es, sintácticamente, el complemento del verbo transitivo *aprovechar* dado que es un complemento directo para dicho verbo.

De este modo, se obtiene un corpus de análisis de 20 ocurrencias de UFE eventivas formadas por la estructura NV+NT que se describen a continuación: (Tabla 4).

¹³ Diccionario de la Real Academia de la Lengua Española.

¹⁴ Inter-Active Terminology for Europe.

Tabla 4. Frecuencia de las ocurrencias de las UFE eventivas con estructura NV+NT

NV	NT	Total
Almacenar	el calor	1
	Energía	1
	la energía	1
Aprovechar	el potencial energético	1
	Saltos	1
Generar	Vapor	1
	Electricidad	1
	Bioetanol	1
Producir	electricidad	3
	biogás / biodiesel	1
	Energía	1
	Electricidad	1
Suministrar	energía eléctrica	1
	Calefacción	1
	Refrigeración	1
	Agua	1
Usar	Gas	1
	alcohol etílico	1
Total		20

Esta organización de la información recogida sirvió para conocer los NV, sus respectivos NT y su frecuencia de aparición para la estructura NV+NT. Debe mencionarse que la frecuencia de ocurrencia de las unidades no es un factor relevante para considerar una combinación léxica como UFE eventiva porque, como lo sustenta Bevilacqua (2004) “el valor especializado de las unidades recogidas está dado por otros factores que no la frecuencia” (p. 44). Entre los factores mencionados se encuentran la relación sintáctica que, establece entre los elementos que conforman la unidad y la correspondencia del NT a un nodo cognitivo del área tratada.

De esta forma, los núcleos identificados muestran, superficialmente, la concentración del conocimiento referente a los procesos y usos de las Energías Renovables. Observamos, entonces que los NT *energía* y *electricidad* tienen un mayor número de ocurrencias y por eso se consideran

los nodos nucleares o representativos respecto de la temática. Los que tienen una frecuencia menor son considerados nodos periféricos porque son aquellos que se sitúan alrededor de los nucleares.

5.2 Metodología de análisis de las UFE eventivas en los productos de los TA

Para analizar los productos de los programas de TA se retomarán aspectos relevantes de algunos métodos para evaluar la TA mencionados anteriormente.

Del sistema de métricas se retoma el uso de una traducción humana como referencia que se compara con los productos de los traductores automáticos con el objetivo de analizar aspectos morfo-sintácticos, léxico-semánticos y contextuales *in situ*, y validar la precisión con que estos programas traducen unidades terminológicas específicamente las 20 UFE eventivas que conforman el corpus de análisis.

En el aspecto morfo-sintáctico se consideran las formas que adoptan las UFE eventivas y si estas se ajustan o no a las modificaciones obligatorias de la lengua meta. Se analiza la conservación o no de las categorías y estructuras gramaticales del TB y de la TR en los productos de los TA y su repercusión con relación a su inteligibilidad.

Así, por ejemplo, se contrastó la traducción de la unidad *usar gas*, traducida en la TR como *using gas* y por Google como *using gas*. Como se observa Google traduce la UFE eventiva utilizando la misma estructura gramatical que la TR, realiza los ajustes requeridos en la lengua meta para que esta pueda ser leída como un texto de esta cultura, y como resultado se mantiene la

transmisión del nodo de conocimiento que contiene la unidad en la traducción de Google. Esto a su vez permite mantener la correspondencia del NT *gas* a un nodo cognitivo del área tratada.

Se debe reiterar que la evaluación de la traducción de las UFE eventivas a partir de una TR no responde a una visión idealizada de la traducción sino que pretende ser un criterio para observar la manera como estas unidades son traducidas.

En el aspecto léxico- semántico se considera la conservación, cambio o restricción del uso del NT de la unidad con relación a la TR y sus implicaciones en la transmisión del nodo cognitivo que las UFE eventivas contienen.

En el aspecto contextual se analizan los elementos del fragmento que contiene la UFE eventiva desde el nivel morfológico, sintáctico y semántico con el objetivo de identificar errores en la traducción que pueden interferir en la inteligibilidad de la unidad y del fragmento mismo y la relación de la unidad especializada con las demás unidades léxicas.

Entre los aspectos considerados se encuentran:

- Las relaciones sintácticas entre los elementos que componen el fragmento.
- Ajustes a las modificaciones obligatorias de la lengua meta.
- Organización de las unidades léxicas en el fragmento.
- Selección de unidades léxicas.

Los aspectos considerados para el análisis de los productos de los traductores automáticos y su respectiva evaluación coinciden con los criterios de análisis de las UFE planteadas por Cabré *et al.*, (1996, p. 71).

A modo de ejemplo se presenta la siguiente UFE eventiva:

Cuadro 4. Ejemplo (1)

TO	TR	Producto Systran
España se está preparando para <i>aprovechar el enorme potencial energético</i> del mar y sumarlo a las restantes tecnologías renovables y así avanzar en el desarrollo de una estrategia energética fiable y segura.	Spain is preparing <i>to benefit from the huge potential</i> of the sea and add it to the rest of its renewables portfolio in order to progress with the development of a reliable and secure energy strategy.	Spain is preparing <i>to take the huge potential energy</i> of the is and the remaining add it renewable technologies and under progress in the development of to reliable and secure energy strategy.

Al comparar la TR y el producto de Systran y revisar aspectos morfo-sintácticos puede observarse que la UFE eventiva, en ambos casos, es traducida con la misma estructura gramatical: un NV *to benefit*, para la TR, y *to take*, para Systran, y un NT *huge potential*, para el primero, y *huge potencial energy*, para el segundo. Debe subrayarse que la TR utiliza *potential* como sustantivo y elimina *energy* al realizar ajustes del texto a las características de la lengua meta y evitar la repetición de información y del mismo término.

Por su parte, Systran sigue el patrón del TB utilizando dos adjetivos *huge potential* para calificar *energy*, así, no se realizan ajustes en la traducción de la unidad y se genera repetición de

información. Como consecuencia, la traducción de la UFE eventiva realizada por el traductor automático conserva la estructura del TB pero no se ajusta a las modificaciones obligatorias de la lengua para que el texto se asemeje a uno de la lengua meta.

Al revisar aspectos léxico-semánticos, se observa que Systran no logra transmitir la idea general del nodo conceptual de la UFE eventiva y tiene dificultades para conservar el nivel de especialidad de la unidad. El NV *aprovechar* es traducido como *to take* que tiene nueve acepciones como verbo en el Oxford dictionaries¹⁵ y ninguna puede considerarse como equivalente del NV expresado en el TB. Esto puede indicar que el lenguaje general es impreciso. Al traducir el NT *potencial energético* como *potential energy* Systran realiza un calco con transposición de la estructura adjetivo+sustantivo equivalente en ambas lenguas pero que, en este caso particular, resulta en la elección de un término que hace referencia a entidades distintas afectando el sentido del texto. El Oxford dictionaries define *potential energy* como “*the energy possessed by a body by virtue of its position relative to others, stresses within itself, electric charge, and other factors.*” Contrario al sentido dado a *potencial* en el TB que hace referencia a los múltiples recursos que en el mar pueden utilizarse para producir energía.

Al analizar aspectos contextuales se ponen de manifiesto errores de traducción como omisión de unidades léxicas (*mar* se traduce como *is*), dificultad para organizarlas de tal forma que cumplan con las exigencias de la lengua meta (*and the remaining add it*) y dificultad con el uso de las preposiciones (*the development of to reliable*).

¹⁵ Se encuentra en la página <http://www.oxforddictionaries.com/definition/english/take>

Además de las dificultades específicas relacionadas con la traducción de la UFE eventiva, los errores a lo largo del fragmento obstaculizan la inserción e inteligibilidad de la unidad en el fragmento que la contiene.

De manera particular, con el análisis de la unidad *aprovechar el potencial energético* en el traductor automático Systran se pudo constatar que:

- La conservación de la estructura sintáctica del TB o de la TR no garantiza la selección de términos equivalentes.
- Con el análisis de este caso se podría pensar que Systran muestra dificultades para desambiguar esta unidad que presentan polisemia. lo que generaría imprecisiones en su producto.
- Los errores en la traducción del contexto que contiene la UFE eventiva son determinantes para garantizar la transmisión del nodo de conocimiento que contiene la unidad y su inteligibilidad.

5.3 Síntesis

En este capítulo se describió como se seleccionaron las herramientas de TA, el corpus textual, el corpus de análisis y la metodología para analizar los productos de los programas de TA.

Google Translator, Systran y Promt fueron los tres traductores automáticos seleccionados por cumplir los criterios establecidos:

- Reconocimiento y permanencia en el espacio virtual.
- Disponibilidad de una versión online gratuita.
- Evidencia de incorporación de diccionarios y/o corpus especializados.
- Participación en estudios similares.

Entre los criterios tenidos en cuenta para la constitución del corpus textual se encuentran:

- Paralelismo
- Nivel de especialidad
- Presencia de UFE eventivas.

A partir de los 26 NE propuestos por Bevilacqua (2004) se identificaron, en el corpus textual conformado por 10 textos bajo el título “Renovables made in Spain, 20 UFE eventivas con estructura subyacente NV+NT. Las unidades se ingresaron a los tres traductores automáticos (Google, Systran y Prompt) con sus respectivos contextos.

