



UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOLOGÍA

GRADO EN LINGÜÍSTICA Y LENGUAS APLICADAS

TRABAJO DE FIN DE GRADO

Estudio de la eficacia de la Traducción Automática de expresiones ambiguas en español y chino

AUTORA: Weirui Zhang

TUTORA: Ana M. Fernández-Pampillón Cesteros

SEPTIEMBRE 2018



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FILOLOGÍA

DEPARTAMENTO DE FILOLOGÍA ROMÁNICA, FILOLOGÍA ESLAVA Y LINGÜÍSTICA GENERAL
ÁREA DE LINGÜÍSTICA GENERAL

GRADO EN LINGÜÍSTICA Y LENGUAS APLICADAS

TRABAJO DE FIN DE GRADO

DECLARACIÓN DE INTEGRIDAD ACADÉMICA

El abajo firmante, D^a. Weirui Zhang con D.N.I. nº Y2399581M

Declaro que el presente trabajo, titulado Estudio de la eficacia de la Traducción Automática de expresiones ambiguas en español y chino

entregado en el tiempo y forma previstos como Trabajo de Fin de Grado del *Grado en Lingüística y Lenguas Aplicadas* es fruto de mi investigación y trabajo personal, y que en él no constan contenidos ni materiales cuyas fuentes no estén claramente identificadas y citadas en el cuerpo del texto o en la bibliografía.

Entiendo, por tanto, que incurriré en plagio si se diera, entre otras, alguna de las siguientes circunstancias:

- Entregar un trabajo ajeno como si fuera propio.
- Copiar un texto literalmente o parafrasearlo sin citar la fuente consultada.
- Entregar un trabajo copiado, en todo o en parte, de fuentes impresas o electrónicas.

Comprendo que el plagio es una grave ofensa académica que puede tener consecuencias negativas en la calificación de mi TFG.

Y para que conste, firmo esta declaración en Madrid, a 07 de 09 de 2017

Firmado:

Índice

Resumen.....	4
1. Introducción.....	5
2. Objetivo.....	6
3. Metodología.....	6
4. Estado de la cuestión en Traducción Automática.....	7
4.1 Qué es la Traducción Automática.....	7
4.2 Métodos de Traducción Automática.....	7
4.2.1 Traducción Automática basada en reglas.....	8
4.2.2 Traducción Automática basada en ejemplos.....	9
4.2.3 Traducción Automática estadística.....	10
4.2.4 Híbrido de métodos.....	11
4.3 Resumen.....	11
5. Traducción Automática basada en Redes Neuronales.....	12
6. Estudio de la ambigüedad aplicada a la Traducción Automática.....	13
6.1 Estudio de la TA chino-español de producciones ambiguas en chino.....	14
6.1.1 Ambigüedad léxica.....	14
6.1.1.1 Homonimia.....	15
6.1.1.2 Polisemia.....	16
6.1.1.3 Ambigüedad de categoría lexical.....	17
6.1.2 Ambigüedad sintáctica.....	18
6.1.2.1 Diferencia de jerarquía estructural.....	18
6.1.2.2 Diferencia de relación estructural.....	20
6.1.3 Ambigüedad semántica.....	21
6.1.3.1 Indefinición de la relación agente-paciente.....	21
6.1.3.2 Indefinición de la dirección semántica.....	22
6.1.3.3 Relación incierta de la posesión.....	22
6.1.3.4 Referencia incierta del pronombre.....	23
6.1.4 Ambigüedad pragmática.....	23
6.2 Estudio de la Traducción Automática español-chino de producciones ambiguas en	

español.....	24
6.2.1 Ambigüedad léxica.....	24
6.2.1.1 Homonimia.....	24
6.2.1.2 Polisemia.....	25
6.2.2 Ambigüedad sintáctica.....	25
6.2.3 Ambigüedad semántica.....	27
6.2.4 Ambigüedad pragmática.....	28
6.2.4.1 Ambigüedades referenciales.....	28
6.2.4.2 Ambigüedades ilocutivas.....	28
7. Discusión de resultados.....	29
8. Conclusión.....	30
Bibliografía.....	32
Anexo: Tabla con los resultados de las traducciones de la muestra de frases.....	34

Resumen

Hoy en día, el uso frecuente de internet hace que el intercambio de ideas y culturas sobrepase en el límite geográfico, pero las diferencias de idioma se convierten en la mayor barrera. La Traducción Automática (en adelante TA) es una manera más eficaz de resolver rápidamente el problema de la traducción de cantidad de información. A pesar de que muchos sistemas de TA entran en el mercado, su calidad de traducción permanece lejos de la satisfacción. Debido a las características de la lingüística, uno de los principales cuellos de botella de TA se enfrenta a la ambigüedad.

En este trabajo nos vamos a centrar en presentar los métodos principales de TA y resumir los tipos diversos de ambigüedades desde una perspectiva lingüística y estudiarlos para el caso particular de la TA entre las lenguas española y china mediante las dos aplicaciones de TA más utilizadas en el mundo: el traductor de Google y el traductor de Bing. Este estudio, además, se lleva a cabo utilizando los dos idiomas más hablados a nivel mundial: el español y el chino (mandarín). El objetivo final es aportar una síntesis de cómo afecta la ambigüedad al rendimiento de los TA en dos casos paradigmáticos de forma pueda ayudar a mejorar la Traducción Automática (TA) y explorar si están resueltos en los sistemas de TA de Google y Bing.

Palabras claves: Traducción Automática (TA), ambigüedad, procesamiento del lenguaje natural (PLN)

1. Introducción

Hoy en día, para comprender la información escrita en los diferentes idiomas, más personas están empezando a usar la TA y muchas personas lo han estado utilizando a menudo como un dispositivo auxiliar en su trabajo o vida cotidiana. Sin embargo, como la TA, ha estado ofreciendo traducciones masivas, especialmente en modo en línea, a las personas, al mismo tiempo también ha recibido muchas críticas por sus versiones de poca calidad. En este sentido, nos preguntamos ¿cuáles son los problemas de la TA que causan esos resultados de calidad insuficiente?

En vista de la limitación de conocimientos, idiomas, tiempo y número de páginas, este trabajo no puede abarcar todos los aspectos de TA. Por lo tanto, desde la perspectiva de la lingüística, vamos a limitar este estudio a los problemas que presenta la ambigüedad en la TA pero en dos lenguas concretas, el español y el chino, que son las lenguas, actualmente, más habladas en el mundo (WEF¹, 2018). Además, se utilizarán como sistemas de TA de evaluación los traductores en línea de Google y de Microsoft, Bing, por ser, ambos, de acceso libre y los más utilizados en español y chino respectivamente.

Así, se puede comprobar, como ejemplo, cuál es la traducción al chino de la oración, “Juan ve a la mujer con un telescopio”, en los dos traductores. El traductor de Google lo traduce a 胡安用望远镜看见那个女人 (Hú ān yòng wàngyuǎnjìng kànjiàn nàgè nǚrén), que significa que el portador del telescopio es Juan. Por su parte, el traductor de Bing ofrece una traducción diferente: 约翰看见那个女人用望远镜 (Yuēhàn kànjiàn nàgè nǚrén yòng wàngyuǎnjìng), que significa que la portadora del telescopio es la mujer.

La oración presenta una ambigüedad sintáctica que, como vemos, es tratada de modo diferente por los dos traductores, generando una única solución cuando debería ofrecer al usuario las dos posibles traducciones o una sola solución en caso de que disponga de contexto para desambiguar (Sonia Vázquez Pérez, 2009). Para poder entender cómo afecta la ambigüedad a la TA nos planteamos las preguntas más específicas: ¿cuáles son los métodos de TA? ¿cómo funcionan los sistemas de TA en la actualidad? ¿qué tipos de ambigüedades existen entre el idioma español y chino?, ¿cómo restan eficiencia los errores de TA por ambigüedad a los sistemas de TA?

¹ World Economic Forum del día 19 de febrero 2018. Recuperado del sitio web de WEF: <https://es.weforum.org/agenda/2018/02/cuales-son-los-idiommas-mas-hablados-en-el-mundo>

2. Objetivo

Para responder a estas preguntas nos planteamos los objetivos siguientes en este Trabajo de Fin de Grado:

1. Resumir los tipos diversos de ambigüedades desde una perspectiva lingüística y estudiarlos para el caso particular de la TA entre las lenguas española y china.
2. Explorar si están resueltos o resueltos parcialmente los errores por ambigüedad resumidos en el objetivo 1 en el sistema de TA ampliamente utilizados a nivel mundial el traductor de Google, *Google Translate*², y el traductor de *Bing Translate*³, este último se ha utilizado porque en China es el más utilizado al estar vetado el uso de *Google Translate*.

3. Metodología

Para llevar a cabo estos dos objetivos se ha seguido los siguientes pasos:

En primer lugar, para comprender mejor el sistema de TA, se ha realizado una revisión y síntesis general de los principales métodos de TA: traducción automática basada en reglas, en ejemplos y traducción automática estadística.

En segundo lugar, se ha revisado cómo funciona el método de TA de Redes Neuronales Recurrentes, *Recurrent Neural Networks* en inglés (en adelante RNN), porque es el método actual utilizado por el sistema de TA de Google, *Google Translate*.

En tercer lugar, se ha revisado y sintetizado el concepto y tipología de ambigüedad en relación con la TA y los lenguajes; español y chino.

