

UNIVERSITAT AUTÒNOMA DE BARCELONA
FACULTAT DE TRADUCCIÓ I INTERPRETACIÓ

MÀSTER EN TRADUMÀTICA: TECNOLOGIES DE LA TRADUCCIÓ
CURS 2017 - 2018

TREBALL DE FI DE MÀSTER

LOS TIPOS DE TRADUCCIÓN AUTOMÁTICA Y SU EVALUACIÓN
MEDIANTE PERÍFRASIS VERBALES Y EXPRESIONES IDIOMÁTICAS
(ALEMÁN-ESPAÑOL)

MÓNICA OLMEDO RUIZ

DIRECTOR:
GÖKHAN DOĞRU

BARCELONA, JUNY DE 2018


Universitat Autònoma
de Barcelona

DADES DEL TFM

Títol (ca): *Els tipus de traducció automàtica i la seva avaluació mitjançant perífrasis verbals i expressions idiomàtiques (alemany-castellà).*

Título (es): *Los tipos de traducción automática y su evaluación mediante perífrasis verbales y expresiones idiomáticas (alemán-español).*

Title (en): *Types of machine translation and their evaluation using verbal periphrases and idiomatic expressions (German-Spanish).*

Autora: Mónica Olmedo Ruiz

Tutor: Gökhan Doğru

Centre: Facultat de Traducció i Interpretació

Estudis: Màster en Tradumàtica: Tecnologies de la Traducció

Curs acadèmic: 2017-2018

PARAULES CLAU

Traducció automàtica, motor de traducció automàtica, problemes de traducció automàtica, perífrasis verbals, expressions idiomàtiques.

RESUM DEL TFM

El present treball de fi de màster té com a objectiu avaluar la traducció automàtica entre el parell de llengües alemany-castellà. Atès que s'hi poder avaluar moltes qüestions idiomàtiques, en aquest treball es duu a terme l'avaluació mitjançant perífrasis verbals i expressions idiomàtiques. S'hi ofereix un marc teòric sobre la traducció automàtica, els tipus de traducció automàtica i els motors escollits per dur a terme l'avaluació. Així mateix, s'hi expliquen amb detall cadascunes de les perífrasis verbals i les diferències que presenten en relació amb la llengua alemanya i les expressions idiomàtiques amb les que es treballa. Una vegada obtinguts tots els resultats dels motors, amb la mètrica BLEU es determina quin motor proporciona millors resultats. Finalment, amb l'anàlisi de les dades, s'arriba a unes determinades conclusions incloent-hi la confirmació / refutació de les hipòtesis de partida.

PALABRAS CLAVE

Traducción automática, motor de traducción automática, problemas de la traducción automática, perífrasis verbales, frases idiomáticas.

RESUMEN DEL TFM

El presente trabajo final de máster tiene como objetivo evaluar la traducción automática entre el par de lenguas alemán-español. Dado que pueden evaluarse muchas cuestiones idiomáticas, en este trabajo se lleva a cabo la evaluación mediante perífrasis verbales y expresiones idiomáticas. A lo largo del trabajo se ofrece un marco teórico relacionado con la traducción automática, los tipos de traducción automática y los motores seleccionados para llevar a cabo la evaluación. Asimismo, se explican con detalle cada una de las perífrasis verbales y sus diferencias en relación con la lengua alemana, así como las expresiones idiomáticas con las que se trabaja. Una vez obtenidos los resultados de los motores, con la métrica BLEU se determina qué motor proporciona mejores resultados. Finalmente, con el análisis de los datos, se obtiene una serie de conclusiones que incluyen la confirmación / refutación de las hipótesis de partida.

KEYWORDS

Machine translation, Machine translation engine, problems of Machine translation, verbal periphrases, idiomatic expressions.

ABSTRACT TFM

The aim of this Master's Degree Dissertation is to evaluate the results of the different translation engines in order to determine which of them offers the best results in relation to verbal periphrases and idiomatic expressions between German and Spanish. A theoretical explanation about machine translation and an account of the existing machine translation systems are provided. Verbal periphrases and their differences in relation to the German language are also explained in detail, and the same procedure is applied to the idiomatic expressions used in the work. Once the results of the engines are obtained, the BLEU automatic quality metric is implemented to determine which engine provides the best results. Finally, with the analysis of the data, a series of conclusions are obtained that include the confirmation / refutation of the starting hypotheses.

Mir kommt Spanisch vor

Me suena a chino

LISTA DE ABREVIATURAS

BLEU	<i>BiLingual Evaluation Understudy</i>
CAT	<i>Computer-Aided Translation</i>
HAMT	<i>Human Aided Machine Translation</i>
LO	Lengua origen
LM	Lengua meta
LD	Lengua de destino
MAHT	<i>Machine Aided Human Translation</i>
MQM	<i>Multidimensional Quality Metrics</i>
MT	<i>Machine Translation</i>
NMT	<i>Neuronal Machine Translation</i>
SMT	<i>Statistical Machine Translation</i>
TA	Traducción Automática
TAE	Traducción Automática Estadística
TAO	Traducción Asistida por Ordenador
TAN	Traducción Automática Neuronal

LISTA DE ABREVIATURAS	V
ÍNDICE DE ILUSTRACIONES - TABLAS Y GRÁFICOS.....	VIII
INTRODUCCIÓN.....	10
OBJETIVOS E HIPÓTESIS DE TRABAJO	11
1. Objetivos	11
2. Hipótesis	12
CAPÍTULO I	14
<i>Marco teórico</i>	14
1. La traducción automática	15
1.1 Historia de la traducción automática	20
1.2 Enfoques tradicionales de la traducción automática.....	24
1.2.1 Sistema de traducción directa	24
1.2.2 Sistema de transferencia.....	25
1.2.3 Sistema de interlengua	28
1.3 Enfoques modernos de la traducción automática	31
1.3.1 La traducción automática basada en reglas	31
1.3.2 La traducción automática basada en corpus y modelos matemáticos	34
1.3.2.1 La traducción automática estadística	34
1.3.2.2 La traducción automática neuronal.....	38
2. La complejidad de la traducción automática.....	40
2.1 Ventajas e inconvenientes	40
2.2 Las limitaciones de los sistemas de la traducción automática.....	40
2.3 Problemas de la traducción automática	41
2.4 Problemas de la traducción automática en la combinación lingüística alemán-español.....	44
2.5 Los motores de traducción automática	44
3. Las perífrasis verbales.....	47
3.1 Concepto de perífrasis verbal	48
3.1.1 Clasificación de las perífrasis verbales.....	50
3.2 Perífrasis verbales de infinitivo.....	50
3.3 perífrasis verbales de participio	53
3.4 perífrasis verbales de gerundio.....	54
3.5 perífrasis verbales modales.....	55
4. Expresiones idiomáticas	57
5. Métodos de evaluación de la calidad de la traducción automática.....	60
CAPÍTULO II	62
<i>Metodología</i>	62

CAPÍTULO III	64
<i>Análisis de los datos</i>	64
1. Resultados de la evaluación de las perífrasis verbales.....	64
1.1 Resultados de las perífrasis verbales de infinitivo	65
1.2 Resultados de las perífrasis verbales de gerundio	67
1.3 Resultados de las perífrasis verbales de participio.....	69
1.4 Resultados de las perífrasis verbales modales	71
1.5 Síntesis general de los datos obtenidos.....	73
2. Resultados de la evaluación de las expresiones idiomáticas.....	76
2.1 Síntesis general de los datos obtenidos.....	82
 CAPÍTULO IV	 84
<i>Conclusiones</i>	84
 CAPÍTULO V	 87
<i>Bibliografía</i>	87
 CAPÍTULO VI	 90
<i>Anexos</i>	90
Anexo 1.....	90
Anexo 2.....	95

ÍNDICE DE ILUSTRACIONES – TABLAS Y GRÁFICOS

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Grados de automatización del proceso de traducción según Hutchins y Somers (1995: 27)...	18
Ilustración 2. Modelos de traducción automática	19
Ilustración 3. Método de traducción directa	25
Ilustración 4. Sistemas de transferencia	26
Ilustración 5. Otro esquema más completo de sistemas de transferencia.....	27
Ilustración 6. Modelo basado en interlingua con seis pares de lenguas	29
Ilustración 7. Gráfico piramidal de transferencia e interlingua	30
Ilustración 8. Triángulo de Vauquois	32
Ilustración 9. Alineación según Koehn.....	36
Ilustración 10. Proceso de traducción en un motor de traducción automática estadística	38
Ilustración 11. Muestra de traducción automática con una hoja de cálculo	45

ÍNDICE DE TABLAS

Tabla 1. Ejemplos de traducción con Google Translate.....	16
Tabla 2. Ejemplos de traducción con Lucy Software	27
Tabla 3. Otros ejemplos de Lucy Software.....	33
Tabla 4. Ejemplos de traducción con Google Translate estadístico	35
Tabla 5. Ejemplos de traducción de los diferentes motores objeto de estudio	39
Tabla 6. Ambigüedad léxica de la palabra ‘Tiempo’	42
Tabla 7. Perífrasis verbales que se evalúan.....	56
Tabla 8. Expresiones idiomáticas que se evalúan	58
Tabla 9. Métrica de BLEU para las perífrasis verbales de infinitivo.....	65
Tabla 10. Métrica de BLEU para las perífrasis verbales de gerundio	67
Tabla 11. Métrica de BLEU para las perífrasis verbales de participio	69
Tabla 12. Métrica de BLEU para las perífrasis verbales modales	71
Tabla 13. Perífrasis verbales de infinitivo reconocidas por los motores de traducción automática	73
Tabla 14. Perífrasis verbales de gerundio reconocidas por los motores de traducción automática	74
Tabla 15. Perífrasis verbales de participio reconocidas por los motores de traducción automática	74
Tabla 16. Perífrasis verbales modales reconocidas por los motores de traducción automática	75
Tabla 17. Resultados de la evaluación de las expresiones idiomáticas	76
Tabla 18. Expresiones idiomáticas idénticas en alemán y en español traducidas por Lucy Software	77
Tabla 19. Expresiones idiomáticas diferentes en alemán y en español traducidas por Lucy Software	77
Tabla 20. Expresiones idiomáticas idénticas en alemán y en español traducidas por Google Translate estadístico.....	78

Tabla 21. Expresiones idiomáticas diferentes en alemán y en español traducidas por Google Translate estadístico.....	78
Tabla 22. Expresiones idiomáticas idénticas en alemán y en español traducidas por Google Translate neuronal.....	79
Tabla 23. Expresiones idiomáticas diferentes en alemán y en español traducidas por Google Translate neuronal.....	79
Tabla 24. Expresiones idiomáticas idénticas en alemán y en español traducidas por DeepL.....	80
Tabla 25. Expresiones idiomáticas diferentes en alemán y en español traducidas por DeepL.....	81
Tabla 26. Expresiones idiomáticas idénticas en alemán y en español traducidas por Microsoft Translator.....	81
Tabla 27. Expresiones idiomáticas diferentes en alemán y en español traducidas por Microsoft Translator.....	81
Tabla 28. Traducciones de las perífrasis verbales de los diferentes motores	90
Tabla 29. Traducciones de las expresiones idiomáticas de los diferentes motores.....	95

ÍNDICE DE GRÁFICOS

Gráfico 1. Gráfico con los resultados de las perífrasis verbales de infinitivo.....	66
Gráfico 2. Gráfico con los resultados de las perífrasis verbales de gerundio.....	69
Gráfico 3. Gráfico con los resultados de las perífrasis verbales de participio.....	71
Gráfico 4. Gráfico con los resultados de las perífrasis verbales modales	73
Gráfico 5. Porcentaje de las perífrasis verbales correctas.....	75
Gráfico 6. Gráfico con los resultados de las expresiones idiomática	82
Gráfico 7. Porcentaje de las expresiones idiomáticas correctas.....	83

INTRODUCCIÓN

Traducir es hacer un viaje por un país extranjero.

GEORGE STEINER

El presente trabajo de investigación constituye el trabajo final de máster de los estudios del Máster de Tradumàtica: Tecnologies de la traducció que ofrece la Universitat Autònoma de Barcelona y se centra en la evaluación de la traducción automática. La idea de ese trabajo surgió a raíz de observar que cada uno de los sistemas de traducción automática funciona de manera diferente y, en consecuencia, en ocasiones, estos proporcionan traducciones diferentes. Hoy en día existen numerosas aplicaciones que permiten consultar o esclarecer el significado de una palabra de forma rápida, fácil y multilingüe, pero, sin embargo, no todas las combinaciones ofrecen resultados fiables, ya que, por ejemplo, no son iguales las traducciones proporcionadas con el par de lenguas inglés-español que alemán-español. Hasta el momento, ninguno de los motores disponibles puede prescindir del humano, puesto que es necesaria su intervención o bien en la preparación de los materiales de referencia, en el entrenamiento de motores, en la evaluación de los diferentes motores o en la posesición de las traducciones. Por ello, todos los motores continúan desarrollándose con el fin de lograr una mayor correlación entre el original y su traducción.

Como estudiante de alemán y conocedora de las arduas diferencias gramaticales entre la lengua española y la alemana, el principal interés es investigar si los traductores automáticos son capaces de reconocer dichas diferencias y reflejarlas en sus traducciones. Dado que dentro de la traducción automática se pueden investigar muchas cuestiones lingüísticas entre el par de lenguas alemán-español, entre el abanico de posibilidades, se han seleccionado las perífrasis verbales y las expresiones idiomáticas. Ambos temas se han escogido al observar que trasladar determinadas construcciones de una lengua origen a una lengua de destino requiere un cambio de estructura si la lengua

de destino posee una estructura distinta, tal como ocurre entre el alemán y el español. Un ejemplo de ello, tal como se expone a lo largo del marco teórico, sería la correspondencia entre la perífrasis verbal ACABAR DE + INFINITIVO por el adverbio *weiter* en alemán.

En definitiva, el tema central de esta investigación son las traducciones que ofrecen los diferentes sistemas de traducción automática, es decir, los basados en reglas, los estadísticos y los neuronales con el fin de evaluarlas mediante la métrica BLEU y determinar qué motor proporciona mejores resultados en relación con la traducción humana.

OBJETIVOS E HIPÓTESIS DE TRABAJO

Translation is that which transforms
everything so that nothing changes.

GÜNTER GRASS

1. OBJETIVOS

De lo expuesto anteriormente se deriva que el principal objetivo de este trabajo sea evaluar los diferentes motores de traducción automática con el fin de determinar cuál de ellos proporciona una traducción de calidad, es decir, igual o muy similar a la traducción humana.

Para alcanzar este objetivo principal se establecerán los siguientes objetivos específicos:

- Elaborar un marco conceptual sobre la traducción automática.
- Describir cuáles son los motores de traducción automática de trabajo.
- Definir qué cuestiones lingüísticas se evalúan: por un lado, las perífrasis verbales y, por otro, las expresiones idiomáticas.

- Evaluar los diferentes motores de traducción en relación con sus traducciones.

De este modo, los resultados que se obtengan de este trabajo podrán ser el punto de partida para una posible investigación con el fin de determinar si es recomendable el uso de la traducción automática en la enseñanza-aprendizaje de lenguas extranjeras.

2. HIPÓTESIS

En lo que respecta a esta investigación, la hipótesis inicial radica en que los motores de traducción automática —esos recursos en línea que ofrecen la posibilidad de traducir desde una simple palabra hasta un texto completo— proporcionan resultados fiables en determinadas búsquedas, como es el caso de la palabra compuesta en alemán *Kinderwagen*; sin embargo, cuanto más idiomática es una palabra, expresión o idea en un idioma origen, más dificultades poseen dichos motores para ofrecer una traducción de calidad en una lengua de destino. Por tanto, en lo que concierne al presente trabajo, la hipótesis inicial arraiga que las traducciones de las perífrasis verbales y expresiones idiomáticas, dado que, por un lado, presentan otra estructura sintáctica totalmente distinta y, por otro lado, no poseen el mismo vocabulario en el caso de algunas expresiones idiomáticas, no presentan la equivalencia correcta de la lengua alemana a la española.

Por ello, de lo expuesto anteriormente se deriva otra hipótesis a favor de uno los motores de traducción automática que son objeto de estudio en este trabajo. En este caso, la hipótesis de la que se parte es que el motor de traducción automática neuronal DeepL (<https://www.deepl.com/translator>) es el que ofrece traducciones más cercanas o iguales a la humana.

Además de estas hipótesis iniciales, también se establece una subhipótesis que radica en que los motores de traducción automática tendrán menos dificultades a la hora de proporcionar la traducción correcta de las perífrasis verbales, ya que estas son las únicas que no presentan ninguna diferencia estructural entre estas construcciones gramaticales.

En conclusión y a modo de resumen, las principales hipótesis son:

- Los motores de traducción automática a evaluar no son capaces de reconocer las diferencias estructurales entre el par de idiomas alemán-

español en lo que respecta a determinadas perífrasis y las expresiones idiomáticas que se evalúan.

- El motor de traducción automática DeepL es el que ofrece las traducciones más cercanas o incluso iguales a la humana para la mayoría de las perífrasis verbales que se estudian.
- La mayoría de los motores de traducción automática reconocen las perífrasis verbales modales y, en consecuencia, ofrecen la traducción tal como se espera.

CAPÍTULO I

MARCO TEÓRICO

El asunto de la traducción, a poco que lo persigamos, nos lleva hasta los arcanos más recónditos del maravilloso fenómeno que es el habla.

J. ORTEGA Y GASSET (1937)

En este capítulo se abordan los conceptos que constituyen el cuerpo teórico de este estudio. En primer lugar, se ofrecerá en términos generales una imagen introductoria de la traducción automática —de ahora en adelante TA—, junto a una breve explicación de su historia, desde sus orígenes hasta la actualidad. Asimismo, se presentarán los tipos de TA que son objeto de estudio en este trabajo, así como las cuestiones lingüísticas escogidas para evaluar a la TA, es decir, las perífrasis verbales y las expresiones idiomáticas entre el par de lenguas alemán-español. La mayoría de trabajos realizados en el ámbito de la TA, independientemente al tipo de motor (basado en reglas, estadístico, neuronal, etc.), analizan el inglés como lengua predominante; por ello, dado que existen pocos trabajos cuya lengua de partida o de destino no sea el inglés, el presente estudio tratará de solventar algunas dudas relacionadas con la TA, en este caso, entre la combinación lingüística alemán-español. Dadas las diferencias entre estas lenguas, es muy probable que los resultados de las traducciones de los motores de TA no sean del todo satisfactorios; con todo, hay que tener presente que se estudiarán tres tipos de motores de TA distintos y, ya que cada uno de ellos funciona de forma diferente, es probable que alguno de ellos origine la traducción esperada.

En segundo lugar, después de abordar los conceptos teóricos relativos a la TA y a los aspectos lingüísticos que son objeto de estudio, se presentarán los motores de TA con los que se dará respuesta a las hipótesis planteadas al inicio del presente trabajo. Para ello, en el marco teórico se expondrán los motores de TA con lo que se trabajará, así como la forma de evaluar los resultados que originan.

Por último, se plantearán los resultados de los diferentes motores de TA con su correspondiente reflexión para dar confirmación o refutar las hipótesis de partida de este trabajo y se finalizará con las correspondientes conclusiones extraídas de todo el proceso realizado.

1. LA TRADUCCIÓN AUTOMÁTICA

La TA se engloba dentro de la disciplina de la lingüística computacional¹. Los inicios en la investigación de la TA se remontan a poco después de estar disponibles los primeros ordenadores en motivo de la Guerra Fría, sobre todo, a la gran necesidad de conocer lo que se escribía en otros idiomas, en especial, entre el inglés y el ruso (cfr. Capítulo I.1.1). A partir de ese momento, la mecanización de la traducción ha sido y es toda una hazaña para aquellos que luchan por lograr una traducción perfecta de un idioma a otro sin intervención humana.

El concepto de TA es conocido también por su denominación en inglés *machine translation* (MT). Berner presenta la siguiente definición:

Machine Translation (MT) is the use of computer software to translate text or speech from one natural language into another. Like translation done by humans, MT does not simply involve substituting words in one language for another, but the application complex linguistic knowledge: morphology, syntax, semantics, and understanding of concepts such as ambiguity (2003: 7).

En una época en la que el volumen de demandas de traducciones es superior a la capacidad de los traductores (Hutchins y Somers, 1995: 26), la ayuda de un sistema de TA cuenta con los apoyos suficientes para que este se desarrolle hasta tal punto que, no es tan importante que ofrezca una traducción de calidad, sino más bien una traducción que se adapte a las circunstancias en las que se ha producido y del destinatario al que va dirigido. En este instante es dónde debe precisarse la combinación idiomática para evaluar las traducciones de la TA y el destinatario al que van dirigidas, ya que no es lo mismo introducir una palabra, frase o texto con la finalidad de hacerse una idea sobre lo

¹ Según Crystal, la lingüística Computacional (LC) es la «rama de la lingüística en la que se emplean conceptos computacionales para la elucidación de problemas lingüísticos y fonéticos» (2000: 345). Entre otros, además de la traducción automática, algunas aplicaciones relacionadas con la LC son los correctores ortográficos y los correctores de sintaxis.

que expresa el texto que para un estudiante de lenguas extranjeras² que para un turista alemán que está de visita en Barcelona, por ejemplo.

En los ejemplos que se exponen más adelante, puede percibirse que, en el caso de que un estudiante de lenguas cuya lengua inicial es la alemana, no debería emplear un motor de traducción automática como Google Translate para obtener el significado de la perífrasis verbal IR + INFINITIVO debido a que las traducciones podrían producir interfecciones³ idiomáticas y ello se transformaría en errores fosilizables⁴.

Dichos errores se originan cuando la lengua materna provoca interferencia en el aprendizaje del estudiante debido a que intenta calcar las estructuras de su lengua inicial a la lengua extranjera. Por ello, los ejemplos (a) y (b) no han sido elegidos al azar, sino que son fruto de la reflexión entre las diferencias de estructuras sintácticas entre el par de lenguas de estudio.

Por un lado, en el ejemplo (a.3) se halla el futuro del verbo 'llover' y la perífrasis verbal IR + INFINITIVO; por otro lado, en el ejemplo (b.1) se halla la misma perífrasis junto con el pronombre personal.

Tabla 1. Ejemplos de traducción con Google Translate

ORACIÓN LENGUA DE ORIGEN	TRADUCCIÓN HUMANA	TRADUCCIÓN DE GOOGLE TRANSLATE
(a.1) Lloverá	<i>Morgen regnet es</i>	<i>Wird regnen</i>
(a.2) Mañana lloverá	<i>Morgen wird es regnen</i>	<i>Morgen wird es regnen</i>
(a.3) Va a llover	<i>Es wird regnen</i>	<i>Es wird regnen</i>
(b.1) Voy a soñar contigo	<i>Ich werde von dir träumen</i>	<i>Ich werde von dir träumen</i>
(b.2) Soñaré contigo	<i>Ich werde von dir träumen</i>	<i>Ich werde mit dir träumen</i>

² Se habla de estudiantes de lenguas extranjeras porque en los estudios de Traducción e Interpretación no se recomienda el uso de la traducción automática ni para aprendizaje de lenguas ni para traducir.

³ Muchos de los errores de expresión que se cometen cuando se está aprendiendo una lengua extranjera vienen dados por el hecho de intentar calcar las estructuras de la lengua nativa. Wotjak afirma que «la lengua materna influye de n modo inconsciente y difícilmente controlable en la lengua extramjera que se pretende aprender (1984: 113).

⁴ Según Sonsoles Fernández (1995: 149), «consideramos como errores fosilizables, o mejor persistentes, aquellos que ofrecen mayor resistencia y que permanecen en sucesivos estadios del aprendizaje. Estos errores se refieren a estructuras que ofrecen una especial dificultad para el aprendiz, dificultad intrínseca, dificultad relacionada con la LM (lengua materna), dificultad percibida por el alumno o dificultad introducida por la metodología».

Tal como se observa en los ejemplos de (a.2) y (a.3), el motor de TA no distingue entre el futuro y el valor perifrástico. En los ejemplos de (b), además de evaluar la perífrasis verbal, cuya traducción es buena, también se evalúa la preposición que acompaña al verbo, ya que mientras en español se dice 'soñar con', en alemán se expresa con la preposición 'de', por lo que, mientras en español se sueña con alguien, en alemán se sueña de alguien (*träumen + von*). Las traducciones que ofrece Google Translate a este respecto, en la oración con la perífrasis verbal ofrece la traducción correcta, pero, sin embargo, en la segunda oración se observa que la preposición que propone es 'mit', que corresponde al 'con' del español. Además, en la oración 'Va a soñar contigo', la persona en la que recae la acción es la 3ª del singular (él / ella), pero el motor de TA traduce la oración con 'du' que equivale a la 2ª persona del singular (tú) en español.

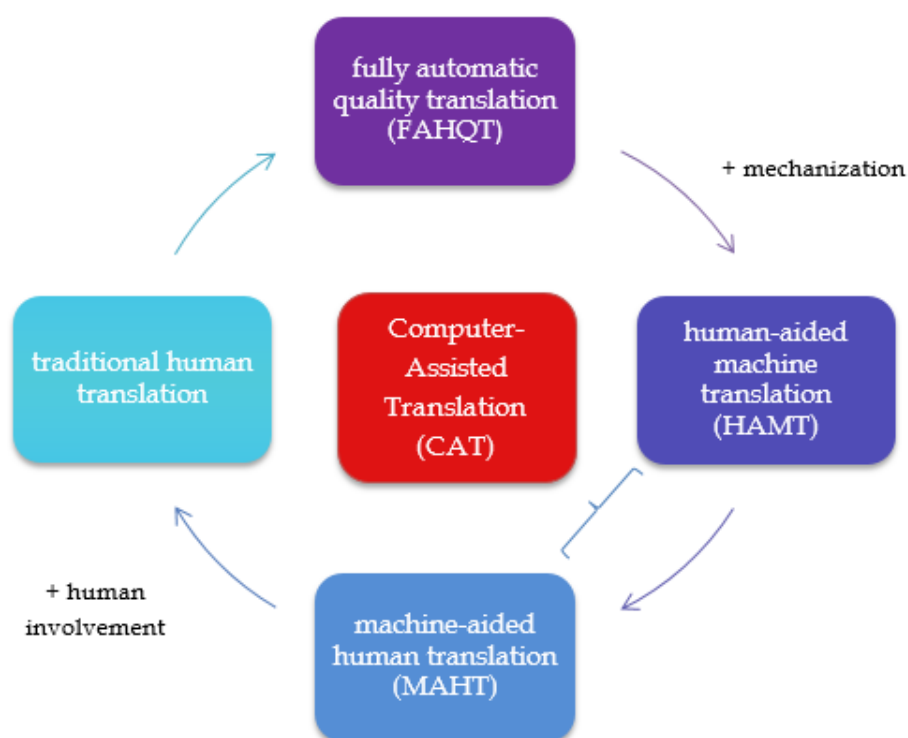
Grosso modo, uno de los objetivos prioritarios de la TA es solventar los principales problemas que tienen los diferentes motores a la hora de traducir (cfr. Capítulo 1.2.3). Algunos de los problemas son de índole lingüística, es decir, de todo aquello que tiene que ver con la extracción del significado de las oraciones o textos a partir de signos escritos. Dichos problemas se pueden generar debido a la ambigüedad léxica, la complejidad sintáctica, las diferencias de vocabulario entre lenguas y las construcciones elípticas o incorrectas gramaticalmente (Hutchins y Somers, 1995: 26). Para solventar estos problemas de procesamiento del lenguaje natural se situaría la lingüística computacional y la inteligencia artificial.