La metodología para analizar los productos de los programas de TA sigue los siguientes pasos:

- Observación de los productos de los tres traductores automáticos de donde emergen los tres aspectos a analizar: aspecto morfo-sintáctico, aspecto léxico-semántico y aspecto contextual.

- Contratación de los productos de los traductores automáticos con la TR para determinar lo que sucede con los aspectos a analizar.
- Evaluación de los productos de los programas de TA respecto a cómo traducen las UFE eventivas.

En el siguiente capítulo se presentan los resultados y las conclusiones obtenidas a partir del análisis de los datos.

CAPITULO VI. ANÁLISIS Y RESULTADOS

Los resultados se analizan y evalúan con respecto al diseño metodológico planteado en el capítulo 4. Primero, se ingresaron a cada uno de los traductores automáticos las 20 UFE eventivas identificadas en el corpus textual con la estructura NV+NT.

Posteriormente, los productos de cada uno de los programas de TA son analizados en dos etapas:

- Observación de los productos de los tres traductores automáticos de donde emergen los tres aspectos a analizar: aspecto morfo-sintáctico, aspecto léxico-semántico y aspecto contextual.
- Contrastación de los productos de los traductores automáticos con la TR para determinar lo que sucede con los aspectos a analizar.

Finalmente, se triangulan los datos generados por cada uno de los traductores automáticos.

Para la presentación de los resultados del análisis cualitativo primero se describen y ejemplifican los datos particulares de cada uno de los programas de TA según los aspectos a analizar seguidamente, se detallan y ejemplifican los resultados de la triangulación de los datos de cada uno de los traductores automáticos.

6.1 Comparación productos de los programas de TA con la TR

6.1.1 Aspectos morfo-sintácticos

6.1.1.1 Google Translator

Al comparar los productos de Google con la TR se observa que el 11% de las UFE eventivas traducidas coinciden en adoptar la misma estructura gramatical de la TR, el 50% de las unidades son traducidas con la misma estructura del TB que coincide en ser la misma de la TR, el 33% restante mantienen la estructura del TB, siendo esta diferente a la estructura adoptada por la TR.

Así, el 89% de las UFE eventivas son traducidas con la misma estructura sintáctica del TB. Puede pensarse que este traductor automático privilegia la traducción literal de las UFE eventivas con el riesgo de no cumplir con las funciones en la cultura meta.

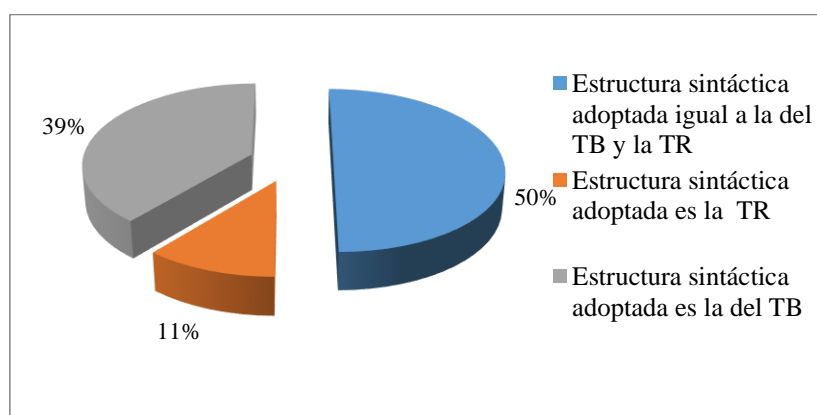


Gráfico 1. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google con la TR

Contrario a los resultados generales de Systran y Prompt, para Google el porcentaje de similitud entre la estructura usada por la TR y los productos de los traductores automáticos respalda

la idea de que la estructura NV+NT es equivalente en ambas lenguas y puede garantizar mejores resultados para la TA.

Del total de UFE eventivas traducidas por Google que comparten la misma estructura de la TR el 34% utilizan el mismo término para traducir el NV y el NT que la TR. El bajo porcentaje de similitud entre la TR y la traducción de Google es un indicativo de que pueden existir dificultades en la transmisión del nodo de conocimiento de la unidad debido a las limitaciones que tiene el traductor automático para desambiguar aquellas que son de uso general en la lengua origen pero que tienen un nivel de especialidad mayor en la lengua meta.

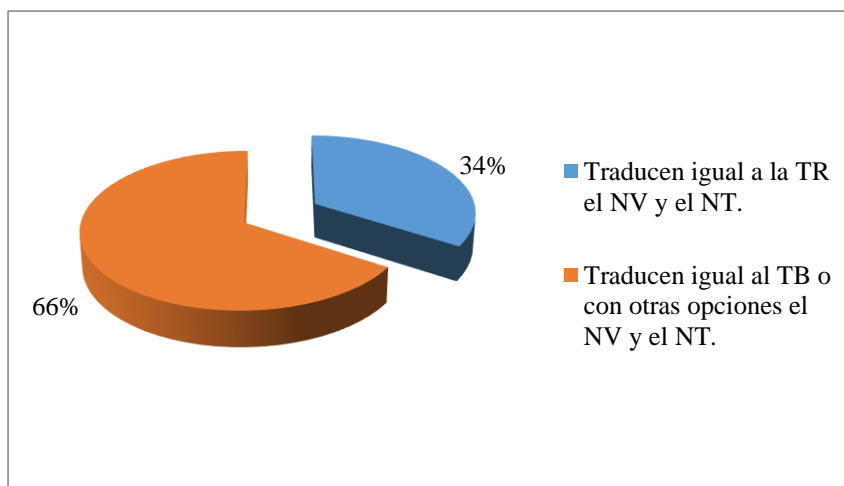


Gráfico 2. Similitud en la traducción de los NV y NT con la TR – Google

Cuadro 5. Ejemplo (2)

TO	TR	Producto Google
El proyecto perseo busca, en concreto, desarrollar una tecnología utilizando los residuos orgánicos de la basura capaz de <i>generar bioetanol</i> a un precio que permita competir con la gasolina sin la exención del impuesto especial de hidrocarburos.	The specific objective of project perseo is to develop a technology <i>to obtain bioethanol</i> from organic waste at a price that can compete with gasoline, without requiring the special hydrocarbons tax exemption.	the perseus project seeks, in particular, develop a technology using organic waste from the garbage can <i>generate ethanol</i> at a price that can compete with gasoline without the waiver of the special tax on hydrocarbons.

En este fragmento se encuentra la UFE eventiva *generar bioetanol*. Google traduce la unidad con la estructura NV+NT del mismo modo que la TR. Sin embargo, la coincidencia en la estructura gramatical no garantiza la selección de equivalentes pertinentes para el campo de especialidad. En este caso, el NV es traducido como *generate*, seguramente, por ser una palabra cognado, es decir, por la cercanía morfológica de esta unidad léxica en la lengua origen y en la meta. No sucede lo mismo con la traducción del NT que es traducido por la TR como *bioethanol* y por Google como *ethanol*.

Químicamente estos dos compuestos son difíciles de distinguir. Su diferencia radica en su origen: la partícula *bio* hace referencia a un origen natural. La traducción de Google resulta ser imprecisa dado que *etanol*, aunque en determinadas ocasiones tiene un significado coloquial o general, en contextos especializados se activan características especializadas¹⁶ mientras *Bioetanol*

¹⁶ RAE, en <http://lema.rae.es/drae/?val=bioetanol>

hace referencia a características más específicas del compuesto. Como consecuencia la traducción de Google logra transmitir el sentido del texto pero no logra conservar el nivel de especialidad del nodo conceptual.

6.1.1.2 Systran

Si se comparan los productos de Systran con la TR se observa que el 11% de las UFE eventivas traducidas coinciden en adoptar la misma estructura gramatical de la TR, el 33% de estas unidades son traducidas con la misma estructura del TB en el TR y el 55% de las unidades traducidas por los programas de TA mantienen la estructura del TB siendo esta diferente a la estructura adoptada por la TR. Este último porcentaje podría suponer que Systran privilegia la traducción literal de las UFE eventivas con la posibilidad de no lograr transmitir el nodo de conocimiento de la unidad y su nivel de especialidad. Aunque el porcentaje de similitud entre la TR y el producto de Systran es menor que el de Google, este resultado favorece la idea de que la estructura NV+NT es equivalente en ambas lenguas y podría garantizar mejores resultados para la TA.

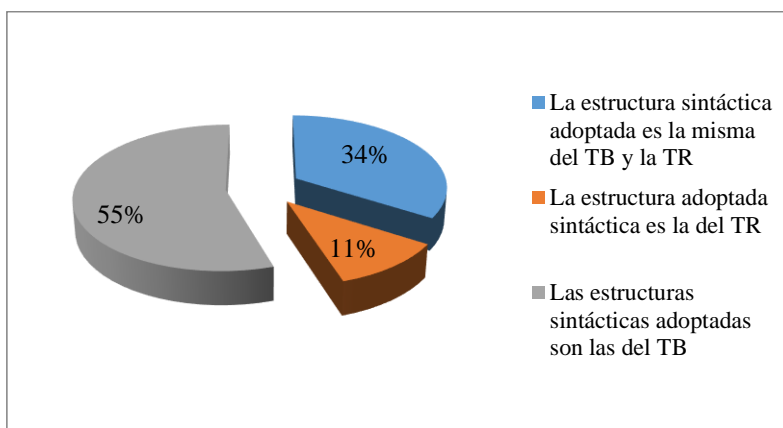


Gráfico 3. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Systran con la TR

Del total de UFE eventivas traducidas por Systran que comparten la misma estructura de la TR el 50% utilizan el mismo término para traducir el NV y el NT que la TR.