En cuarto lugar, se ha realizado un estudio sobre el tratamiento de las ambigüedades en español y chino por parte de los TA de Google y Bing. El objetivo de este paso es, utilizando un caso específico (TA bilingüe español-chino, chino-español) entender con mayor profundidad la cuestión

² El Traductor de Google (del inglés Google Translate) es un sistema multilingüe de traducción automática, desarrollado y proporcionado por Google, para traducir texto, voz, imágenes o video en tiempo real de un idioma a otro. Ofrece interfaz web, así como interfaces móviles para iOS y Android, y una API, que los desarrolladores pueden utilizar para construir extensiones de navegador, aplicaciones, y otros softwares. Recuperado del sitio web Wikipedia.URL: <https://translate.google.com/?hl=es>

³ Bing Translator (anteriormente Live Search Translator y Windows Live Translator) es un servicio proporcionado por Microsoft como parte de sus servicios Bing que permiten a los usuarios traducir textos o páginas web completas en diferentes idiomas. Todas las traducciones son con tecnología de Microsoft Translator (anteriormente SYSTRAN), desarrollado por Microsoft Research. Dos pares de transliteración (entre chino tradicional y chino simplificado) son proporcionados por Microsoft Windows International team.¹ Textos relacionados con la informática son traducidos por la propia traducción automática estadística basado en sintaxis de la tecnología de Microsoft. Recuperado del sitio web Wikipedia.URL: <https://www.bing.com/translator>

que se plantea en este trabajo sobre la calidad de los TA en el caso de producciones ambiguas. Para ello hemos creado una muestra de frases ambiguas escritas en español y chino recopilada de estudios, documentos y materiales previos⁴. La muestra abarca los distintos tipos de ambigüedad establecidos en el paso tercero, ambigüedad léxica, sintáctica, semántica y pragmática.

Cada ejemplo de la muestra se ha traducido manualmente (por la autora) y automáticamente con el traductor Google (caso español-chino) y traductor Bing (caso chino-español). En el anexo se muestra los resultados de la traducción contrastiva. En último lugar, se han comparado los resultados de la traducción humana con los resultados de TA.

Finalmente, en quinto lugar, con la síntesis hecha en los pasos uno a tres y los resultados del estudio de caso del paso cuarto, se han elaborado las conclusiones del trabajo.

4. Estado de la cuestión en TA

4.1. Qué es la TA

La traducción automática consiste en utilizar ordenadores para automatizar algunos o todos los procesos que intervienen en la traducción de una lengua a otra (Jurafsky y Martin, 2009).

4.2. Métodos de TA

En general, la TA incorpora los métodos del Procesamiento del Lenguaje Natural (PLN, en adelante). Casi todos los resultados de investigación en el PLN se pueden aplicar directa o indirectamente a la TA. Los métodos de TA se pueden dividir en tres categorías: la primera, basada en reglas (Rule-Based Machine Translation, RBMT), la segunda, basada en ejemplos (Example-Based Machine Translation, EBMT) y la tercera, en estadística (Statistical Machine Translation, SMT). Las dos últimas categorías también se pueden englobar en una denominada TA basada en corpus (Corpus-based Machine Translation, CBMT) (Mikov, 2005).

⁴ estudios, documentos y materiales previos: se han sacado desde las bibliografías

Jianming, Lu. *Modern Chinese Grammar Research Course (Gramática china moderna)*. Peking University Press, 2003.

Pérez, Sonia Vázquez. *Resolución de la ambigüedad semántica mediante métodos basados en conocimiento y su aportación a tareas de PLN*, página 3-6. Universidad de Alicante, 2009.

Salvador, Gutierrez Ordoñez. *De pragmática y semántica*. ARCO/LIBROS, S.L., 2002. *Introducción a La Semántica Funcional*. SINTESIS, 1996.

Weijun, Gong. *The research of Modern Chinese Ambiguity and the utilization in Chinese Courses for Foreigners*. University of Shandong, 2010.

Wenmin, Zhao. *Ambiguity in Modern-Chinese*. 2012.

Yanyun, Song. *Modern Chinese Ambiguity Research*. Northwest University China, 2007

Yingxian, Cui. *Modern Chinese Grammar Learning and Research Introduction, Introduction to the Learning and Research of Modern Chinese Grammar*. Tsinghua University Press, 2004.

4.2.1 Traducción Automática basada en reglas⁵

Un sistema de TA basado en reglas consiste en la recopilación de reglas llamadas reglas gramaticales, léxico y programas de software para procesar las reglas. Es extensible y mantenible, es decir, fácil de escalar y fácil de modificar. Las reglas están escritas con el conocimiento lingüístico de los lingüistas. Las reglas juegan un papel importante en las diversas etapas de la traducción: procesamiento sintáctico, interpretación semántica y procesamiento contextual del lenguaje (Arnold, D.J, 1993).

La estructura del árbol se usa para representar la estructura de la oración. Una oración típica en español consta de dos partes principales: sintagma nominal (NP) y sintagma verbal (VP). Estas dos partes se pueden dividir más según la estructura de la oración. Solo la oración con la estructura correcta puede conducir a la traducción correcta. Las siguientes son las reglas para representar una gramática simple.

S -> NP VP

VP -> V NP

NP -> Nombre

NP -> ART N

S significa oración, V para verbo, N para sustantivo y ART para artículo. Una gramática puede derivar una oración si hay una secuencia de reglas para reescribir el símbolo de inicio, S, en una oración.

La forma lógica se usa comúnmente en la interpretación semántica. Por ejemplo, la oración, Carlos estaba feliz, se puede escribir en forma lógica como:

(<PAST FELIZ> (NOMBRE 1 "Carlos"))

PAST representa el tiempo pasado. La interpretación semántica es un proceso de composición en el que las interpretaciones se pueden construir de forma incremental a partir de las interpretaciones de las subfrases. El léxico juega un papel importante en la interpretación semántica. Las reglas de gramática se utilizan para calcular la forma lógica de la oración dada.

⁵ Esta sección se ha sacado desde Wikipedia y la bibliografía: Arnold, D.J. Machine Translation: an Introductory Guide, 1993.

4.2.2 Traducción automática basada en ejemplos⁶

La idea básica de la TA basada en ejemplos (EBMT) es reutilizar ejemplos de traducciones ya existentes como base para una nueva traducción. El proceso de EBMT se divide en tres etapas (Mikov, 2005):

1. Encaje (matching)
2. Alineamiento (alignment)
3. Recombinación (recombination)

La etapa de matching en la TA basada en ejemplos encuentra ejemplos que contribuirán a la traducción basada en su similitud con la entrada. La forma en que se debe implementar la etapa de matching se basa en cómo se almacenan los ejemplos.

La entrada y los ejemplos se pueden match carácter por carácter. Este proceso se llama comparación de secuencia. El alineamiento y la recombinación serán difíciles si se utiliza este enfoque. Los ejemplos pueden ser anotados con etiquetas. Varios ejemplos simples se pueden combinar en un solo ejemplo más general que contiene variables.

El alineamiento se usa para identificar qué partes de la traducción correspondiente se reutilizarán. El alineamiento se hace usando un diccionario bilingüe o comparándolo con otros ejemplos. El proceso de alineamiento en la TA basada en ejemplos debe ser automático.

La recombinación es la fase final en la TA basada en ejemplos. La recombinación asegura que las partes reutilizables identificadas en el ejemplo se conformen de forma legítima. Toma las oraciones del lenguaje fuente y un conjunto de patrones de traducción como entradas y produce oraciones en el idioma de destino como salidas. El diseño de la estrategia de recombinación depende de las fases de matching y alineamiento previas.

Por ejemplo, el ejemplo inglés traduce al japonés (Sato y Nagao, 1990):

- a. He buys a book on international politics.
- b. He buys a notebook. *Kare wa noto o kau.*
- c. I read a book on international politics. *Watashi wa kokusai seiji nitsuite kakareta hon o yomu.*

La etapa de matching identificará varios ejemplos adecuados, cada ejemplo contiene partes del texto

⁶ Esta sección se ha sacado desde la bibliografía: Mitkov, Ruslan. *The Oxford Handbook of Computational Linguistics*. Página 514-516, OUP Oxford, 2005.

a traducir. Si quisiéramos traducir (a) basada en ejemplos recuperados (b, c), deberíamos ser capaces de identificar qué partes de los equivalentes japoneses corresponden al texto subrayado en los ejemplos (b,c).

4.2.3 Traducción automática estadística

La esencia del método de TA basado en estadísticas es primero alinear frases, grupos de palabras y palabras individuales de los textos, y luego calcular las probabilidades de que cualquier palabra en una oración del lenguaje corresponda a una palabra o palabras en la oración traducida con la que está alineado en el otro lenguaje. Una característica esencial es la disponibilidad de un gran corpus bilingüe adecuado de traducción fiable (Mikov, 2005). Este método a menudo se considera "anti-lingüística" (Brown et al., 1990),

Por ejemplo, si tomamos la oración como la entrada A, (Mikov, 2005), entonces las posibles traducciones se encuentran A1 y A2.