La mayoría de las traducciones que ofrecen los motores de TA pasan por una fase de corrección con la finalidad de que estas alcancen un nivel de calidad aceptable en relación con si tiene o no una funcionalidad específica, puesto que, cualquier persona que acostumbre a realizar búsquedas en los motores de TA, puede quedarse con el producto final del traductor automático, sea una traducción de calidad o no. Esto en algunas ocasiones funciona, pero en otras no tanto debido a que los motores de TA no saben reconocer, por ejemplo, las diferencias de significado entre las lenguas que se traducen.

Según Hutchins y Somers (1995: 27), pueden distinguirse diferentes niveles de automatización que van desde la traducción completamente automática de alta calidad (FAHQT o *Fully Automatic High Quality Translation*), hasta la traducción tradicional (*Tradicional Human Translation*). Asimismo, dependiendo del papel que la tecnología

posea en la traducción, se puede distinguir entre *Human Aided Machine Translation* (HAMT) y *Machine Aided Human Translation* (MAHT). En el caso de los procesos de HAMT, en caso de ambigüedad, será necesaria la intervención humana para garantizar una buena traducción. La principal diferencia entre *machine translation* y *human translation* recae en quién realiza la traducción: la máquina o el humano (cfr. Bowker, 2002: 4). Con todo, el núcleo de la TA misma lo constituye la automatización del proceso de traducción en su totalidad (Hutchins y Somers, 1995: 27).

Ilustración 1. Grados de automatización del proceso de traducción según Hutchins y Somers (1995: 27)



A continuación, se presentan los tipos de TA que han ido surgiendo con el paso de los años y que varían dependiendo de su funcionamiento. Existen tres sistemas principales de traducción automática:

- los basados en reglas TABR (o RBMT, por sus reglas en inglés *Rule-based machine translation*);
- los basados en corpus que pueden ser estadísticos TAE (o SMT, *Statistical Machine Translation*).
- basados en ejemplos TAFE (o EBMT, *Example Based Machine Translation*) y,

- finalmente, los basados en la generación de sistemas neuronales TAN (o NMT, *Neuronal Machine Translation*).

Para el presente trabajo, únicamente se tendrán en cuenta los basados en reglas, los estadísticos y los neuronales. El hecho de que se haya optado por tres sistemas de TA y no solo uno para evaluar las traducciones se debe a que el interés que suscita este trabajo es comparar los resultados de diferentes motores de TA con los que se trabaja para determinar si alguno de ellos ofrece mejores resultados a los otros. Asimismo, se pretende investigar si la traducción automática neuronal ofrece mejores resultados que cualquier otro sistema por el hecho de ser más novedosa.

Ilustración 2. Modelos de traducción automática⁵



El hecho de que las traducciones que nos proporcionan los sistemas de TA mencionados con anterioridad no sean del todo satisfactorias se debe a que parten del lenguaje natural⁶, es decir, además de palabras (léxico), formas (morfología), orden

⁵ Fuente: adaptación de Lavanchy (2010: 2).

⁶ El término Lengua natural designa una variedad lingüística o de forma de lenguaje humano. Las lenguas naturales están basadas en signos sonoros o gestuales. El Procesamiento de Lenguajes Naturales –PLN o

(sintaxis) y significado (semántica), parte de la presuposición de un conocimiento del mundo, un contexto cultural que, hasta el día de hoy, solo los humanos son capaces de reconocer.

Para entender el funcionamiento de la TA deben tenerse en cuenta las interpretaciones que esta puede tener. Por un lado, existe la interpretación de la TA como herramienta para traducir oraciones de un texto a otro de forma independiente, es decir, sin tener en cuenta el contexto en que se encuentre. De este modo, traducir frases implica, por un lado, construir el significado de la oración en la lengua origen —LO de ahora en adelante— y, por otro lado, construir una oración en la lengua meta —LM de ahora en adelante—.

Otra premisa de la que parte la TA es la idea de que no es necesario construir una interpretación de significado completa de una frase para traducirla, sino que únicamente es necesario transformar la estructura de la lengua origen en una estructura válida en la lengua meta, aunque no se garantice el sentido real de la frase. Esta premisa, partiendo de la base de los objetivos y las hipótesis del presente estudio, no funcionaría para el planteamiento del uso de la TA en los aspectos lingüísticos escogidos para evaluar la TA.

1.1 HISTORIA DE LA TRADUCCIÓN AUTOMÁTICA

La TA es una de las primeras tareas computacionales que se desarrollaron una vez estuvieron disponibles los primeros ordenadores (Oliver 2016: 160). Según expone Trujillo (1999), la historia de la TA ha estado muy influida por la política, la ciencia y la economía de los diferentes periodos de la historia moderna. A continuación, se detallarán los principales periodos en la historia de la traducción automática:

Precursores y pioneros (etapa anterior a los ordenadores). Desde el siglo XVII, varios científicos y filósofos propusieron representaciones del significado independientes a la lengua con el objetivo de superar las barreras lingüísticas. Descartes y Leibniz (1629) son algunos de los que desarrollaron códigos que relacionaban las palabras entre diferentes lenguas. Las primeras propuestas de TA datan del 1933, cuando el ruso Petr Smirnov-Troyanski y el franco-armenio Georges Artsrouni presentaron las primeras patentes con su máquina de traducción automática.

NLP el inglés *Natural Language Processing*— es una subdisciplina de la Inteligencia artificial y la lingüística computacional. *Víd.* nota 1.

Fases iniciales (1946 - 1966). A partir del uso de ordenadores para descifrar mensajes cifrados durante la Segunda Guerra Mundial, empezaron a utilizarse técnicas numéricas para investigar la traducción automática. Andrew Booth y Warren Weaver⁷ –de profesión criptógrafo–, en 1946 y 1947, fueron los que consiguieron despertar el interés por la TA al observar que podían utilizarse los sistemas que descifraban código inventados durante la Segunda Guerra Mundial en aplicaciones no numéricas como, por ejemplo, el caso de la traducción. En aquella época, se empleaba el término de ‘traducción mecánica’ (*mechanical translation*).

En 1948, Andrew Booth se encargó de trabajar en el análisis morfológico para un diccionario mecánico. Un año más tarde, en 1949, Warren Weaver puso las bases para el tratamiento del problema de la ambigüedad semántica. A partir de este momento, se inicia la investigación de la TA tanto en Estados Unidos como en otros países del mundo, entre otros, Japón e Inglaterra. En 1954 se realizó una demostración pública del primer sistema de TA con el par de lenguas inglés-ruso. Dicha demostración consistía en la traducción de 49 frases del ruso al inglés utilizando un diccionario de 250 palabras y seis reglas gramaticales. No fue hasta la década de 1956 y 1966 cuando hubo más investigación sobre el tema por parte de Estados Unidos y la antigua Unión Soviética debido a la Guerra Fría, pero, dado que los resultados no fueron satisfactorios, se redujo considerablemente la investigación en este campo y no fue hasta los años ochenta cuando volvió a surgir el interés por esta área.

El informe ALPAC⁸ (1966). El optimismo inicial llegó a su fin en Estados Unidos con este informe en el que se afirmaba que la TA era más lenta, menos precisa y el doble de costosa que la traducción humana y, por ello, no había expectativas inmediatas o futuras de que resultase útil, lo que provocó un efecto negativo en la financiación de la TA, sobre todo, en Estados Unidos.

Según explica Somers (2006: 140), los primeros sistemas de TA estaban diseñados conforme a técnicas de *dictionary-based direct replacement* (sustitución directa basada en diccionarios). Este sistema consiste, en primer lugar, en analizar el texto original; en

⁷ En palabras de Weaver (1947), la TA podría compararse con las técnicas criptográficas de la Segunda guerra mundial:

I have a text in front of me which is written in Russian, but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the next.

⁸ Automatic Language Processing Advisory Committee.

segundo lugar, en la búsqueda de diccionarios bilingües para sustituir las palabras de la lengua origen por las de la lengua meta; finalmente, en un reajuste del orden de la frase según las palabras consultadas.

Los optimistas que sobrevivieron a las críticas formuladas en el informe ALPAC se encargaron de revisar las técnicas básicas de los motores de traducción y de desarrollar un método indirecto; probablemente la figura más destacada de esta época fue Bernard Vauquois⁹. Este método se elaboró desde los dos enfoques que se mencionarán a lo largo de los próximos epígrafes (cfr. Capítulo 1.1.2.2 y 1.2.3): el de *transferencia* y el de *interlingua*. En ambos enfoques, primero se analiza el texto origen y después se genera el texto meta, aunque difieren en la etapa intermedia del proceso de traducción. En el enfoque de transferencia, existe un trasvase entre una lengua y otra, mientras que en el enfoque interlingua se hace una representación abstracta de la información del texto origen y se traslada directamente al texto meta.

Los años 1970 y los sistemas de TA operativos (1966-1976). En Estados Unidos, la investigación se centró en la traducción al inglés de textos científicos rusos. Algunos de los éxitos que surgieron unos años después fueron, por un lado, Systran: un sistema híbrido¹⁰ de TA desarrollado en 1970 por Peter Toma en California para la combinación ruso-inglés. Por otro lado, Météo, desarrollado en 1976 por el grupo TAUM¹¹ en Montreal para traducir boletines meteorológicos del inglés al francés. Los problemas para cubrir las traducciones de Europa provocaron que Systran implantara la versión inglés-francés, lo cual fomentó el inicio de las negociaciones y a la financiación del proyecto Eurotra. Gracias a estos avances, se recuperaron las investigaciones en el campo de la TA. De nuevo se celebraron conferencias, se crearon los primeros sistemas comerciales (aunque eran caros, contenían fallos), y se publicaron revistas especializadas en la materia. La dirección que se tomó fue que ordenador y humano trabajaran conjuntamente.

Los años finales de 1990 y MAT (*Machine Aided Translation*). En los años noventa, aparecen diferentes sistemas de traducción potentes y se generaliza el uso de

⁹ Bernard de Vauquois (1929 - 1985) era matemático e informático. Su fama se debe a que fue un líder influyente en la investigación de la TA en la década de 1970 en la Universidad de Grenoble, uno de los centros de investigación más importantes de la TA basada en reglas.

¹⁰ Para más información, véase HUTCHINS y SOMERS, 1995.

¹¹ Traduction Automatique de l'Université de Montréal

memorias de traducción junto con programas de ayuda para traductores. Hay un gran interés por la TA basada en ejemplos. En Europa se financió el mayor proyecto de TA de todos los tiempos: Eurotra. Disponía de 150 investigadores de veinte lugares de Europa e incluía todas las combinaciones de las nueve¹² lenguas representadas en la Comisión Europea en ese momento. Otras empresas que financiaron proyectos de TA fueron Siemens y Philips. En Japón, casi todas las empresas de electrónica desarrollaron sistemas de TA y apoyaron investigaciones en el sector público. A mediados de los años noventa se empezaron a comercializar otros sistemas, como los de traducción asistida por ordenador¹³. Asimismo, surgió el interés por emplear corpus y estadísticas para traducir los textos, en lugar de utilizar normas lingüísticas. En este caso, con base en probabilidades estadísticas, se extraían equivalentes de traducción del léxico y la sintaxis a partir de corpus.

La TA en internet. En internet, toda la información es de acceso rápido. En la era de las tecnologías, la demanda de traducciones no ha dejado indiferente a aquellos interesados en mecanizar la traducción. En el año 1997 se publicó la primera aplicación gratuita en línea, Babelfish, un sistema basado en reglas. Dicha aplicación ofrece una versión de Systran para traducir del inglés al francés, al alemán y al español, y viceversa. Unos años más tarde, en 2006, llegó el Traductor Google, anteriormente era estadístico pero que, desde hace muy poco tiempo, pasó a ser neuronal.

En resumen, en la actualidad se dispone de numerosos sistemas de TA que ofrecen traducciones de buena calidad, es decir, parecidas a la traducción humana, para determinadas combinaciones idiomáticas; unos ejemplos sería Microsoft Translator (<https://www.bing.com/translator>), DeepL o el anteriormente mencionado Google Translate (<https://translate.google.com/?hl=es>). Con todo, del mismo que se ha expuesto en a lo largo de los epígrafes, es importante considerar la finalidad de la traducción y los destinatarios, así como el par de idiomas que es objeto de evaluación, dado que no es lo mismo obtener una traducción del inglés al alemán que del alemán al japonés, por ejemplo.

¹² Por aquel entonces, las lenguas de la Unión Europea eran las de los seis fundadores en los años cincuenta (Alemania, Bélgica, Francia, Italia, Luxemburgo y los Países Bajos); las de Dinamarca, Irlanda y el Reino Unido, que se incorporaron en 1973; la de Grecia, en 1981 y, por último, las de España y Portugal en 1986. Las lenguas eran: danés, holandés, inglés, francés, alemán, italiano, griego, portugués y español.

¹³ Bajo el nombre de *Traducción asistida por ordenador* (TAO), se agrupa una serie de recursos informáticos que hoy en día se utilizan en los procesos de traducción. Estos recursos incluyen diccionarios en línea, correctores ortográficos y gramaticales, procesadores de textos, etc.

1.2 ENFOQUES TRADICIONALES DE LA TRADUCCIÓN AUTOMÁTICA

Antes de empezar con la explicación del funcionamiento de los diferentes motores de TA, es importante tener en cuenta que existen motores para un par determinado de lenguas (sistemas bilingües), para más de dos lenguas (sistemas multilingües) o que traducen en una única dirección (sistemas unidireccionales) o en ambas direcciones (sistemas bidireccionales) (Hutchins y Somers, 1995: 28). En este caso, los sistemas de TA en los que se basa este estudio son los multilingües, es decir, los que ofrecen la posibilidad de realizar traducciones de una LO a varias LM como, por ejemplo: español-inglés, español-alemán, español-catalán, etc.

El diseño de un sistema de TA combina elementos de la lexicografía, la lingüística, la lingüística computacional y la inteligencia artificial.

1.2.1 SISTEMA DE TRADUCCIÓN DIRECTA

Son aquellos en los que la traducción se realiza reemplazando una palabra de lengua origen por una palabra de la lengua de destino. Después, a través de un componente sintáctico, se reordenan estas palabras según la estructura del idioma meta. La traducción se realiza en un solo paso: se consulta el significado de las palabras de origen en un diccionario y directamente se generan sus equivalentes en lengua meta. Después, se aplican las reglas de flexión, concordancia y reordenamiento para solucionar los principales problemas de la traducción. El vocabulario y la sintaxis de los textos de la lengua de origen no se analizan. La calidad de estos sistemas depende en gran medida del par de lenguas con el que se trabaja.

Ilustración 3. Método de traducción directa¹⁴



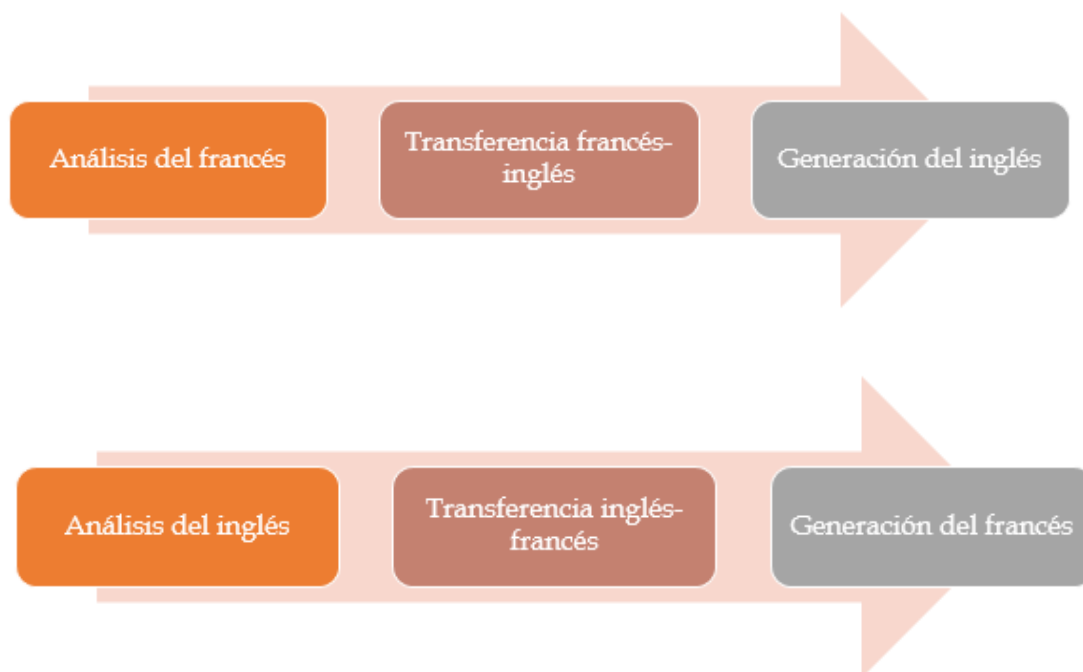
El producto final que se obtiene en este sistema puede caracterizarse como una traducción *palabra por palabra*, aunque se aplique las reglas de reordenamiento. A pesar de que hoy en día se siga empleando este método en programas bilingües unidireccionales, el fracaso de estos sistemas condujo al desarrollo de métodos lingüísticos de traducción más perfeccionados.

1.2.2 SISTEMA DE TRANSFERENCIA

De alguna manera, todos los sistemas de traducción incluyen este proceso, ya que la finalidad es convertir un texto de origen a un texto meta. Con todo, a diferencia de los programas de interlingua (cfr. Capítulo 1.1.2.3), estos son dependientes de la lengua que caracterizan, es decir, el resultado del análisis es una representación abstracta del texto origen y el punto de partida para la generación es una representación abstracta del texto meta. Esta técnica propone dos representaciones intermedias, una por cada idioma de la combinación. La traducción se realiza en tres fases: en primer lugar, se realiza el análisis, en segundo lugar, la transferencia y, por último, la generación. La transferencia se puede realizar tanto a nivel léxico, sintáctico o semántico.

¹⁴ Fuente: Adaptación de ZAPATA, I. (1995). «Un recorrido por algunos principios de Traducción Automática». En: Germán Ruipérez. *Enseñanza de lenguas y traducción con ordenadores*. Madrid: ediciones pedagógicas, p. 94.

Ilustración 4. Sistemas de transferencia¹⁵



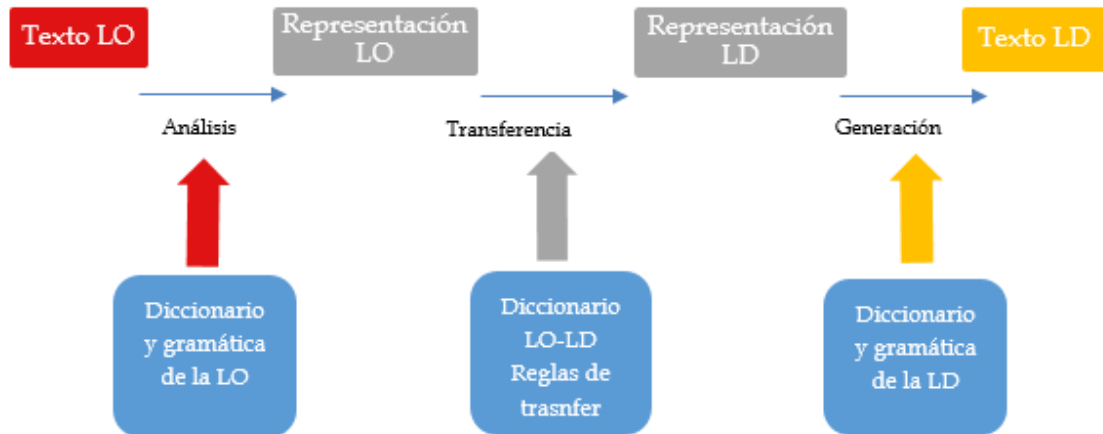
En este sistema, la función de los módulos bilingües no es otra que la de convertir las representaciones intermedias de la lengua origen en representaciones de la LM.

A través del análisis morfológico que se realiza en la primera fase, se consigue identificar los componentes del texto original y clasificarlos según su función; seguidamente, se realiza la transferencia léxica, donde se analiza el contenido y se resuelven las ambigüedades y, por último, se lleva a cabo la generación, donde se realiza la transferencia léxica. Aquí a cada palabra del texto original se le asigna un equivalente en el idioma meta y todo ello depende de cómo se obtenga la información en la fase de transferencia.

Además de la transferencia léxica, también es posible que realice la transferencia estructural donde se analiza el texto teniendo en cuenta segmentos mayores con el fin de encontrar expresiones que requieran un tratamiento especial como podría ser el caso de las metáforas, las expresiones idiomáticas, etc.

¹⁵ Fuente: adaptación de ZAPATA I., *op. cit.*, pág. 96.

Ilustración 5. Otro esquema más completo de sistemas de transferencia¹⁶



El sistema de TA que se enfoca en la transferencia y, además, con el que se trabaja a lo largo del trabajo es Lucy Translation (cfr. Capítulo 1.2.5).

En los sistemas de transferencia, el principal problema es el de la transferencia léxica¹⁷ y la transferencia estructura. Por un lado, con relación a la transferencia léxica, este sistema no plantea problemas cuando solo hay un término equivalente de la lengua meta o este se asemeja tanto en la LO como en la LM como, por ejemplo, la palabra ‘país’ cuya traducción es *pays*, en francés o *paese*, en italiano. Una de las palabras que se traduce de una forma similar en varios idiomas es ‘sur’ en español (*Süden* en alemán, *south* en inglés o *sud* en francés); con todo, la complejidad concurre cuando existen diferencias en la estructura sintáctica o cuando un término léxico posee numerosas traducciones.

De lo expuesto anteriormente, a continuación, se detallan las traducciones de Lucy Software para observar si hace algún tipo de distinción para las siguientes oraciones¹⁸: (a) ‘Tengo sueño’ y (b) ‘Tengo un sueño’.

Tabla 2. Ejemplos de traducción con Lucy Software

ORACIÓN	TRADUCCIÓN DE LUCY SOFTWARE
(a) Tengo sueño	<i>Ich bin müde</i>
(b) Tengo un sueño	<i>Ich habe einen Traum Schlaf</i>

¹⁶ Fuente: adaptación de ZAPATA I., *op. cit.*, pág. 96.

¹⁷ La transferencia léxica consiste en sustituir un componente léxico de la lengua origen por otro de la lengua meta.

¹⁸ En alemán hay dos palabras para este mismo concepto en español: *Die Schläfrigkeit* y *Der Traum*.

Las traducciones de los ejemplos (a) y (b) son correctas. Si en el caso de que en el ejemplo (b), en lugar de ofrecer las dos posibilidades (*Traum* | *Schlaf*), hubiera puesto *Schlaf*, la traducción sería incorrecta porque la diferencia entre estas dos palabras en alemán radica en el significado y significante, es decir, una se refiere al acto de dormir (*Schlaf*) y la otra al sueño como fantasía (*Traum*). Por otro lado, la oración (a), el motor de TA ha reconocido que el significado de la expresión 'Tengo sueño' en referencia a 'Estar cansado'; de ahí que la traducción sea correcta.

El principal problema que presenta este tipo de sistema y que no parece relevante para dar respuesta a una de las hipótesis es que los motores de TA puedan pasar por alto diferencias en el orden gramatical de las oraciones. Un claro ejemplo sería el caso de las lenguas que son objeto de estudio en este trabajo, dado que, en cuanto a estructura gramatical, el alemán y el español son idiomas completamente diferentes.

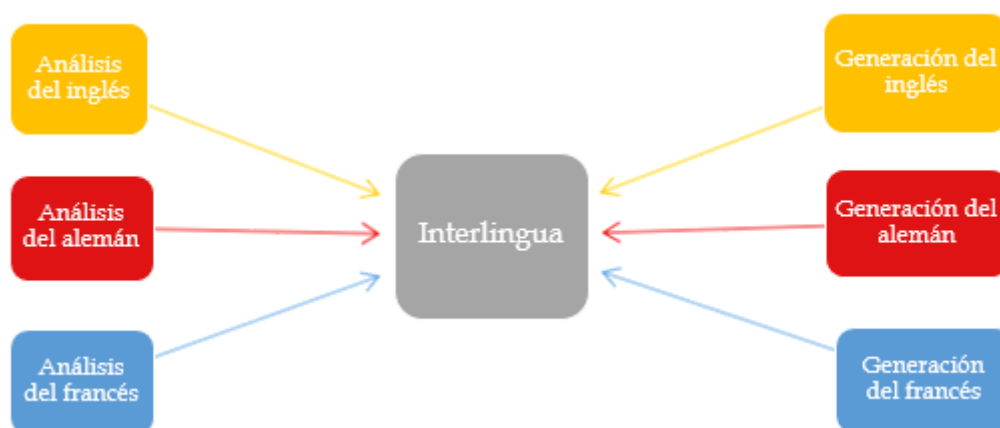
1.2.3 SISTEMA DE INTERLENGUA

Los sistemas de traducción basados en este sistema asumen que es posible convertir un texto de LO en representaciones sintácticas y semánticas para más de una lengua. En este sistema hay dos componentes monolingües: por un lado, el análisis de la lengua origen a interlingua y, por otro lado, la generación de interlingua a la lengua de destino. En este caso, se prescinde del método directo y de transferencia, mencionados en los epígrafes anteriores.

El funcionamiento de dicho sistema sería, en primer lugar, asignar una estructura al texto origen y, posteriormente, partiendo de esta estructura, generar el texto meta sin tener en cuenta cuál es la lengua origen. El objetivo es «que cualquier texto en cualquier lengua origen se represente con interlingua y que este lenguaje sea lo suficientemente potente como para representar cualquier tipo de información relevante para traducir» (Zapata, 1995: 97).