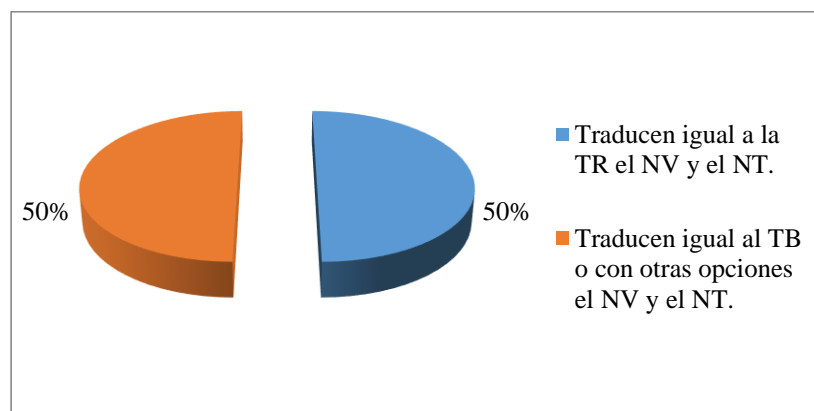


Gráfico 4. Similitud en la traducción de los NV y NT con la TR - Systran

Contrario a Google, Systran tiene un mayor porcentaje de similitud con la TR. Puede suponerse que los corpus y diccionarios de este traductor automático tienen incorporados un mayor número de términos pertenecientes al área de las energías renovables o temáticas afines demás, que el producto es más preciso respecto a la traducción de la terminología.

Cuadro 6. Ejemplo (3)

TB	TR	Systran
La primera fuente de energía que utilizó el hombre para <i>producir electricidad</i> fue el agua.	The first source of energy used by man to <i>produce electricity</i> was water.	The first energy source that the man used to <i>produce electricity</i> was the water.

En este fragmento se encuentra la UFE eventiva *producir electricidad*. Systran traduce la unidad con la estructura NV+NT del mismo modo que la TR y coinciden totalmente en la selección de los términos. Esta similitud tiene origen en la equivalencia de la estructura gramatical y en la similitud morfológica de las unidades que conforman la UFE eventiva en ambas lenguas. La correspondencia entre estos dos aspectos lleva al traductor automático a resolver de manera conveniente la traducción de la unidad y a transmitir el nodo cognitivo del término y su nivel de especialidad.

6.1.1.3 Promt

Al contrastar los productos de Promt con la TR se observa que el 11% de las UFE eventivas traducidas concuerdan en el uso de la misma estructura sintáctica de la TR (NV+NT), es decir, coinciden en la estructura gramatical sin considerar aspectos relacionados con la transmisión del significado; el 23% de estas unidades son traducidas con la misma estructura del TB en el TR y el 66% de las unidades traducidas por los programas de TA mantienen la estructura del TB siendo esta diferente a la estructura adoptada por la TR.

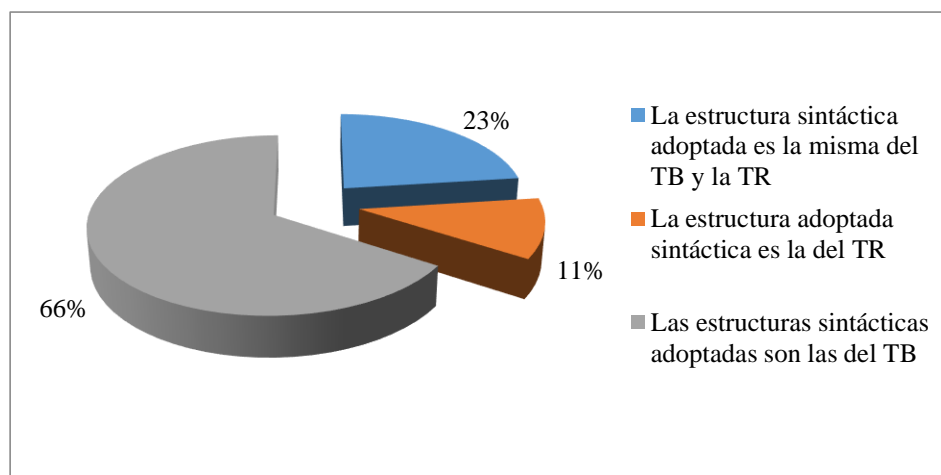


Gráfico 5. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Prompt con la TR

Prompt presenta porcentajes bajos de similitud con la estructura gramatical de la TR y el TB si se compara con Google y Systran. Peinado (2004) afirma que la proximidad de la TA con la de un traductor humano profesional puede considerarse una mejor traducción por lo que se generan inquietudes respecto a la precisión con que este traductor automático traduce las UFE eventivas.

Lo que se observa es el uso de estructuras gramaticales alternas, como el uso del *-ing*, que analizadas en contexto no responden a las modificaciones obligatorias de la lengua meta y generan pérdida de sentido y nivel de especialidad del texto.

Del total de UFE eventivas traducidas por Prompt que comparten la misma estructura de la TR el 50% son traducidas con el mismo NV y NT que la TR. Como se observa el porcentaje en la similitud del uso de los términos es mayor comparado con la similitud en el uso de las estructuras gramaticales entre Prompt y la TR. Estos datos muestran que la fortaleza de este programa de TA se

encuentra en las unidades léxicas que contienen sus diccionarios pero no en su método de desambiguación sintáctico y morfológico.

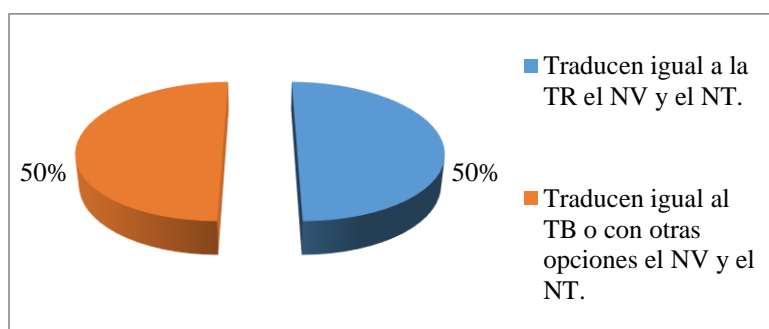


Gráfico 6. Similitud en la traducción de los NV y NT con la TR - Promt

Cuadro 7. Ejemplo (4)

TB	TR	Prompt
Sistema de sales, que permite <i>almacenar la energía</i> sobrante del campo solar en las horas de máxima radiación, y posibilita generar electricidad tras la puesta de sol.	Molten salts, enabling surplus <i>energy</i> in the solar field during the period of maximum radiation <i>to be stored</i> to generate electricity after sunset.	System of salt, which allows <i>to store</i> the remaining <i>energy</i> of the solar field in the hours of maximum radiation, and makes possible to generate electricity after the putting sun.

En este fragmento se encuentra la UFE eventiva *almacenar la energía*. Prompt traduce la unidad con la estructura NV+NT igual al TB pero diferente al TR que utiliza la estructura pasiva usada con frecuencia en la lengua meta en el discurso especializado.

El traductor automático realiza acomodaciones en la unidad buscando ajustarse a la lengua meta, así en la traducción el adjetivo *sobrante* es desplazado antes del sustantivo *energy* como esta

lengua lo requiere. La traducción humana y la automática coinciden en la elección de los términos *store* y *energy* lo que llevaría a la transmisión del nodo conceptual.

Es posible que el producto de Promt no se lea como un texto de la cultura meta pero sí logra comunicar el mensaje del TB.

6.1.1.4 Resultados de la triangulación

En el contraste de los productos de los programas de TA – Google, Systran, Promt- con la TR se evidencia una baja similitud en las estructuras sintácticas adoptadas, el 48% comparado con el 52% de similitud con el TB.

Por el contrario se presenta una alta correspondencia entre la traducción del NV (54%) y, sobre todo, con el NT (73%) que realiza el traductor automático con la TR.

Esto puede indicar que la estructura NV+NT, equivalente en ambas lenguas, no garantiza mejores resultados para la TA dado que para lograr una traducción inteligible es necesario realizar adaptaciones a las normas y convenciones de la cultura meta (valoración de aspectos lingüísticos, semánticos y pragmáticos) y no simplemente usar la misma estructura gramatical.