A. President Lincoln was a god lawyer.

A1. *Le président Lincoln était un bon avocat.* ('President Lincoln was a god lawyer')

A2. *Le matin je me brosse les dents.* ('In the morning I brush my teeth')

La probabilidad de que A1 sea una buena traducción es muy alta, mientras que la probabilidad de A2 es muy baja. Entonces para cada par de oraciones S y T hay una probabilidad $P(T|S)$, es decir, la probabilidad de que T sea la oración objetivo, dado que S es la fuente. El procedimiento de traducción es una cuestión de encontrar el mejor valor para $P(T|S)$

Este cálculo de probabilidad depende de otras dos medidas. La primera es la probabilidad de que las palabras en S conducen a las palabras en T, se denomina el modelo de traducción. El segundo es la probabilidad de que estas palabras se combinen correctamente en el lenguaje de destino, se denomina el modelo de lenguaje (el lenguaje de destino). El modelo de traducción es el vínculo entre dos lenguajes, que describe la posibilidad de que la cadena de un lenguaje se traduzca a la cadena correspondiente de otro lenguaje, generalmente expresada como una probabilidad condicional. El modelo de lenguaje solo es relacionado con el lenguaje de destino y refleja la razonabilidad de una cadena en el lenguaje de destino, generalmente expresada como un modelo de N-grama del lenguaje de destino. Lo que se llama N-grama⁷, significa que es suficiente para

⁷ En Inteligencia Artificial, procesamiento de lenguaje natural, se llama n-grama a una subsecuencia de n elementos consecutivos en una secuencia dada. Si $n=2$ se denominan bigramas; $n=3$, trigramas. Los n-gramas son de gran utilidad

calcular las probabilidades de secuencias de palabras relativamente cortas. N es el número de palabras en la secuencia. Bigrama son conjuntos de pares de palabras, los trigramas consisten en secuencias de tres palabras.

4.2.4 Híbrido de métodos

Los métodos anteriormente descritos tienen ventajas que, de aunarse en un solo método, podrían ofrecer mejores traducciones. Entre las ventajas que sería deseable tener conjuntamente podemos destacar: de los métodos estadísticos que ofrecen resultados razonablemente buenos de traducción y por su capacidad de aprendizaje no supervisado, etc.; de los métodos basados en reglas es destacable su buena capacidad de generalización y descripción, y, el método basado en ejemplos es adecuado en dominios restringidos, cuando se dispone de una buena base de datos de oraciones similares.

Por ello, el desarrollo de estrategias de TA híbridas que combinen dos o más es una dirección importante en la investigación y desarrollo de aplicaciones de TA. Por ejemplo, los modelos basados en las reglas incorporan estadística para seleccionar las reglas más probables en contextos determinados; Groves y Way (2005), por su parte, proponen los conceptos de TA basada en ejemplos de estadística y TA estadística basada en ejemplos.

Otra aproximación híbrida de TA se basa en la premisa de utilizar una única estrategia de traducción como el cuerpo principal y complementarla con otras estrategias. Diferentes tecnologías se utilizan en diferentes módulos del sistema. Por ejemplo, los sistemas híbridos combinan el análisis basado en reglas y la generación con transferencia basada en ejemplos. Otra combinación parece particularmente adecuada para el espinoso problema de traducción oral, por ejemplo, los elementos de la parte de análisis pueden basarse más en el análisis estadístico, mientras que la transferencia y la generación son más adecuadas un método basado en reglas (Mikov, 2005).

4.3 Resumen

Los métodos de TA, basada en reglas, ejemplos y estadísticos se reconocen actualmente para clasificar los métodos de TA. En general, los métodos estadísticos y los métodos de ejemplos son métodos basados en datos lingüísticos, normalmente corpus. El método de la TA estadístico utiliza un corpus bilingüe en la fase de entrenamiento para obtener un modelo estadístico, y un modelo

en el tratamiento de textos, la determinación del lenguaje de documentos, el procesamiento estadístico de los mismos o el descubrimiento o predicción de genes en tanto estos son subsecuencias dadas dentro del material genético.

estadístico pre-entrenado para la traducción. El método de TA, basado en los ejemplos es esencialmente un método de razonamiento fundamentado en ejemplos, que utiliza ejemplos de la traducción directamente durante la fase de traducción. A diferencia de los métodos basados en los datos, en los métodos tradicionales de TA basada en las reglas, a menudo, estas reglas se describen mediante reglas de escritura manual. Entonces, ¿cuál es la relación entre los tres métodos de TA? Esencialmente, estos tres métodos no son mutuamente excluyentes, sino que se enfocan en diferentes puntos y enfatizan diferentes aspectos del problema desde diferentes perspectivas. El método estadístico enfatiza la capacidad de modelo estadístico desde un punto de vista matemático. El método de ejemplos destaca el grado de abstracción de los ejemplos de traducción desde la perspectiva del aprendizaje máquina, mientras que el método de reglas describe las leyes, cómo formar la composición del lenguaje desde la perspectiva de los fenómenos del lenguaje. Por lo tanto, los modelos estadísticos rigurosos, el gran número de ejemplos de traducción y la composición multinivel pueden ser completamente simbióticos.

5. Traducción Automática basada en Redes Neuronales⁸

Un nuevo método de traducción basado en Redes Neuronales. Este nuevo método de traducción fue presentado por Google en el año 2016. Utiliza un conjunto de algoritmos basados en un tipo de Red Neuronal llamada Redes Neuronales Recurrentes (Recurrent Neural Networks, en adelante RNN) (Bahdanau & Bengio ,2014). Las redes neuronales analizan una gran cantidad de datos del corpus, y, automáticamente, aprenden las reglas gramaticales correspondientes a los ejemplos de los corpus.

Las redes neuronales recurrentes son redes con bucles que permiten que la información que se obtiene como salida de un proceso sirva de realimentación y se combine con la información de entrada (figura 1). Este método utiliza las técnicas más avanzadas disponibles en la actualidad para lograr la mayor mejora en la calidad de la traducción automática hasta la fecha.

⁸ Consultado en sitio web. URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

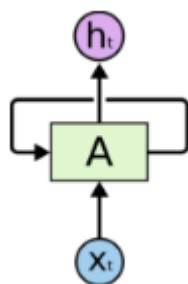


Figura 1. Las redes neuronales recurrentes tienen bucles.

En el diagrama de arriba, un fragmento de la red neuronal A, recibe una entrada x_t y emite un valor h_t . Un bucle permite pasar información desde un paso anterior de la red al siguiente. Estos bucles no son tan diferentes a una red neuronal normal. Una red neuronal recurrente, si se “desarrolla”, se puede considerar como copias múltiples de la misma red, cada una pasando un mensaje a un sucesor como muestra la figura 2:

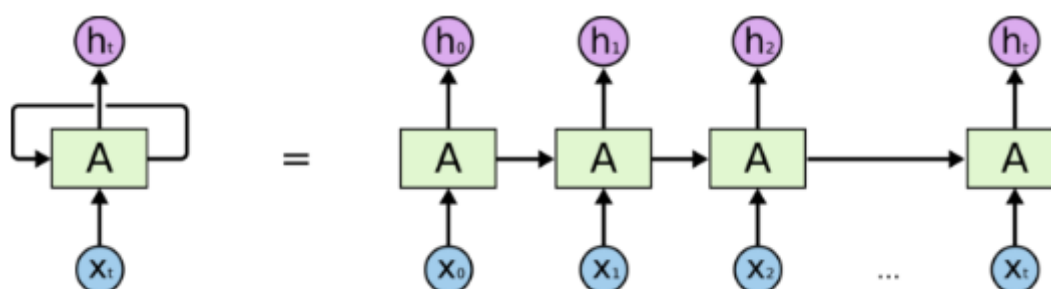


Figura 2. Una red neuronal recurrente desenrollada.

La clave de una red neuronal recurrente reside en la palabra “recurrente”. El sistema “recordará” la última salida y la usará para determinar la siguiente salida. Con los conceptos de “la última vez” y “la próxima vez”, las redes neuronales no tratan la información de entrada y salida como independiente, sino como series de tiempo interrelacionadas. De esta manera, se puede adivinar la palabra que aparecerá en la siguiente secuencia a través de la asociación de secuencia previa.

Al traducir, RNN toma el idioma fuente como la secuencia de entrada y el idioma de destino como la secuencia de salida. Como cada salida se referirá al resultado de la última salida, la TA es holística en vez de ser una traducción de palabras simplemente.

6. Estudio de la ambigüedad aplicada a la TA

Según la definición de Gutiérrez Ordoñez (1996)⁹, “la ambigüedad es una característica notable del

⁹ Salvador, Gutiérrez Ordoñez. *De pragmática y semántica*. ARCO/LIBROS, S.L., *Introducción a La Semántica Funcional*.

lenguaje natural. El receptor se encuentra a veces ante una perplejidad interpretativa: no sabe qué sentido elegir entre varios posibles. Esta situación ha sido designada con el término ambigüedad. En lingüística definimos un mensaje ambiguo como la secuencia de signos a la que, fuera de todo contexto, es posible asignar dos o más interpretaciones.”

Dado que la ambigüedad es un problema complicado en el lenguaje natural, varias personas clasifican las ambigüedades en diferentes tipos según diferentes estándares. Como afirma Gutiérrez (1996), la tipología de ambigüedad en este estudio tendrá cuatro clases que son: la ambigüedad léxica, la ambigüedad sintáctica, la ambigüedad semántica y la ambigüedad pragmática.

En esta sección, a partir de una muestra de ejemplos, en chino y en español, de cada tipo de ambigüedad considerada, léxica, sintáctica, semántica y pragmática, se estudia la calidad de sus TA bilingües español-chino/chino-español. Para ello se comparará el resultado de traducir la muestra de ejemplos un traductor humano, que se toma como la traducción de referencia o traducción correcta. El objetivo es valorar la eficacia de la TA, especificada en las aplicaciones *Google Translator* y *Bing Translator*.

La muestra de ejemplos procede de estudios, documentos y materiales previos¹⁰ y se trata sólo textos escritos, no literarios.

6.1 Estudio de la TA chino-español de producciones ambiguas en chino

En este estudio, la TA chino-español se lleva a cabo con el traductor *Bing Translator*, debido a que en China está vetado el uso de la aplicación *Google Translator*

6.1.1 Ambigüedad léxica

La ambigüedad léxica generalmente se da cuando una palabra tiene dos o más significados. Dividimos las ambigüedades léxicas del chino en tres subtipos: homonimia, polisemia y ambigüedad de categoría léxica.