La ventaja de este enfoque reside en que si se incorpora una lengua al sistema se crean dos módulos: una gramática destinada al análisis y una gramática destinada a la generación (Ilustración 5.). En el caso de tener un sistema de interlingua del inglés y el francés, mediante la incorporación, por ejemplo, de una gramática destinada al análisis del alemán, el número de sentidos al que se puede llevar a cabo una traducción se incrementa de dos (del alemán al francés y del alemán al inglés). Además, la integración de un módulo de generación, una gramática destinada a la generación del alemán, en este caso añadiría dos nuevos pares (del inglés al alemán y del francés al alemán).

Ilustración 6. Modelo basado en interlingua con seis pares de lenguas¹⁹



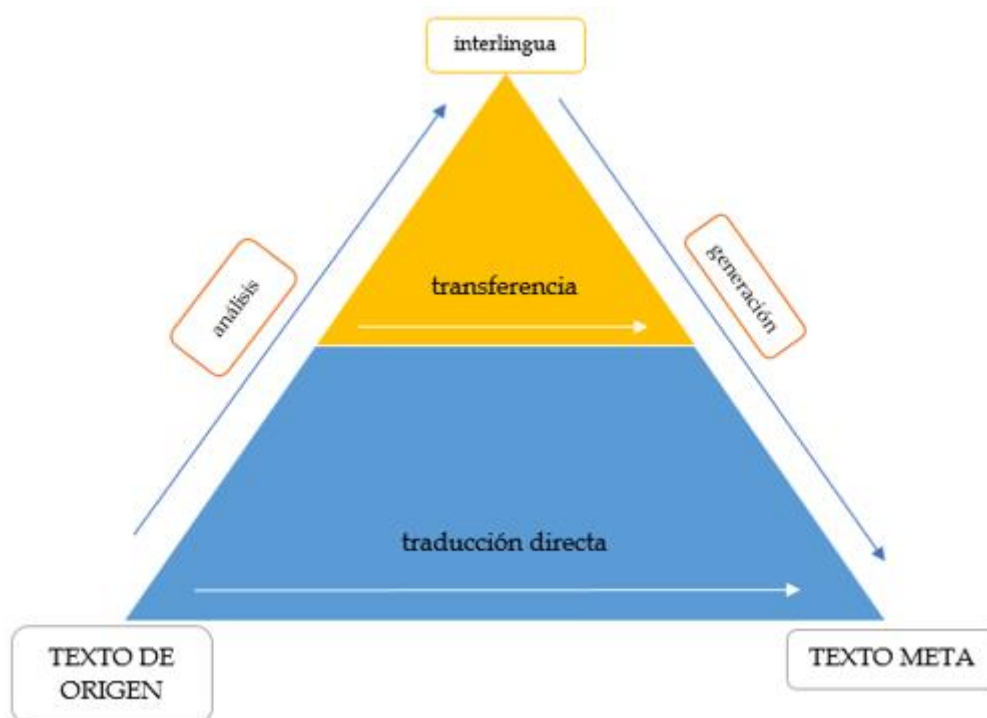
Mediante este sistema, la traducción es de calidad solo si se traduce el significado de las frases, algo poco probable si solo se realiza la transformación de la estructura superficial del texto sin comprender su significado.

En el sistema interlingua, con relación a la representación estructural, el orden y el significado de palabras ayudan a distinguir las funciones gramaticales que posee en la oración, como la función de sujeto o las de objeto.

Respecto a la representación de léxico, en los sistemas interlingua, dado que son multilingües, los problemas emergen cuando una de las palabras en cuestión tiene dos o más formas posibles para una única palabra de otra de las lenguas. Un ejemplo de ello es la palabra 'tiempo'. En alemán, se podría traducir como *Zeit* o como *Uhr*. Por un lado, la palabra *Wetter* describe el tiempo atmosférico (si hace sol, si llueve o nieva, etc.); por otro, la palabra *Zeit* describe la dimensión temporal o una época histórica (cfr. Capítulo 1.2.3 - Tabla 6).

¹⁹ Fuente: Adaptación de HUTCHINS y SOMERS (1995).

Ilustración 7. Gráfico piramidal de transferencia e interlingua²⁰



En la ilustración, puede observarse cómo el triángulo de Vauquois se basa en las diferencias de los tres componentes de la traducción: análisis, transferencia y generación. Dicha ilustración muestra el análisis del texto de origen en dirección ascendente y la generación en la lengua meta en dirección descendente. En la cima de la pirámide se representa el método interlingua, conseguido mediante el análisis monolingüe y su posterior generación. Con todo, también se puede dejar de lado el análisis monolingüe y recurrir a la fase de transferencia bilingüe, tal como se observa en el lado intermedio de la pirámide, así como acudir a la traducción directa.

El método de transferencia presenta ciertas desventajas si se compara con el de interlingua. En lo que respecta a un programa multilingüe, la principal dificultad para el sistema interlingua radica en que la incorporación de una nueva lengua conlleva, además incluir dos módulos de análisis y de generación, también deben incluirse nuevos módulos de transferencia, cuyo número puede variar en función del número de lenguas con que cuente el programa. Por ello, si un programa está diseñado para dos lenguas, una tercera implicaría la adición de cuatro nuevos módulos de transferencia.

²⁰ Fuente: HUTCHINS Y SOMERS, 1995: 161 (aunque fue utilizado antes por Vauquois en 1968).

1.3 ENFOQUES MODERNOS DE LA TRADUCCIÓN AUTOMÁTICA

Los enfoques modernos están muy relacionados con los enfoques tradicionales. De ellos, parten los sistemas que a continuación se describen y que son objeto de este estudio. Por un lado, los motores de traducción automática basados en reglas se basan en la utilización de información lingüística o cultural y una serie de reglas que explicitan cómo deber ser gestionado este conocimiento. En este caso, dicho motor utilizaría los tres enfoques tradicionales descritos en el epígrafe anterior. Sin embargo, por otro lado, los motores de traducción automática estadística y neuronal entraría dentro del enfoque empírico, donde hay una ruptura con los enfoques tradicionales y, en este caso, lo que se persigue es aplicar estadísticas y probabilidades, junto a redes neuronales para alcanzar una traducción de calidad.

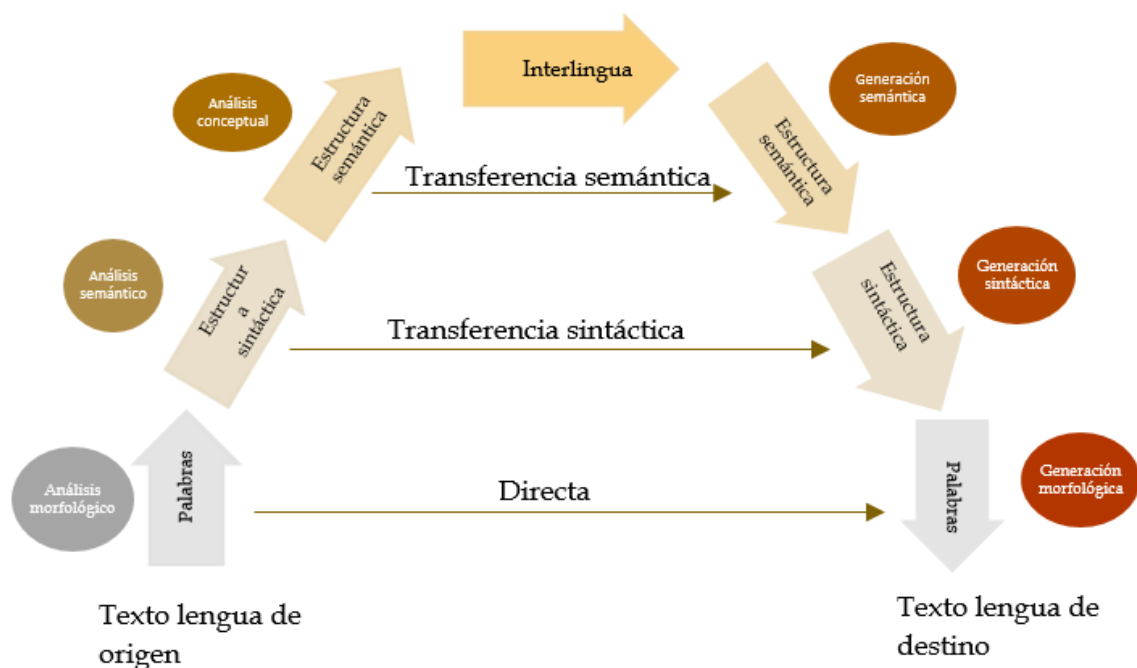
1.3.1 LA TRADUCCIÓN AUTOMÁTICA BASADA EN REGLAS

Los inicios de la investigación en el área de la traducción automática se centraron en los sistemas basados en reglas donde emplearon técnicas como la traducción directa, la transferencia y la traducción basada en una lengua interlingua (Koehn 2010: 15-16) (cfr. Capítulo 1.1.2). En el caso de la traducción automática basada en reglas —de ahora en adelante TABR—, el sistema de TA se basa en reglas lingüísticas integradas y en millones de diccionarios bilingües para cada par de idiomas.

Estos diccionarios computacionales, tanto monolingües como bilingües, junto con las gramáticas computacionales con reglas morfológicas y sintácticas correspondientes, derivan una traducción de texto de una lengua origen a una lengua meta. Para alcanzar la traducción, la idea es transferir al motor de traducción el conocimiento lingüístico por medio de reglas. Por tanto, mediante el procesamiento sintáctico, la interpretación semántica, el procesamiento contextual de la lengua y el uso de millones de glosarios, estos sistemas encuentran la equivalencia semántica.

A continuación, tal como se ha mencionado al inicio de este epígrafe, se puede observar un árbol en el que se puede visualizar las técnicas que emplean los sistemas TABR, es decir, las técnicas de traducción directa, transferencia e interlingua.

Ilustración 8. Triángulo de Vauquois²¹



En la ilustración, puede observarse como ascendiendo en el triángulo de Vauquois se mantiene el contenido léxico de las palabras (nombres, verbos, adjetivos y adverbios). La forma más ingenua de traducción es simplemente traducir palabra-por-palabra, es decir, buscar una palabra en un léxico bilingüe (parte inferior del triángulo). En la traducción directa, generalmente, hay una etapa de análisis morfológico y, después, vendría la traducción palabra-por-palabra mediante léxicos bilingües. Seguidamente, un conjunto de reglas se encarga de reordenar la traducción resultante y, finalmente, un generador morfológico produce la información morfológica correcta de cada palabra, es decir, todo aquello relacionado con el género, número, tiempo, aspecto verbal, etc. y, por último, se eliminan las palabras funcionales o gramaticales como conjunciones o artículos.

Si se asciende al siguiente nivel, en el enfoque de transferencia sintáctica, se analiza la sintaxis de la oración de la lengua origen y la estructura sintáctica resultante se asigna mediante reglas a una nueva estructura sintáctica en el idioma de destino. Las palabras se traducen como en el enfoque de traducción directa, a través de la búsqueda en un léxico. La principal ventaja de este sistema es que la sintaxis y el orden de las palabras correspondientes en el idioma de destino se construyen de forma gramaticalmente correcta.

²¹ Fuente: VAN GOMPEL, 2009: 16.

Un paso más arriba, la transferencia semántica funciona de una forma muy similar a la transferencia sintáctica, en este caso, intenta analizar la estructura semántica de la oración de origen y usa reglas para asignarlas a una estructura semántica en el idioma de destino. Ello se consigue, además de con el análisis sintáctico, con roles semánticos.

Finalmente, en la cumbre del triángulo se haya el sistema interlingua, que parte de la idea de solucionar los impedimentos para traducir de una lengua a varias, es decir, no solo hay dos idiomas de origen y de destino, sino que hay múltiples idiomas que se relacionan entre sí para evitar reconstruir los conjuntos de reglas de cada par de idiomas.

Básicamente, un motor de este tipo requiere:

- Un diccionario que asocie cada palabra del idioma origen al idioma de destino.
- Reglas que representen la estructura del idioma origen.
- Reglas que representen la estructura del idioma de destino.
- Reglas que permitan relacionar a cada uno de los idiomas relacionar estas dos estructuras.

Desde el punto de punto de vista de la sintaxis, la TABR ofrece una mejor traducción. Este tipo de sistemas puede emplearse, sobre todo, en estructuras sintácticas simples. A continuación, en los ejemplos (a) - (b) y (c), se observan dos frases con su correspondencia traducción con el motor de TABR Lucy²².

Tabla 3. Otros ejemplos de Lucy Software

ALEMÁN	ESPAÑOL	TRADUCCIÓN DE LUCY SOFTWARE
a. <i>Als wir jung waren</i>	Cuando éramos jóvenes	Cuando éramos estábamos jóvenes
b. <i>Ich habe keinen Wagen</i>	No tengo coche	No tengo ningún coche
c. <i>Ich höre gern Musik</i>	Me gusta escuchar música	Oigo con gusto música

Por un lado, en la frase (a), la traducción que ofrece es correcta, pero, sin embargo, no distingue entre los verbos *ser* y *estar*; por tanto, en el caso de esta traducción, debería haber posesión²³. Por otro lado, en la frase (b), la traducción está bien, a excepción de la doble negación, que en español no se emplea. Finalmente, en el ejemplo (c), el traductor reconoce el adverbio *gern*, pero en este caso, no reconoce la función que posee

²² Recuérdese que, para evaluar los resultados de la traducción automática basada en reglas, se eligió el motor de TA Lucy Software.

²³ Con todo, cabe añadir que algunas variedades del español no hacen distinción entre el *ser* y *estar*.

en la mayoría de los casos, ya que, junto a un verbo, dicho adverbio actúa como la perífrasis verbal del español GUSTAR + INFINITIVO.

Estos ejemplos, en el caso de la primera frase, se ha escogido porque en alemán hay distinción entre las conjunciones *als* y *wenn*. *Als* se utiliza en acciones pasadas que no volverán a suceder, de ahí el verbo *war* (imperfecto del verbo 'sein') y *wenn* para momentos que se repiten, acciones futuros o irreales. En el caso del segundo ejemplo, se ha escogido por la doble negación del alemán. Por último, el ejemplo se ha planteado, por un lado, por el adverbio *gern* y, por otro lado, por el verbo reflexivo en español, dado que en alemán no lo es.

1.3.2 LA TRADUCCIÓN AUTOMÁTICA BASADA EN CORPUS Y MODELOS MATEMÁTICOS

Los sistemas de TA que se expondrán en los siguientes epígrafes son inductivos y no funcionan con conocimientos lingüísticos, a diferencia de la TABR, sino mediante corpus y modelos matemáticos.

1.3.2.1 LA TRADUCCIÓN AUTOMÁTICA ESTADÍSTICA

La traducción automática estadística mediante métodos estadísticos está alcanzando importantes logros y presenta un futuro prometedor (Ney *et al.*, 2000). Este tipo de TA trata de resolver el problema de la traducción con un enfoque inductivo, sin utilizar conocimientos lingüísticos, a diferencia de la TABR. La aproximación clásica en traducción automática estadística —TAE de ahora en adelante— utiliza modelos de traducción del tipo palabra por palabra. Con todo, hoy en día, con tal de solventar los problemas que presenta este tipo de traducción, se trabaja con *phrasen* y no con palabras. *Phrasen* (del inglés) hace referencia a los datos de entrenamiento donde las unidades de traducción se dividen en grupos con un tamaño y una secuencia determinada; de este modo, se intentan solventar los problemas que se pueden originar con el sistema palabra por palabra como, por ejemplo, falta de contexto y, en consecuencia, una mala traducción de concordancia.

En el caso del presente trabajo, para evaluar un motor de TAE se utiliza Google Translate. En la actualidad, dicho motor es neuronal, pero mediante una fórmula y una hoja de cálculo, puede obtenerse los resultados que proporcionaba como motor estadístico. De este modo, puede valorarse si la evolución de estadístico a neuronal ofrece mejores resultados.

Tabla 4. Ejemplos de traducción con Google Translate estadístico

PALABRA	TRADUCCIÓN DE REFERENCIA	TRADUCCIÓN DE GOOGLE TRANSLATE ESTADÍSTICO
<i>Die Steuer</i>	El impuesto	El control
<i>Das Steuer</i>	El volante	El control
<i>Der Schild</i>	El escudo	El escudo
<i>Das Schild</i>	La señal de tráfico	El escudo

En la tabla se observa cómo Google Translate no es capaz de distinguir estas dos palabras que, aunque la unidad léxica sea la misma, no poseen el mismo artículo. En alemán es muy importante distinguir el artículo para diferenciar los significados frente a casos como este donde puede haber ambigüedad. Si se realiza la prueba sin el artículo, la palabra *Schild*, la traduce como ‘Impuesto’, mientras que *Schild*, la traduce como ‘blindar’.

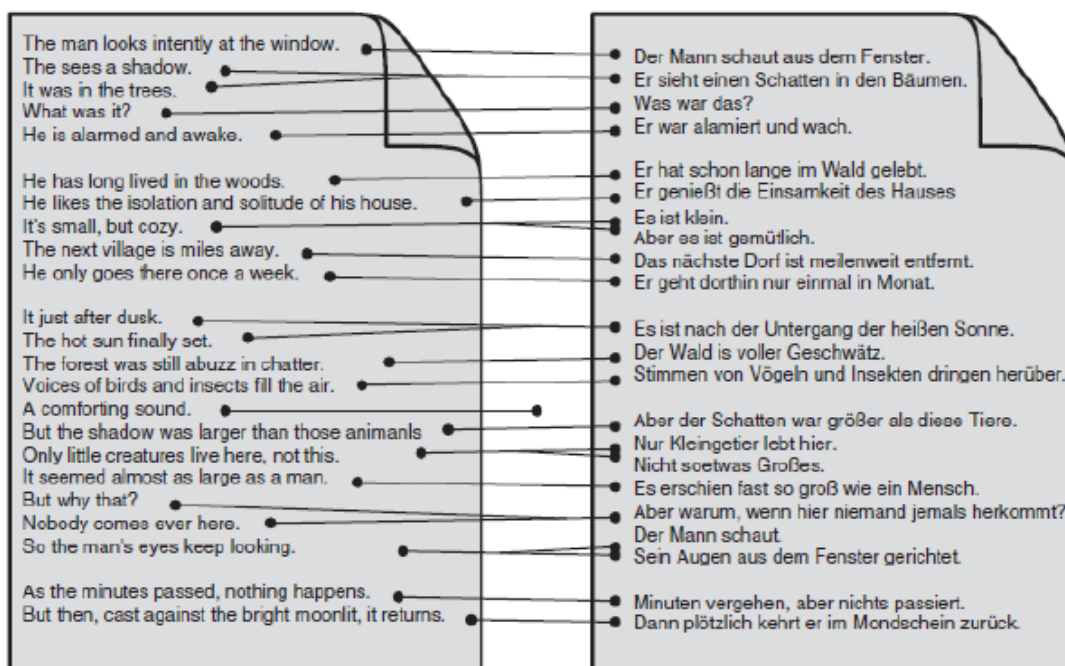
Volviendo al funcionamiento de este motor de TA, el principal funcionamiento es el de alinear frases, grupos de palabras o palabras individuales de textos paralelos y calcular las probabilidades²⁴ para que una palabra expuesta individualmente o en una frase de cualquier lengua se corresponda con una palabra en una frase de una traducción con la que esté alineada. Para ello, los motores se entrenan para traducir automáticamente a partir de traducciones existentes (*bitextos*). El procedimiento que sigue la TA se divide en tres pasos: *training*, *tuning* y *decoding*. El primer paso implica el uso de corpus monolingües y bilingües; el segundo, consiste en conocer la relevancia de cada modelo cuando se trata de combinarlo y, por último, el tercer paso consiste en encontrar la mejor traducción para una palabra, frase o texto. El objetivo de este tipo de motores es la de configurar el proceso de traducción en términos de probabilidades estadísticas.

Para lograr las traducciones, los textos deben segmentarse en oraciones y las frases deben alinearse. Para el alineamiento de las palabras es necesario saber qué palabras son traducción mutua, es decir, la palabra en lengua original y su correspondiente equivalencia en lengua meta. A partir de los alineamientos a nivel de palabra se puede

²⁴ La probabilidad es siempre mayor o igual que cero y menor o igual que uno. La suma de las probabilidades es siempre igual a uno; es decir, si la frase en la LD consta de cinco palabras, estas sumarán uno.

obtener un diccionario bilingüe probabilístico (modelo de traducción de palabras)²⁵. Por tanto, mediante el alineado de oraciones se fija qué oraciones se traducen de cuáles; mediante el alineado de palabras y las probabilidades de traducción se establece qué palabras se traducen de cuáles y, finalmente, mediante el modelo de lenguaje²⁶ se obtiene la frase correcta en la lengua de destino.

Ilustración 9. Alineación según Koehn²⁷



Los motores TAE son entrenados con extensos corpus bilingües paralelos de los cuales se derivan probabilidades estadísticas para abordar la traducción de textos nuevos. No se requiere ningún conocimiento lingüístico para entrenar un motor de este tipo, sino únicamente conocimientos informáticos y estadísticos. La disponibilidad de un gran corpus de textos bilingüe en la traducción automática estadística es una

²⁵ El modelo de traducción de palabras trata de estimar la probabilidad de que una palabra o frase sea la traducción de otra, es decir, estima la probabilidad de que un segmento en una lengua origen corresponda a otro en la lengua meta. Estos modelos se aprenden a partir de textos paralelos.

²⁶ Un modelo de lengua es un modelo que expresa cómo de verosímil es una oración en el idioma de destino. Dicho modelo se basa en *n-gramas*, es decir, en segmentos más pequeños que la oración. El producto final se obtiene de forma automática mediante grandes cantidades de textos en la LD. Mediante un modelo de lengua se puede averiguar qué oración es más verosímil en alemán de *Ich träume mit dir* o *Ich träume von dir*. En este caso, la frase correcta sería la segunda (cfr. Capítulo I.1).

²⁷ Fuente: Koehn, 2010: 56.

característica esencial para que este tipo de motores generen traducciones fiables y de calidad.

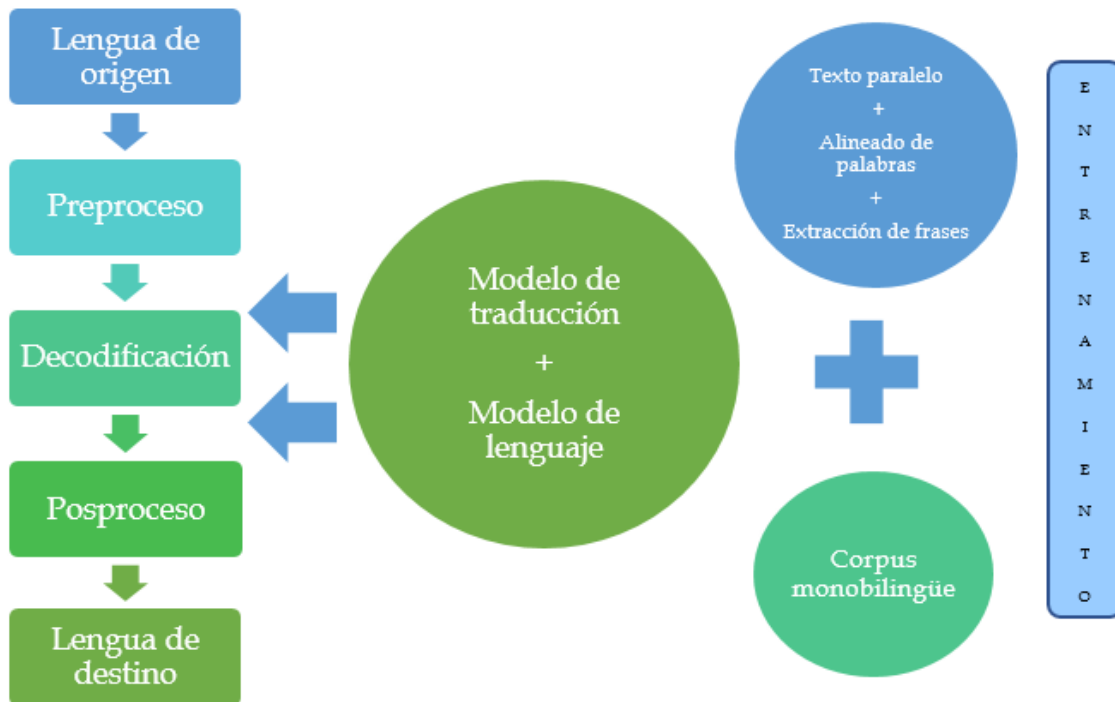
Los sistemas basados en estadísticas parten de la premisa de que al traducir una oración de un idioma origen e a un idioma meta f , todas las oraciones del idioma meta f son potencialmente la traducción correcta de la oración en cuestión en el idioma origen. Sin embargo, algunas oraciones tienen una probabilidad más alta de ser la traducción correcta. Para hallar esta posibilidad, el sistema utiliza textos bilingües o corpus alineados y mediciones de probabilidad. Para realizar el cálculo probabilístico, los sistemas tienen que ser entrenados con traducciones existentes (textos bilingües o corpus paralelos).

Con $p(e)$ se formula la probabilidad de que e suceda, es decir, por ejemplo, si e representa en alemán *Haus*, la probabilidad de que una persona emplee esta opción es más probable a que utilice otra. Con $p(f|e)$ se describe una probabilidad condicional. Para calcular dicha probabilidad se emplean algoritmos y se aplican cuando se desea calcular la probabilidad de un evento que ocurrió primero dependiendo de lo que ocurrió después. Sin embargo, si se utiliza siempre el equivalente más probable proporcionado por un diccionario lingüístico de probabilidades se traduce palabra por palabra y no se tiene en cuenta el contexto. Por ello, para alcanzar una traducción de calidad se recurre al modelo de la lengua meta²⁸.

En los sistemas estadísticos se emplean dos modelos: por un lado, el de traducción y, por otro lado, el del lenguaje. El modelo de traducción calcula la frecuencia en que se dan las traducciones de ciertas secuencias de palabras en los corpus; el modelo de lengua valora la legibilidad de las traducciones comparándolas con un corpus en la lengua meta.

²⁸ A diferencia del modelo de traducción, donde se utilizan corpus bilingües para hacer el cálculo de probabilidades, en este modelo, se calcula la probabilidad de que la traducción de una oración sea correcta y fluida en la LM.

Ilustración 10. Proceso de traducción en un motor de traducción automática estadística



Con todo, la TAE presenta dificultades de traducción con muchos idiomas, ya sean considerados cercanos (inglés-español) o lejanos (alemán-serbio). Por ello, hay ciertas dudas con respecto a la calidad del producto final y, en consecuencia, muchos traductores se resisten a trabajar con estos motores. Para este trabajo, como motor TAE se evaluarán los resultados ofrecidos por Google Translate (cfr. Capítulo 1.2.5).

1.3.2.2 LA TRADUCCIÓN AUTOMÁTICA NEURONAL

La traducción automática neuronal —TAN de ahora en adelante— es una aproximación a la TA basada en corpus, es decir, la estadística, y en muchas ocasiones, proporciona mejores resultados.