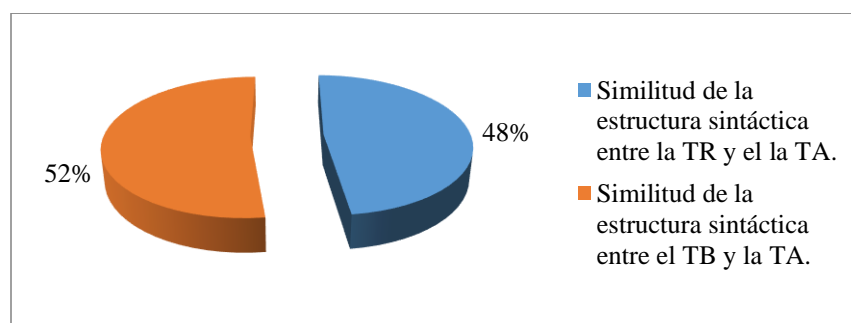


Gráfico 7. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google, Systran, Prompt con la TR y el TB

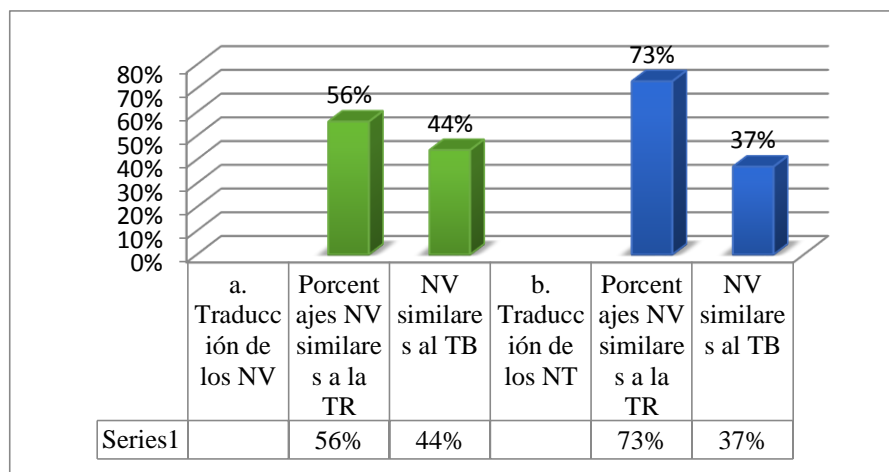


Gráfico 8. Similitud en la traducción de las UFE eventivas entre la TR y la TA

La diferencia entre el porcentaje de coincidencia entre el NV y el NT puede deberse al hecho que las UFE que se forman con un NV representan mayores problemas al ser traducidas “por la no consideración de especializado que tiene un verbo y porque no suelen estar incluidos en los diccionarios del dominio” (Aguado, 2007, p. 6) así, se asigna a los NV equivalentes del lenguaje general que no corresponden a la precisión requerida en el ámbito especializado en el que se encuentran inscritos. El alto porcentaje de similitud en la traducción de NT puede deberse a que los diccionarios y corpus de los programas de TA contienen estas unidades terminológicas, además, de indicar el nivel de precisión con que las UFE eventivas son traducidas por estos programas ya que el NT corresponde a un nodo cognitivo en el ámbito tratado (Bevilacqua, 2004) y que los traductores automáticos puede reconocer o al menos mantener la relación sintáctica y semántica entre los elementos que conforman la UFE eventiva.

Es importante mencionar que el 62% de las unidades traducidas por los programas de TA mantienen la estructura del TB siendo esta diferente a la estructura adoptada por la TR. Solamente, el 11% de las unidades traducidas corresponden a la estructura de la TR; este porcentaje es igual para los tres programas de traducción automática. El 27% coinciden en tener la misma estructura en el TB y la TR, es decir, los traductores automáticos privilegian la traducción literal de las UFE eventivas. No debe atribuirse una idea negativa a la traducción literal pues, como lo afirma Nord (2010) esta no debe descartarse porque su utilidad depende de la función del texto meta. Sin embargo, la autora advierte que una traducción que usa estructuras análogas a las del TB corre más peligro de no lograr las funciones pretendidas para los receptores de la cultura meta.

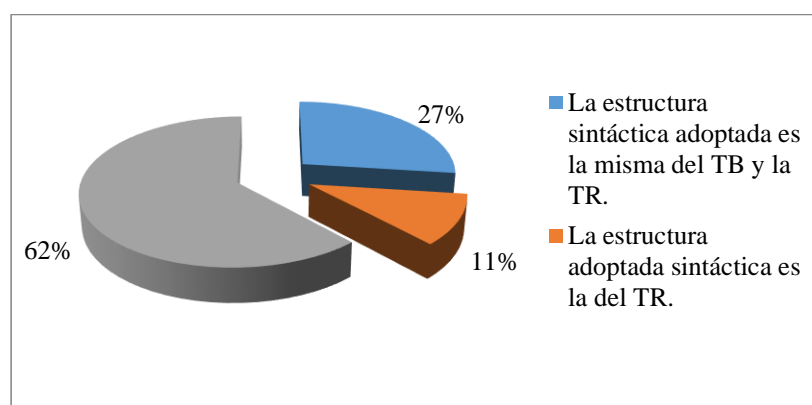


Gráfico 9. Similitud en la estructura sintáctica de las UFE eventivas traducidas por Google, Systran, Prompt con la TR

6.1.2 Aspectos léxico – semánticos

El análisis de los aspectos léxico-semánticos está centrado en los NT que conforman las UFE eventivas. Su relevancia, como lo explica Bevilacqua (2004), se debe a que estos representan el nodo de conocimiento en la estructura o mapa conceptual del ámbito especializado, es decir, los

NT transmiten y vehiculan el conocimiento del área temática a la que pertenecen y su evaluación permite conocer la calidad global de la traducción evaluada. Además como lo explica la misma autora, el NV puede tener carácter terminológico o no lo que posibilita el uso de unidades léxicas del lenguaje general como equivalentes de este núcleo sin que pierdan su valor semántico.

6.1.2.1 Google

Google traduce el 60% de las UFE eventivas utilizando el mismo NT que la TR. Adicionalmente, este traductor automático muestra que sus métodos estadísticos tienden a traducir de manera coherente y sistemática los NT puesto que un mismo término es traducido de la misma forma en cada una de sus apariciones, tal es el caso de *electricity* y *ethanol*. La coherencia proporciona una idea de la precisión de Google al traducir las UFE eventivas; la sistematicidad puede responder al privilegio que dan sus métodos estadísticos de valorar unidades que tengan un porcentaje mayor de uso sin que, necesariamente, el término seleccionado corresponda a las necesidades contextuales en las que se enmarca.

El 40% de las UFE eventivas restantes utilizan un NT diferente al de la TR dado que siguen el patrón indicado por el TB sin considerar aspectos contextuales que validan otras alternativas, es decir, si en el TB el NT es *electricidad* el programa de TA traducirá *electricity*. Al igual que en el ejemplo (1), Google traduce *potencial energético* por *potencial energy* realizando modificaciones obligatorias de la lengua meta (inversión del adjetivo y el sustantivo) pero al seguir el patrón del

TB cambia el sentido de la UFE eventiva puesto que *potential energy* hace referencia a una entidad diferente a la indicada en el TB.

Aunque la coherencia y sistematicidad al traducir los NT puede considerarse un aspecto positivo en la TA, también causa dificultades cuando su uso es generalizado. Así por ejemplo, Google traduce como *ethanol* todos los combustibles obtenidos a partir de la fermentación de una materia prima (la caña, la remolacha) sin considerar que existen términos específicos para indicar la materia de la que estos se derivan, entre ellos, *bioethanol* y *ethyl alcohol*.

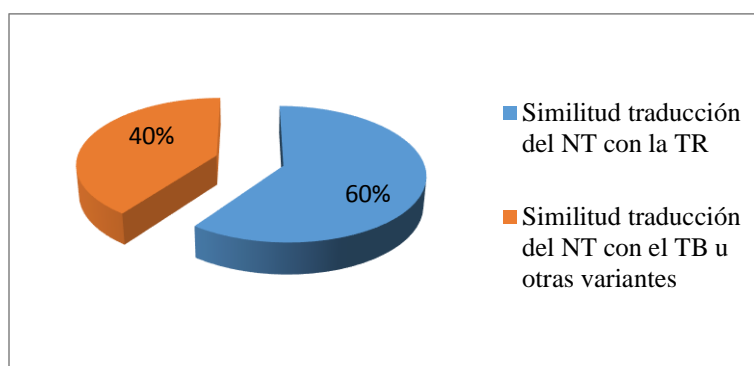


Gráfico 10. Traducción NT – Google

6.1.2.2 Systran

Al igual que Google, Systran traduce el 60% de las UFE eventivas utilizando el mismo NT que la TR lo que indica que, a pesar de utilizar arquitecturas distintas sus corpus y diccionarios contienen estas unidades terminológicas. Systran demuestra su coherencia y sistematicidad léxica

al traducir *electricity* y *energy* de igual forma en todos los contextos en que tienen aparición sin que esta generalización comprometa contextos donde otras alternativas son posibles.

Siguiendo la tendencia de Google, el 40% de las UFE eventivas restantes utilizan un NT diferente al de la TR dado que siguen el patrón del TB sin considerar aspectos contextuales que validan otras alternativas. En este mismo sentido, Systran generaliza las reglas de la lengua origen e intenta ajustarlas a la lengua meta, es el caso de *alcohol étílico*. El traductor automático reconoce que, en la lengua origen, *etílico* cumple la función de calificar el sustantivo, sin embargo, en la lengua meta, este es un nombre compuesto por dos sustantivos. Por tal razón, al traducir el NT Systran utiliza *ethylic alcohol*, adjetivo +sustantivo, contrario al equivalente dado por la TR *ethyl alcohol*. Aunque se logra transmitir la idea general del TB estas imprecisiones causan pérdida de precisión y sentido en la traducción de la unidad afectando la inteligibilidad de la misma y la posibilidad de ser leída como un texto de la lengua meta.