¹⁰ estudios, documentos y materiales previos: se han sacado desde las bibliografías

Jianming, Lu. *Modern Chinese Grammar Research Textbook (Gramática china moderna)*. Peking University Press, 2003.

Pérez, Sonia Vázquez. *Resolución de la ambigüedad semántica mediante métodos basados en conocimiento y su aportación a tareas de PLN*, página 3-6. Universidad de Alicante, 2009.

Salvador, Gutierrez Ordoñez. *De pragmática y semántica*. ARCO/LIBROS, S.L., 2002. *Introducción a La Semántica Funcional*. SINTESIS, 1996.

Weijun, Gong. *The research of Modern Chinese Ambiguity and the utilization in Chinese Courses for Foreigners*. University of Shandong, 2010.

Wenmin, Zhao. *Ambiguity in Modern-Chinese*. 2012.

Yanyun, Song. *Modern Chinese Ambiguity Research*. Northwest University China, 2007

Yingxian, Cui. *Modern Chinese Grammar Learning and Research Introduction, Introduction to Learning and Research of Modern Chinese Grammar*. Tsinghua University Press, 2004.

6.1.1.1 Homonimia

La homonimia hace referencia a dos palabras que son homófonas u homógrafas, pero que tienen diferentes significados. Este es un fenómeno muy común en chino. Las dos palabras homófonas que tienen el mismo sonido y que no están conectadas o distantes en el significado son fácilmente incomprensibles, pero la ambigüedad causada por los homófonos generalmente solo ocurre en la comunicación oral. Una vez que estas formas orales ambiguas se convierten en formas escritas, la mayoría de ellas desaparecerán, por lo que no las discutiremos. Aquí solo vamos a investigar las palabras homógrafas y homónimas, que son dos palabras iguales que tienen la misma pronunciación, pero con significados diferentes.

(1) 山上到处是杜鹃 [Shānshàng dào chù shì dùjuān]

La montaña llena de rododendros. (Flor) / La montaña llena de cucos. (Ave)

(traducción humana)

La montaña está llena de cucos. (Google)

Las montañas están llenas de rododendros. (Bing)

Esta oración es ambigua. La escritura y la pronunciación de " dùjuān " son las mismas. Se puede entender como " La montaña llena de rododendros.". También se puede entender como " La montaña llena de cucos". Esta ambigüedad se debe al hecho de que el término " dùjuān " es un homónimo, tiene la misma pronunciación y escritura, ya que puede referirse a las flores como rododendros o a las aves como cucos.

Podemos ver en la traducción de Google que " dùjuān " ha sido traducido al "pájaro cuco". En la traducción Bing ha sido traducido al " rododendros".

(2) 注意仪表 [Zhùyì yíbiǎo]

Preste atención a la apariencia. / Preste atención al panel. (traducción humana)

Medidor de atención. (Google)

Cuidado con el taxímetro. (Bing)

La palabra " yíbiǎo " también es una homonimia, porque puede entenderse como "apariencia de una persona" y también como "instrumento de medición que se usa para comparar magnitudes físicas mediante un proceso de medición". Google y Bing consideran " yíbiǎo " como un instrumento de medición.

Existen más ambigüedades causadas por homografía, como:

(3) 他还欠款 2 万 [Tā hái qiàn kuǎn 2 wàn] / [Tā huán qiàn kuǎn 2 wàn]

Él todavía debe 20,000 yuanes / Él devuelve la deuda 20,000 yuanes

(traducción humana)

Él todavía debe 20,000 yuanes. (Google)

Aún debe 20.000. (Bing)

Si "还" pronuncia "hái", significa que aún debe 20,000 yuanes; si "还" pronuncia "huán", ya ha devuelto 20,000 yuanes. Tanto Google como Bing tratan "还" como "hái".

6.1.1.2 Polisemia

La polisemia consiste en que una palabra tiene dos o más significados asociados entre sí. En general, una vez que la polisemia entra en un contexto específico, tiene un único significado. Si todavía hay múltiples significados en un contexto específico, puede ser malentendido y conduciría a la ambigüedad.

(4) 昨天, 老板给这家商店配备了一些机动车辆

[Zuótiān, lǎobǎn gěi zhè jiā shāngdiàn pèibèile yīxiē jīdòng chēliàng]

Ayer, el propietario equipó a esta tienda con unos vehículos motorizados. / Ayer, el propietario equipó a esta tienda con unos vehículos para todo. (traducción humana)

Ayer, el propietario equipó esta tienda con algunos vehículos de motor. (Google)

Ayer, el jefe le dio a la tienda algunos vehículos motorizados. (Bing)

La palabra "jīdòng" puede entenderse como "equipar con una máquina" o "preparado para un uso flexible". Para excluir tales ambigüedades causadas por la polisemia, se deben agregar ciertas condiciones restrictivas para que el significado de estas palabras polisémicas sea único. Y aquí las dos versiones de traducción son diferentes, pero ambas traducen el significado de "jīdòng" a "motorizar".

(5) 他已经走了两个小时了 [Tā yǐjīng zǒule liǎng gè xiǎoshíliǎo]

Él caminó durante dos horas / Se fue hace dos horas. (traducción humana)

Él ha estado caminando durante dos horas. (Google)

Hace dos horas que se fue. (Bing)

La palabra "走" es polisémica, en este caso, "走" puede interpretarse como el verbo "caminar", un movimiento continuo. Es decir, ha estado caminando durante dos horas. El resultado de Google pertenece a esta versión. El verbo "走" también tiene otra interpretación, puede significar "irse".

Es decir, se fue hace dos horas, y el resultado de Bing tiende a esta versión.

6.1.1.3 Ambigüedad de categoría lexical

Palabras con dos o más categorías, a menudo, debido a sus diferentes categorías causan ambigüedades.

(6) 这间教室的门没有锁。 [Zhè jiān jiàoshì de mén méiyǒu suǒ]

La puerta del aula no tiene cerradura. / La puerta del aula no está cerrada. (traducción humana)

La puerta de este aula no está cerrada. (Google)

La puerta de esta habitación está abierta. (Bing)

La palabra "suǒ" se puede usar como sustantivo, lo que significa "mecanismo de metal que se fija en puertas, tapas de cofres, arcas, cajones, etc., y sirve para cerrarlos por medio de uno o más pestillos que se hacen jugar con la llave" (RAE) y también se puede usar como verbo para indicar que "asegurar con cerradura, pasador, pestillo, tranca u otro instrumento, una puerta, ventana, tapa, etc., para impedir que se abra" (RAE). Claramente, ambos traductores automáticos tienden a pensar en ella como un verbo.

(7) 他原来住在这儿 [tā yuánlái zhù zài zhè]

Él vivió aquí antes. / No esperaba que él viviera aquí. (traducción humana)

Originalmente se quedó aquí. (Google)

Solía vivir aquí. (Bing)

La palabra "yuánlái" puede considerarse como un adverbio de tiempo, significa *antes*, en el pasado, y también se puede considerar como un adverbio de tono, significa que el hablante no esperaba que él viviera aquí.

En este ejemplo, el *Google Translator* utiliza el pretérito perfecto simple del indicativo, para expresar el verbo continuo "quedar", que muestra que vivió aquí en el pasado, durante un tiempo, pero que ahora no vive aquí. El *Bing Translator* usa el indicativo, pretérito imperfecto del verbo "soler" para reflejar el "yuánlái", lo que indica el fenómeno habitual del pasado.

En China, los lingüistas consideran que el sistema gramatical está compuesto por tres niveles: nivel sintáctico, semántico y pragmático. Jianming Lu (2003) señaló que cualquier estructura sintáctica que consta de palabras conceptuales siempre coexiste con dos tipos de relaciones estructurales: relaciones de estructura sintáctica y relaciones de estructura semántica.

6.1.2 Ambigüedad sintáctica

La ambigüedad sintáctica o ambigüedad estructural ocurre cuando una oración tiene más de un análisis de constituyentes plausible. Por lo general, la estructura se representa mediante el árbol de análisis sintáctico y, en las oraciones ambiguas sintácticamente ocurre que la estructura tiene dos o más árboles sintácticos posibles. En esta sección, dividiremos esta ambigüedad en dos subclases: diferencia de jerarquía estructural y diferencia de relación estructural (Weijun Gong, 2010).

6.1.2.1 Diferencia de jerarquía estructural

En el análisis sintáctico del chino, podemos especificar la categoría de palabras y el tipo o formato general de palabras, frases y oraciones en el constituyente de una oración a través, de los elementos sintácticos en la estructura sintáctica. En este proceso, habrá varias jerarquías estructurales. A veces la misma oración puede analizar dos o más niveles estructurales.

(8) 关心孩子的母亲 [Guānxīn háizi de mǔqīn]

la madre que cuida a los niños / cuida a la madre de los niños. (traducción humana)

Cuidando a la madre del niño. (Google)

Cuidado de la madre del niño. (Bing)

Podemos ver que esta frase puede tener dos estructuras diferentes. Esta ambigüedad es causada por el indeterminado de la extensión que el verbo modifica. Ambos resultados de TA tienden a la segunda estructura, el objeto de preocupación es la madre del niño.

(9) 五个学校的学生参加了活动。 [wǔ gè xuéxiào de xuéshēng cānjiāle huódòng]

Los estudiantes de cinco escuelas participan en la actividad. / Cinco alumnos participan en la actividad. (traducción humana)

Cinco estudiantes de escuela participaron en el evento. (Google)

Estudiantes de cinco escuelas asistieron al evento. (Bing)

Esta frase tiene dos estructuras sintácticas diferentes.