Una característica de la TA radica en que las palabras y las frases son representadas de forma numérica mediante vectores (Bengio *et al.*, 2003). El concepto de red neuronal artificial consiste en imitar el funcionamiento de las redes neuronales del cerebro humano. Una neurona artificial es un elemento simple de procesado que recibe señales de un conjunto de neuronas, seguidamente las combina con pesos y produce una señal de salida a la que se somete a una función de activación con el fin de transmitirla a un conjunto de neuronas.

La traducción de la red neuronal funciona en dos etapas: por una parte, en la primera etapa modela la palabra que necesita traducirse basándose en el contexto de esta palabra y sus posibles traducciones. No importa la longitud del contexto, es decir, la frase puede ser tanto de 6 palabras como de 20. Por otra parte, en la segunda etapa traduce este modelo de palabra dentro del contexto de la oración en la lengua de destino. Dicha traducción no tiene que ser la palabra en sí, sino la red contextual que se ha realizado en la primera etapa.

Este tipo de motores ofrece mejores traducciones por tres motivos: en primer lugar, tiene en cuenta la frase completa y no solo las palabras individualmente; en segundo lugar, maneja el lenguaje a través del reconocimiento de patrones similares a los del cerebro y, en tercer lugar, aprende las especificaciones de cada idioma, como el caso del género.

A lo largo del trabajo se trabaja con tres motores de TAN: Google Translate; DeepL y Microsoft Translator. A modo de resumen, y para concluir este apartado sobre los diferentes motores de TA objeto de estudio, es importante tener presente las principales diferencias de funcionamiento que poseen con el fin de ofrecer la mejor traducción. En primer lugar, la TABR funciona mediante reglas lingüísticas; en segundo lugar, la TAE lo hace mediante estadísticas y probabilidades y, en tercer y último lugar, la TAN funciona a través de redes neuronales artificiales.

Para concluir este apartado de los diferentes sistemas de TA que son objeto de este estudio, a continuación, se presentan unos ejemplos para considerar el producto final de tres motores de TA distintos y ver si la traducción que ofrece se puede considerar de calidad o, si por el contrario, uno proporciona una traducción más próxima respecto a la traducción humana.

Tabla 5. Ejemplos de traducción de los diferentes motores objeto de estudio

TRADUCCIÓN	
LUCY SOFTWARE	<i>Zu sein zum Papagei</i>
GOOGLE TRANSLATE ESTADÍSTICO	<i>Sein, der Papagei</i>
GOOGLE TRANSLATE NEURONAL	<i>Sei Papagei</i>
DEEPL	<i>Seetaucher</i>
MICROSOFT TRANSLATOR	<i>Sein auf dem Papagei</i>

En ninguna de las traducciones de los diferentes motores que se evalúan se observa la expresión idiomática 'Estar al loro'.

2. LA COMPLEJIDAD DE LA TRADUCCIÓN AUTOMÁTICA

2.1 VENTAJAS E INCONVENIENTES

La principal ventaja que posee la TA es la de traducir grandes volúmenes de textos en un corto plazo de tiempo. Asimismo, permite la comunicación entre dos personas que no hablan la misma lengua; por ello, cualquier persona que desee enviar un mensaje a otra puede hacerlo mediante la TA. Es probable que se pierda el significado contextual del texto, pero es indudable que la posibilidad de que dos o más personas se comuniquen existe. Además, no debe olvidarse que el uso de la TA es barato, e incluso gratuito en muchos casos. En ocasiones, dado que el contenido web está aumentando cada vez más, hay áreas a las que los traductores humanos no pueden acceder como, por ejemplo, las traducciones automáticas de Facebook o Twitter.

Por otro lado, a pesar de que haya gente reticente a emplear los motores de traducción automática, estos con el paso de los años han ido cobrando más importancia y es innegable que recurrir a la TA para obtener la traducción de una palabra, frase o texto pueda ser contraproducente para la finalidad con la que realizamos la búsqueda.

Por ello es importante saber qué motor de TA proporciona la traducción esperada; en otras palabras, el uso de la TA para los aspectos lingüísticos escogidos entre el par de lenguas alemán-español puede estar determinado por la traducción que ofrezca cada motor, ya que puede haber grandes diferencias de resultado entre los basados en reglas, los estadísticos y los neuronales y de ello depende que el uso de la TA sea una ventaja o un inconveniente.

2.2 LAS LIMITACIONES DE LOS SISTEMAS DE LA TRADUCCIÓN AUTOMÁTICA

La principal limitación de un sistema de TA es la calidad de la traducción, cuya evaluación se realiza con base a los parámetros de inteligibilidad, fidelidad, precisión y estilo. Con relación a la inteligibilidad y la fidelidad, la comprensión de una frase está motivado por el conocimiento en la LO completo, unos mecanismos que procesen la información lingüística y del conocimiento del contenido de la frase y una planificación inteligente de cómo se generará la frase en la lengua de destino (Oliver, Moré y Climent,

2008). Tal como se ha mencionado en apartados anteriores, para los motores de TA reconocer todo el conocimiento lingüístico y cultural de todas las lenguas es imposible, por lo que, en numerosas ocasiones, la calidad de la traducción será mejor entre los pares de lenguas que posean la misma raíz, como es el caso del catalán y el español, mientras que con el par de lenguas que son objeto de estudio en el presente estudio no ocurre lo mismo.

La fidelidad en relación con la TA se refiere a que cuando un determinado motor de TA nos proporciona una traducción, no tiene en cuenta el contexto de la palabra, frase o texto, por lo que se pierde el hilo argumental.

Las limitaciones que afectan al estilo se basan en que en la traducción de un motor de TA se pierde el deseo del autor de expresar una opinión enfatizada o de subrayar algo a través de sus palabras. Por ello, si esto el motor no lo reconoce en la lengua de origen, tampoco lo hará en la de llegada (Oliver, Moré y Climent, 2008).

2.3 PROBLEMAS DE LA TRADUCCIÓN AUTOMÁTICA

Tal como señalan Sitman y Cruz Piñol (1999), el problema fundamental de la TA remite a la naturaleza del lenguaje natural, ya que las máquinas no son capaces de reconocer los factores extralingüísticos que puede presentar una lengua. En estos factores se incluye la intención o el contexto situacional, histórico o cultural. En estos casos, no pueden tenerse en cuenta las ambigüedades del lenguaje, como en el caso del ejemplo (a) y (b), donde se expone una ambigüedad gramática. Por ello, el traductor automático, carece de sentido común y no puede respaldarse en su experiencia o conocimiento del mundo para solventar los problemas. En el apartado 1.2.3, se mencionó el problema de ambigüedad léxica de la palabra 'Tiempo' en alemán; a continuación, se exponen las traducciones de los diferentes motores para observar cómo traducen la palabra 'Tiempo', ya que, en el alemán, en función del contexto, se emplee una u otra opción. Las oraciones que se traducen son: (a) '¿Tienes tiempo?' y (b) 'Hoy hace buen tiempo', cuya traducción al alemán es: (a) *Hast du Zeit* y (b) *Heute ist das Wetter schön*.

Tabla 6. Ambigüedad léxica de la palabra ‘Tiempo’

	TRADUCCIONES
LUCY SOFTWARE	(a) <i>Heute vor guter Zeit Wetter</i> (b) <i>Hast du Zeti Wetter?</i>
GOOGLE TRANSLATE ESTADÍSTICO	(a) <i>Haben Sie Zeit?</i> (b) <i>Schönes Wetter heute</i>
GOOGLE TRANSLATE NEURONAL	(a) <i>Hast du Zeit?</i> (b) <i>Heute ist schönes Wetter</i>
DEEPL	(a) <i>Haben Sie Zeit?</i> (b) <i>Es ist schönes Wetter heute</i>
MICROSOFT TRANSLATOR	(a) <i>Guter Wetter Heute</i> (b) <i>Haben Sie Zeit?</i>

Tal como se observa en la tabla, la ambigüedad léxica que existe con la palabra ‘Tiempo’ al traducirla en alemán, no es un problema para los motores de TA a la hora de traducir. Sin embargo, las traducciones que se ofrecen presentan errores de conjugación, puesto que en la oración *Hast du Zeit?*, la persona que realiza la acción es la 2ª del singular, mientras que en las traducciones, se observa la 3ª del singular (forma de cortesía). Por tanto, mientras que el problema de ambigüedad podría darse como resuelto en este caso, siguen existiendo otros problemas de gramática.

Hutchins y Somers (1995), en lo que concierne a la TA, plantean diversos problemas de análisis y de transferencia e interlengua (cfr. Capítulo 1.1.2.2 y 1.2.3). Por un lado, los problemas de análisis tratan las dificultades monolingües que surgen en el texto de la lengua origen a la hora de ser traducido; por otro lado, los problemas de transferencia e interlingua se centra en la generación del texto en la lengua meta.

En lo que concierne al presente estudio, los problemas entre las diferencias sintácticas en el par de lenguas alemán-español. El principal problema con el que se puede encontrar un motor de TA es con el hecho de que, en español, la estructura sintáctica es SUJETO + VERBO + OBJETO, mientras que en alemán, a pesar de que el orden típico de la oración simple es (a) SUJETO + VERBO + OBJETO, también otras dos estructuras sintácticas tales como (b) SUJETO + OBJETO + VERBO, en el caso de las subordinadas, y (c) VERBO + SUJETO + OBJETO cuando se anteponen adverbios, oraciones subordinadas o sintagmas preposicionales.

- | | |
|--|--------------------------------|
| a. <i>Ich kaufe ein Buch</i> | Compro un libro |
| b. <i>Weil ich ein Buch habe, bin ich froh</i> | Como tengo un libro, soy feliz |
| c. <i>Nachts lese ich das Buch</i> | Por las noches leo el libro |

Si, además de este problema, se suma el hecho de que la expresión de anterioridad, simultaneidad y posterioridad de las perífrasis verbales en español, en alemán se formulan mediante un (d) adverbio en lugar de un verbo auxiliar junto con un verbo en infinitivo, gerundio o participio, es muy probable que los motores de TA no nos proporcionen la traducción esperada. Con todo, con el presente estudio se pretende dar respuesta a las hipótesis de las cuales parte y se ha originado la idea de realizarlo.

- | | |
|------------------------------------|----------------------|
| d. <i>Hast du wieder geraucht?</i> | ¿Has vuelto a fumar? |
|------------------------------------|----------------------|

En el ejemplo (d), el adverbio *wieder* actúa con el mismo significado que la perífrasis verbal VOLVER A + INFINITIVO, (cfr. Capítulo I. 3.2).

Por ahora, estos son los problemas con los que se pueden encontrar los motores de TA en relación con las perífrasis verbales; sin embargo, otro de los problemas con los que se encuentra la TA es el de las expresiones idiomáticas, debido a que con frecuencia modifican su estructura de una lengua a otra o no se expresan de la misma manera. Este es el caso, por ejemplo, de la cita con la que inicia este trabajo (e), puesto que lo que para los hispanohablantes nos suena a chino cuando algo es totalmente desconocido o no se entiende, para los 'suenan a español'.

TRADUCCIÓN

- | | | |
|-----------------------------------|--------|---|
| e. <i>Mir kommt Spanisch vor.</i> | DEEPL | 1. <i>Klingt für mich wie Chinesisch</i> |
| | GOOGLE | 2. <i>Ich höre mich an wie Chinesisch</i> |

En los resultados 1. y 2. se observa cómo la TA no reconoce la expresión idiomática, sino que traduce la frase literalmente. Este es un ejemplo de los muchos que se puede exponer tal como se verá más adelante (cfr. Capítulo VI; **Error! No se encuentra el origen de la referencia.**).

2.4 PROBLEMAS DE LA TRADUCCIÓN AUTOMÁTICA EN LA COMBINACIÓN LINGÜÍSTICA ALEMÁN-ESPAÑOL

En la TA con esta combinación lingüística, dado que son dos lenguas totalmente distintas gramaticalmente, es normal que se hallen problemas de equivalencia. Según García Yebra (1984) no hay casi interferencias entre estas lenguas porque entre ellas existe una distancia considerable, más que la que existente, por ejemplo, entre el francés y el castellano. En gran medida esto es debido a que el número de traducciones del alemán al castellano hasta el momento no son muy notable, en comparación con el inglés. Sin embargo, estos problemas son muy probables que con el paso de los años se solventen, ya que, en los últimos años, el contacto con la lengua alemana se ha intensificado debido al potencial económico que posee el país germano en la Unión Europea, es decir, debido a la crisis económica que ha sufrido España en estos últimos años y la fuerte demanda de trabajo de Alemania, el estudio de la lengua alemana se ha intensificado, lo cual ha provocado que haya un mayor contacto entre dichas.

2.5 LOS MOTORES DE TRADUCCIÓN AUTOMÁTICA

Una vez establecidos los objetivos y las hipótesis del trabajo, es hora de establecer los motores de traducción con los que se va a evaluar a la TA para observar si, en este caso, es recomendable el uso de algunos de estos motores en el par de lenguas de estudio.

Para nuestra investigación, en primer lugar, se han seleccionado aquellos motores que ofrecen traducción del alemán al español. En el caso de Google Translate, hoy en día ha pasado de ser TAE a ser TAN; sin embargo, dado que es uno de los motores más utilizados por los informantes, mediante una Hoja de cálculo de Google y la fórmula que se expone a continuación, se obtendrá la traducción que ofrecía cuando era traductor automático estadístico. De este modo, se investigará si los avances ofrecidos por la TAN se cumplen y, de algún modo, se determinará si se producen avances en el campo de la traducción automática.

- ❖ **Fórmula:** =GOOGLETRANSLATE(columnname; "source text abbreviation"; "target text abbreviation")
- ❖ **Ejemplo:** =GOOGLETRANSLATE(a1; "de"; "es")

A continuación, se expone un ejemplo para visualizar cuál es la intención de este ejercicio de evaluación del producto final de este motor con diferentes directrices de funcionamiento.

Ilustración 11. Muestra de traducción automática con una hoja de cálculo

	A	B	C
1	Oración	Traducción automática neuronal	Traducción automática estadística
2	Ich besuche eine Freundin hier in Nürnberg	Visito a un amigo aquí en Nuremberg	Me visita una novia aquí en Nuremberg

Las traducciones que se observan en la ilustración 11 no se pueden considerar como válidas por diversos motivos. En primer lugar, antes de empezar con el análisis de las traducciones, una posible traducción de esta oración podría ser ‘Estoy visitando a una amiga aquí en Núremberg’. A continuación, se exponen los errores de traducción de ambos motores, por un lado, a pesar de que el topónimo esté bien traducido, le falta el acento; por otro lado, en ninguna de las traducciones, la palabra *Freundin* está traducida de forma correcta. Este hecho se debe a que en alemán no hay distinción entre ‘novio /-a’ - ‘Amigo /-a’. *Freundin* con el artículo indeterminado ‘un’, se refiere a ‘una amiga’. La palabra ‘amigo’ en alemán es *Freund*. Un último error es la persona que realiza la acción del verbo. *Ich* en alemán es el equivalente al ‘Yo’ en español, por lo que solo está bien traducido en la oración de la columna b. Como reflexión final, cabe añadir que la traducción de Google como neuronal es mejor que la de Google estadístico.

A modo de resumen, Google Translate servirá como motor de TA estadístico y neuronal; Microsoft Translator y DeepL se emplearán como motores de TAN y, finalmente, Lucy es el elegido para ver las traducciones que ofrece como motor de TA basada en reglas.

- **GOOGLE TRANSLATE:** Actualmente este motor es capaz de traducir a más de 100 idiomas, desde el afrikáans al yiddish. Para conseguir una traducción basta con entrar en su página web²⁹, introducir la palabra, frase o texto en la lengua origen, seleccionar la lengua de destino y pulsar ‘Traducir’. En la actualidad, el traductor de Google ha pasado de ser estadístico a neuronal para algunos idiomas³⁰, lo que implica que las traducciones que ofrece para estos idiomas son más fiables. Para la mayoría de los idiomas cuya LO o LD no sea el inglés, para garantizar una traducción de calidad, el mecanismo que sigue este motor de TA es el siguiente:

²⁹ El traductor automático de Google está disponible de manera gratuita en <https://translate.google.com/?hl=es>

³⁰ Los idiomas son: alemán, chino, español, francés, inglés, japonés, portugués y turco.

en primer lugar, analiza la oración de la lengua de origen; en segundo lugar, la traduce al inglés y, finalmente, presenta la traducción en la LD.

Asimismo, el traductor de Google dispone de una aplicación móvil compatible con 90 idiomas, se puede traducir a 37 idiomas a través de una foto y a 32 a través de la voz en el modo conversación.

- **DEEPL:** El 29 de agosto de 2017, DeepL, una empresa europea con sede en Colonia (Alemania) sacó a la luz esta nueva herramienta de TA. Basada en redes neuronales artificiales y algoritmos para mejorar en los ámbitos de la inteligencia artificial, basta con entrar en su página web³¹, introducir la palabra, frase o texto en la lengua origen y después elegir a qué lengua meta se desea traducir. Este motor de TA ofrece 42 combinaciones posibles entre español, inglés, alemán, francés, italiano, polaco y neerlandés, pero estas poseen distintos niveles de efectividad según el par de lenguas, es decir, inglés-español por ahora proporcionará mejores resultados que español-neerlandés. Con todo, se está trabajando en incorporar el ruso, el chino y el japonés.

Con todo, DeepL lucha por conseguir la calidad de sus traducciones y, para ello, en lugar de manejar grandes cantidades de datos, incorpora a su base de datos de su ordenador ubicado en Islandia únicamente aquellas traducciones que están validadas por el ser humano. Por este motivo, hoy en día, es considerado como uno de los motores que ofrece las traducciones más precisas del mundo.

Este motor TAN está asociado al buscador de traducciones Linguee, cuya compañía fue fundada en 2009. Para la encuesta realizada, otro de los motores utilizados para evaluar las traducciones ha sido DeepL.

- **BING TRANSLATOR (MICROSOFT):** El 3 de junio de 2009 se lanzó el motor de TA Bing Translator. En la actualidad traduce 60 idiomas diferentes. Del mismo modo que los otros motores, a través de su página web³², de forma gratuita hasta los 5000 caracteres, puede obtenerse la traducción de cualquier palabra, frase o texto de una LO a una LD.

Del mismo modo que ocurre con el traductor de Google, a pesar de que este motor sea TAN, también recurre a la TAE para ofrecer sus traducciones dependiendo de la lengua.

³¹ El traductor automático de DeepL está disponible de manera gratuita en www.deepl.com/translator

³² El traductor automático de Microsoft está disponible de manera gratuita en <https://www.bing.com/translator>

Se puede integrar en otras aplicaciones Microsoft como Microsoft Office, Skype y Visual Studio mediante la aplicación Microsoft Translator. A través de su página web únicamente pueden traducirse textos y páginas web, pero, con la aplicación puede traducirse tanto texto como voz en tiempo real a 10 idiomas³³.

- **LUCY SOFTWARE:** Este motor de TA basado en reglas se fundó en 2006 por la compañía Lucy Software and Services GmbH a partir del programa desarrollado anteriormente, Comprendium. Se fundó con la idea de ofrecer servicios lingüísticos de calidad a empresas que se enfrentan a proyectos multilingües. Actualmente, el sistema Lucy LT traduce de 9 lenguas³⁴ de origen a diez lenguas de destino y, debido a que posee una arquitectura abierta, se pueden añadir direcciones de traducción utilizando, por ejemplos, el sistema Moses de código abierto u otros sistemas de TA.

A pesar de que a través de su página web³⁵ pueden traducirse palabras, frases o textos cortos, debido a es un servicio de pago. En el momento de introducir lo que se desee traducir, puede escogerse entre las pestañas *Vocabulario general*; *Vocabulario socio general* y *Vocabulario técnico general*.

3. LAS PERÍFRASIS VERBALES

Tal como se ha mencionado en epígrafes anteriores, la cuestión lingüística seleccionada para evaluar los sistemas de TA es el de las perífrasis verbales en castellano y su correspondencia en alemán. A pesar de que en la enseñanza-aprendizaje de una lengua extranjera siempre se persigue hallar las reglas que posibiliten la adquisición de las normas lingüísticas de la lengua que se estudia en cuestión, en el caso de las perífrasis verbales, no existe una regla para todos los casos.

El uso de las perífrasis verbales es muy frecuente en las lenguas románicas, a diferencia de lo que ocurre en otras lenguas indoeuropeas como el alemán, donde existe un sistema de verbos modales mucho más rico que el español, así como el uso de determinados adverbios que equivalen a las perífrasis en español. Por ejemplo, una perífrasis de infinitivo en español no puede ser traducida siempre por una perífrasis de

³³ Estos idiomas son: alemán, árabe, chino simplificado, chino tradicional, español, francés, inglés, italiano, japonés, portugués y ruso.

³⁴ Entre ellas destacan el vasco, el catalán y el gallego.

³⁵ El traductor automático de Lucy está disponible de manera gratuita en <http://www.lucysoftware.com/espanol/traduccion-automatizada/kwik-translator-/>

infinitivo en alemán debido a que la lengua alemana recurre a otras categorías gramaticales para expresar dicho valor. Por ello, ante las particularidades que presentan cada una de las lenguas, es muy probable que un motor de TA, tal como se ha explicado en epígrafes anteriores (cfr. Capítulo 1.2.3), no reconozca dichas diferencias entre el par de lenguas alemán-español.

3.1 CONCEPTO DE PERÍFRASIS VERBAL

A pesar de que no sea de gran relevancia entrar en el concepto de qué es una perífrasis verbal, a continuación, se muestran las definiciones ofrecidas por Roca Pons (1958), Gómez Torrego (1999) y Fernández de Castro (1999).

Según Gómez Torrego, una perífrasis es «la unión de dos o más verbos que sintácticamente constituyen *un solo núcleo del predicado*. De ese conjunto verbal, el verbo principal o auxiliado debe ser una forma no personal: infinitivo, gerundio o participio; el otro (u otros) verbo es siempre un mero auxiliar, pero será la forma que podrá conjugarse en todas o algunas formas de las formas de conjugación» (1999: 9).

Asimismo, Roca Pons³⁶ en su estudio sobre perífrasis verbales propone la siguiente definición «se entenderá por perífrasis verbal la unión de un verbo auxiliar – más o menos auxiliar o gramaticalizado (...)– con una forma nominal de un verbo conceptual» (1958: 10). Finalmente, Fernández Castro las define como:

[...] nos ha llevado a hablar tradicionalmente de perífrasis verbal, como término técnico con el que se designan los complejos verbales que no pueden ser considerados disjuntos (es decir integrados por un núcleo y su adyacente), sino que funcionan unitariamente como un solo núcleo oracional (1999: 16).

En el caso del verbo auxiliar, a pesar de que existen excepciones, para que un verbo funcione como auxiliar debe estar parcial o totalmente gramaticalizado, es decir, que haya perdido todo o parte de su significado. El verbo principal o auxiliado de una perífrasis verbal debe aparecer en infinitivo, gerundio o participio. Por tanto, tal como manifestó Alarcos (1994: 259) con «el núcleo oracional puede consistir en una combinación de unidades que funciona en conjunto como lo hace un solo verbo», se entiende que en una perífrasis verbal se precisan dos elementos que, a su vez, se necesitan mutuamente.

³⁶ J. ROCA PONS (1958). *Estudios sobre perífrasis verbales en español*. Madrid: C.S.I.C, pág. 58.

La elaboración del listado de las perífrasis verbales de este trabajo se fundamenta con base al propuesto por Andreu Castell (2008) y, que contiene un total de 11 y se distinguen entre las de infinitivo, participio y gerundio. A esta lista se añadirán las aportaciones 13 de Brigitte y Roberto Corcoll (2014). Este listado representa la correspondencia del infinitivo, gerundio y participio en la lengua alemana debido a que, en la mayoría de los casos, la perífrasis verbal está representada por otro carácter gramatical totalmente diferente, el adverbio.

Otro de los casos importantes en nuestra investigación es el de los verbos modales en las perífrasis verbales. Roca Pons³⁷ en su estudio sobre la gramaticalización expone que los verbos modales «no han sufrido propiamente ningún proceso de gramaticalización» (1958: 58). Este tipo de verbos conlleva una doble dificultad de traducción en el par de lenguas alemán-español porque se les puede atribuir una doble interpretación: la modalidad objetiva y la modalidad subjetiva. Por ello, la distinción y correcta interpretación de estos verbos es esencial para una buena comprensión del texto de origen. Con todo, es importante reconocer qué verbos modales de la lengua de destino corresponden al de origen. En este caso, dado que se parte de la hipótesis de que los motores de traducción automática son capaces de ofrecer la equivalencia de algunos de los verbos catalogados como modales.

Como verbo modal, en alemán, se trata de verbos que aparecen a modo de auxiliares de otro verbo con el infinitivo sin *zu*, con el que conforman un predicado complejo (Castell, 2008: 162). Los verbos modales³⁸ que expresan la modalidad objetiva, en la mayoría de los casos también expresan la modalidad subjetiva. Sin embargo, mientras que la modalidad objetiva expresa «cómo concibe o cómo debe interpretar el sujeto de la oración el enunciado: como una necesidad u obligación, una intención o deseo, una capacidad o posibilidad, permiso o prohibición, una exhortación o un consejo» (Castell, 2008: 163), la modalidad subjetiva expresa el grado de veracidad o certeza que el hablante otorga a lo que dice.

³⁷ J. ROCA PONS, *op. cit.*, pág. 10.

³⁸ Según Von Wright (1951, citado en Ridruejo, 1999: 3212) existen cuatro grupos de conceptos modales: En primer lugar, la modalidad alética, que expresa lo necesario, posible e imposible. En segundo lugar, la modalidad epistémica, que expresa lo sabido como cierto, indeciso y sabido como falso. En tercer lugar, la modalidad deóntica, que expresa lo obligatorio, permitido y prohibido y, finalmente, la modalidad existencia, que expresa lo universal, existente y nulo.

En el caso de la lengua española, los verbos modales se asocian a los auxiliares, dada la gramaticalización y la pérdida total o parcial del significado, así como su carácter de instrumento gramatical³⁹. Respecto a este trabajo, las construcciones perifrásticas que suscitan interés son las de infinitivo que expresan modalidad objetiva. Por tanto, del mismo modo que con las perífrasis verbales de infinitivo, gerundio y participio, se partirá de una lista basada en Gómez Torrego⁴⁰ (1999) y a García Fernández (2006):

- DEBER + INFINITIVO
- DEBER DE + INFINITIVO
- TENER QUE + INFINITIVO

3.1.1 CLASIFICACIÓN DE LAS PERÍFRASIS VERBALES

La lengua española es una lengua rica en perífrasis verbales. Mediante valores temporales, aspectuales, modales, el verbo auxiliar modifica la idea verbal expresada por el verbo principal. Para el presente estudio no es demasiado relevante cuántas perífrasis verbales posee la lengua española, sino más bien las diferencias existentes entre la combinación lingüística elegida, dado que si en castellano se emplea un verbo auxiliar + (un nexo) + un verbo en forma no personal, en alemán en muchas ocasiones, tal como se ha mencionado anteriormente, se recurre a un adverbio para expresar los mismos valores.