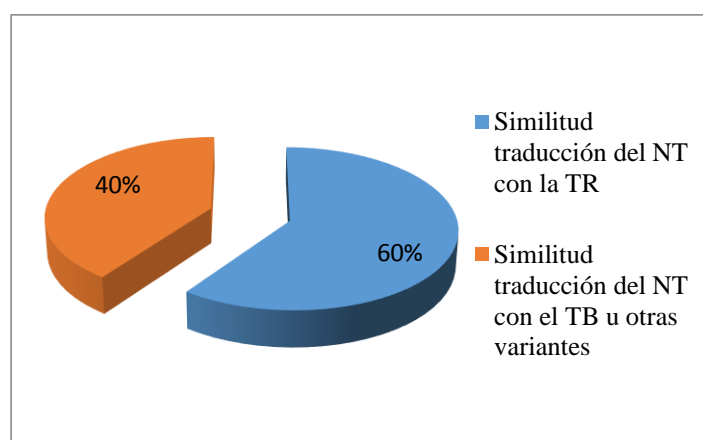


Gráfico 11. Traducción NT – Systran

6.1.2.3 Promt

Prompt traduce el 50% de los NT de manera similar al TR y el otro 50% adopta formas diversas, esto hace referencia a la selección de una unidad similar tanto morfológica como semánticamente a la TR. La observación de este fenómeno en contexto lleva a plantear que el traductor automático no cuenta con una arquitectura que le permita alcanzar un nivel aceptable respecto a la traducción de unidades léxicas especializadas.

A modo de ejemplo se puede observar la UFE eventiva *producir biogás/biodiésel*. Al traducirla, Prompt conserva los acentos ortográficos correspondientes a la lengua origen lo que no genera dificultades en la comprensión del texto en la lengua meta pero sí se afectan aspectos de forma ya que la lengua meta no utiliza acentos ortográficos.

En la traducción de la UFE eventiva *usar alcohol etílico* el traductor automático solo traduce *ethyl* y omite *alcohol* teniendo como consecuencia la imprecisión en la traducción del NT, la afectación en su nivel de especialidad y en la comprensión del mismo.

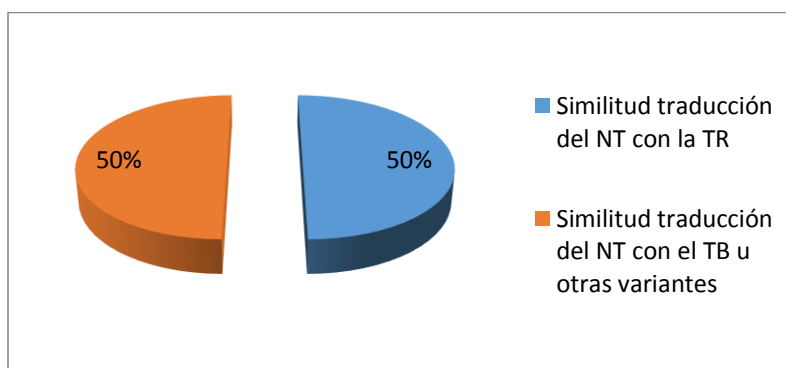


Gráfico 12. Traducción NT – Prompt

6.1.2.4 Resultados triangulación

El análisis particular de los NT que constituyen las UFE eventivas analizadas se fundamenta en la importancia de su función como transmisores del nodo cognitivo dentro del campo de especialidad estudiado dado que es en ellos que se condensa el conocimiento del ámbito especializado (Cabré, 2000). La traducción acertada de los NT posibilita, en parte, la exactitud de los productos de los programas de TA. Debe considerarse que las imprecisiones de los productos de TA, en la mayoría de los casos, se debe a la dificultad que estos programas tienen para desambiguar la información lingüística, sintáctica, semántica y pragmática del TB y no solamente a la exactitud con que traducen las unidades terminológicas.

Al comparar las traducciones de los NT realizadas por los traductores automáticos con la TR se observa que el 57% son traducidos de manera similar siendo Systran y Google los que mayor porcentaje de similitud presentan. Con esto puede respaldarse la afirmación de Peinado (2004), cuando plantea que la proximidad de la traducción automática con la traducción de un humano profesional de referencia es indicativa de la producción de una mejor traducción por parte del programa de TA.

El 43% de los NT se traducen de manera diferente a la TR presentando dificultades para la transmisión del conocimiento especializado que contienen y la inteligibilidad del texto debido a la falta de precisión con que se traducen los términos. Entre los traductores automáticos no se identifica una sistematicidad entre los NT que son traducidos de manera distinta a la TR, es decir, estos NT son diferentes para cada programa de TA. Sin embargo, es común para Google, Systran y Promt tener dificultades para traducir el NT *aprovechar el potencial energético y aprovechar saltos*. Como ejemplo presentamos:

Cuadro 8. Ejemplo (5)

TB	TR	Google	Systran	Prompt
<p>permiten aprovechar saltos de un metro gracias a turbinas hidráulicas de elevada velocidad específica.</p>	<p><i>one-metre heads can even be exploited</i> if high specific-speed hydraulic turbines are used.</p>	<p>They <i>made use of a meter jumps</i> through hydraulic turbines high specific speed.</p>	<p>They allow to <i>take advantage of jumps</i> a meter thanks to hydraulic turbines of high specific speed.</p>	<p>They allow to <i>make use of one meter jumps</i> thanks to water turbines of high specific speed.</p>

Para traducir el NT *saltos* los traductores automáticos eligen la misma unidad léxica. Esta unidad proviene del lenguaje general¹⁷ y no logra transmitir el conocimiento especializado que contiene. Según IATE¹⁸ la propuesta de la TR es equivalente al NT indicado en el TB.

Es posible que a partir de esta traducción el usuario pueda inferir a lo que refiere *jumps* pero es esto mismo lo que genera falta de confianza en el producto por su inexactitud y dificultad para ser comprendido como un texto de la lengua meta.

Fomicheva *et al.*, (2013) indican que debido a los problemas de ambigüedad o falta de cobertura del vocabulario, los traductores automáticos ofrecen variantes que no pertenecen al dominio de especialidad. La traducción de estos NT exige que los programas de TA cuenten con sistemas de desambiguación más precisos y diccionarios o corpus con una amplia cobertura con la que, al parecer, no cuentan los traductores automáticos analizados.

¹⁷ Ver en Internet: <http://www.oxforddictionaries.com/definition/english/jump>.

¹⁸ Ver en Internet: <http://iate.europa.eu/SearchByQuery.do>.

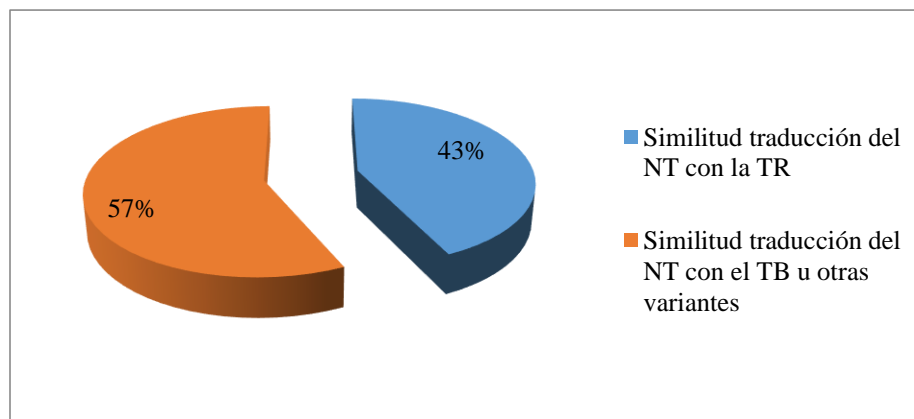


Gráfico 13. Triangulación Traducción NT – Google, Systran, Prompt

6.1.3 Aspectos contextuales

En el análisis de los aspectos contextuales se pretende estudiar cómo son traducidos los elementos que conforman el fragmento en el que se realiza la UFE eventiva con el objetivo de señalar cómo los errores que allí se presentan interfieren en la comprensión y precisión de la unidad objeto de estudio. Entre los aspectos considerados se encuentran:

- Las relaciones sintácticas entre los elementos que componen el fragmento.
- Ajustes a las modificaciones obligatorias de la lengua meta.
- Selección de unidades léxicas.
- Organización de las unidades léxicas en el fragmento.