Esta ambigüedad radica en los "estudiantes de las cinco escuelas", que pueden entenderse como "cinco estudiantes de la misma escuela" o "estudiantes de cinco escuelas diferentes". Aunque los dos tienen las mismas relaciones estructurales, causa diferencias en el significado debido a las diferencias de jerarquías estructural. Esta ambigüedad es causada por indeterminación de la extensión modificado de atributo. En esta frase, "cinco" puede modificar "escuela" y también "estudiante", y cuando las palabras "escuela" y "estudiante" pueden formar una relación

modificativa, puede surgir ambigüedad. Los dos resultados de TA corresponden por separado a dos posibles estructuras sintácticas.

(10) 西藏和青海的部分地区。 [xīzàng hé qīnghǎi de bùfèn dìqū]

Algunas partes de Tíbet y la provincia Qinghai. / Tíbet y algunas partes de la provincia Qinghai.

(traducción humana)

Parte de Tíbet y Qinghai. (Google)

Algunas áreas del Tíbet y Qinghai. (Bing)

Esta frase también puede tener dos estructuras sintácticas diferentes.

Esta ambigüedad es causada por la indeterminación de la extensión conjuntivo de la conjunción "hé". Ambas traducciones tienden a la primera estructura.

(11) 对于服务员的意见 [duì fúwùyuán de yìjiàn]

Para la opinión del camarero / La opinión sobre el camarero (traducción humana)

Para la opinión del camarero (Google)

Para la opinión del camarero (Bing)

Esta frase tiene dos estructuras sintácticas.

A. (esta opinión proviene de camarero). B. (no sabemos la opinión de quién, esta opinión es sobre el camarero). Ambas traducciones son idénticas y pertenecen a la primera estructura.

Esta ambigüedad se debe a la incertidumbre de la jurisdicción de las preposiciones.

(12) 老板要求我们办好发布会 [lǎobǎn yāoqiú wǒmen bàn hǎo fābù huì]

El jefe nos pidió que organizáramos bien la conferencia. / El jefe nos pidió que organizáramos una buena conferencia. (traducción humana)

El jefe nos pidió que organizáramos la conferencia. (Google)

El jefe nos pidió que hagamos la Conferencia de prensa. (Bing)

Esta oración puede tener dos significados: primero, "el jefe nos pidió que organicemos la conferencia correctamente", y el otro es "el jefe nos pidió que organicemos una buena conferencia".

Entre ellos, "bàn hǎo fābù huì" es la causa de la ambigüedad. Debido a las diferencias de jerarquía estructural de la oración, los diferentes significados son causados. Podemos categorizar "bàn hǎo fābù huì" como un tipo de oración de "V + A + N". Cuando V, A y N satisfacen las siguientes condiciones al mismo tiempo, surge la ambigüedad de la oración: A se puede usar como el complemento de V; A se puede usar como el atributo de N; y N puede ser el objeto de V.

Podemos ver que los dos resultados de TA solo difieren en el tiempo gramatical de oración subordinada y la traducción de la " fābù huì", y el resultado de TA no refleja la clave " bàn hǎo fābù huì" de ambigüedad.

6.1.2.2 Diferencia de relación estructural

Cuando una palabra se combina con otra palabra es inevitable que exista una relación estructural, y por lo tanto, la diferencia de la relación estructural también producirá ambigüedad. (Jianming Lu, 2003)

(13) 我们老师工作很认真。 [wǒmen lǎoshī gōngzuò hěn rènzhēn]

Nuestros profesores trabajan con seriedad. / Los profesores, nosotros trabajamos con seriedad. (Traducción humana)

Nuestro maestro está trabajando muy duro. (Google)

Nuestro maestro trabaja muy duro. (Bing)

Hay dos interpretaciones para esta oración. Uno es: "Nuestro maestro trabaja muy duro" y el otro es "Trabajamos muy duro. Somos maestros". La razón de tal ambigüedad es el sujeto "wǒmen" (nosotros) como el sustantivo es considerado el atributo para modificar el sustantivo siguiente "lǎoshī" (maestro), también se puede considerar que "lǎoshī" (maestro) es la aposición de "wǒmen" (nosotros). Ambos resultados traducen que "wǒmen" (nosotros) es un atributo para modificar "lǎoshī" (maestro).

(14) 学习精神 [xuéxí jīngshén]

Espíritu de aprendizaje / Estudia el espíritu (traducción humana)

Espíritu de aprendizaje (Google)

Espíritu de aprendizaje (Bing)

Este ejemplo tiene solo una jerarquía estructural, pero también crea ambigüedades debido a las diferencias de relaciones estructurales. La razón de esta frase que produce una relación estructural diferente es que la palabra "xuéxí" puede verse como un verbo y un sustantivo. Cuando "xuéxí" se usa como verbo, significa "aprender algún espíritu"; y cuando se usa como sustantivo, significa "algún espíritu de aprendizaje". Ambos resultados de TA traducen "xuéxí" como un sustantivo.

(15) 同学们一起商量决定 [tóngxuémen yīqǐ shāngliáng juéding]

Los estudiantes discuten y deciden juntos. / Los estudiantes discuten la decisión juntos.

(traducción humana)

Los estudiantes discuten la decisión juntos. (Google)

Los estudiantes discuten la decisión juntos. (Bing)

Esta oración también tiene dos interpretaciones. La primera es "los estudiantes discuten y toman decisiones juntos" y la otra es "los estudiantes discuten la decisión juntos". El problema es que aunque la "shāngliáng juédìng" tiene la misma jerarquía estructural, la palabra "juédìng" se puede usar aquí como verbo "decidir" y forma una relación paralela con el verbo "shāngliáng" (discutir), también se puede usar un sustantivo, como un objeto del verbo "shāngliáng" (discutir), en consecuencia causa la ambigüedad. Por lo cual, en esta estructura "Verbo 1 + Verbo 2", cuando Verbo 2 se puede usar como ambas categorías, verbo o sustantivo, por lo que podemos observar que fácilmente se genera ambigüedad. Obviamente, ambas traducciones traducen la "juédìng" en un sustantivo.

6.1.3 Ambigüedad semántica

El caso de la ambigüedad semántica ocurre cuando dos conceptos tienen dos o más relaciones semánticas posibles en la misma construcción, es decir, hay más de una interpretación de su significado entre los dos conceptos. En esta sección, seguiremos enfocando nuestra discusión en la ambigüedad causada por incertidumbre de la relación agente-paciente, indefinición de la dirección semántica, relación incierta de la posesión, referencia incierta del pronombre.

6.1.3.1 Indefinición de la relación agente-paciente

(16) 鱼不吃了 [Yú bù chīle]

El pez no come / Yo no como el pez. (traducción humana)

Los peces no comen. (Google)

Los peces no comen. (Bing)

Este es un ejemplo típico para analizar la ambigüedad semántica. Generalmente se refleja en las diferentes relaciones semánticas entre nombres y verbos. La generación de dicha ambigüedad no tiene nada que ver con la jerarquía estructural de las oraciones. La mayoría de ellas se deben a los diferentes significados de los componentes nominales y la flexibilidad de los componentes verbales en la oración, lo que da lugar a una relación agente-paciente poco clara entre los componentes nominales y verbales. Por ejemplo, en este caso, el componente nominal "pez" se puede usar tanto como el agente de la acción "comer" como el paciente de la acción "comer". Del mismo modo, el componente verbal "comer" puede vincularse al sujeto explícito de "pez" y entenderse

como "pez come comida". Al mismo tiempo, también puede relacionarse con el sujeto implícito de "persona" y comprender que "la gente come pez". Por lo tanto, ha surgido la ambigüedad.

Para los resultados de traducción de Google y Bing, podemos ver que sus resultados son mismos, lo que indica que están más inclinados a usar "pez" como agente de la acción, no se ha considerado el sujeto implícito "persona" que no aparece en el ejemplo del lenguaje fuente.

6.1.3.2 Indefinición de la dirección semántica

(17) 她在小船上画画儿。 [tā zài xiǎo chuánshàng huà huà er.]

Ella pinta en el bote. / El bote fue pintado por ella. (traducción humana)

Ella pinta en un pequeño bote. (Google)

Pinta en el bote. (Bing)

"zài xiǎo chuánshàng" (en el bote) es un complemento circunstancial de lugar. Cuando su semántica apunta al sujeto "ella", significa la localidad de "ella". El significado de la oración se puede interpretar como "ella estaba dibujando en el bote"; si "zài xiǎo chuánshàng" (en el bote) su semántica apunta a "huà huà er" (pintar), entonces la oración significa "el bote fue pintado por ella". Por eso, debido a las diferentes orientaciones semánticas, la posición de las personas y las cosas, y el lugar donde se realizan las actividades ha cambiado, surge la ambigüedad. Los dos resultados de TA son casi idénticas y apuntan al sujeto "ella", que presenta la localidad.

6.1.3.3 Relación incierta de la posesión

(18) 这是梵高的画 [Zhè shì fàn gāo de huà]

Esta es la pintura de Van Gogh. / Esta es la pintura que Van Gogh pinta.

(traducción humana)

Esta es la pintura de Van Gogh. (Google)

Esta es la pintura de van Gogh. (Bing)

Para esta oración, no hay ambigüedad en su jerarquía estructural ni relaciones estructurales, pero ha surgido ambigüedad debido a la relación incierta de la posesión. Específicamente, "fàn gāo" (Van Gogh) en la oración se usa como el atributo de "pintura", que explica al poseedor de la palabra central "pintura" y forma la relación de posesión. Sin embargo, el autor de la pintura posiblemente no sea Van Gogh; también se interpreta que el autor de la pintura es Van Gogh. Ambas traducciones solo determinan la relación de la posesión.