Por tanto, la estructura de una perífrasis verbal en español y en alemán se formularía de la siguiente manera:

❖ **PERÍFRASIS VERBAL EN ESPAÑOL**

VERBO AUXILIAR + {NEXO} + INFINITIVO/PARTICIPIO/GERUNDIO⁴¹

❖ **PERÍFRASIS VERBAL EN ALEMÁN**

VERBO + ADVERBIO – VERBO + CONJUNCIONES – ORACIONES SUBORDINADAS

A continuación, se expondrán las perífrasis verbales de infinitivo y se mostrarán las diferencias existentes entre ambas lenguas.

3.2 PERÍFRASIS VERBALES DE INFINITIVO

³⁹ Cfr. MORERA (1991, 29 y 35).

⁴⁰ Otros verbos que expresan modalidad, pero no se incluyen en la lista de Gómez Torrego (1999) serían *necesitar, desear, gustar*, etc.

⁴¹ Existen ciertas vacilaciones por parte de algunos autores para considerar algunas construcciones como perifrásticas. En este estudio, se tendrá en cuenta el criterio establecido por Gómez Torrego (1988).

Cuando en un texto alemán aparece una o varias expresiones que, si bien en alemán no muestran ningún distintivo especial, al trasladarlas al castellano, se podría mejorar a nivel idiomático su carácter durativo, incoativo, etc., mediante una perífrasis verbal, ya que la lengua alemana carece de un sistema tan amplio de perífrasis y otras construcciones verbales con infinitivo, participio o gerundio como es el caso del castellano.

A continuación, se exponen las perífrasis verbales de infinitivo que presentan una mayor relevancia para este estudio, ya que son las que en la lengua alemana se recurre al adverbio para expresar la misma idea que en español de anterioridad, simultaneidad o posterioridad.

- **ACABAR + INFINITIVO:** Para esta perífrasis verbal de posterioridad inmediata, el adverbio *gerade* es el encargado de trasladar al alemán el significado perifrástico. Significa (en este/aquel preciso instante) y únicamente se formula con el presente y el pretérito imperfecto;

Por tanto, por un lado, cuando en español se emplea el presente, en alemán recurre al *Perfekt*.

a. *Ich bin gerade aufgewacht* – Acabo de levantarme

Por otro lado, cuando en español se emplea el pretérito imperfecto, la lengua alemana recurre al *Plusquamperfekt*.

b. *Ich war gerade aufgewacht* – Acababa de levantarme

- **IR + INFINITIVO:** Para esta perífrasis, Castell (2008) y Corcoll (2010) exponen que la correspondencia de esta perífrasis depende del tiempo en que aparezca el verbo 'ir' en español.

Cuando este aparece en presente, equivale en alemán al presente o el futuro I del verbo que en español aparece en infinitivo.

c. *Wir sprechen morgen darüber* – Vamos a hablar de ello mañana (Presente)

d. *Ich werde mir einen Buch kaufen* – Voy a comprarme un libro (Futuro I)

El primer ejemplo podría traducirse con significado procedimental y composicional, procedimental sería la equivalencia al ejemplo que se expone y composicional sería traducir el ejemplo como *Mañana hablamos de ello*.

Tal como ocurre con este ejemplo de Castell (2008), ocurre con el que nos ofrece Corcoll (2010), ya que una de las posibilidades que ofrecen para trasladar el significado perifrástico al alemán es utilizar el adverbio *heute*.

e. *Juan kommt heute um 12 Uhr* – Juan va a venir hoy/Juan viene hoy

En el ejemplo (e), cabría la posibilidad de emplear ambas traducciones.

- **NACHSEHEN:** Esta es el único verbo que presenta una excepción en lengua alemana; este verbo, a pesar de que en su naturaleza significa ‘corregir’, ‘revisar’, una de las acepciones que también tiene es que significa ‘(ir a) ver’.

Por tanto, en el siguiente ejemplo sí se entiende el significado de la perífrasis.

f. *Wir sehen mal nach, was da los ist* – Vamos a ver qué pasa por ahí

- **PONERSE + INFINITIVO:** Del mismo modo que en español se usa la perífrasis *EMPEZAR/COMENZAR A + INFINITIVO*, en alemán *PONERSE + INFINITIVO* correspondería a los verbos *anfangen/beginnen* ‘empezar/comenzar’ seguidos de un infinitivo con *zu*. Por su parte, en la gramática de Corcoll (2010), además de estos verbos, también se propone *sich daran machen*:

g. *Um 10 Uhr abends machte ich mich daran/begann ich/fing ich an zu lernen* – A las 10 de la noche me puse a estudiar.

h. *Er hat angefangen zu regnen* – Se ha puesto a llover.

i. *Dann begann er zu schreien* – Entonces se puso a gritar.

- **SOLER + INFINITIVO:** Para la perífrasis *soler + infinitivo*, Castell (2008) propone dos maneras de trasladar el significado de la perífrasis en español al alemán.

Por un lado, el alemán recurre a la perífrasis *PFLEGEN + INFINITIVO* con *zu* cuando se expresa que el sujeto animado acostumbra, como acto volitivo, a hacer determinada cosa

j. *Er pflegt nach dem Abendessen spazieren zu gehen* – Suele ir a pasear después de cenar.

El verbo *pflegen*, en su naturaleza, se traduciría como ‘cuidar de/atender de’, pero otra de las acepciones que posee es ‘soler hacer algo’ (*etw zu pflegen*).

En caso de que no se trate de que un sujeto animado acostumbra a hacer determinada cosa, la perífrasis equivaldría al verbo conjugado acompañado de locuciones adverbiales como (*für*) *gewöhnlich* ‘normalmente’, *im Allgemeinen* ‘por regla general’, *meistens* ‘la mayoría de las veces

k. *Er ist (für) gewöhnlich netter* – Suele ser más simpático.

l. *Der Bus hatte meistens Verspätung* – El autobús solía llevar retraso.

- **VOLVER + INFINITIVO:** Para esta perífrasis también se recurre al adverbio para encontrar la correspondencia de significado en alemán. Los adverbios que a los que se puede recurrir para expresar el significado perifrástico son los siguientes:
 - *wieder* 'de nuevo'
 - *nochmal* 'otra vez'

Para esta perífrasis, tal como ocurre en otras descritas anteriormente, Corcoll (2010) añade como nexos la preposición *a* entre el verbo auxiliar y el infinitivo, además de añadir otras posibilidades equivalentes: *erneut* y *nie*.

m. *Juan komm wieder zu spät* – Juan volvió a llegar tarde

n. *Der Chef sagte mir erneut, dass ich pünktlich sein sollte* - El jefe me volvió a decir que fuera puntual.

o. *Hast du wieder gegessen?* – ¿Has vuelto a comer?

p. *Als ich wieder ins Haus ging...* – cuando volví a entrar en casa..

3.3 PERÍFRASIS VERBALES DE PARTICIPIO⁴²

En comparación con la lengua española, la alemana carece de recursos para expresar una acción perfectiva. En consecuencia, a continuación, únicamente⁴³ se exponen las que, entre las diversas traducciones existentes en alemán, realizan su correspondencia con el adverbio *bereits* o la preposición *seit*.

- **TENER + PARTICIPIO:** Se realiza con el verbo *haben* + *Partizip II*. La estructura oracional alemana exige que frente a un verbo auxiliar como *haben*, el verbo auxiliado se halle al final de la frase.

a. *Ich habe bereits 20 Seiten geschrieben* – Tengo escritas ya 20 páginas.

- **LLEVAR + PARTICIPIO:** Dicha perífrasis puede realizarse mediante el adverbio *bereit* o la preposición *seit*. En el primero de los casos, tendría la misma traducción que en el ejemplo anterior. En el caso de la preposición, *seit* equivale en español a 'desde'.

b. *Mads Nielsen seit einer Woche vermisst* – Mads Nielsen lleva desaparecido una semana⁴⁴.

⁴² Debido al carácter predicativo del participio, las perífrasis de participio presentan ciertas dificultades de adscripción para algunos teóricos.

⁴³ En la gramática de Corcoll (2014) se exponen otras perífrasis verbales de participio tales como «Ir / andar + participio»; «Seguir + participio»; «Quedar + participio»; «Dejar + participio»; «Dar(se) por + participio».

⁴⁴ Este ejemplo se ha extraído de la serie *Dark* (2017).

3.4 PERÍFRASIS VERBALES DE GERUNDIO

En este epígrafe, se exponen algunas de las perífrasis verbales de gerundio que presentan una mayor relevancia para este estudio, ya que son las que en la lengua alemana se recurre al adverbio para expresar la misma idea que en español de anterioridad, simultaneidad o posterioridad.

En el caso de las perífrasis de gerundio, en la lengua alemana se prefiere el uso de adverbios tal como sucede en el caso de las perífrasis de infinitivo. Sin embargo, también se puede recurrir al uso de subordinadas.

- **ESTAR + GERUNDIO:** Cuando el verbo está conjugado en presente o con el verbo en pasado junto al adverbio *gerade*, en tales casos, se traduce por una perífrasis en español.

a. *Ich lese gerade ein Buch* – Estoy leyendo una carta.

b. *Ich habe gerade ein Buch gelesen* – He estado leyendo un libro.

Además del adverbio *gerade*, cuando el gerundio exprese una acción no concluida, también puede recurrirse al verbo *sein*, el adverbio preposicional *dabei* y una estructura oracional de infinitivo, así como a la estructura *im Begriff sein*.

c. *Wir waren gerade dabei, die Koffer zu machen* – Estábamos haciendo las maletas.

Con todo, el aspecto durativo en algunos casos se pierde. Obsérvese el ejemplo (b), en el que mediante el adverbio ‘ganz’, cuyo significado es ‘todo’, se realiza la perífrasis verbal con acción terminativa.

a. Ayer te estuve buscando todo el día – *Gestern habe ich dich den ganzen Tag gesucht*.

- **IR + GERUNDIO:** Esta perífrasis posee varias posibilidades de traducción al alemán, entre ellas: *allmählich*, *mehr und mehr*, *langsam* y *zusehends*.

e. El tiempo va mejorando – *Das Wetter wird allmählich besser*.

- **SEGUIR / CONTINUAR + GERUNDIO:** En lo que respecta a esta perífrasis, podría traducirse añadiendo el prefijo *weiter-* al verbo o bien con el adverbio *immer noch*.

f. Sigue lloviendo – *Es regnet immer noch*.

- **VENIR + GERUNDIO:** En lo que concierne a la traducción de esta perífrasis, puede realizarse mediante *ununterbrochen*, *solange*, *bis* y *seit*.

g. De la forma en que viene haciéndose hasta nuestros días – *So, wie mn es bis zum heutigen Tag macht*.

- **EMPEZAR + GERUNDIO:** Esta perífrasis puede realizarse mediante *Am Anfang* o *zuerst*.

h. Empezó trabajando de camarero – *Am anfang arbeitete er als Kellner*

3.5 PERÍFRASIS VERBALES MODALES

- **KÖNNEN:** Se recurre a este verbo cuando ‘poder’ expresa una posibilidad objetiva. Asimismo, si el infinitivo hacer referencia a una capacidad física o intelectual, se usa *können* + infinitivo sin *zu*.
 - a. *Hier kann man nicht parken. Die Straße ist zu eng.* – Aquí no se puede aparcar, la calle es demasiado estrecha.
 - b. *Sie kann sehr gut schwimmen* – Sabe nadar muy bien.
- **DÜRFEN:** Si lo que se desea es transmitir un permiso o prohibición, en lugar de utilizar *können*, se recurre a *dürfen*. Compárese el ejemplo (c) con el (a) del ejemplo expuesto en el punto anterior.
 - c. *Hier darf man nicht parken. Es ist verboten* – Aquí no se puede aparcar. Está prohibido.
- **MÜSSEN:** Este verbo se emplea cuando su traducción en español corresponde a ‘deber’ para expresar una necesidad o una obligación.
 - d. *Ich muss gehen, sonst verpasse ich den Zug* – Me tengo que ir, sino perderé el tren.
 - e. *Es ist schon spät. Wir müssen uns beeilen* – Ya es tarde- Tenemos que darnos prisa.
- **SOLLEN:** Se emplea para expresar una obligación si se quiere transmitir al interlocutor que se trata de una obligación impuesta por otra persona. Asimismo, también se emplea con el sentido de ‘consejo u orden de otra persona’.
 - f. *Ich soll morgen nach Paris* – Mañana tengo que ir a París.
 - g. *Ich soll mehr Sport treiben. Das hat der Arzt gesagt* – Debo hacer más deporte. Lo ha dicho el médico.

Las perífrasis verbales que se evalúan forman parte de diferentes libros de gramática contrastiva entre el alemán y el español⁴⁵, algunos de ellos se han expuesto a lo largo de las explicaciones. Asimismo, también se plantean ejemplos presentados por la autora de este trabajo. A continuación, se presentan las 50 perífrasis verbales que son objeto de evaluación de la TA.

⁴⁵ Andreu Castell (2008); Corcoll (2010) y Astrid Schmidhofer (2014), así como de la autora del presente trabajo.

Tabla 7. Perífrasis verbales que se evalúan

PERÍFRASIS VERBAL EN ESPAÑOL	EQUIVALENTE EN ALEMÁN
¿Has vuelto a comer?	<i>Hast du wieder gegessen?</i>
Estoy cantando	<i>Ich singe gerade</i>
Sigo leyendo	<i>Ich lese weiter</i>
Estoy leyendo el periódico	<i>Ich bin gerade dabei, Zeitung zu lesen</i>
Cocinamos cantando	<i>Während wir kochen, singen wir</i>
Está escribiendo una carta	<i>Er ist dabei, einen Brief zu schreiben</i>
Voy a ir yendo	<i>'Ich werde gehen'</i>
Se ha puesto a llover	<i>Er hat angefangen zu regnen</i>
Suele ir a pasear después de cenar	<i>Er pflegt nach dem Abendessen spazieren zu gehen</i>
Me pongo a trabajar	<i>Ich beginne jetzt zu arbeiten</i>
Tienes que comer más	<i>Du musst mehr essen</i>
Mañana debe de llover	<i>Es kann sein, dass es morgen regnet</i>
El cocinero está preparando la comida	<i>Der Koch bereitet gerade das Essen zu</i>
Mañana volvemos a vernos	<i>Morgen sehen wir uns wieder</i>
Tus malas notas este año me vienen preocupando	<i>Deine schlechten Noten dieses Jahr machen wir große Sorgen</i>
Mads Nielsen lleva una semana desaparecido	<i>Mads Nielsen seit einer Woche vermisst</i>
Ayer volvió a nevar	<i>Gestern schneite es wieder</i>
El cartero suele pasar a eso de las 9 horas	<i>Meistens kommt der Briefträger gegen 9 Uhr</i>
Acabó por escribir la carta	<i>Schließlich schrieb er den Brief</i>
Tengo que volver a llamar	<i>Ich muss wieder anrufen</i>
Te lo voy a explicar	<i>Ich werde es dir erklären</i>
Iba a decírselo, pero...	<i>Ich wollte es ihm sagen, aber</i>
El autobús solía llevar retraso	<i>Der Bus hatte meistens Verspätung</i>
Él ha vuelto a fumar	<i>Er hat wieder geraucht</i>
Entonces se puso a gritar	<i>Dann begann er zu schreien</i>
Juan va a venir hoy a las 12 horas	<i>Juan kommt heute um 12 Uhr</i>
Acabo de comer	<i>Ich habe gerade gegessen</i>
Estoy comiendo	<i>Ich esse gerade</i>

Acabo de ver a Klaus	<i>Ich habe gerade Klaus gesehen</i>
Mañana va a llover	<i>Morgen wird es regnen</i>
Sigue lloviendo	<i>Es regnet (immer) noch</i>
Mi padre se va haciendo viejo	<i>Mein Vater wird langsam älter</i>
Esto lo vengo haciendo así desde hace años	<i>Das mache ich schon seit Jahren so</i>
Estamos haciendo las maletas	<i>Wir sind (gerade) dabei, die Koffer zu packen</i>
¡No vuelvas a llegar tarde!	<i>Komm nie wieder später!</i>
Suele levantarse a las 6	<i>Er steht meistens um 6 Uhr auf</i>
No llego a comprender lo que dices	<i>Ich verstehe nicht ganz, was du sagst</i>
María se echó a llorar	<i>Maria brach in Tränen</i>
Vamos a ver qué pasa por ahí	<i>Wir sehen mal nach, was da los ist</i>
Viene a ganar 1800€ al mes	<i>Er verdient etwa 1800€ im Monat</i>
El tiempo va mejorando	<i>Das Wetter wird allmählich besser</i>
Le encontré durmiendo	<i>Ich fand ihn schlafend</i>
Él volvió cantando	<i>Singend kam er zurück</i>
Tengo escritas ya 20 páginas	<i>Ich habe bereits 20 Seiten geschrieben</i>
Acabó estudiando derecho	<i>Am Anfang arbeitete er als Kellner</i>
Él debe venir hoy	<i>Er soll heute kommen</i>
Mañana tengo que ir a París	<i>Ich soll morgen nach Paris</i>
Mañana tengo que ir a París	<i>Ich muss morgen nach Paris</i>
Tengo que irme, sino perderé el tren	<i>Ich muss gehen, sonst verpasse ich den Zug</i>
No debes comer chocolate si quieres adelgazar	<i>Du darfst keine Schokolade esse, wenn du abnehmen willst</i>

4. EXPRESIONES IDIOMÁTICAS

Las expresiones idiomáticas, objeto de estudio de la fraseología, son secuencias de palabras cuyo significado es compositivo, es decir, su significado no se deriva del significado de sus componentes, sino del conjunto que estos forman. La comparación interlingüística de expresiones idiomáticas ha sido objeto de estudio en numerosos proyectos⁴⁶ debido a que algunas presentan pequeñas diferencias en la estructura

⁴⁶ Entre otros: Awwad, 1990; Roberts, 1998; Mellado 2000 y Corpas 2000ab, 2003.

morfosintáctica. La expresión en alemán de ‘Andarse con cuidado’, se traduce como *auf der Hut sein*; *Hut* en este caso no tiene nada que ver con sombrero, sino que significa ‘protección’.

La dificultad que existe generalmente a la hora de traducir determinadas expresiones se debe a que estas admiten una interpretación literal y otra figurada; en el caso de la literal, el riesgo que hay es que se malinterprete el significado, por ello, la interpretación preferente suele ser la figurada.

Un ejemplo que podría tener una interpretación literal sería el de *etwas unter die Lupe nehmen*, cuya traducción en español es ‘Mirar con lupa’. Este ejemplo se emplearía para alguien que busque observar algo detenidamente o entenderlo en profundidad. Sin embargo, el siguiente ejemplo no tiene alternativa de interpretación literal debido a que ninguno de los componentes representa algo en común en ambas lenguas, es decir, mientras que en alemán se dice *wer nicht hören will, muss fühlen* ‘Quien no quiere escuchar, tiene que sentir’, en español se expresa ‘La letra con sangre entra’.

Las expresiones idiomáticas que se evalúan forman parte del libro *Du kannst du laut sagen* de Jorge Sece y Stefanie Wimmer, ambos profesores del departamento de Traducción e Interpretación de la Universitat Autònoma de Barcelona. En él se hallan 1000 expresiones en alemán para hablar como un nativo; en este caso, el libro está pensado para estudiantes de alemán como lengua extranjera, pero, dado que a lo largo de los capítulos se encuentran tanto las expresiones en alemán como en español, se puede tener en cuenta como diccionario de expresiones idiomáticas. A continuación, se presenta el listado con las 50 expresiones⁴⁷ que son objeto de evaluación de la TA. En algunos casos, algunas de las expresiones poseen la misma composición en alemán y en español; sin embargo, en otras poseen una estructura totalmente distinta. Esto puede generar que los resultados sean favorables o desfavorables en función del entrenamiento del motor de TA.

Tabla 8. Expresiones idiomáticas que se evalúan

EXPRESIONES EN ALEMÁN	EXPRESIONES EN ESPAÑOL
<i>mit (dem) Feuer spielen</i>	Jugar con fuego
<i>seinem Herzen einen Stoß geben</i>	Hacer de tripas corazón

⁴⁷ Las expresiones idiomáticas han sido seleccionadas según el criterio de la autora de este trabajo.

<i>um jeden Preis</i>	A toda costa
<i>auf dem Spiel stehen</i>	Estar en juego
<i>ein heißer Ofen</i>	Estar como un tren
<i>ein Mauerblümchen sein</i>	Ser un patito feo
<i>eine graue Maus sein</i>	Ser del montón
<i>in guten Händen sein</i>	Estar en buenas manos
<i>eine Zicke sein</i>	Ser una bruja
<i>jemanden vor den Kopf stoßen</i>	Hacer un feo
<i>eine Null sein</i>	Ser un desastre
<i>tote Hose</i>	Ser un muermo
<i>die Nerven verlieren</i>	Perder los nervios
<i>jemanden auf die Folter spannen</i>	Tener a alguien en vilo
<i>viel um die Ohren haben</i>	Estar muy liado
<i>hin oder her⁴⁸</i>	más o menos
<i>Pi mal Daumen</i>	A ojo de buen cubero
<i>in Hülle und Fülle</i>	A porrillo / a montones
<i>um Haaresbreite / um ein Haar breit</i>	Por los pelos
<i>von A bis Z</i>	De pe a pa
<i>Von vorne bis hinten</i>	De cabo a rabo
<i>auf ex</i>	De un trago
<i>Blau sein</i>	Ir / estar: pedo / como una cuba
<i>einen Kater haben</i>	Tener resaca
<i>im Wein liegr die Wahrheit</i>	Los borrachos y los niños siempre dicen la verdad
<i>ganz Ohr sein</i>	Ser todo oídos
<i>jemandem spanisch vorkommen</i>	Sonar a chino
<i>das Geld zum Fenster hinaus werfen</i>	Tirar la casa por la ventana
<i>Pleite sein</i>	Estar sin blanca
<i>ohne Wenn und Aber</i>	Sin peros
<i>nicht aus Holz sein</i>	No ser de piedra
<i>auf geht's!</i>	¡Vamos!
<i>Fix und fertig</i>	Estar listo

⁴⁸ En caso de no emplear dicha expresión idiomática, también se dice *mehr oder weniger* que equivaldría a 'más o menos', tal como se observa en la columna de la derecha.

<i>alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene su final
<i>Berge versetzen</i>	Mover montañas
<i>Himmel und Hölle in Bewegung setzen</i>	Remover cielo y tierra
<i>jemandem die Meinung geigen</i>	Cantar las cuarenta a alguien
<i>mit dem linken Fuß aufstehen</i>	Levantarse con el pie izquierdo
<i>die Nase voll haben</i>	Estar hasta las narices
<i>die Zeit totschiagen</i>	Matar el tiempo
<i>jemanden sausen lassen</i>	Dejar plantado a alguien
<i>jemanden zappeln lassen</i>	Tener a alguien en vilo
<i>in den Tag hinein leben</i>	Vivir al día
<i>meine Rede!</i>	¡Eso digo yo!
<i>von wegen!</i>	¡ni hablar!
<i>bis in die Puppen</i>	Hasta la tantas
<i>von früh bis spät</i>	De sol a sol
<i>es ist nicht alles Gold, was glänzt</i>	No es oro todo lo que reluce
<i>jede Wette</i>	Pongo la mano en el fuego
<i>Mist bauen</i>	Meter la pata

5. MÉTODOS DE EVALUACIÓN DE LA CALIDAD DE LA TRADUCCIÓN AUTOMÁTICA

Los métodos de control de calidad servirán para comparar los diferentes motores de TA para determinar si uno ofrece una traducción mejor que el otro. En este caso, los errores de traducción se evaluarán, por un lado, mediante métricas automáticas de estimación de calidad como BLEU (*Bilingual Evaluation Understudy*) (Papineni *et al.*, 2002). BLEU⁴⁹ fue una de las primeras métricas utilizada y fue implementada por IBM. La idea de estas métricas es encontrar la métrica automática mediante traducciones humanas, es decir, compara el número de coincidencias de *n-gramas* (palabras) entre el resultado del sistema de TA y una o más traducciones de referencia (traducciones humanas). Por tanto, no se compara la traducción con el original, sino con la referencia y puede calcularse utilizando más de una traducción como referencia, así como determinar cuán cerca se encuentra la traducción del motor de TA en relación con la traducción de referencia. La

⁴⁹ En el caso del presente trabajo, se empleará el sistema BLEU de: TILDE. Tilde Custom Machine Translation [en línea]. <https://www.letsmt.eu/Bleu.aspx>

métrica de BLEU es independiente de las lenguas de origen y de destino y las puntuaciones se corresponden a los juicios realizados por el humano.

BLEU combina medidas de precisión calculadas para segmentos de distinta longitud. El resultado de la calidad es siempre mayor o igual que cero y menor o igual que uno. Cuanto más se acerque al uno, de mejor calidad estará catalogada. La traducción automática debe coincidir con la traducción de referencia en la selección de palabras, el orden de palabras y la longitud para obtener una buena evaluación. Con todo, puede haber alguna crítica en las medidas de la evaluación de la TA: en primer lugar, trata a todas las palabras por igual; en segundo lugar, no operan a nivel sintáctico, sino que se basan en comparar cadenas de caracteres y, finalmente, las traducciones humanas pueden recibir puntuaciones bajas.

En el caso de este trabajo, el sistema BLEU de TILDE permite introducir tanto el archivo con las frases en idioma de origen, el archivo con las frases traducidas por un humano y otro archivo con las traducciones de los distintos motores que se evalúan. Asimismo, de manera opcional, también puede añadirse otro archivo con las traducciones de un segundo motor. De esta manera, podrían evaluarse dos motores de TA a la vez. Con todo, para la evaluación llevada a cabo en este trabajo, únicamente se ha seleccionado un archivo por motor en cada paso, es decir, primero se ha evaluado un motor, después otro y así sucesivamente hasta evaluar los cinco motores.

Tal como se observará, en múltiples casos, la precisión de *n-gramas* penaliza las palabras de la LD que no aparecen en la de referencia, es decir, en la traducción humana y en la longitud de la traducción. Dicho en otras palabras, tanto la longitud de la traducción de referencia como la longitud en la lengua de destino deben ser similares. No obstante, estas puntuaciones no se tendrán demasiado en cuenta debido a que la estructura gramatical de cada una de las lenguas de trabajo es totalmente distinta.