6.1.3.1 Google

Al comparar la traducción que de los fragmentos que contienen las UFE eventivas hace Google con la TR y el TB se observa que el 45% tienen un número cercano de palabras a la TR y 55% al TB. Esto puede indicar que los productos del traductor automático son más cercanos a la lengua origen afirmación que toma fuerza al corroborar que el 100% de los productos de Google siguen patrones de la estructura del TB y que las modificaciones obligatorias de la lengua meta que realiza son básicas, entre ellas, inversión del adjetivo y el sustantivo y concordancia de número. Con menos frecuencia, el traductor automático utiliza recursos estilísticos cercanos a la lengua meta, como la voz pasiva, uso de posesivos, seguimiento de referentes, lo que genera un texto que no se lee como uno de la cultura meta. Google no solo sigue el patrón del TB en su estructura gramatical sino en la selección de las unidades léxicas que componen el fragmento (el 50%) ofreciendo variantes que pertenecen al lenguaje general pero que tienen un equivalente especializado en la lengua meta. Este fenómeno también se presenta debido a que algunas unidades de la lengua origen son cognados de la lengua meta sin que estas últimas cumplan la misma función en el área especializada que las primeras. Como consecuencia, Google tiene dificultades para organizar las unidades que forman el fragmento, lo que afecta el nivel de especialidad, la precisión y la inteligibilidad del texto y de la UFE eventiva que contiene.

6.1.3.2 Systran

Por su parte Systran presenta un porcentaje mayor de coincidencia entre el número de palabras del TB, el 80%. Al igual que en Google los fragmentos en que las estructuras sintácticas de la lengua origen y la meta son equivalentes se logran mejores resultados. Por el contrario, los contextos que presentan estructuras sintácticas no equivalentes muestran dificultades de precisión,

transmisión del nivel de especialidad e inteligibilidad debido a que el traductor automático tiende a seguir el patrón de la lengua origen y no a ajustarse a los patrones de la lengua meta.

Para Systran el porcentaje de unidades léxicas del fragmento que presentan dificultades también es mayor, el 65%. Estas dificultades tienen el mismo origen que las presentadas por Google: selección de unidades del lenguaje general cuando en la lengua meta existe un equivalente especializado distinto y adopción de cognados que no son, semánticamente, equivalentes en ambas lenguas.

De manera similar a Google, el mayor obstáculo de Systran radica en la dificultad para transmitir el mensaje del fragmento. Este traductor automático no logra organizar las unidades léxicas de manera que se ajusten a los requerimientos de la lengua meta y producir un texto preciso e inteligible debido a su tendencia a conservar el patrón sintáctico de la lengua origen.

Cuadro 9. Ejemplo (6)

TB	TR	Google	Systran
Sistema de sales, que permite almacenar la energía sobrante del campo solar en las horas de máxima radiación, y posibilita generar electricidad tras la puesta de sol.	Molten salts, enabling surplus <i>energy</i> in the solar field during the period of maximum radiation <i>to be stored</i> to generate electricity after sunset.	Salt system, which can <i>store</i> the excess <i>energy</i> of the solar field in the hours of maximum radiation, and allows to generate electricity after sunset.	Salt system, that allows to <i>store</i> the leftover <i>energy</i> of the solar field in the hours of maximum radiation, and makes possible to generate electricity after the sun putting.

El ejemplo expuesto en el cuadro permite ilustrar la manera como Google y Systran traducen el contexto de la UFE eventiva *almacenar la energía*. Se evidencia que el producto de los traductores automáticos siguen el patrón del TB en la voz activa que adoptan mientras la TR adopta la voz pasiva característica de los textos especializados escritos en la lengua meta; adicionalmente, se conservan los signos de puntuación del TB cumpliendo las mismas funciones que en él.

Así mismo, se observan dificultades con la selección de las unidades léxicas que componen el fragmento, tal es el caso de *sistema de sales*. IATE¹⁹ reconoce como término equivalente el usado por la TR *molten salt* pero no el usado por los traductores automáticos *salt system*. La variación propuesta por los programas de TA para la traducción del término responde al seguimiento del patrón del TB haciendo ajustes obligatorios en la lengua meta como lo es la inversión del adjetivo *salt* con el sustantivo *system*. En esta misma dirección, se presenta la traducción de *energía sobrante* donde *energía* es el NT de la UFE eventiva. Google propone como traducción *excess* y Systran *leftover*. Ambos tienen una connotación negativa que no se evidencia en el TB además, estas unidades léxicas no tienen el mismo nivel de especialización que la UFE eventiva y restan precisión a la misma.

Al analizar globalmente el contexto donde se encuentra la UFE eventiva y la traducción de esta, en particular, puede decirse que el fragmento logra transmitir el mensaje del TB pero no con el nivel de precisión, especialidad e inteligibilidad requerido.

¹⁹ Disponible en Internet: <http://iate.europa.eu/SearchByQuery.do>.

6.1.3.3 Prompt

Los productos de Prompt presentan características muy similares a las de Google y Systran respecto a la traducción del fragmento donde se encuentra la UFE eventiva, sin embargo, de manera particular, este traductor automático siempre tiene un número mayor de palabras si se compara con las de la TR y las del TB. En los productos de Prompt se evidencia un intento del programa de TA por ajustarse a las características sintácticas de la lengua origen sin dejar de lado el patrón estructural del TB. Como resultado, se presentan dificultades en la selección del 70% de las unidades léxicas debido a que se omiten, se conservan en la lengua origen o se seleccionan unidades no equivalentes semánticamente. Además, se adicionan preposiciones y otras unidades léxicas innecesarias. Es por esto que, al igual que Google y Systran, el origen de estas dificultades radica en el uso de unidades léxicas del lenguaje general que no son equivalentes en el área especializada y en la elección de unidades transparentes pero no equivalentes en ambas lenguas.

Así, Prompt tiene un 50% de similitud con relación a la TR y el TB acerca de su estructura sintáctica y la elección de unidades léxicas lo que puede indicar que solo la mitad de los productos alcanza un nivel aceptable de precisión, inteligibilidad y transmisión del nivel de especialidad del fragmento.

Cuadro 10. Ejemplo (7)

TB	TR	Prompt
España se está preparando para <i>aprovechar el enorme potencial energético</i> del mar y	Spain is preparing <i>to benefit from the huge potential of the sea</i> and add it to the rest of its	Spain is preparing itself <i>to make use of the enormous energy potential of the sea</i> and

sumarlo a las restantes tecnologías renovables y así avanzar en el desarrollo de una estrategia energética fiable y segura.	renewables portfolio in order to progress with the development of a reliable and secure energy strategy.	it to add the remaining renewable technologies and this way to advance in the development of a trustworthy energy strategy and segura.
---	--	--

En el cuadro se expone la traducción de la UFE eventiva *aprovechar el potencial energético*. Prompt realiza un intento de ajustarse a la lengua meta pero tiene la tendencia a conservar el patrón del TB por tanto, su producto resulta con un número adicional de palabras o con dificultades en la organización de las unidades léxicas que terminan siendo un obstáculo para la comprensión del texto. Se presentan dificultades para organizar el referente *it* en la frase *the sea and it to add the remaining renewable technologies*. El traductor automático reconoce la necesidad del referente, acierta en la selección de uno que cumpla con las reglas de la lengua meta pero lo expresa según las características de la lengua origen lo que genera dificultades en la comprensión del fragmento.

Respecto a la traducción de las unidades léxicas se pone de manifiesto que Prompt selecciona unidades que morfológicamente sean similares en ambas lenguas. Por ejemplo *renewable technologies* corresponde a la traducción de *tecnologías renovables*. Es probable que esta unidad léxica permita la transmisión del mensaje del TB pero resta precisión dado que no se habla solo de tecnologías renovables sino de todos los recursos renovables con los que cuenta España.

Como resultado se obtiene un texto que transmite el mensaje general del TB con múltiples imprecisiones tanto a nivel morfo-sintáctico como léxico-semántico.

6.1.3.4 Resultados triangulación

El contraste de los productos de los traductores automáticos con la TR y el TB indica que el 40% de estos tiene un número similar de palabras al TR y el 60% al TB, siendo Systran el que más similitudes tiene al respecto con el TB. Adicionalmente, el 90% de las traducciones de Google, Systran y Promt siguen la estructura sintáctica del TB y solo realizan algunas modificaciones obligatorias de la lengua meta como la inversión del adjetivo y el sustantivo, el uso de la *s* en la tercera persona del singular, la concordancia de número y el uso acertado de algunos posesivos y referentes. Sin embargo, el mayor inconveniente se evidencia en la organización morfo-sintáctica de las unidades que conforman el fragmento. La tendencia a seguir el patrón del TB lleva a la omisión, adición, no traducción de unidades, entre otros, que generan sin sentidos y dificultan la transmisión del nodo conceptual de la UFE eventiva.

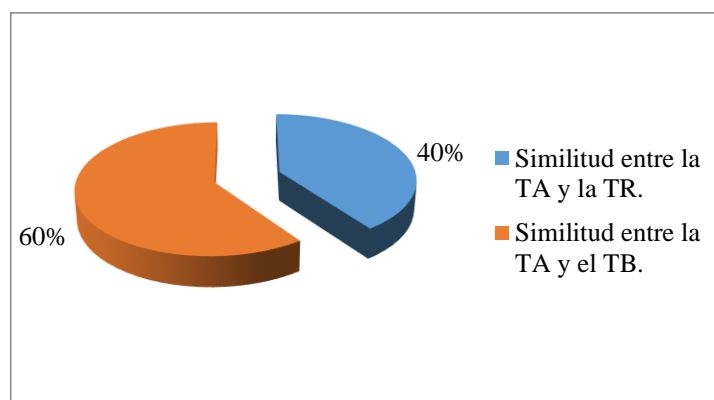


Gráfico 14. Similitud en el número de palabras – Google, Systran, Promt

Con relación a la traducción de las unidades léxicas que constituyen el fragmento, el 61% de las unidades no son equivalentes de las mencionadas en el TB, siendo Promt el traductor que

más dificultades presenta al respecto. Estas dificultades tienen origen en el uso de unidades del lenguaje general que no son equivalentes en el área especializada o a la adopción de cognados de la lengua origen que no cumplen las mismas funciones o no son equivalentes en la lengua meta.