6.1.3.4 Referencia incierta del pronombre

(19) 玛丽雅看见老师正在和她的妈妈说话。

[mǎlǐ yǎ kànjiàn lǎoshī zhèngzài hé tā de māmā shuōhuà]

María vio al maestro hablando con su madre. / María vio al maestro hablando con la madre de María. (traducción humana)

María vio a la maestra hablando con su madre. (Google)

María vio al maestro hablando con su madre. (Bing)

En este ejemplo, la ambigüedad surge porque el pronombre personal "ella" puede referirse tanto a "María" como a "maestro". La condición que crea esta ambigüedad es que los pronombres en una oración pueden referirse a dos cosas al mismo tiempo, y que no importa a qué cosa se refiera, igual la oración seguiría teniendo sentido. Podemos ver que los dos resultados de TA se traducen según el modo narrativo de la lengua de fuente, y aún no indican explícitamente la referencia específica del pronombre.

6.1.4 Ambigüedad pragmática

La ambigüedad pragmática se refiere a aquellas oraciones que en sí mismas no son ambiguas, pero se vuelven ambiguas o producen ambigüedad en algún contexto especial. Aquí tomamos una visión estrecha de la ambigüedad pragmática. Es decir, solo tomamos aquellas que tienen dos o más entidades posibles de forma escrita, y sus interpretaciones correctas (o una de ellas) tienen que depender del contexto. No se considera la ambigüedad pragmática causada por factores fonéticos o coloquiales, como el tono, la voz, el acento, la pausa, etc.

(20) 她有个儿子，在银行工作。 [tā yǒu gè er zi, zài yínháng gōngzuò.]

Ella tiene un hijo y trabaja en el banco. / Ella tiene un hijo que trabaja en un banco.

(traducción humana)

Ella tiene un hijo y trabaja en el banco. (Google)

Tiene un hijo que trabaja en un banco. (Bing)

La cláusula posterior de esta oración omite el sujeto, y este sujeto omitido puede ser "ella" o "hijo". Si pensamos que el sujeto omitido es "ella", entonces la oración significa que "ella trabaja en el banco"; si el sujeto omitido es "hijo", entonces significa que "ella tiene un hijo y su hijo trabaja en el banco". A menudo, las personas omiten palabras en el proceso de comunicación. Si se abandona un contexto específico, habrá múltiples interpretaciones. Los dos resultados de TA corresponden a posibles interpretaciones de la ambigüedad.

(21) 爸爸明天要做手术。 [Bàba míngtiān yào kāidào.]

Mañana papá efectuará una operación. / Papá será operado mañana. (traducción humana)

Papá hará cirugía mañana. (Google)

Papá va a tener una operación mañana. (Bing)

Esta oración debe analizarse junto con el entorno específico y el antecedente familiar, si el padre es médico, entonces, esta frase representa que el padre realizará la cirugía al paciente. Si el papá está enfermo, El padre, como paciente, espera que el médico le haga una cirugía. El resultado de Google trata al padre como médico. El resultado de Bing está más cerca de la descripción del lenguaje de fuente.

6.2 Estudio de la TA español-chino de producciones ambiguas en español

La clasificación de la ambigüedad en español que tomamos es el mismo que hemos tomado para analizar las ambigüedades del idioma chino. Es decir, las ambigüedades españolas también tendrán cuatro tipos de niveles: léxico, sintáctico, semántico y pragmático. En este estudio, la TA español-chino se hace con *Google Translator* y *Bing Translator*.

6.2.1 Ambigüedad léxica

La homonimia de un signo léxico puede ser la responsable de la ambigüedad de ciertas construcciones que reciben una misma descripción fonológica y sintagmática. Y, además, la palabra polisémica también produce ambigüedad léxica.

6.2.1.1 Homonimia

(22) Lo bajo

我把它放低。 / 比较低的那个。 (traducción humana)

我降低它 (Google)

它下 (Bing)

La palabra “bajo” es un homógrafo, pertenece a diferentes categorías, puede ser la conjugación de la primera persona del presente del modo indicativo, sobre el verbo “bajar” y también puede ser adjetivo, sustantivo, adverbio o preposición. El resultado de Google fija el uso de “bajo” como un verbo, “lo” como pronombre masculino singular. Para el Traductor de Bing, “lo” funciona como un artículo que se antepone a un adjetivo “bajo”.

La palabra "para" también pertenece a esta situación. Puede ser la conjugación de la tercera persona del presente del modo indicativo del verbo “parar”, o la tercera persona del presente del modo

subjuntivo del verbo “parir”. También puede referirse a una preposición, o a un prefijo. Pero tal ambigüedad puede evitarse poniéndola en oraciones concretas. En este trabajo, no discutimos los homófonos porque ellos se pronuncian igual pero su escritura y significado son diferentes.

6.2.1.2 Polisemia

(23) Le gusta la heroína

他喜欢海洛因。 / 他喜欢女英雄。 (traducción humana)

他喜欢海洛因 (Google)

他喜欢海洛因 (Bing)

La palabra “heroína” puede referirse tanto a la protagonista de una obra de ficción como a la droga adictiva obtenida de la morfina (RAE). Los dos resultados de TA son iguales y ambas traducen “heroína” a droga.

(24) Me lo cargo todo.

我把所有东西都装走。 / 我负责所有的事。 (traducción humana)

我把这一切都拿走了。 (Google)

我负责这一切。 (Bing)

Este es un anuncio de una furgoneta que usa una palabra polisémica "cargar", un juego de palabras, que puede interpretarse como "poner en la furgoneta mercancías para transportarlas" (RAE) y también puede interpretarse como "tomar o tener sobre sí alguna obligación o cuidado" (RAE). El resultado de Google se traduce literalmente, enfatizando la característica de la furgoneta, es decir, cargar y transportar cosas; La traducción de Bing se traduce en otro significado de anuncio, que enfatiza que la furgoneta no solo puede cargar y transportar cosas, sino también garantizar la seguridad.

6.2.2 Ambigüedad sintáctica

Como en el idioma chino, la ambigüedad sintáctica española también se llama ambigüedad estructural. Debido a la forma en que se asocian los distintos constituyentes de una oración, podemos interpretarla de varias formas distintas. Una oración muy típica es:

(25) Carlos vio a la mujer con un telescopio.

胡安用望远镜看见那个女人。 / 胡安看见一个拿着望远镜的女人。(traducción humana)

胡安用望远镜看见那个女人。 (Google)

约翰看见那个女人用望远镜。 (Bing)

Esta oración tiene dos interpretaciones de estructura sintáctica. Por lo cual, usamos el diagrama de árbol para explicarla.

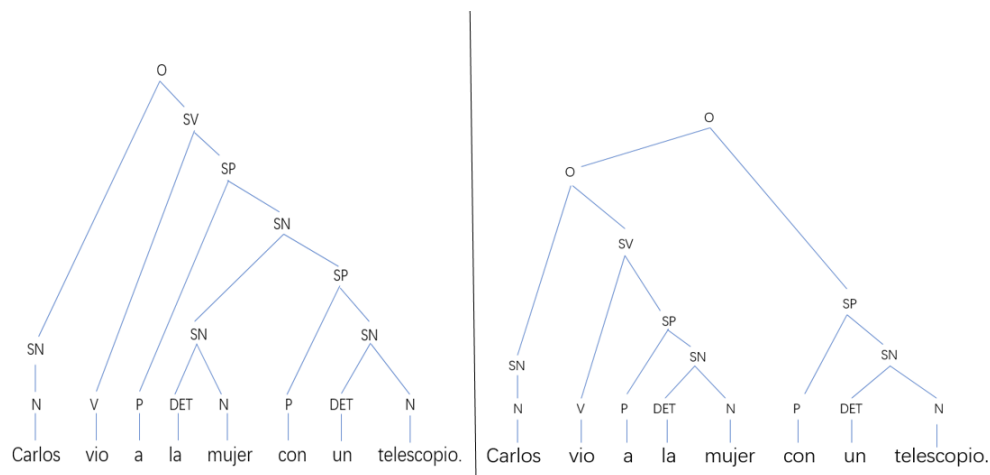


Figura 3. El diagrama de árbol

¿Carlos usó el telescopio para ver a la mujer o Carlos vio que la mujer tenía un telescopio?

Ambos resultados de TA se traducen literalmente. Además de las diferencias del nombre traducido “Carlos” y órdenes de palabras, la ambigüedad aún existe.

(26) La prima *de la* que me hablaste.

你跟我说的那个人的表妹。 / 你跟我说的表妹的那件事。 (traducción humana)

你告诉过我的表弟。 (Google)

你跟我说的那个表弟。 (Bing)

La doble función del sintagma en cursiva “de la” es la responsable de la ambigüedad. Se interpreta: a. X que es prima de Y, de la que me hablaste; b. me hablaste de la prima. El sintagma “de la” son transpositores o índice funcional y artículo.

Ambas traducciones tienden a la primera interpretación, pero todas ignoran el papel de "de la". Solo se traduce la palabra “prima” y no se traduce que esta prima es de la persona que son conocidas por el hablante y el receptor, es decir, la prima es de la persona a quien se refiere el pronombre “la”. Además, la palabra "prima" debe traducirse como femenino. Ambas traducciones lo traducen al masculino.

(27) Llegará *el día de Pascua*.