CAPÍTULO II

METODOLOGÍA

Translation is that which transforms
everything so that nothing changes.

GÜNTER GRASS

El presente capítulo muestra la metodología que se va a seguir para la investigación del alcance de la traducción automática en torno a las cuestiones lingüísticas que se definen en el marco teórico (cfr. Capítulo I).

Así pues, en relación con el marco teórico, los objetivos y la hipótesis que se han descrito en el capítulo anterior, el marco metodológico está dividido en 2 partes. En primer lugar, se expondrán los ejemplos que servirán para evaluar los diferentes motores de TA seleccionados. Para ello, se mostrará una lista con cada una de las 50 perífrasis verbales seleccionadas mediante los diferentes manuales y gramáticas contrastivas entre el alemán y el español, así como las 50 expresiones idiomáticas junto a las respectivas traducciones ofrecidas por dichos motores. El criterio de selección de las perífrasis verbales y las expresiones idiomáticas se ha realizado bajo criterio de la autora de este trabajo y su tutor. Algunas de las expresiones idiomáticas, con el fin de obtener buenos resultados, mantienen la misma estructura en alemán y en español.

Por otro lado, la selección de los motores de TA se debe a la finalidad de obtener qué motor ofrece mejores resultados según las traducciones que ofrece para la combinación lingüística alemán-español. Por ello, se ha seleccionado uno basado en reglas, otro estadístico y varios neuronales. El hecho de que se haya escogido varios neuronales se debe a que hoy en día se considera que ofrecen mejores resultados que los basados en reglas o los estadísticos.

La metodología para evaluar la TA se realiza mediante la plataforma TILDE, donde mediante las oraciones en lengua origen, la traducción de referencia y las traducciones de los diferentes motores, se establecerá qué motor es mejor.

Así pues, el procedimiento será el siguiente: por un lado, en un documento .txt se copiarán las frases en lengua origen. Estas deben colocarse una debajo de la otra para que la plataforma pueda alinearlas con los otros documentos. Por otro lado, en otro documento .txt se reproducirán las frases traducidas por un humano y, finalmente, en otro documento .txt se transcribirán las traducciones de los diferentes motores de forma independiente, es decir, habrá un documento para las traducciones de Lucy Software, otro para Google Translate como estadístico y otro para Google Translate como neuronal, otro para las de DeepL y finalmente, otro para las de Microsoft Translator. Además de que las frases deben estar alineadas entre sí, para garantizar que TILDE pueda realizar la evaluación de la TA, la codificación de los archivos .txt debe ser UTF-8⁵⁰.

Los resultados se calculan del 0 al 100. Se consideran óptimos aquellos resultados que oscilan entre el 50 y el 60. La evaluación se realiza mediante la traducción de referencia, es decir, la humana y la traducción de cada uno de los diferentes motores.

Con todo, debido a la extensión exigida para este trabajo de investigación, tal como se ha mencionado en epígrafes anteriores, únicamente se plantean los ejemplos de las perífrasis verbales y algunas expresiones idiomáticas con los que se va a evaluar la TA, así como el sistema de evaluación basado en BLEU. Sin embargo, en el mundo de la TA existen muchas cuestiones a tratar entre el mismo par de lenguas o con otras combinaciones además de la que existen otros sistemas con los que evaluar la TA, por lo que estas últimas cuestiones podrían llevarse a cabo en una tesis doctoral.

⁵⁰ Recuérdese que, por defecto, la codificación que aparece es AINSI.

En el presente capítulo se ofrece la descripción⁵¹ de los datos; es decir, por un lado, en cada uno de los epígrafes que se presentan a continuación, de forma general, se describen aquellas traducciones ofrecidas por los diferentes motores de TA objeto de estudio en este trabajo que presentan mayor relevancia para dar respuesta a las hipótesis de las que se parte; por otro lado, se muestran los resultados obtenidos de la evaluación de los motores mediante la métrica BLEU.

1. RESULTADOS DE LA EVALUACIÓN DE LAS PERÍFRASIS VERBALES

Antes de profundizar en el análisis de la evaluación de la TA ofrecido por la métrica BLEU de TILDE, es conveniente dedicar una parte de este epígrafe a la reflexión personal sobre las traducciones obtenidas. En primer lugar, por lo general, en lo que respecta a las perífrasis verbales, independientemente de si son de infinitivo, gerundio, participio o modales, los resultados no son los esperados. Cabe añadir, que las hipótesis de partida muestran que antes de empezar con la evaluación, no se confiaba demasiado en los motores de traducción automática en lo que respecta a la combinación idiomática alemán-español ni en las cuestiones lingüísticas por las que se optó para evaluar la TA. Sin más, a continuación, se detallan los resultados de la evaluación de los diferentes motores junto a una reflexión personal sobre las traducciones obtenidas.

Recuérdese que la métrica se calcula del 0 al 1 y es favorable cuando el resultado supera el 50; además, si en alguno de los casos, el resultado oscila entre el 60 y el 80, puede decirse que el resultado es favorable para recomendar el uso de la TA. Sin más, a continuación, se detallan cada uno de los resultados por motor⁵².

⁵¹ Debido a que las frases que se evalúan contienen un grado de complejidad mayor a cualquier otra oración, la gran mayoría de las frases que se han seleccionado son de longitud corta.

⁵² Debido a la extensión de este trabajo, únicamente se comentarán las traducciones que generen mayor controversia para determinar si el resultado del motor de TA es favorable o desfavorable.

1.1 RESULTADOS DE LAS PERÍFRASIS VERBALES DE INFINITIVO

A continuación, se exponen los resultados obtenidos de la métrica BLEU en relación con la evaluación de cada uno de los motores para los distintos tipos de motores de TA.

Tabla 9. Métrica de BLEU para las perífrasis verbales de infinitivo

LUCY	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT
30.35	17.67	43.71	35.77	33.58

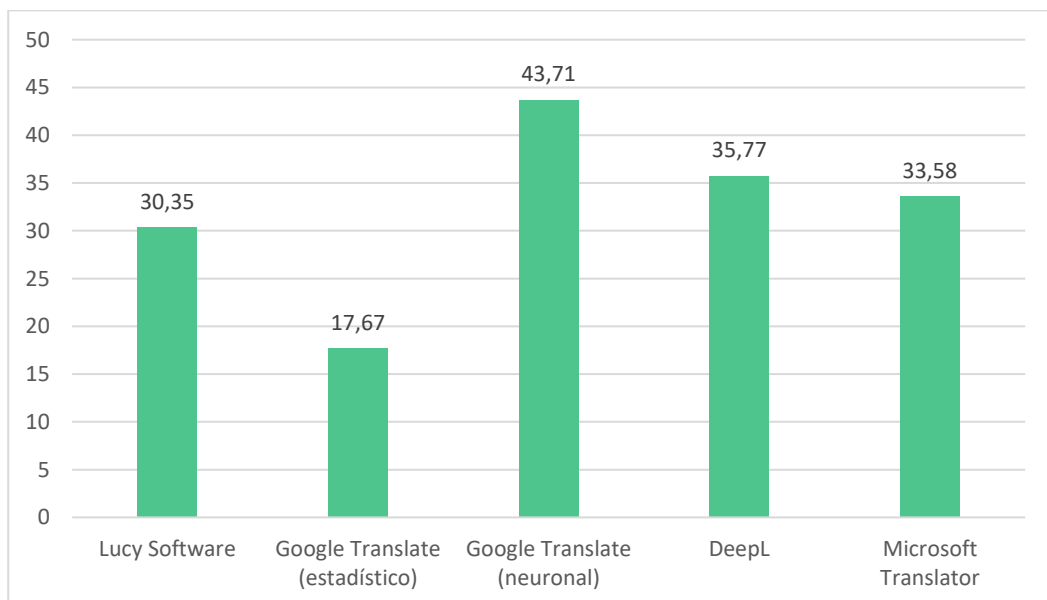
- **RESULTADOS DE LUCY SOFTWARE:** En lo que respecta a la evaluación de este motor ha obtenido la segunda mayor puntuación. En total 9 de 23 oraciones las ha traducido de forma correcta; asimismo, cabe destacar que en varias oraciones se emplean los adverbios *wieder* y *gerade*. Una de las traducciones que resulta interesante mencionar es la de ‘Acabo de comer’, ya que tal como se explicó en el apartado de las perífrasis verbales, en alemán, además de recurrirse al adverbio *gerade*, también cambia de tiempo verbal, es decir, en lugar de presente se emplea el *Perfekt*, que equivale al pasado en ambas lenguas.
- **RESULTADOS DE GOOGLE TRANSLATE ESTADÍSTICO:** Este motor, únicamente ha recibido un 100 en la oración ‘Ayer volvió a nevar’, *Gestern schneite es wieder*, por lo que nuevamente se observa el uso del adverbio *wieder*. Las otras traducciones poseen estructuras totalmente diferentes a la de referencia, de ahí la métrica tan baja. Por ejemplo, en la oración *Juan wird kommen heute um 12 Uhr*, un nativo de alemán observaría que el verbo *kommen* debe ir al final de la oración. Sin embargo, si se realiza la traducción a la inversa, es decir, traducir la oración que se ha obtenido al español, la traducción que proporciona Google Translate es la perífrasis verbal de participio, mientras que, si realizamos la búsqueda con la estructura gramatical correcta *Juan wird heute um 12 Uhr kommen*, la traducción que realiza dicho motor es en futuro y no con la perífrasis verbal.
- **RESULTADOS DE GOOGLE TRANSLATE NEURONAL:** En el caso de las perífrasis verbales, Google Translate es el que mayor puntuación ha obtenido. Con un total de 11 oraciones⁵³ traducidas como la de referencia, Google Translate ha demostrado

⁵³ En algunos casos, se observan diferencias insignificantes, por lo que se han contado como correctas.

que ofrece mejores resultados de los esperados. En varias traducciones emplea el uso de los adverbios *wieder* y *gerade* para dar valor perifrástico a la oración.

- **RESULTADOS DE DEEPL:** En lo que concierne a este motor, 6 de 23 oraciones están traducidas como las de referencia. Asimismo, 3 de ellas, en lo que respecta a las perífrasis verbales están bien traducidas; sin embargo, los errores que ha cometido el motor a la hora de traducir son de estructura gramatical o de la persona sobre la cual recae la acción verbal.
- **RESULTADOS DE MICROSOFT TRANSLATOR:** Para este motor de TA, cabe mencionar que se esperaban mejores resultados a los obtenidos. Con todo, en algunas oraciones se observa el uso de los adverbios *wieder* y *gerade*. Asimismo, algunas penalizaciones que ha obtenido se deben a que en lugar de usar el mismo verbo que la oración de referencia, emplea otro distinto. Dado que en algunas ocasiones el uso de un verbo u otro no implica un cambio de significado, algunas de las traducciones de Microsoft Translator se pueden dar como válidas, por ejemplo, en el caso de *Morgen sehen wir uns wieder* (traducción de referencia) y *Morgen treffen wir uns wieder* (traducción del motor de TA).

Gráfico 1. Gráfico con los resultados de las perífrasis verbales de infinitivo



1.2 RESULTADOS DE LAS PERÍFRASIS VERBALES DE GERUNDIO

A continuación, se detalla la métrica de evaluación de cada uno de los motores de TA y, posteriormente, se expondrá si se aconseja o no el uso de la TA para las perífrasis verbales de gerundio. Antes de empezar con los resultados, es interesante mencionar que no se esperan grandes expectativas en cuanto a los resultados de la evaluación de las traducciones de estas perífrasis debido a que en alemán hay diversas formas para expresar el gerundio en español.

Tabla 10. Métrica de BLEU para las perífrasis verbales de gerundio

LUCY	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT
9.69	11.83	15.30	25.46	15.40

- **RESULTADOS DE LUCY SOFTWARE:** En relación con la evaluación de las perífrasis verbales de gerundio, tal como se observa en la tabla, es el motor que peor puntuación ha obtenido. Con todo, Lucy Software, en algunos casos, ha tenido en consideración los adverbios a los que recurre la lengua alemana para expresar el valor perifrástico en español. Un ejemplo de ello es la oración *Ich lese weiter*. Sin embargo, en tres casos se precisaba el uso del adverbio *gerade* y dicho motor no ha recurrido a él, simplemente ha traducido la frase al alemán en presente. Si se hace la prueba con *Ich esse gerade / Ich singe gerade*, se observa que la traducción al español que ofrece tampoco refleja el valor perifrástico; es decir, las traducciones quedarían ‘Yo como precisamente / Yo canto precisamente’.

Sin embargo, cabe destacar el siguiente ejemplo, ya que a pesar de que la estructura varíe, la traducción es totalmente válida a pesar de que la puntuación BLEU sea baja. Compárese el siguiente ejemplo: *Das mache ich schon seit Jahren so* vs. *Ich mache das schon seit Jahren*. La única diferencia es la posición del sujeto y del pronombre demostrativo.

Asimismo, es importante destacar que, en una de las traducciones de expresiones idiomáticas, Lucy Software proporciona dos opciones con el fin de escoger la traducción correcta en función del contexto. El ejemplo al que se está haciendo referencia es el de *Zeit und Wetter* (cfr. 1.2.3).

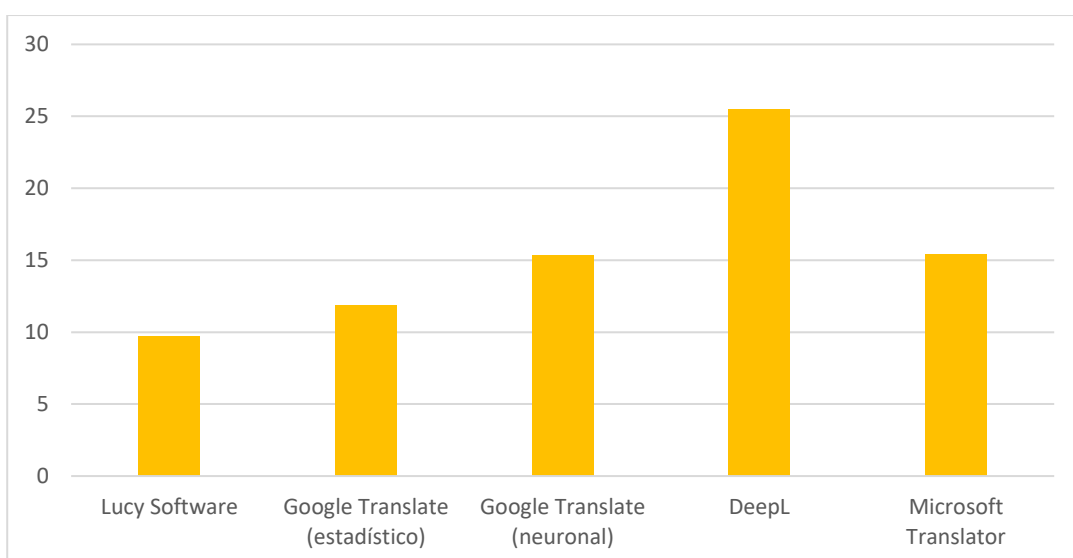
También puede darse como válida la traducción de la oración 'Juan va a venir hoy a las 12 horas' debido a que la única diferencia que existe es que a la traducción se le ha añadido el verbo *werde*, es decir, la frase queda así: *Juan wird heute um 12 Uhr kommt*. Sin embargo, la presencia de este verbo genera un problema, y es que dicho verbo se emplea para expresar el Futur I en alemán y, por ello, si se realiza la prueba a la inversa, la traducción que se obtiene es: Juan vendrá hoy a las 12.

- **RESULTADOS DE GOOGLE TRANSLATE ESTADÍSTICO:** En lo que respecta a Google Translate como estadístico, de las 16 oraciones que se han evaluado, únicamente una se ha obtenido igual que la traducción humana. La oración es *Ich fand ihn schlafend*, y tal como se aprecia, no se presenta con ninguno de los adverbios mencionados en el apartado 3, sino que aparece en Partizip I cuando el gerundio cumple la función de complemento predicativo de un nombre y refiriéndose a una acción simultánea expresada por el verbo conjugado (Castell, 2008: 175).
- **RESULTADOS DE GOOGLE TRANSLATE NEURONAL:** En lo que concierne a este motor, en comparación con los resultados obtenidos de Google Translate estadístico, se observa que las traducciones son mejores. En este caso, dos casos que se evalúan han recibido una puntuación de 100. Es interesante el caso de *Es regnet immer noch*, por un lado, e *Ich lese weiter*, por otro. Ambos poseen un adverbio que hace que no se pierda el sentido perifrástico de la oración en español. Con todo, en ninguno de los casos se ha observado el uso del adverbio *gerade*.
- **RESULTADOS DE DEEPL:** En lo que concierne a este motor, tal como se aprecia en la tabla, es el motor de TA que mejor valoración ha obtenido. Esto se debe a que 3 de las 16 frases están bien traducidas. Asimismo, cabe destacar que es el único motor que ha empleado el adverbio *gerade* en la traducción de la oración 'Estoy comiendo'. Para la traducción de la frase 'El tiempo va mejorando' se observa que el motor DeepL no ha hecho distinción entre *Wetter* y *Zeit*, por lo que sería un error de traducción.
- **RESULTADOS DE MICROSOFT TRANSLATOR:** Si se observa detenidamente la evaluación de este motor y, además, se comparan las traducciones que se han evaluado, puede observarse que algunas oraciones han sido penalizadas por el orden estructural de sus componentes. En este caso, se hace referencia al mismo ejemplo que al motor Lucy Software *Das mache ich schon seit Jahren* vs. *Ich mache das schon seit Jahren*.

Asimismo, la oración ‘Sigo leyendo’ y ‘Sigue lloviendo’, se recurre al correspondiente adverbio para expresar el sentido perifrástico, es decir, *Ich lese weiter* y *Es regnet immer noch*.

Con todo, un fallo que se observa en uno de los ejemplos es la traducción de la palabra ‘Tiempo’, ya que no ha sabido diferenciar entre *Wetter* y *Zeit*. La oración de origen es ‘El tiempo va mejorando’ y, claramente, se percibe que se refiere al tiempo meteorológico.

Gráfico 2. Gráfico con los resultados de las perífrasis verbales de gerundio



1.3 RESULTADOS DE LAS PERÍFRASIS VERBALES DE PARTICIPIO

Antes de iniciar el análisis de la evaluación de la TA, tal como se ha realizado en los anteriores apartados, a continuación, se detalla la métrica con la que se determinará si se aconseja el uso de la TA o si por el contrario, su uso no es recomendable.

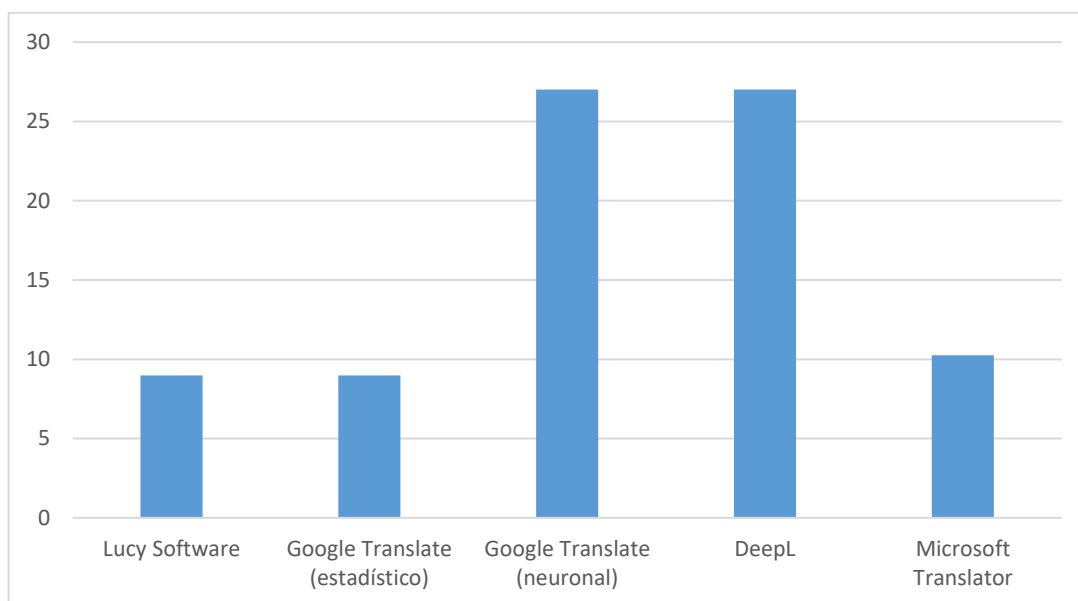
Tabla 11. Métrica de BLEU para las perífrasis verbales de participio

LUCY	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT
8.98	8.98	27.02	27.02	10.26

- **RESULTADOS DE LUCY SOFTWARE:** Este motor es el que peores resultados ha generado. Por tanto, podría concluirse que este motor no se recomienda para traducir perífrasis verbales.

- **RESULTADOS DE GOOGLE TRANSLATE ESTADÍSTICO:** En lo que respecta a este motor, ninguna de las traducciones puede darse como válida debido a que presentan varios errores de traducción debido a las diferencias de estructura existentes entre el alemán y el español.
- **RESULTADOS DE GOOGLE TRANSLATE NEURONAL:** Por lo que respecta a este motor, en comparación con Google Translate estadístico, los resultados que ofrece son mejores. Las tres traducciones que presenta son válidas como perífrasis verbal de participio y, únicamente en la oración *Mads Nielsen wird seit einer Woche vermisst* se observa la diferencia de que se añade el verbo *werden* como auxiliar del verbo *vermissen*. Si se realiza la búsqueda a la inversa, es decir, del alemán al español, se observa que la traducción no contiene el sentido perifrástico de la oración, sino que se traduce como 'Mads Nielsen ha estado desaparecido durante una semana'. En consecuencia, en lugar de reflejar que ha desaparecido y todavía está desaparecido, la traducción muestra que ha estado desaparecido durante una semana, pero que ahora ya no lo está.
- **RESULTADOS DE DEEPL:** En lo que concierne a este motor, es importante mencionar que las tres traducciones que se evaluaban están traducidas de forma correcta. El hecho de que no se haya obtenido una métrica del 100.00 se debe a que en la traducción de la oración 'Mads Nielsen lleva desaparecido una semana', se ha añadido el verbo auxiliar *werden* al auxiliado *verschwinden*. Este hecho no perjudica en la traducción, ya que, si se realiza la prueba a la inversa, es decir, del alemán al español, la traducción que se obtiene contiene la perífrasis verbal de participio. Por tanto, podría decirse que no hay problemas a la hora de traducir las perífrasis verbales de participio con este motor.
- **RESULTADOS DE MICROSOFT TRANSLATOR:** En lo que respecta a este motor, cabe destacar que dos de las tres perífrasis verbales de participio que se han evaluado, las ha traducido igual que la versión de referencia, es decir, la traducción hecha por un humano. La única oración que no se ha traducido correctamente es *Mads Nielsen wurde für eine Woche vermisst*. En este caso, la traducción que se haría de la oración en alemán sería 'Mads Nielsen ha estado desaparecido una semana', por lo que el sentido perifrástico no se mantiene, es decir, en lugar de expresar que Mads Nielsen lleva una semana desaparecido y en la actualidad sigue desaparecido, lo que se traduce es que ha estado desaparecido una semana y ya ha aparecido.

Gráfico 3. Gráfico con los resultados de las perífrasis verbales de participio



1.4 RESULTADOS DE LAS PERÍFRASIS VERBALES MODALES

Antes de iniciar el análisis de la evaluación de la TA, tal como se ha realizado en los anteriores apartados, a continuación, se detalla la métrica con la que se determinará si se aconseja o desaconseja el uso de la TA.

Tabla 12. Métrica de BLEU para las perífrasis verbales modales

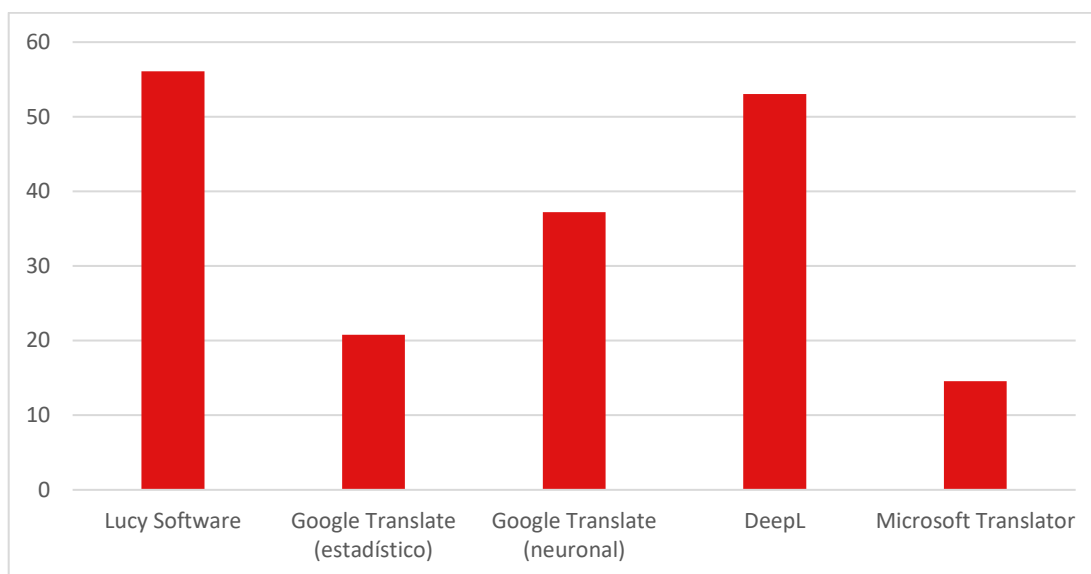
LUCY	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT
56.09	20.79	37.21	53.03	14.57

- **RESULTADOS DE LUCY SOFTWARE:** En el caso de estas perífrasis, se observa que Lucy es de los cinco motores evaluados el que mejores resultados presenta. En total, 4 de las 6 perífrasis que se evaluaban se han traducido de forma correcta. En el caso de la oración ‘Tengo que volver a llamar’, la penalización que ha sufrido para no obtener el 100.00 se debe a que en la traducción de referencia se emplea el verbo *anrufen* y en la traducción automática se opta por el verbo *rufen*. Ambos verbos significan ‘llamar’, la única diferencia entre ellos es que uno es transitivo y el otro intransitivo. Por ello, dicha traducción se ha considerado como válida.