En la observación de los datos se pudo constatar que aquellos fragmentos de corta extensión y que tienen estructuras sintácticas equivalentes en ambas lenguas permite a los programas de TA traducciones más precisas e inteligibles.

CAPÍTULO VII. CONCLUSIONES, RECOMENDACIONES Y LIMITACIONES

7.1 Conclusiones

Se ha presentado un análisis descriptivo – cualitativo – por medio del que se evaluó la manera como los traductores automáticos Google, Systran y Promt traducen las UFE eventivas. Esto se pudo alcanzar mediante la definición de tres aspectos de análisis –aspecto morfo-sintáctico, aspecto léxico semántico y aspecto contextual- que emergieron de una primera observación de los productos de los programas de TA. Los resultados de este análisis se validaron a través de su comparación con la TR del corpus textual.

Los aspectos morfo-sintáctico y léxico-semántico estaban centrados en el estudio de cómo son traducidas las UFE eventivas específicamente y el aspecto contextual se centró en el análisis del contexto donde tenía aparición la UFE eventiva para determinar cómo las dificultades que allí se presentaban podían afectar la precisión y la inteligibilidad de la unidad.

Con respecto al objetivo general de evaluar tres traductores automáticos cuando traducen UFE eventivas se observó:

La interpretación de los datos, desde un punto de vista cualitativo sugieren que:

- Los tres traductores automáticos tienden a seguir el patrón sintáctico del TB para traducir la UFE eventiva, es decir, la estructura NV+NT aunque el TR adopte una estructura diferente.

- A pesar de que más de la mitad de las UFE eventivas traducidas por los programas de TA seleccionaron el mismo NT de la TR no logran alcanzar resultados precisos, inteligibles y que transmitan el nivel de especialidad de los términos. Esto contradice lo planteado por *Peinado (2004)* quien afirma que *cuanto más se aproxime una traducción automática a la de un traductor humano profesional, mejor traducción será.*
- Los programas de TA analizados presentan dificultades importantes al traducir los contextos donde tienen su aparición las UFE eventivas. Estas dificultades varían de acuerdo a la extensión del fragmento, es decir, a mayor extensión, mayores dificultades. También varían con relación a la estructura sintáctica del fragmento, cuanto más cercana sea la estructura en ambas lenguas se presentarán menos dificultades en su traducción. Las dificultades que se presenten en la traducción del fragmento tienen repercusiones en la comprensión de la UFE eventiva, su precisión, en la transmisión del nodo conceptual y del mensaje en general.

La interpretación de los datos, desde un punto de vista cuantitativo sugieren que:

- El 61% de los productos de TA traducen las UFE eventivas con la misma estructura que el TR. De este porcentaje, solo el 11% corresponde a una estructura sintáctica diferente a NV+NT. Estas cifras confirman que los programas de TA siguen el patrón sintáctico del TB para traducir. Si la estructura sintáctica es equivalente en ambas lenguas los resultados serán favorables para la traducción automática de lo contrario, se presentan dificultades en la transmisión del mensaje y en la precisión de las unidades terminológicas.

- Respecto a la traducción de los NT, el 57% son similares a los propuestos por la TR lo que sería un indicador de precisión, inteligibilidad y transmisión del nodo cognitivo de las unidades.
- Con relación a la traducción de unidades léxicas que se encuentran en el contexto, el 61% no se traducen acertadamente debido al uso de cognados no equivalentes en ambas lenguas, niveles de especialidad diferentes a los expresados en el TB y dificultades en su organización sintáctica dentro del contexto.

Se concluiría entonces que las UFE eventivas por estar compuestas por palabras de uso general que toman una connotación especializada al inscribirse en un área específica del conocimiento son apropiadas para ingresar a los programas de TA.

Esta afirmación es válida para la estructura analizada en este trabajo NV+NT dado que estas unidades también pueden manifestarse como nombre de verbal *–consumo de energía–* o participio *–energía consumida–*. Se haría necesario un estudio de las otras estructuras en que se manifiestan las UFE eventivas para generalizar los resultados aquí planteados.

Con respecto al primer objetivo específico de describir y analizar las principales dificultades con las que se enfrentan los tres programas de TA cuando se traducen las UFE eventivas se identificaron 7 obstáculos estos son:

- Tendencia de los traductores automáticos a conservar la estructura sintáctica del TB en sus productos, cuando la lengua meta cuenta con estructuras de uso frecuente en los textos especializados.
- Selección de unidades léxicas del lenguaje general que no cumplen con los requerimientos semánticos como equivalentes de términos especializados, generando imprecisiones en las unidades traducidas.
- Uso de cognados que no son equivalentes semánticamente en ambas lenguas.
- Conservación de UFE eventivas en lengua origen.
- Conservación de símbolos ortográficos (signos de puntuación, tildes, etc) de la lengua origen en el TM.
- Generalización y aplicación de las reglas gramaticales de la lengua origen en los TM.
- Uso sistemático del mismo término para traducir la misma UFE eventiva en diferentes contextos. Esto puede tener dos consecuencias, una negativa que lleva a una generalización del uso del término que no contempla aspectos contextuales que pueden requerir o habilitar el uso de otro término más preciso y una positiva como muestra de la coherencia a nivel léxico que tiene el traductor automático.

Con respecto al segundo objetivo específico de determinar si el uso de una estructura equivalente en las lenguas en cuestión garantiza un mejor desempeño de los programas de TA, puede evidenciarse que no solo la equivalencia en la estructura sintáctica es suficiente para optimizar los resultados de la TA dado que para lograr una traducción inteligible y precisa es necesario realizar adaptaciones a las normas y convenciones de la cultura meta (valoración de aspectos lingüísticos, semánticos y pragmáticos) y no simplemente usar la misma estructura gramatical. Sin embargo, el uso de una estructura sintáctica equivalente en ambas lenguas sí aumenta las posibilidades para que el traductor automático realice una traducción en la que se transmita el mensaje del TB especialmente si es un texto corto y compuesto por unidades léxicas de uso general o no muy especializado.

Con respecto al tercer objetivo específico de determinar cuál de los tres programas de TA es más efectivo cuando traduce UFE eventivas, se concluye que Google y Systran ofrecen mejores resultados al traducir las unidades en estudio. El 60% de sus productos manifiestan una traducción de los NT similar a la TR. Lo anterior podría significar que sus arquitecturas internas (estadística e híbrida entre estadística y diccionarios respectivamente) contienen unidades terminológicas que les permiten obtener mejores resultados al traducir las UFE eventivas. Aunque un el hecho de que Systran logre un 45% de similitud con la TR al traducir el NV+NT comparado con un 34% de Google, le da una ventaja a este traductor automático. Se puede pensar que al utilizar tanto métodos estadísticos como diccionarios, Systran logra seleccionar con mayor precisión las unidades terminológicas y minimizar las dificultades que pueda presentar cada estructura.

De igual manera debe considerarse que los traductores automáticos analizados no generan traducciones de alta calidad. Estos productos deben ser revisados y corregidos por un traductor humano o, en su defecto, por el usuario del programa. Las versiones que ofrecen estos programas permiten, en el mejor de los casos, comprender el mensaje general del TB con limitaciones considerables.

Los programas de TA analizados pueden traducir de manera aceptable las UFE eventivas con estructura subyacente NV+NT pero debe considerarse que las imprecisiones de los productos de TA, en la mayoría de los casos, se debe a la dificultad que estos programas tienen para desambiguar la información lingüística, sintáctica, semántica y pragmática del TB y no solamente a la exactitud con que traducen las unidades terminológicas.

7.2 Limitaciones

Si bien a través de este estudio descriptivo en el que se analizaron 20 UFE eventivas con estructura subyacente NV+NT, no se han podido formular generalizaciones respecto a la traducción de las UFE eventivas, ni a los resultados que genera cada una de las arquitecturas internas de los traductores automáticos analizados, los resultados han permitido hacer un abordaje inicial de cómo se desempeñan estos programas de traducción automática al traducir unidades terminológicas. De otra parte, la metodología de análisis propuesta se intentó abordar el estudio de la traducción automática desde la descripción de sus productos y de las UFE eventivas en contexto, con el ánimo de abordar no solo aspectos sintácticos sino también semánticos y, de esta manera, evidenciar las

consecuencias que las variaciones de éstos tienen en la precisión, transmisión del nivel de especialidad de la unidad y la inteligibilidad de la misma.

Sin embargo, en un estudio a gran escala, se sugiere contemplar el análisis de un número mayor de UFE eventivas y, quizás, incluir todas las estructuras en que éstas se manifiestan para compararlas. Esto implicaría considerar una metodología de análisis distinta, dado que la evaluación manual demanda mucho tiempo y por consiguiente se hace más compleja.