复活节将到来。 / 他将在复活节到来。(traducción humana)

复活节将到来。 (Google)

复活节将会到来。 (Bing)

El sintagma en cursiva “el día de Pascua” puede ser sujeto o aditamento. Ambos resultados lo traducen como sujeto.

6.2.3 Ambigüedad semántica

Una misma estructura sintáctica puede interpretar diferentes significados, en consecuencia, ocurren ambigüedades semánticas.

(28) La elección del alcalde”

市长选别人/市长当选 (traducción humana)

市长选举 (Google)

市长选举 (Bing)

(29) La condena del juez”

法官给别人定罪/法官被定罪 (traducción humana)

法官的信念 (Google)

法官的定罪 (Bing)

(30) La defensa del PSOE”

PSOE 为自己辩护/ PSOE 被捍卫 (traducción humana)

PSOE 的防御 (Google)

PSOE 的辩护 (Bing)

(31) La crítica del crítico”

评论家批评别人/评论家被批评 (traducción humana)

评论家的批评 (Google)

评论家的批评 (Bing)

Estos ejemplos poseen una misma descripción sintáctica y fonológica, presentan doble sentido. La clave de su ambivalencia se encuentra en la duplicidad de relaciones semánticas que evalúa entre los componentes de tales construcciones. Estos ejemplos, pertenecen al modelo del genitivo subjetivo y del genitivo objetivo: el alcalde elige / el alcalde es elegido; el juez condena / el juez es condenado; el PSOE se defiende / el PSOE es defendido; la crítica critica / la crítica es criticada. No importa el tipo de TA, los resultados de TA se traducen literalmente según el idioma de fuente y no ofrece ninguna inclinación por las interpretaciones posibles de ambigüedades, por lo que, los resultados de TA aún son ambiguos.

(32) Le compró flores.

他给她买了花。/他买了花给卖给她的人。(traducción humana)

他买了她的花。 (Google)

给她买了花。 (Bing)

La clave de ambigüedad reside en el destinatario de las flores. La palabra “le” es un complemento indirecto, puede ser una persona a la que le regalan flores por ser su cumpleaños, o bien la persona que las vende (el florista).

El resultado de Google puede interpretarse como "el compró sus flores", que no refleja la palabra "le" en absoluto, que está lejos del significado de la oración. En comparación con el resultado de Google, Bing traduce más palabra por palabra.

6.2.4 Ambigüedad pragmática

Decimos que un enunciado X es pragmáticamente ambiguo cuando en un contexto dado el receptor Y puede asignarle dos o más interpretaciones. Las ambigüedades pragmáticas no son hechos de significación, sino de sentido. Están ligadas a los agentes del discurso (emisor y receptor, primordialmente). (Salvador Gutiérrez Ordóñez, 2002)

Vamos a presentar dos formas de declarar de la ambigüedad pragmática: ambigüedades referenciales, ambigüedades ilocutivas.

6.2.4.1 Ambigüedades referenciales

Esta ambigüedad puede situar al receptor en situaciones de tener que escoger entre dos o más referentes.

(33) Vivió varios años en Santiago.

他在圣地亚哥住了几年。 (traducción humana)

他在圣地亚哥住了几年。 (Google)

他在圣地亚哥住了几年。 (Bing)

Podemos ver que el resultado es el mismo ya sea una traducción humana o una traducción automática. Sin embargo, sabemos que existe varias ciudades llamadas Santiago, y no sabemos a cuál nos referimos aquí. Esta ambigüedad ocurre cuando el contexto no es suficientemente claro.

6.2.4.2 Ambigüedades ilocutivas

(34) ¿Tienes hora?

你有一个小时? /你有时间吗?

你有一个小时吗? (Google)

你有时间吗? (Bing)

Entre la interpretación literal y la indirecta, los actos de habla indirectos pueden exponer ambigüedad pragmática. El valor lingüístico de esta oración es interrogativo y el valor ilocutivo puede ser petición o pregunta.

Los resultados de Google se traducen literalmente según el idioma de fuente (¿Tienes una hora?), no refleja el valor ilocutivo, sino que solamente indica la pregunta. El resultado de Bing tiende a expresar la petición (¿Tienes tiempo libre?) para hacer más peticiones más tarde.

Además, en las ambigüedades pragmáticas las ironías y antífrasis de los usos figurados es más conocido. En algunos contextos, el discurso necesita semantizarse de acuerdo con las intenciones del hablante, es decir, darle un significado especial. Por ejemplo:

(35) Tu jefe no se vende por un puñado de dólares.

你的老板不会因为这点钱就出卖自己。(traducción humana)

你的老板不会卖几块钱。(Google)

你的老板不卖一小块钱。(Bing)

El hablante es posiblemente muy honesto al decir esta frase para alabar al jefe, o su insinuación es que el jefe piensa que no hay suficiente dinero, y él cobrará más sobornos. En la misma oración, hay una gran diferencia semántica. Los resultados de TA son "Tu jefe no vende por un puñado de dólares", no traducen con precisión el verbo pronominal "venderse" y no muestran sentido irónico. La mayoría de las oraciones de ambigüedad pragmática no son ambiguas, sin embargo, la ambigüedad puede ocurrir en el contexto o la comunicación. Si se pone una misma oración en un contexto diferente, el significado será diferente.

7. Discusión de resultados

Del análisis anterior sobre las ambigüedades, podemos ver que hay algunos fenómenos compartidos por el español y el chino, como la ambigüedad léxica y la ambigüedad sintáctica. Al observar detenidamente los ejemplos, podemos ver que algunos casos son igualmente ambiguos en ambos idiomas. Dado que este no es un estudio comparativo entre los idiomas chino y español, no enumeraremos más acerca de sus similitudes o diferencias dentro del alcance de la ambigüedad. Sin embargo, no hay duda de que es importante reconocer las ambigüedades en los idiomas al observar y analizar sus fenómenos gramaticales y las estructuras lingüísticas profundamente y

exhaustivamente. Por otro lado, el estudio de las ambigüedades se considera que es la forma más eficiente de investigar las complicadas relaciones entre la forma y el significado de un idioma, sin duda, es una gran ayuda para avanzar en el desarrollo de la lingüística y desambiguación del sistema de TA.

Además, a partir de los resultados de los sistemas de TA de Google y Bing, podemos ver que la TA ya ha resuelto algunas ambigüedades, aunque a veces ocurre que el mismo problema de ambigüedad es identificado por Google pero no por Bing, y viceversa. Para el traductor Google, la coherencia y la flexibilidad son mejores, puede traducir frases comunes y las expresiones idiomáticas con exactitud. Puede efectivamente resolver los materiales de la traducción. Del chino a español y del español a chino no hay diferencia significativa en la calidad de traducción. Para la traducción del adverbio en los dos traductores es inexacta o directamente ignorada cuando el español traduce al chino. Los campos de la traducción son más globales e incluye un rico corpus.

Sobre el traductor Bing, la cantidad del vocabulario es suficiente, pero la estructura de la traducción es muy pobre, aunque adecuada para la búsqueda de palabras. Sobre el grado de precisión, la calidad de traducción del español al chino es mejor que del chino al español. En la traducción del español al chino los cuantificadores chinos son más precisos.

8. Conclusión

En resumen, este TFG ha abordado los principales métodos de TA con el fin de entender cómo se lleva a cabo. Además se ha resumido el método de TA utilizado por el traductor de Google. Finalmente, se ha analizado respecto a cada tipo de ambigüedad cuáles son los resultados que ofrece el traductor de Google y de Bing con el fin de analizar su calidad y detectar dónde existen problemas de traducción.

La conclusión es que ambos traductores tienen una capacidad analítica limitada. Hay una falta de flexibilidad en la selección del léxico, el análisis del tiempo verbal, el análisis de la estructura de oraciones, errores en el orden de palabras, la modificación de componentes de oraciones. Ésta no puede reflejar el significado interno de la oración, ya que conducen fácilmente a una traducción incompleta y, en consecuencia, a malentendidos. En general, la TA puede ayudar inicialmente al traductor que, después, debe perfeccionar las traducciones y resolver los problemas de traducción, pero respecto a la flexibilidad de traducción, en el análisis todavía encontramos defectos.

Bibliografía

Arnold, D.J. (1993). *Machine Translation: An Introductory Guide*

Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural machine translation by jointly learning to align and translate

Consultado en: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Declan Groves, Andy Way. (2005). Hybrid Example-Based SMT: The Best of Both Worlds? *In Proceedings of ACL Workshop on Building and Using Parallel Texts: Data-Driven Machines Translation and Beyond*, pages 183-190. Ann Arbor, Michigan.

Hua Wu, Haifeng Wang, Kai Tang. (2005). Improving Translation Memory with Word Alignment Information. *In proceedings of MT SUMMIT X*, pages 313-320. Phuket Island, Thailand

Indurkha, Nitin, & Fred J Damerau.(2010). *Handbook of Natural Language Processing*. Second edition. Chapman & Hall/CRC Press.

James, Allen. (1995). *Natural language understanding*. Benjamin-Cummings Publishing Co., Inc. Redwood City, CA, USA

Jianming, Lu. (2003). *现代汉语语法研究教程*. Peking University Press.

Jurafsky, D., & J.H. Martin. (2009). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd Edition. PEARSON.

Mitkov, Ruslan. (2005). *The Oxford Handbook of Computational Linguistics*. OUP Oxford.