- **RESULTADOS DE GOOGLE TRANSLATE ESTADÍSTICO:** En lo que respecta a los resultados de la evaluación de este motor, cabe decir que ninguna de las perífrasis se ha traducido de forma correcta, lo que significa que no reconoce bien la diferencia entre los verbos modales y, en consecuencia, es el que peores resultados ofrece en relación con las perífrasis verbales modales de todos los motores evaluados.
- **RESULTADOS DE GOOGLE TRANSLATE NEURONAL:** En relación con los datos obtenidos, puede decirse que en este caso ha habido una evolución de pasar de estadístico a neuronal. Con todo, únicamente dos frases se han traducido como las de referencia.
- **RESULTADOS DE DEEPL:** Después de Lucy Software, DeepL es el que ha obtenido unos mejores resultados en relación con la evaluación de las traducciones evaluadas. Dos de las seis traducciones evaluadas han obtenido un 100.00, y el resto está entre el 30 y el 50. En el caso de la traducción de ‘Tengo que volver a llamar’, la traducción se ha realizado con el verbo *zrückrufen*, que significa ‘volver a llamar’. Esta traducción se podría dar por válida a menos que se tomase como ejemplo en alemán y se quisiera traducir al español, ya que, si se realiza la prueba, la traducción que genera es ‘Tengo que llamarte luego’, por lo que se pierde el significado de la perífrasis verbal.
- **RESULTADOS DE MICROSOFT TRANSLATOR:** En lo que respecta a la evaluación de este motor, cabe destacar que, de los cinco motores evaluados, es el que peor puntuación ha obtenido. En el siguiente ejemplo, la TA se halla con el problema de no reconocer la persona que realiza la acción, es decir, mientras en la oración de origen ‘No debes comer chocolate si quieres adelgazar’, la acción recae sobre la 2ª persona del singular, la TA realiza la traducción con la conjugación de la 3ª persona del plural, que se emplea también para dirigirse a cualquier persona de usted⁵⁴.

⁵⁴ En la lengua alemana es más frecuente tratar a alguien de usted que en español. Sin embargo, en caso de la TA, esta no debería tener problemas de traducción de este tipo, ya que no existe ningún tipo de confusión entre las conjugaciones de las personas que realizan la acción del verbo.

Gráfico 4. Gráfico con los resultados de las perífrasis verbales modales



1.5 SÍNTESIS GENERAL DE LOS DATOS OBTENIDOS

A modo resumen, a continuación, se detallan las peculiaridades de las traducciones que se han observado de los motores de TA evaluados, así como las perífrasis verbales que los cinco motores han sido capaces de reconocer. De esta manera, se determinará en qué punto se hallan las hipótesis de partida.

En primer lugar, para las oraciones de infinitivo se observó que había un mayor uso de los adverbios *wieder* y *gerade* para estas perífrasis que no para otras como, por ejemplo, las de gerundio. A continuación, se detallan las 10 perífrasis que, de alguna manera u otra se recogen en las distintas evaluaciones.

Tabla 13. Perífrasis verbales de infinitivo reconocidas por los motores de traducción automática

PERÍFRASIS VERBAL EN ESPAÑOL	PERÍFRASIS VERBAL EN ALEMÁN
¿Has vuelto a comer?	<i>Hast du wieder gegessen?</i>
Se ha puesto a llover	<i>Es hat angefangen zu regnen</i>
Mañana volvemos a vernos	<i>Morgen sehen wir uns wieder</i>
Ayer volvió a nevar	<i>Gestern schneite es wieder</i>
Te lo voy a explicar	<i>Ich werde es dir erklären</i>

Él ha vuelto fumar	<i>Er hat wieder geraucht</i>
Acabo de comer	<i>Ich habe gerade gegessen</i>
Acabo de ver a Klaus	<i>Ich habe gerade Klaus gesehen</i>
Mañana va a llover	<i>Morgen wird es regnen</i>
Iba a decírselo, pero...	<i>Ich wollte es ihm sagen, aber...</i>
Juan va a venir hoy	<i>Juan kommt heute um 12 Uhr</i>

En segundo lugar, para las perífrasis verbales de gerundio, cabe destacar que, a pesar de en ninguno de los casos se haya dado por válida la traducción, la oración *Mein Vater wird langsam älter*, todos los motores han traducido bien el verbo. Lo único que no se ha traducido correctamente ha sido el adjetivo superlativo *älter*, ya que se ha traducido como forma simple del adjetivo, es decir, *alt*.

Tabla 14. Perífrasis verbales de gerundio reconocidas por los motores de traducción automática

PERÍFRASIS VERBAL EN ESPAÑOL	PERÍFRASIS VERBAL EN ALEMÁN
Sigo leyendo	<i>Ich lese weiter</i>
Le encontré durmiendo	<i>Ich fand ihn schlafend</i>
Sigue lloviendo	<i>Es regnet immer noch</i>
Estoy comiendo	<i>Ich esse gerade</i>

En tercer lugar, para las perífrasis verbales de participio, contando que solo se evaluaban dos oraciones, es interesante mencionar que, en uno de los casos, uno de los motores de TA ofrece la traducción exacta a la que requería. A continuación, se muestra el ejemplo del que se hace mención.

Tabla 15. Perífrasis verbales de participio reconocidas por los motores de traducción automática

PERÍFRASIS VERBAL EN ESPAÑOL	PERÍFRASIS VERBAL EN ALEMÁN
Tengo escritas ya 20 páginas	<i>Ich habe bereits 20 Seiten geschrieben</i>

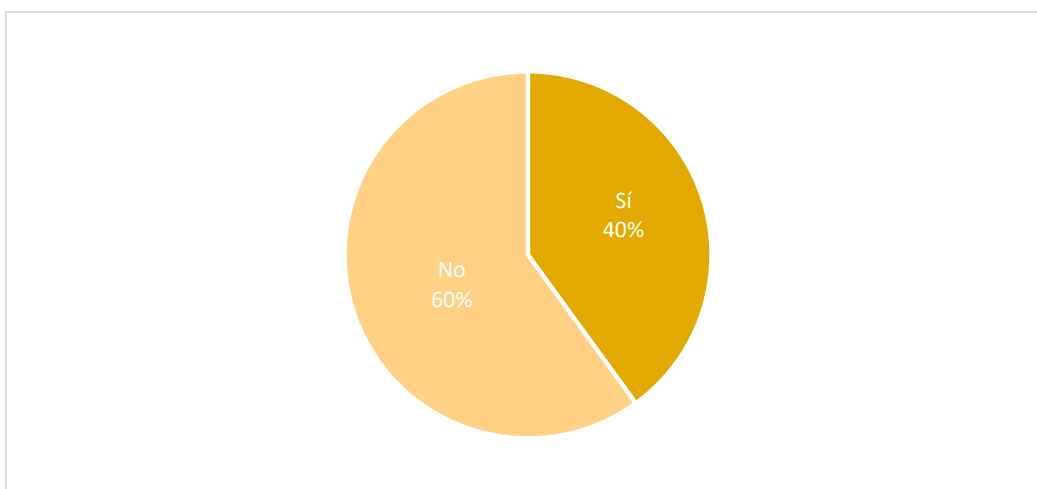
En cuarto y último lugar, para las perífrasis verbales modales, los motores de TA no reconocen entre la modalidad objetiva y la subjetiva (cfr. Capítulo 1.3.5). En muchos casos, la diferencia entre el verbo *müssen* y el verbo *sollen* no se observa.

Tabla 16. Perífrasis verbales modales reconocidas por los motores de traducción automática

PERÍFRASIS VERBAL EN ESPAÑOL	PERÍFRASIS VERBAL EN ALEMÁN
Tienes que comer más	<i>Du musst mehr essen</i>
Tengo que volver a llamar	<i>Ich muss wieder rufen</i>
Él debe venir hoy	<i>Er soll heute kommen</i>
No debes comer chocolate si quieres adelgazar	<i>Du darfst keine Schokolade essen, wenn du abnehmen willst</i>

Finalmente, y como conclusión a este epígrafe, cabe señalar que los motores de traducción automática neuronal no garantizan una traducción de calidad a pesar de que en la actualidad se apueste por ellos. Con todo, cabe destacar los resultados de Google Translate como neuronal, ya que en este caso han superado con creces los resultados que ofrecía Google Translate como estadístico, por lo que puede decirse que ha habido mejoras. Asimismo, a pesar de que sea complicado establecer qué motor de TA ha generado mejores traducciones, si se revisan las métricas de evaluación, se observa que DeepL es el que mejores traducciones ha ofrecido.

Gráfico 5. Porcentaje de las perífrasis verbales correctas



El gráfico muestra que una gran parte de las perífrasis verbales que se han evaluado no se han traducido tal como se espera. Con todo, si se tiene en cuenta el par de idiomas de trabajo, los resultados no son tan desfavorables. Además, a modo resumen de este epígrafe y de los datos obtenidos, cabe añadir que para obtener este porcentaje se han tenido en cuenta todos y cada uno de los resultados, por lo que se recomienda emplear más de un motor de TA y contrastar las traducciones que ofrecen.

2. RESULTADOS DE LA EVALUACIÓN DE LAS EXPRESIONES IDIOMÁTICAS

Antes de profundizar en el análisis de la evaluación de la TA de cada uno de los motores de TA, es conveniente dedicar una parte de este epígrafe a la reflexión personal sobre los resultados obtenidos. De los cinco motores que se evalúan, el que peores traducciones ha ofrecido ha sido Microsoft Translator con un 7.17; seguido de DeepL con un 9.22 y Lucy Software con un 12.63; Google Translate, tanto estadístico como neuronal es el que ha conseguido mejores resultados con un 13.47 y 20.76 respectivamente. Con todo, cabe señalar que el 20.76 pertenece a Google Translate como motor estadístico. Por ello, es interesante constatar que, por lo que respecta a las expresiones idiomáticas, dicho motor no ha mejorado al pasar de estadístico a neuronal, sino todo lo contrario, ya que ofrece peores traducciones. A continuación, se detallarán las traducciones junto con las diferencias y similitudes que han presentado los motores de TA.

Tabla 17. Resultados de la evaluación de las expresiones idiomáticas

LUCY	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT
12.63	20.76	13.47	10.50	7.17

- **RESULTADOS DE LUCY SOFTWARE:** En lo que concierne a este motor, de un total de 50 expresiones para evaluar, solo 9 ha reconocido con su significado composicional. De esta cifra, cabe señalar que 6 de ellas poseen la misma estructura oracional para expresar el sentido de la expresión idiomática. Con todo, las 3 expresiones restantes no poseen la misma estructura y, aun así, el motor de TA ha reconocido dichas expresiones. Asimismo, para la expresión 'VON WEGEN' no ha realizado ninguna traducción, sino que ha dejado el texto en la lengua de origen.

**Tabla 18. Expresiones idiomáticas idénticas en alemán y en español
traducidas por Lucy Software**

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE LUCY
<i>Mit dem Feuer spielen</i>	Jugar con fuego
<i>In guten Händen sein</i>	Estar en buenas manos
<i>Die Nerven verlieren</i>	Perder los nervios
<i>Alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene un final
<i>Mit dem linken Fuß austehen</i>	Levantarse con el pie izquierdo
<i>Die Zeit vorschlagen</i>	Matar el tiempo

Sin embargo, a pesar de que 5 de las expresiones obtenidas poseen diferente terminología y estructura sintáctica en ambas lenguas, el motor de traducción ha reconocido la expresión idiomática y la ha traducido correctamente. A continuación, en la siguiente tabla, se detallan dichas expresiones.

**Tabla 19. Expresiones idiomáticas diferentes en alemán y en español
traducidas por Lucy Software**

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE LUCY
<i>In Hülle und Fülle</i>	En abundancia ⁵⁵
<i>Die Nase voll haben</i>	Estar harto
<i>In den Tag hinein leben</i>	Vivir al día

En resumen, partiendo de lo expuesto anteriormente, puede decirse que, en lo que respecta al uso de este motor con el fin de obtener el significado de una determinada expresión idiomática, no se recomienda acudir a él para dicha consulta debido a que la mayoría de las consultas realizadas se traducen de manera literal.

- **RESULTADOS DE GOOGLE TRANSLATE ESTADÍSTICO:** En relación con las traducciones obtenidas por este motor, cabe admitir que los resultados no han sido tan desfavorables como se esperaban. De todos los motores que se han evaluado es el que ha ofrecido un mayor número de traducciones de las expresiones idiomáticas como se esperaban, es decir, su equivalencia de un idioma al otro sin traducir literalmente. En total, el número de expresiones que se han traducido de forma

⁵⁵ La traducción que se esperaba era: 'A PORRILLOS'. Con todo, se ha considerado que la traducción proporcionada refleja el sentido real de la expresión idiomática.

correcta han sido 13; de esta cifra, 5 poseen la misma estructura oracional y 8 poseen o bien una estructura oracional o bien una terminología diferente.

Tabla 20. Expresiones idiomáticas idénticas en alemán y en español traducidas por Google Translate estadístico

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE GOOGLE TRANSLATE
<i>Mit dem Feuer spielen</i>	Jugar con fuego
<i>Alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene un final
<i>Bergen versetzen</i>	Mover montañas
<i>Die Zeit totschiagen</i>	Matar el tiempo
<i>Es ist nicht alles Gold, was glänzt</i>	No es oro todo lo que reluce

Sin embargo, a pesar de que 4 de las expresiones obtenidas poseen diferente terminología y estructura sintáctica en ambas lenguas, el motor de traducción ha reconocido la expresión idiomática y la ha traducido correctamente. A continuación, se detallan dichas expresiones.

Tabla 21. Expresiones idiomáticas diferentes en alemán y en español traducidas por Google Translate estadístico

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE GOOGLE TRANSLATE
<i>Um jeden Preis</i>	A toda costa
<i>In Hülle und Fülle</i>	En abundancia
<i>Einen Kater haben</i>	Tener resaca
<i>Ohne Wenn und Aber</i>	No hay peros
<i>Auf geht's!</i>	¡Vamos!
<i>Himmel und Hölle in Bewegung setzen</i>	Mover cielo y tierra
<i>In den Tag hinein leben</i>	Vivir al día a día
<i>Von früh bis spät</i>	De sol a sol

En resumen, partiendo de lo observado, puede decirse que, en relación con las expresiones idiomáticas, Google Translate como estadístico ofrece mejores traducciones que los otros motores. Con todo, con respecto al uso de este motor con el fin de obtener la traducción de una determinada expresión idiomática, tampoco se aconseja su uso.

- **RESULTADOS DE GOOGLE TRANSLATE NEURONAL:** En relación con las traducciones obtenidas de Google Translate neuronal, cabe mencionar que, del mismo modo que ha ocurrido con Google Translate estadístico, los resultados obtenidos no han sido tan desfavorables como se esperaban. De todos los motores que se han evaluado es el segundo que ha ofrecido un mayor número de traducciones de las expresiones idiomáticas como se esperaban. En total, el número de expresiones que se han traducido de forma correcta han sido 12; de esta cifra, 6 poseen la misma estructura oracional y 6 poseen o bien una estructura oracional o bien una terminología diferente.

Tabla 22. Expresiones idiomáticas idénticas en alemán y en español traducidas por Google Translate neuronal

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE GOOGLE TRANSLATE
<i>Auf dem Spiel stehen</i>	Estar en juego
<i>In guten Händen sein</i>	Estar en buenas manos
<i>Alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene un final
<i>Bergen versetzen</i>	Mover montañas
<i>Mit dem link Fuß aufstehen</i>	Levantarse con el pie izquierdo
<i>Die Zeit totschiagen</i>	Matar el tiempo

Sin embargo, a pesar de que 4 de las expresiones obtenidas poseen diferente terminología y estructura sintáctica en ambas lenguas, el motor de traducción ha reconocido la expresión idiomática y la ha traducido correctamente. A continuación, se detallan dichas expresiones.

Tabla 23. Expresiones idiomáticas diferentes en alemán y en español traducidas por Google Translate neuronal

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE GOOGLE TRANSLATE
<i>Um jeden Preis</i>	A toda costa
<i>In Hülle und Fülle</i>	En abundancia
<i>Einen Kater haben</i>	Tener resaca
<i>Pleite sein</i>	Estar sin dinero
<i>Auf geht's!</i>	¡Vamos!
<i>Von wegen!</i>	De ninguna manera

En resumen, partiendo de lo expuesto, puede decirse que, en relación con las expresiones idiomáticas, Google Translate como estadístico ofrece mejores traducciones que los otros motores. Con todo, con respecto al uso de este motor con el fin de obtener la traducción de una determinada expresión idiomática, tampoco se aconseja su uso.

- **RESULTADOS DE DEEPL:** Por lo que respecta a este motor, cabe mencionar que se esperaban mejores resultados de los que ha generado. De un total de 50 expresiones, únicamente 11 se han traducido tal como se tenían que traducir con el significado que posee como expresión idiomática. De esta última cifra cabe destacar que 7 poseen la misma terminología, es decir, el significado de la expresión y los elementos que la componen son los mismos. A continuación, se muestran las expresiones a las que se está haciendo referencia.

Tabla 24. Expresiones idiomáticas idénticas en alemán y en español traducidas por DeepL

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE DEEPL
<i>Mit dem Feuer spielen</i>	Jugar con fuego
<i>In guten Händen sein</i>	Estar en buenas manos
<i>Die Nerven verlieren</i>	Perder los nervios
<i>Ganz Ohr sein</i>	Ser todo oído
<i>Alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene un final
<i>Die Zeit vorschlagen</i>	Matar el tiempo
<i>Berge versetzen</i>	Mover montañas

Sin embargo, a pesar de que 4 de las expresiones obtenidas poseen diferente terminología y estructura sintáctica en ambas lenguas, el motor de traducción ha reconocido la expresión idiomática y la ha traducido correctamente. A continuación, se detallan dichas expresiones.

Tabla 25. Expresiones idiomáticas diferentes en alemán y en español traducidas por DeepL

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE DEEPL
<i>Um jeden Preis</i>	A toda costa
<i>Einen Kater haben</i>	Tener resaca
<i>Auf geht's!</i>	¡Vamos!
<i>Die Nase voll haben</i>	Estar harto

En resumen, partiendo de lo observado, puede decirse que, en relación con las expresiones idiomáticas, DeepL ofrece traducciones demasiado literales, que no reflejan el sentido real de la expresión, sino que traduce palabra por palabra y ahí es donde se desaconseja el uso de DeepL para realizar este tipo de búsquedas en la combinación lingüística de trabajo.

- **RESULTADOS DE MICROSOFT TRANSLATOR:** En relación con este motor, cabe señalar que de los cinco que se han evaluado, este ha sido el que peores resultados ha ofrecido. De las 50 frases que se han evaluado, únicamente 5 de las traducciones pueden darse como válidas. A continuación, se muestran las expresiones idiomáticas que poseen la misma estructura oracional o terminológica.

Tabla 26. Expresiones idiomáticas idénticas en alemán y en español traducidas por Microsoft Translator

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE MICROSOFT
<i>Die Nerven verlieren</i>	Perder los nervios
<i>Alles hat ein Ende, nur die Wurst hat zwei</i>	Todo tiene un final
<i>Berge versetzen</i>	Mover montañas

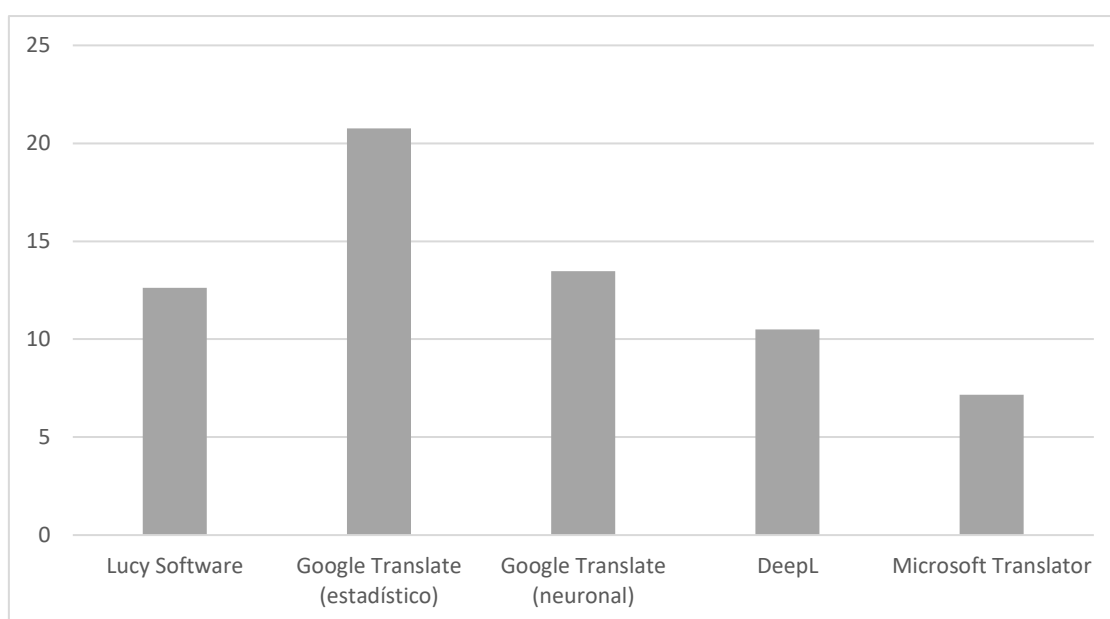
Además de estas tres expresiones idiomáticas, también ha reconocido 2 que poseen otra estructura oracional. A continuación, se muestran con el fin de poder extraer los datos finales de este epígrafe.

Tabla 27. Expresiones idiomáticas diferentes en alemán y en español traducidas por Microsoft Translator

EXPRESIÓN IDIOMÁTICA EN ALEMÁN	TRADUCCIÓN DE DEEPL
<i>Die Nerven verlieren</i>	Perder los nervios
<i>Einen Kater haben</i>	Tener (una) resaca
<i>Auf geht's!</i>	¡Vamos!

En resumen, partiendo de lo expuesto, se puede decir que, para consultar el significado de una determinada expresión idiomática entre el par de lenguas de trabajo, es decir, alemán-español, no se recomienda el uso de Microsoft Translator.

Gráfico 6. Gráfico con los resultados de las expresiones idiomática

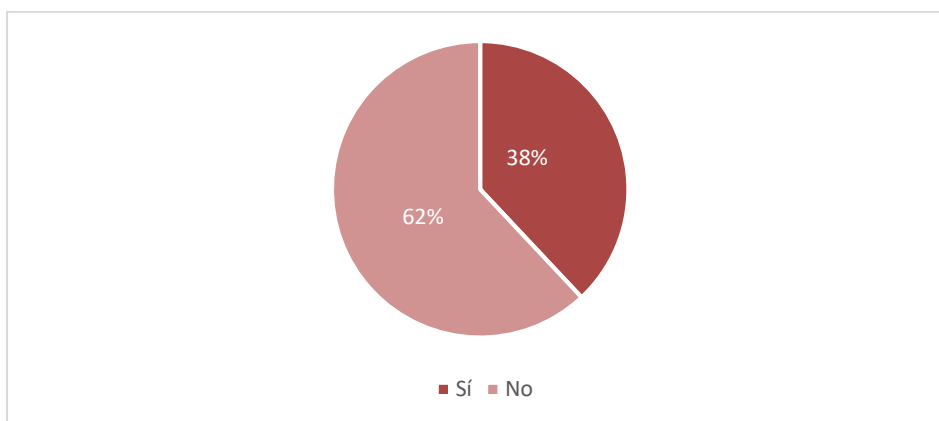


2.1 SÍNTESIS GENERAL DE LOS DATOS OBTENIDOS

A modo de resumen, en el presente epígrafe se sintetizan los datos expuestos en los anteriores apartados con el fin de determinar si las hipótesis de partida se confirman o se refutan. Mediante la siguiente tabla se resumen las expresiones idiomáticas expuestas en los apartados anteriores y que sintetizan los datos hasta ahora expuestos.

En lo que concierne a las expresiones idiomáticas, es interesante constatar que los resultados no son tan desfavorables para la TA como se esperaba en un inicio. Tal como se ha mencionado a lo largo del trabajo, con un total de 50 expresiones idiomáticas a evaluar, solo 19 han sido identificadas correctamente por los motores de TA.

Gráfico 7. Porcentaje de las expresiones idiomáticas correctas



En el gráfico anterior se observa el porcentaje de expresiones idiomáticas que se han obtenido como correctas, es decir, que reflejan el sentido composicional de las diferentes expresiones seleccionadas para evaluar la TA. Para presentar este dato, se han recopilado todas las expresiones obtenidas de los diferentes motores —recuérdese que las traducciones válidas de cada uno de los motores se recogen de forma independiente en los apartados anteriores—. A pesar de que los datos presentados puedan parecer contraproducentes para determinar si es o no es aconsejable el uso de la TA, es importante mencionar que, en este caso, se refuta una de las hipótesis de partida, ya que en ella se decía que los motores de TA no son capaces de reconocer estructuras como las presentadas en los anteriores apartados y, tal como se ha mostrado, en algunos casos, sí son capaces de reconocerlas.

Por último, y en conclusión a este epígrafe, cabe señalar que, a pesar de que únicamente 19 de las 50 expresiones que se han evaluado han sido reconocidas por los motores de TA, tal como se mencionó al inicio del presente trabajo, es importante tener presente la combinación idiomática, ya que los motores de TA alemán-español no están tan entrenados como los que poseen el inglés como lengua de origen o de destino, por lo que en numerosas traducciones, los resultados han sido favorables.

CAPÍTULO IV

CONCLUSIONES

A lo largo del presente trabajo se han expuesto los conceptos teóricos relativos a la traducción automática y los diferentes motores que se han evaluado. El hecho de que se escogiera la combinación idiomática alemán-español parte de la observación de las grandes diferencias entre ambas lenguas, sobre todo en lo que respecta a la gramática y las estructuras oracionales. Por ello y por los problemas que posee la traducción automática hoy en día, se decidió dar respuesta a las hipótesis.

Así pues, la principal hipótesis de partida de este trabajo se confirma, ya que, tal como se ha observado en la evaluación de las perífrasis verbales y las expresiones idiomáticas, las traducciones que los motores de traducción automática han generado no pueden catalogarse de calidad, sino que la mayoría de los resultados presentan dificultades para traducir de una lengua a otra. Asimismo, no hay un motor que predomine sobre otro, sino que el conjunto de todos hace que las traducciones obtenidas puedan compararse y así se ha llegado a la conclusión de que no hay un sistema mejor que el otro, sino que cada uno proporciona las traducciones de la manera que ha sido entrenado. Lo recomendable en estos casos es emplear diferentes motores de traducción automática para establecer, bajo criterio personal, qué traducción es mejor y, en caso necesario, hacer su posesión. Con respecto a DeepL, cabe añadir que ha ofrecido mejores resultados para las perífrasis verbales. Asimismo, a grandes rasgos, es el que mejores resultados ha ofrecido en comparación con otros. En relación con los resultados obtenidos, en la mayoría de los casos, los motores de TA confunden los verbos que varían en función del contexto.