7.3 Recomendaciones

A lo que se refiere a los alcances de esta investigación, el futuro de este estudio puede ampliarse a través del análisis de la traducción de las otras dos estructuras subyacentes en que se manifiestan las UFE eventivas - nombre de verbal *–consumo de energía–* o participio *–energía consumida.*

Aún queda mucho por explorar con respecto a los programas de TA. Es importante plantear estudios que lleven a comparar cronológicamente sus productos para poder establecer sus progresos en el tiempo.

Otro aspecto relevante para profundizar es la creación de métodos para evaluar manualmente aspectos lingüísticos de los productos de los traductores automáticos desde el punto de vista del usuario/traductor.

Sería importante que se profundizara en el estudio de las UFE eventivas en inglés para conocer cómo se forman, sus estructuras subyacentes, etc, y así poder realizar un estudio que deleve más detalles sobre estas unidades.

CAPÍTULO VIII. REFERENCIAS BIBLIOGRÁFICAS

- Aguado de Cea, G. (2007). *Las lenguas profesionales y académicas. La fraseología en las lenguas de especialidad*. España: Ariel.
- Alcina, C. (2011). *Principales traductores automáticos. El español en el mundo > Anuario 2010-2011 >*. Obtenido de http://cvc.cervantes.es/lengua/anuario/anuario_10-11/alcina/p04.htm
- Amigó; Gimenez & Verdejo. (2009). Procesamiento lingüístico en métricas de evaluación automática de traductores. *Procesamiento del lenguaje natural*, 43, 215-222.
- Anderson, W. (2006). *The Phraseology of Administrative French: A Corpus-Based Study*. Amsterdam: Editions Rodopi B.V.
- Arnold, A.;Balkan,L.; Meijer, S.;Humphreys, R. y Sadler, L. (1994). *Machine Translation an Introductory Guide. NCC / Blackwell, Manchester / Oxford*. Manchester, Oxford: NCC, Blackwell.
- Baker, M. (1995). Corpora in translation studies: an overview and some suggestions for future research. *Target*, 7(2), 223-243.
- Benotti, L., & Roger, S. (2004). *Dimensiones y estrategias. In VI Workshop de Investigadores en Ciencias de la Computación*.
- Berner, S. (2003). Lost in Translation: Cross-lingual Communication and Virtual Academic Communities. *Proceedings of the 5th Annual Conference on the World Wide Web Applications. Durban, South Africa*. .
- Bevilacqua, C. (2004). *Unidades fraseológicas especializadas eventivas: descripción y reglas de formación en el ámbito de la energía solar. Tesis doctoral*. Obtenido de Universidad Pompeu Fabra, Bcelona: <http://www.seer.ufrgs.br/riterm/article/download/21294/12279>

- Bevilacqua, C. R. (1999). *Unidades Fraseológicas Especializadas: estado de la cuestión y perspectivas*. Barcelona: Universidad Pompeu Fabra, Instituto Universitario de Lingüística Aplicada.
- Blackwood, G. (2010). *Lattice rescoring methods for statistical machine translation*. Tesis Doctoral. Cambridge, Inglaterra: Universidad de Cambridge y Clare College.
- Botella, M. (2008). *El tratamiento de los fraseologismos en los traductores automáticos*. Tesis doctoral. Alicante, España: Universidad Alicante.
- Cabré, M. T. (2000). *El traductor y la terminología: necesidad y compromiso*. Panacea@, Vol. 1, No. 2. Obtenido de http://medtrad.org/panacea/IndiceGeneral/n2_EditorialCabre.pdf
- Cabré, M. T.; Estopà, R. & Lorente, M. (1996). Criterios de reconocimiento de la fraseología a partir del análisis de corpus. *Criterios del V Simposio Iberoamericano de Terminología*, (págs. 67-81). México.
- Cabré, M.T. (1999). *La terminología. Representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. publicaciones del instituto universitario de lingüística aplicada. Barcelona: Instituto universitario de lingüística aplicada, IULA. Universidad Pompeu Fabra Barcelona .
- Cabré, M.T. (2005). *La terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro*. Obtenido de Universitat pompeu fabra, españa en revista: debate terminológico No 1 : <http://seer.ufrgs.br/index.php/riterm/issue/view/No.%200>
- Christianson, C.; Olive, J.P. & McCary, J. (2011). *Handbook of natural language processing and machine translation*. Springer.
- Cowie, A. (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.

- Dieguez & Riedemann. (1998). *Análisis del error en la traducción automática: algunos ejemplos de las formas -ing del inglés al español*. Obtenido de www.onomazein.net/Articulos/3/12_Dieguez.pdf
- Diéguez, M. (2001). *Aciertos y errores en la traducción automática: metodología de la enseñanza-aprendizaje de la traducción humana*. Obtenido de Instituto de Letras, Pontificia Universidad Católica de Chile: <http://www.onomazein.net/6/aciertos.pdf>
- Estopà , R. (1999). Eficiencia de la extracción automática de terminología. *Perspectives : studies in traductology*, 7(2), 227-286.
- Fomicheva, M.; Cunha, I. & Sierra, G. (2013). *Análisis del tratamiento de la terminología en la traducción automática: Implicaciones para la evaluación*. *Debate Terminológico*. No 9, Nov. 2013; pp. 16-36. Obtenido de <http://www.upf.edu/pdi/iula/iri>
- García, M. (2001). Estructura definicional terminográfica en el subdominio de la oncología clínica. *Estudios de lingüística del español*, 14.
- González, M. (2010). *Estudio comparativo de traductores automáticos En línea: systran, reverso y google*. Obtenido de Revista Núcleo, 27, 187-216: <http://www.scielo.org.ve/pdf/nu/v22n27/art08.pdf>
- Granger, S. & Meunier, F. (2008). *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins Company.
- Guadec, D. (1997). *Terminologie & phraséologie pour traduire. Le concordancier du traducteur*.
- Gouadec, D. (1990). *Le traducteur, la traduction et l'entreprise*. París: Collection AFNOR gestión.

Guerbefort, A. (2010). *Exploring Machine Translation on the Web*. *Revista Tradumática*, 8, 1-6.

Obtenido de <http://www.fti.uab.es/tradumatica/revista/num8/articles/03/03central.htm>

Hovy, E., King, M., & Popescu-Belis, A. . (2002). An introduction to MT evaluation.

Proceedings of Machine Translation Evaluation: Human Evaluators meet Automated Metrics. Workshop en la Conferencia LREC 2002. Las Palmas, España, 1-7.

Hutchins & Somers. (1992). *An Introduction to Machine Translation*. Obtenido de

<http://www.hutchinsweb.me.uk/IntroMT-TOC.htm>

Hutchins, J. (1986). *Machine translation: past, present, future* . Reino Unido: Ellis Horwood.

Hutchins, J. (2003). Has machine translation improved? Some historical comparisons.

Proceedings of the 9th MT Summit, 181-188.

Kit, C., Pan, H. & Webster, J. (2002). *Example-Based Machine Translation: A New Paradigm*,

Translation and Information Technology. Chinese University of HK press.

López, V. (2002). *Posibilidades y realidades de la traducción automática*. Obtenido de La

Linterna del Traductor, 3: http://traduccion.rediris.es/3/tr_au.htm

Moreno, A. (2000). *Diseño e implementación de un lexicón computacional para lexicografía y*

traducción automática. Obtenido de Estudios de lingüística del español vol 9 :

<http://elies.rediris.es/elies9/index.htm> 11 de abril de 2014

Nord, C. (2010). *Las funciones comunicativas en el proceso de traducción: Un modelo*

cuatrifuncional. Obtenido de http://www.scielo.org.ve/scielo.php?pid=s0798-97842010000100010&script=sci_arttext

Orente, M. (2001). *Terminología y fraseología especializada: del léxico a la sintáxis* .

Peinado, V. (2004). *Métricas de evaluación automática*. Obtenido de

<http://nlp.uned.es/~victor/docs/presentacion-MetricasEvaluacionAutomatica.pdf>

- Porta, C. & Hampshire, S. (2010). *Traductores automáticos como medio interpretativo del concepto percepción: dirección única versus multidimensionalidad significativa. Espéculo. Revista de estudios literarios*. Obtenido de Universidad Complutense de Madrid:
<http://www.ucm.es/info/especulo/numero44/tradauto.html>
- Santamaria, I. (2006). *La terminología: definición, funciones y aplicaciones*. Obtenido de Biblioteca de recursos electrónicos de humanidades E-excelence:
http://books.google.com.co/books?id=ZShn6fmIAmYC&printsec=frontcover&hl=es&source=gb_s_ge_sum
- Seoane. (2008). Obtenido de
<http://informatio.eubca.edu.uy/ojs/index.php/Infor/article/view/83/148>
- Suárez de la T., M. (2004). *Análisis contrastivo de la variación denominativa en textos especializados: del texto original al texto meta*. Obtenido de Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra :
<http://www.tdx.cbuc.es/TDX-0217105-130025/index.html>
- Tertoolen, R. (2010). *El alcance de la traducción automática: Un estudio de la traducción automática del par de lenguas español – neerlandés. Tesis de maestría*. Países Bajos: Universidad de Utrecht.
- Toledo, M. (2009). *El resumen automático como recurso documental para la traducción de artículos de investigación del ámbito jurídico-tecnológico (español-inglés-francés) Tesis doctoral Universidad de Málaga*. Obtenido de <http://www.biblioteca.uma.es/b>