M. Nagao. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, In: A. Elithorn, R. Banerji (eds), *Artificial and Human Intelligence*, pages 173-180. NATO publications

Pérez, Sonia Vázquez. (2009). Resolución de la ambigüedad semántica mediante métodos basados en conocimiento y su aportación a tareas de PLN, página 3-6. Universidad de Alicante

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, Paul S. Roossin. (1990). A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79-85

Salvador, Gutierrez Ordoñez. (1996). *De pragmática y semántica*. ARCO/LIBROS, S.L., *Introducción a La Semántica Funcional*. SINTESIS

Weijun, Gong. (2010). *The research of Modern Chinese Ambiguity and the utilization in Chinese Courses for Foreigners*. University of Shandong

Wenmin, Zhao. (2012). *Ambiguity in Modern-Chinese*

World Economic Forum del día 19 de febrero 2018. Recuperado del sitio web de WEF: <https://es.weforum.org/agenda/2018/02/cuales-son-los-idiommas-mas-hablados-en-el-mundo>

Yanyun, Song. (2007). *现代汉语歧义研究*, Northwest University China.

Yingxian, Cui. (2004). 现代汉语语法学习与研究入门. Tsinghua University Press.

Yonghui, Wu, Schuster Mike, Chen Zhifeng & V Quoc. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

Zhanyi Liu, Haifeng Wang, Hua Wu. (2006). Example-Based Machine Translation Based on Tree-string Correspondence and Statistical Generation. *Machines Translation*, 20(1): 25-41

Anexo: Tabla con los resultados de las traducciones de la muestra de frases

Los ejemplos chinos traducen al español

Ejemplo (Chino)	Posible traducción (Traducción humano)	Resultado Google	Resultado Bing
山上到处是杜鹃 [Shānshàng dào chù shì dùjuān]	La montaña llena de rododendros. (Flor) / La montaña llena de cucos. (Ave)	La montaña está llena de cucos	Las montañas están llenas de rododendros.
注意仪表 [Zhùyì yíbiǎo]	Preste atención a la apariencia / Preste atención al panel	Medidor de atención	Cuidado con el taxímetro.
他还欠款 2 万 [Tā hái qiǎn kuǎn 2 wàn] [Tā huán qiǎn kuǎn 2 wàn]	Él todavía debe 20,000 yuanes / Él devuelve la deuda 20,000 yuanes	Él todavía debe 20,000 yuanes	Aún debe 20.000.
昨天, 老板给这家商店配备了一些机动车辆 [Zuótiān, lǎobǎn gěi zhè jiā shāngdiàn pèibèi yìxiē jīdòng chēliàng]	Ayer, el propietario equipó a esta tienda con unos vehículos motorizados Ayer, el propietario equipó a esta tienda con unos vehículos para todo	Ayer, el propietario equipó esta tienda con algunos vehículos de motor	Ayer, el jefe le dio a la tienda algunos vehículos motorizados.
他已经走了两个小时了 [Tā yǐjīng zǒule liǎng gè xiǎoshíliǎo]	Él caminó durante dos horas / Se fue hace dos horas.	Él ha estado caminando durante dos horas	Hace dos horas que se fue.
这间教室的门没有锁 [Zhè jiān jiàoshì de mén méiyǒu suǒ]	La puerta del aula no tiene cerradura. La puerta del aula no está cerrada.	La puerta de este aula no está cerrada	La puerta de esta habitación está abierta.
他原来住在这 [tā yuánlái zhù zài zhè]	Él vivió aquí antes / No esperaba que él viviera aquí	Originalmente se quedó aquí [原本他留在这里]	Solía vivir aquí. [我曾经住在这里。]
关心孩子的母亲 [Guānxīn hái zǐ de mǔqīn]	la madre que cuida a los niños / cuida a la madre de los niños	Cuidando a la madre del niño	Cuidado de la madre del niño
五个学校的学生参加了活动 [wǔ gè xuéxiào de xuéshēng cānjiā le huódòng]	Los estudiantes de cinco escuelas participan en la actividad. Cinco alumnos participan en la actividad	Cinco estudiantes de escuela participaron en el evento	Estudiantes de cinco escuelas asistieron al evento
西藏和青海的部分地区 [xīzàng hé qīnghǎi de bùfèn dìqū]	Algunas partes de Tíbet y la provincia Qinghai Tíbet y algunas partes de la provincia Qinghai	Parte de Tíbet y Qinghai	Algunas áreas del Tíbet y Qinghai

对于服务员的意见 [duì fúwùyuán de yìjiàn] àn]	Para la opinión del camarero La opinión sobre el camarero	Para la opinión del camarero	Para la opinión del camarero
老板要求我们办好发布会 [lǎobǎn yāoqiú wǒmen bàn hǎo fābù huì] huì]	El jefe nos pidió que organizáramos bien la conferencia El jefe nos pidió que organizáramos una buena conferencia	El jefe nos pidió que organizáramos la conferencia	El jefe nos pidió que hagamos la Conferencia de prensa.
我们老师工作很认真 [wǒmen lǎoshī gōngzuò hěn rènzhēn] hěn rènzhēn]	Nuestros profesores trabajan con seriedad. Los profesores, nosotros trabajamos con seriedad.	Nuestro maestro está trabajando muy duro	Nuestro maestro trabaja muy duro.
学习精神 [xuéxí jīngshén] ngshén]	Espíritu de aprendizaje / Estudia el espíritu	Espíritu de aprendizaje	Espíritu de aprendizaje
同学们一起商量决定 [tóngxuémen yìqǐ shāngliáng juédìng] ngliáng juédìng]	Los estudiantes discuten y deciden juntos. Los estudiantes discuten la decisión juntos.	Los estudiantes discuten la decisión juntos	Los estudiantes discuten la decisión juntos.
鱼不吃了 [Yú bù chīle] le]	El pez no comen / Yo no como el pez.	Los peces no comen	Los peces no comen.
他在小船上画画儿。 [tā zài xiǎo chuánshàng huà huà er.] ng huà huà er.]	Ella pinta en el bote / Ella pinta el bote	Él pinta en un pequeño bote.	Pinta en el bote.
这是梵高的画 [Zhè shì fàn gāo de huà] ì fàn gāo de huà]	Esta es la pintura de Van Gogh	Esta es la pintura de Van Gogh	Esta es la pintura de van Gogh.
玛丽雅看见老师正在和她的妈妈说话 [mǎlì yǎ kànjiàn lǎoshī zhèngzài hé tā de māma shuōhuà] ī zhèngzài hé tā de māma shuōhuà]	María vio al maestro hablando con su madre María vio al maestro hablando con la madre de María	Mariah vio a la maestra hablando con su madre	María vio al maestro hablando con su madre
她有个儿子，在银行工作。 [tā yǒu gè er zi, zài yínháng gōngzuò.] nháng gōngzuò.]	Ella tiene un hijo y trabaja en el banco. Ella tiene un hijo que trabaja en un banco.	Ella tiene un hijo y trabaja en el banco.	Tiene un hijo que trabaja en un banco.
爸爸明天要开刀。 [Bàba míngtiān yào kāidāo.] idāo.]	Mañana papá efectuará una operación. Papá será operado mañana	Papá hará cirugía a mañana.	Papá va a tener una operación mañana.

Los ejemplos españoles traducen al chino

Ejemplo (Español)	Posible traducción (Traducción humano)	Resultado Google	Resultado Bing
Lo bajo	底下的	我降低它	它下
para	为了	至	为
Le gusta la heroína	他喜欢海洛因 他喜欢女英雄	他喜欢海洛因	他喜欢海洛因
Me lo cargo todo	我把所有东西都装走了 我负责所有的事	我把这一切都拿走了	我负责这一切
Juan ve a la mujer con un telescopio	胡安用望远镜看见那个女人 胡安看见一个拿着望远镜的女人	胡安用望远镜看见那个女人	约翰看见那个女人用望远镜
La prima de la que me hablaste	你跟我说的那个表妹 你跟我说的关于表妹的那件事	你告诉过我的表弟	你跟我说的那个表弟
Llegará el día de Pascua.	复活节将到来 他将在复活节到来	复活节将到来	复活节将会到来
La elección del alcalde	市长选举 (人们选择别人当市长) 市长的选择 (市长做出选择)	市长选举	市长选举
La condena del juez	法官的判决 (法官判决别人) 法官的判决 (别人判决法官)	法官的信念	法官的定罪
La defensa del PSOE	PSOE 的辩护 (PSOE 为自己辩护) PSOE 的辩护 (别的对 PSOE 的辩护)	PSOE 的防御	PSOE 的辩护
La crítica del crítico	评论家的批评 (评论家批评别人) 评论家的批评 (别人对评论家的批评)	评论家的批评	评论家的批评
Le compró flores.	他跟她买了花。/他买了花给她。	他买了她的花	给她买了花
Vivió varios años en Santiago.	他在圣地亚哥住了几年。	他在圣地亚哥住了几年。	他在圣地亚哥住了几年。
Como complemento, léanse el libro de Pottier	作为补充, 请阅读 Pottier 的书	作为补充, 请阅读 Pottier 的书	作为补充, 读欧仁·鲍狄埃书
Policía: Dígame la verdad Interrogado: 《La verdad》	警察: 告诉我真相 嫌疑人: “真相	警察: 告诉我真相 质疑: “真相	警察: 告诉我真相 质疑: 真相

Edipo dijo que quería casarse con su madre	俄狄浦斯说他想和他的母亲结婚。	俄狄浦斯说他想和他的母亲结婚	俄狄浦斯说他想和他母亲结婚
Tu perro es muy cariñoso	你的狗真可爱	你的狗非常亲热	你的狗很深情
Tu jefe no se vende por un puñado de dólares	你的老板不会为了这点小钱出卖良心的。	你的老板不会卖几块钱	你的老板不卖一小块钱。
Eres un genio	你真是个天才。	你是一个天才	你是个天才