Por todo lo mencionado anteriormente, las principales conclusiones sobre la evaluación de la traducción automática en relación con el par de lenguas de trabajo a las que se llegan son: en primer lugar, se confirman las hipótesis de partida: la mayoría de los motores de TA no reconocen las diferencias estructurales entre las perífrasis verbales y las expresiones idiomáticas, DeepL es el motor de TA que ofrece traducciones más cercanas a las de referencia, es decir, la humana y, finalmente, la mayoría de los motores ofrece y reconocen las perífrasis verbales modales. En segundo lugar, se refuta la idea de

que el par de lenguas alemán-español no genera traducciones de calidad, ya que, tal como se ha ido observando a lo largo de las explicaciones, algunas traducciones eran válidas y se pueden tener en consideración. Lo único que se espera para que se adapte a la calidad del mercado es que la mano del humano intervenga para garantizar la calidad adaptada a las finalidades del texto. En cada uno de los resultados obtenidos mediante la evaluación BLEU se ha observado que alguno de los diferentes motores proporcionaba traducciones mejores que los otros, es decir, en el caso de las perífrasis verbales modales, Lucy Software ofrecía los mejores resultados, para las perífrasis de infinitivo, Google Translate (neuronal) y, para las perífrasis verbales de infinitivo, DeepL es el que mejores resultados ofrece.

Asimismo, cabe señalar que, para el caso de las expresiones idiomáticas, muchas de las seleccionadas son parecidas entre el par de lenguas seleccionado. Un ejemplo de ello es 'Perder los nervios' y *Die Nerven verlieren*. Por último, en relación con la posible evolución de Google Translate de estadístico a neuronal, a través de los resultados ofrecidos, es importante remarcar que este motor no ofrece mejores como neuronal, ya que, en el caso de las expresiones idiomáticas, como estadístico proporciona mejores traducciones que como neuronal.

En cuanto a posibles líneas de investigación, es imprescindible tener presente que en la actualidad el campo de la traducción automática presenta múltiples vías de investigación debido a la falta de trabajos que determinen si algún día la traducción automática podrá garantizar una traducción de calidad que pueda parecerse a la traducción humana. En el caso del par de lenguas alemán-español, existen numerosas posibilidades con las que investigar la traducción automática; un ejemplo de ello sería observar si los diferentes motores que han sido objeto de estudio en este trabajo están entrenados con terminología deportiva. En caso contrario, un posible de tesis doctoral podría ser este junto a la posibilidad de entregar un motor de traducción automática con terminología relacionada con los deportes entre la combinación idiomática alemán-español. Asimismo, podría evaluarse la posible incursión de los motores de TA en la enseñanza-aprendizaje de lenguas extranjeras.

En definitiva, para concluir este capítulo y, por consiguiente, el presente trabajo, es importante tener presente que la traducción automática no ha llegado a ser lo que algunos teóricos esperaban, es decir, no existe un *fully automatic high quality translation* y la profesión del traductor todavía está vigente para encargarse de la posesión de las

traducciones que ofrecen los motores automáticos. Por tanto, hoy en día, falta mucha investigación para determinar si algún día los traductores automáticos suplantarán al traductor humano.

Du bist dran – Es tu turno

CAPÍTULO V

BIBLIOGRAFÍA

- ALARCOS LLORACH, E. (1994). *Gramática de la lengua española*. Madrid: Gredos.
- ALPAC (1996). *Languages and Machines: computers in translation and linguistics*. Washington: D.C. National Academy of Sciences. National Research Council.
- ARNOLD, D. (2003). "Why translation is difficult for computers". E: Somers, H. *Computers and Translation: A translator's guide*. Amsterdam, John Benjamins, pp. 119-142.
- AWWAD M. "Equivalence and translatability of English an Arabic idioms", *Papers and Studies in Contrastive Linguistics* 26 (1990), 57-67.
- BENGIO, Y.; DUCHARME, R.; VINCENT, P.; JANVIN, C. (2003). «A neuronal probabilistic Language model». En: *Journal of Machine Learning Research*, vol. 3, pp. 1137-1155 [PDF].
- BERNARD, V. (2017). En Wikipedia. Recuperado el 20 de mayo de 2018. https://fr.wikipedia.org/wiki/Bernard_Vauquois [Consulta realizada el 22 de mayo de 2018]
- BERNER, S. (2003). *Lost in translation: Cross-Lingual communication, and virtual academic communities* [PDF].
- BO ODAR, B.; FRIESE, J. (Productores).(2017). *Dark* [serie de televisión]. Lugar: Alemania. Netflix.
- BOWKER, L. (2002). *Computer-Aided Translation Technology*. Ottawa: University of Ottawa Press.
- CASACUBERTA, F.; PERIS, A. (2017). *Traducción automática neuronal*, Revista Tradumàtica, núm. 15.
- CASTELL, A. (2008). *Gramática de la llengua alemana. Explicaciones y ejemplos*. Barcelona: Editorial Idiomas, Hueber.
- CLIMENT, S.; MORÉ, J.; OLIVER, A. (2008). *Traducción y tecnologías*. editorial UOC: Barcelona.
- CORCOLL, B.; CORCOLL, R. (2010). *Gramática Programm. Alemán para hispanohablantes*. Barcelona: Herder Editorial.
- CORPAS G. (2003). *Diez años de investigación en fraseología: Análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid Iberoamericana.
- CRUZ PIÑOL, M. Y SITMAN, R. (1999). «who needs to learn spanish anyway? el profesor de E/LE frente a la traducción automática». ASELE. ACTAS X. Centro virtual cervantes.

- CRYSTAL, D. (2000). *Diccionario de lingüística y fonética*. Traducción y adaptación de Xavier Villalba. Barcelona: Octaedro.
- ELENA, P. (1993) “La modalidad objetiva y subjetiva de los verbos alemanes y sus correspondencias en español”. *Revista de filología alemana*, 1, pp. 131-143. [en línea] Consulta realizada el 18 de abril de 2018].
<<http://dialnet.unirioja.es/servlet/articulo?codigo=838994>>).
- ELMAN, J. L. (1990). “Finding structure in time”. *Cognitive Science*, v. 14, n. 2, p. 179-211. <<http://psych.colorado.edu/~kimlab/Elman1990.pdf>>. [Consulta realizada el 18 de marzo de 2018].
- FERNÁNDEZ DE CASTRO, F. (1990). *Las perífrasis verbales en español*. Oviedo: Departamento de Filología Española.
- FERNÁNDEZ DE CASTRO, F. (1999). *Las perífrasis verbales en español actual*. Madrid: Gredos.
- FERNÁNDEZ, S. (1996): “Errores de desarrollo y errores fosilizables en el aprendizaje del E/LE. Tratamiento didáctico”. En: M. Rueda et al. (eds.), *Actas del VI Congreso Internacional de ASELE (León 5-7 de octubre de 1995): Actuales tendencias en la enseñanza del español como lengua extranjera II* (págs. 147-154). León: Universidad de León.
- GARCÍA, V. (1984). «La interferencia lingüística», en *Teoría y práctica de la traducción*, 2ª edición. Biblioteca Románica Hispánica, III. Madrid, Gredos.
- HURTADO, A. (2014). *Traducción y Traductología. Introducción a la traductología*. 7.ª edición. Madrid: Ediciones Cátedra.
- HUTCHINS, W. J.; SOMERS H. L. (1995). *Introducción a la traducción automática*. Madrid: Visor distribuciones.
- HUTCHINS, J. (2005). «The history of machine translation in a nutshell». *Hutchinsweb*. <<http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>> (Consulta realizada el 09 de marzo de 2018).
- JORDAN, M.I. (1990). “Attractor Dynamics and Parallelism in a Connectionist Sequential Machine”. EN: Diederich, J. (ed.). *Artificial neural networks*. IEEE Press, p. 112-127.
- KOEHN, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- LAVANCHY, C. (2010). «Statistische Maschinelle Übersetzung». [PDF]. <<http://kitt.ifi.uzh.ch/clab/smt/smt.pdf>> (Consulta realizada el 16 de marzo de 2018).
- LENGUAJE NATURAL. (2018). En Wikipedia. Recuperado el 5 de mayo de 2018. https://es.wikipedia.org/wiki/Lenguaje_natural (Consulta realizada el 22 de mayo de 2018).

- MELLADO, C. «Convergencia idiomática en alemán y español desde una perspectiva cognitivista en Luque J de Dios y Pamies A (eds.) *La creatividad en el lenguaje: Colocaciones idiomáticas y fraseológicas*. Granada Lingüística, 2005, pp. 73-96.
- MORERA, M. (1991). *Diccionario crítico de las perífrasis verbales del español*. Puerto del Rosario: Cabildo Insular de Fuerteventura.
- NEY, H.; F. J. OCH; S. VOGEL (2000). «Statistical Translation of Spoken Dialogues in the VerbMobil System». En: *Proceedings of Workshop on Multi-lingual Speech Communication*, Kyoto: Japan pp. 69-74.
- OLIVER, A. (2016). *Herramientas tecnológicas para traductores*. Barcelona: Editorial UOC. ISBN: 978-84-9116-048-9.
- RIDRUEJO, E. (1999). "Modo y modalidad. El modo en las subordinadas sustantivas", en I. Bosque y V. Demonte (dirs.), *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, vol. 2, pp. 3209-3251.
- ROBERTS, R. "Phraseology an translation en Fernández Nistal P. y Bravo Gozalo J.M (coords.) J.M. *La traducción: orientaciones lingüísticas y culturales*. Valladolid Servicio de Publicaciones de la Universidad.
- ROCA PONS, J. (1958). *Estudios sobre perífrasis verbales en español*. Madrid: CSIC.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. (1986). «Learning Internat Representations by Error Propagation». En Rumelhart, D. E.; McClelland, James. J.; PDP Research Group, *Parallel distributed processin: Explorations in the microstructure of cognition*. Cambridge, MIT Press. (Computational models of cognition and perception), v. 1, pp. 318-362.
- SCHMIDHOFER, A. (2014). *Alles klar? Dificultades de alemán para hispanohablantes* Barcelona: Pons Idiomas. ISBN: 978-84-8443-292-0
- SOMERS, H. (2003). *Computers and Translation: A translator's guide*. Amsterdam/Philadelphia, John Benjamins Publishing Company.
- SOMERS, H. (2006). «Machine translation». En Baker, M.; Malmkjær, K. (eds.) *Routledge Encyclopedia of Translation Studies*. rep ed. London: Routledge. 136-149.
- STEIN, D. (2009). «Mashinelle Übersetzung – ein Überblick». *JLCL* 24 (3), pp. 5-18.
- VAN GOMPEL, M. (2009). «Phrase-based memory-based machine translation». Tilburg University».
- WARSCHAUER, M. (1996). *Computer-assisted language learning: An introduction*. Multimedia Language teaching, p. 3-20.
- WOTJAK, G. (1984). "Kongruenzen und Divergrenzen im spanischen und deutschen Wortschatz", *Beiträge zur Romanischen Philologie*, XXIII/ H. 1, pp. 109-152.
- ZAPATA, I. (1995). «Un recorrido por algunos principios de Traducción Automática». En: *Enseñanza de lenguas y traducción con ordenadores*. Germán Ruipérez. Madrid: Ediciones pedagógicas, p. 94-97.

ANEXO 1

Tabla 28. Traducciones de las perífrasis verbales de los diferentes motores⁵⁶

LUCY SOFTWARE	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT TRANSLATOR
Hast du wieder gegessen?	Sind Sie schon in Essen zurückgekehrt?	Bist du zurückgekehrt, um zu essen?	Hast du wieder gegessen?	Hast du schon wieder gegessen?
Ich singe	Ich singe	Ich singe	Ich singe	Ich singe
Ich lese weiter	Ich halte das Lesen	Ich lese weiter	Ich lese immer noch	Ich lese weiter
Ich lese die Zeitung	Ich lese die Zeitung	Ich lese die Zeitung	Ich lese die Zeitung / Ich lese gerade die Zeitung	Ich lese das Papier
Wir kochen singend	Koch Singen	Wir kochen singen	Wir kochen mit Gesang	Wir kochen singen
Er schreibt einen Brief	Er schreibt einen Brief	Du schreibst einen Brief	Er schreibt einen Brief / Er schreibt gerade einen Brief	schreibt einen Brief
Ich werde weiter gehen	Ich werde gehen	Ich werde gehen	Ich werde da rüber gehen	Ich werde gehen
Er hat angefangen zu regnen	Es begann zu regnen	Es hat angefangen zu regnen	Es hat angefangen zu regnen	Es regnet
Nachdem er zu Abend gegessen hat, pflegt er spazieren zu gehen	Normalerweise geht für einen Spaziergang nach dem Abendessen	Normalerweise gehen Sie nach dem Abendessen spazieren	Normalerweise geht er nach dem Essen spazieren	In der Regel für einen Spaziergang nach dem Abendessen gehen

⁵⁶ Para observar el listado con las perífrasis verbales con las que se ha evaluado la traducción automática (cfr. Capítulo VI.¡Error! No se encuentra el origen de la referencia.).

Ich fange an zu arbeiten	Ich an der Arbeit	Ich gehe zur Arbeit	Ich gehe an die Arbeit	Ich an die Arbeit
Du musst mehr essen	Sie haben mehr zu essen	Du musst mehr essen	Du musst mehr essen	Du musst mehr essen
Morgen soll es regnen	Morgen sollte regnen	Morgen muss es regnen	Morgen muss es regnen	Es muss morgen regnen
Der Koch bereitet das Essen vor	Der Koch bereitet Essen	Der Koch bereitet Essen zu	Der Koch bereitet das Essen zu	Der Koch bereitet das Essen
Morgen sehen wir uns wieder	Morgen treffen wir uns wieder	Morgen sehen wir uns wieder	Morgen sehen wir uns wieder	Morgen treffen wir uns wieder
Deine schlimmen Noten Notizen kommen mir dieses Jahr sorgend	Ihre schlechte Noten in diesem Jahr sind mir Sorgen	Ihre schlechten Noten in diesem Jahr machen mir Sorgen	Deine schlechten Noten dieses Jahr haben mich beunruhigt	Ihre schlechten Noten in diesem Jahr sind beunruhigend mich
Mads Nielsen hat eine Woche verschwunden	Mads Nielsen verschwunden trägt eine Woche	Mads Nielsen wird seit einer Woche vermisst	Mads Nielsen wird seit einer Woche vermisst	Mads Nielsen wurde für eine Woche vermisst
Gestern schneite es wieder	Gestern schneite es wieder	Gestern hat es wieder geschneit	Der Postbote kommt normalerweise gegen 9:00 Uhr vorbei	Gestern schneite es wieder
Der Postbote pflügt zu dem 9 Uhr zu passieren	Der Postbote geschieht in der Regel bei etwa 09.00 Uhr	Der Briefträger passiert normalerweise gegen 9 Uhr	Gestern hat es wieder geschneit	Der Postbote verbringt in der Regel ca. 9 Stunden
Er endete, weil er den Brief schrieb	Er landete den Brief zu schreiben	Am Ende schrieb er den Brief	Am Ende schrieb er den Brief	Er beendete mit dem Schreiben des Briefes
Ich muss wieder rufen	Ich habe wieder anrufen	Ich muss wieder anrufen	Ich muss zurückrufen	Ich muss zurückrufen
Ich werde es dir erklären	Ich werde Ich werde erklären	Ich werde es dir erklären	Ich werde es Ihnen erklären	Ich erkläre es dir
Er war dabei, sich ihm zu sagen, aber...	Er wollte ihm sagen, aber ...	Ich wollte es ihm sagen, aber ...	Ich wollte es dir sagen, aber...	Ich wollte es ihm sagen, aber...

Der Bus pflegte Verspätung zu bringen haben	Der Bus verwendet Verzögerung zu tragen	Der Bus war früher zu spät	Der Bus kam mal zu spät	Der Bus verwendet, um Verzögerung zu nehmen
Er hat wieder geraucht	Er hat damit begonnen, wieder zu rauchen	Er ist zum Rauchen zurückgekehrt	Er hat wieder geraucht	Er raucht wieder
Dann fing er an zu schreien	Dann begann ich zu schreien	Dann fing er an zu schreien	Dann fing er an zu schreien	Dann er an zu schreien
Juan wird heute um 12 Uhr kommen	Juan wird kommen heute um 12 Uhr	Juan kommt / kommt heute um 12 Uhr	Juan kommt heute um 12:00 Uhr	Juan kommt heute um 12 Uhr.
Ich habe gerade gegessen	Ich aß nur	Ich habe gerade gegessen	Ich habe gerade gegessen	Ich gerade
Ich esse	Ich esse	Ich esse	Ich esse / Ich esse gerade	Ich esse
Ich habe gerade Klausgesehen	Ich habe gerade gesehen, Klaus	Ich habe gerade Klaus gesehen	Ich habe gerade Klaus gesehen	Ich habe gerade Klaus gesehen.
Morgen wird es regnen	Morgen wird es regnen	Morgen wird es regnen	Morgen wird es regnen	Morgen wird es regnen
Es regnet weiter	noch es regnet	Es regnet immer noch	Es regnet immer noch	In der Regel für einen Spaziergang nach dem Abendessen gehen
Mein Vater werdend alt	Mein Vater wird alt	Mein Vater wird alt	Mein Vater wird alt	Mein Vater wird alt
Dies räche ich, seit Jahren so machend	Das habe ich getan, so seit Jahren	Das mache ich seit Jahren	Ich mache das schon seit Jahren	Ich mache das schon seit Jahren
Wir machen die Koffer	Wir packen	Wir packen	Wir packen unsere Sachen	Wir verpacken
Komm spät nicht wieder an!	Nicht immer zu spät!	Komm nicht wieder zu spät!	Komm nicht wieder zu spät!	Kommen Sie nie wieder zu spät!
Um 6 pflegt er sich zu erheben	Normalerweise stehe um 6	Normalerweise steht um 6 auf	Er steht normalerweise um 6:00 Uhr morgens auf	Normalerweise steht er bei 6

Ich erreiche nicht zu verstehen, was du sagst behauptest	Ich kann nicht verstehen, was Sie sagen,	Ich verstehe nicht, was du sagst	Ich verstehe nicht ganz, was Sie sagen.	Ich verstehe nicht, was du sagst.
María begann zu weinen	Mary begann zu trauern	Maria brach in Tränen aus	Maria begann zu weinen	Mary brach in Tränen aus
Wir werden sehen, was er durch da passiert	Mal sehen, was dort geschieht aus	Mal sehen, was da passiert	Mal sehen, was da drüben los ist	Mal sehen, was da draußen passiert
Er kommt dazu, pro Monat 1800€ zu gewinnen	Es kommt 1800 € pro Monat gewinnen	Es kommt zu gewinnen 1800 € pro Monat	Kommt zu verdienen 1800€ pro Monat	Kommt zu verdienen €1800 Monat
Die Zeit Wetter verbessernd	Das Wetter bessert sich	Das Wetter verbessert sich	Das Wetter wird immer besser	Die Zeit wird verbessert
Ich traf ihn schlafend	Ich fand ihn schlafend	Ich habe ihn schlafend gefunden	Ich fand ihn schlafend	Ich fand ihn schlafen
Er kehrte zurück singend	Er kehrte zu singen	Er kam zurück und sang	Er kam zurück und sang	Er kam zurück singen
Ich habe schriftlich schon 20 Seiten	Ich habe bereits 20 Seiten geschrieben	Ich habe bereits 20 Seiten geschrieben	Ich habe bereits 20 Seiten geschrieben	Ich habe bereits 20 Seiten geschrieben
Er fing an, als Kellner zu arbeiten	Er begann als Kellner	Er begann als Kellner zu arbeiten	Er fing an, als Kellner zu arbeiten	Er begann als Kellner zu arbeiten
Er soll heute kommen	Er muss kommen heute	Er muss heute kommen	Er muss heute kommen	Er muss heute kommen
Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris	Morgen habe ich nach Paris zu gehen
Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris gehen	Morgen muss ich nach Paris	Morgen habe ich nach Paris zu gehen
Ich muss weggehen, sondern ich werde den Zug verlieren	Ich muss gehen, aber den Zug verpassen	Ich muss es, sonst werde ich den Zug verpassen	Ich muss, sonst verpasse ich den Zug	Ich werde den Zug vermissen.

Du darfst keine Schokolade essen, wenn du abnehmen willst	Sie sollten nicht Schokolade essen, wenn Sie Gewicht verlieren wollen	Sie sollten keine Schokolade essen, wenn Sie abnehmen wollen	Sie müssen keine Schokolade essen, wenn Sie abnehmen wollen	Sie dürfen nicht essen Schokolade, wenn Sie Gewicht verlieren wollen
---	---	--	---	--

ANEXO 2

Tabla 29. Traducciones de las expresiones idiomáticas de los diferentes motores⁵⁷

LUCY SOFTWARE	GOOGLE TRANSLATE (ESTADÍSTICO)	GOOGLE TRANSLATE (NEURONAL)	DEEPL	MICROSOFT TRANSLATOR
jugar con fuego	jugar con el (los) fuego	jugar con (el) fuego	jugar con (el) fuego	Con (el) juego de fuego
darle un golpe a su Herzen	entregar su corazón un choque	darle un empujón a su corazón	golpear el corazón	Dar a sus corazones un empujón
a todo precio	a toda costa	a toda costa	a cualquier precio / a toda costa	A cualquier costo
estar en el juego	están en juego	estar en juego	estar en juego	En juego
un horno caliente	un horno caliente	Una estufa caliente	un horno caliente	Un horno caliente
ser un pavo	comer pavo	ser un alhelí	ser una flor de pared	Ser un floret de la pared
ser un ratón gris	ser un ratón gris	ser un ratón gris	ser un ratón gris	Ser un ratón gris
estar en manos buenas	en buenas manos	estar en buenas manos	estar en buenas manos	Estar en buen tirador
ser un cabrito	ser una perra	ser una perra	ser una zorra	Ser un zig
alguien delante de la cabeza stroßen	Stopes a alguien en la cabeza	delante de la cabeza	ofender a alguien	Alguien delante de la cabeza Stroßen
ser un cero	ser un cero	ser cero	ser un cero	Ser un cero
pantalón muerto	pantalones muertos	bolsa tubular	pantalones muertos	Pantalones muertos
perder los nervios	perder su nervio	pierde tu nervio	perder los nervios	Perder los nervios
extender tensar a alguien en la tortura	La tensión alguien en suspenso	alguien en la tortura	poner a alguien en el potro de tortura	Alguien en el estiramiento de la tortura
tener mucho sobre las orejas	tienen mucho en su	tener mucho alrededor de las orejas	tener muchas cosas en el plato / tener mucho que hacer	Mucho que tener los oídos

⁵⁷ Para observar el listado con las expresiones idiomáticas con las que se ha evaluado la traducción automática (cfr. Capítulo VI. Anexo 2).

allí o allí	o no	o no	aquí o allá	Atrás o adelante
Pi alguna vez pulgar	La regla de oro	Bad Pi	aproximadamente	PI veces pulgar
en abundancia	en abundancia	en abundancia	en abundancia	Galore
por un pelo / por los pelos ancho amplio	/ Espesor de un cabello hairbreadth	por el ancho de un cabello / por el ancho de un cabello	por la anchura de un pelo	Alrededor de la anchura del pelo/alrededor de un pelo ancho
de A hasta Z	La A a la Z	de la A a la Z	de la A a la Z	De la a a la Z
De hasta delante detrás	De principio a fin	De adelante hacia atrás	De adelante hacia atrás	De frente hacia atrás
en ex	el ex	en ex	en un santiamén / en un abrir y cerrar de ojos	En ex
Ser azul	ser azul	Sé azul	ser azul	Ser azul
tener un gato	tener una resaca	tener una resaca	tener resaca	Tener una resaca
en el vino liegr la verdad	el vino liegr la verdad	La verdad estaba en el vino	en el vino yace la verdad	En el vino Liegr la verdad
ser completamente oreja	todo oídos	estar completamente escuchando	ser todo oídos	Ser todos los oídos
alguien español ocurren	ocurrir a alguien español	alguien español	le parezca español a alguien	Alguien Español
tirar el dinero a la ventana	tirar el dinero por el desagüe	Tira el dinero por la ventana	dejar que el dinero se agote	Arrojando dinero por la ventana
Estar arruinado	se rompió	Estar sin dinero	estar fuera del negocio	Se rompió
sin si y Aber	no hay peros	sin ifs y peros	de forma incondicional	Sin si y pero
ser estar no de madera	no ser de madera	no debe ser de madera	no ser de madera	No estar hecho de madera
¡en va ello!	vamos!	Vamos!	¡vamos!	¡Vamos!
Rápidamente y listamente	Todo listo	Arreglar y listo	Listo para partir	Arreglado y listo

todo tiene un final, sólo la salchicha tiene dos	todo tiene un final, sólo la salchicha tiene dos	todo tiene un final, solo la salchicha tiene dos	todo tiene un final, sólo la salchicha tiene dos	Todo tiene su fin, sólo la salchicha tiene dos
Trasladar montañas	mover montañas	Mover montañas	mover montañas	Mover montañas
Poner en marcha cielo e infierno	mover cielo y tierra	Pon el cielo y el infierno en movimiento	Poniendo el Cielo y el Infierno en Movimiento	Poniendo el cielo y el infierno en movimiento
geigen la opinión a alguien	jugando la opinión de alguien	dar a alguien la opinión	violín la opinión de alguien	La opinión de alguien violines
levantarse con el pie izquierdo	con el pie izquierdo	levantarse con el pie izquierdo	levantarse en el lado equivocado de la cama	Párate con el pie izquierdo
estar harto	son alimentados hasta	harto	estar harto	Tener la nariz llena
matar el tiempo	matar el tiempo	matar el tiempo	matar el tiempo	Matando el tiempo
ser dejado zumbar a alguien	alguien buches	deja ir a alguien	echar de menos a alguien	Deja que alguien Zumba
ser dejado agitarse a alguien	dejar que la platija	deja que alguien se inquiete	retorcerse con las notas	Deja que alguien inquieto
vivir al día	vivir el día a día	vivir en el día	vive por el momento	Vivir en el día
¡mi discurso!	mi discurso!	mi discurso!	mi discurso!	¡mi discurso!
¡von wegen!	por causa!	de ninguna manera!	mi culo!	Yo no
hasta las muñecas	a la muñeca	a las muñecas	hasta que las muñecas	En las muñecas
de temprano hasta tardío	de sol a sol	de la mañana a la noche	desde la mañana hasta la noche	Desde temprano hasta tarde
no es todo el oro, lo que brilla	no es oro todo lo que reluce	no todo es oro que brilla	todo lo que brilla no es oro	No es todo el oro que brilla
cada apuesta	apuesta	cada apuesta	cada apuesta	Cualquier apuesta
Construir estiércol	teta	Construir basura	estropear las cosas	Construcción de basura

