

Literalidad y traducción

Diego Uribe Agundis

And since you know you cannot see
yourself,
so well as by reflection, I, your glass,
will modestly discover to yourself,
that of yourself which you yet know
not of.
- *William Shakespeare*

Diego Uribe Agundis
(Torreón, Coahuila, 1960). Ph.D.
Computer Science. University of
Shefiel. Exbecario del CONACYT.
diego.uribe@lag.uia.mx

En este mundo actual en el que gran cantidad de procesos -en prácticamente todo tipo de actividades- se realizan de manera automática, hemos considerado pertinente preguntarnos si las necesidades de traducción que experimentamos en nuestras actividades cotidianas (ya sea tareas académicas, labores de servicios, o bien en el uso del tiempo libre al conocer nuevos amigos) pueden ser satisfechas con tan sólo hacer uso de las herramientas tecnológicas existentes. Dicho de otra manera, planteamos la siguiente pregunta: ¿las herramientas de traducción existentes nos permiten afrontar exitosamente las actividades que implican el uso de un lenguaje diferente al nuestro?

En vez de dar una respuesta categórica a dicha pregunta, nuestra intención será ilustrar porqué el procesamiento automatizado del lenguaje natural es una ardua tarea. Para ello será necesario delimitar con precisión el problema que nos atañe, siendo éste la transformación de un documento de texto en un texto equivalente pero descrito en términos de un lenguaje distinto al lenguaje del texto original. Esto es, hacemos a un lado los problemas concernientes al reconocimiento de voz para enfocar

nuestra atención en el análisis del texto original que deberá producir, como resultado de dicho análisis, un texto equivalente al original, pero descrito en términos de un lenguaje diferente.

Una vez acotado el problema a abordar, hacemos uso de la estructura piramidal (Figura 1) con la cual Manning (1999) ilustra diferentes alternativas para automatizar el proceso de traducción. Dicha estructura muestra en la parte inferior de la pirámide las estrategias más elementales que sirven de apoyo a estrategias más avanzadas que se observan a medida que se aleja de la base piramidal. Es así como la primera y más obvia estrategia en el proceso de traducción es aplicar un enfoque basado en el análisis de las palabras que conforman el texto –después de todo, las palabras son los elementos atómicos que nos permiten transmitir nuestros pensamientos. En este caso, el uso de un diccionario juega un papel crucial ya que reemplazar cada palabra del texto original con su correspondiente significado nos permite encarar el proceso de traducción desde la perspectiva del mas elemental sentido común –es muy simple diseñar un programa computacional que solo observe los términos y

su correspondiente definición en un diccionario bilingüe. Sin embargo, en vista de que la mayor parte de las palabras representan más de un significado, este enfoque exhibe serias deficiencias. Por ejemplo, cuando buscamos en el diccionario el significado de la palabra *goal*, podemos observar los siguientes significados:

Definition

goal 1. (aim) objetivo m, meta f; to achieve a ~ conseguir un objetivo; to pursue a ~ perseguir un fin; to set a ~ fijar un objetivo **2.** SPORTS (scoring area) portería f; to keep ~ defender la portería; to play in ~ Brit ser portero **3.** SPORTS (point) gol m; to score a ~ marcar un gol; a penalty ~ un gol de penalty (from Diccionario Cambridge Klett Compact)

Como podemos apreciar en la definición, debemos tomar una decisión respecto a cual de los tres significados es el más pertinente. De esta manera,

experimentamos un problema no trivial conocido como problema de ambigüedad léxica: debemos resolver la confusión que una palabra común como *goal* ha generado. Un simple programa computacional, como el que señalamos líneas arriba, también experimentará dicha confusión a menos que el programador implemente algunas reglas que le permitan dilucidar la alternativa más adecuada. Sin embargo, la implementación de dichas reglas requiere plantear la siguiente pregunta: ¿qué hacemos los humanos para determinar el significado correcto de la palabra al hacer uso de un diccionario en el proceso de traducción? La respuesta es nada más y nada menos que la observancia del contexto. Esto es, el texto que circunda la palabra de particular interés –el texto que sigue y precede a la palabra– es lo que nos permite determinar su significado correcto. Por lo tanto, observar el contexto es esencial para la automatización del proceso de traducción.

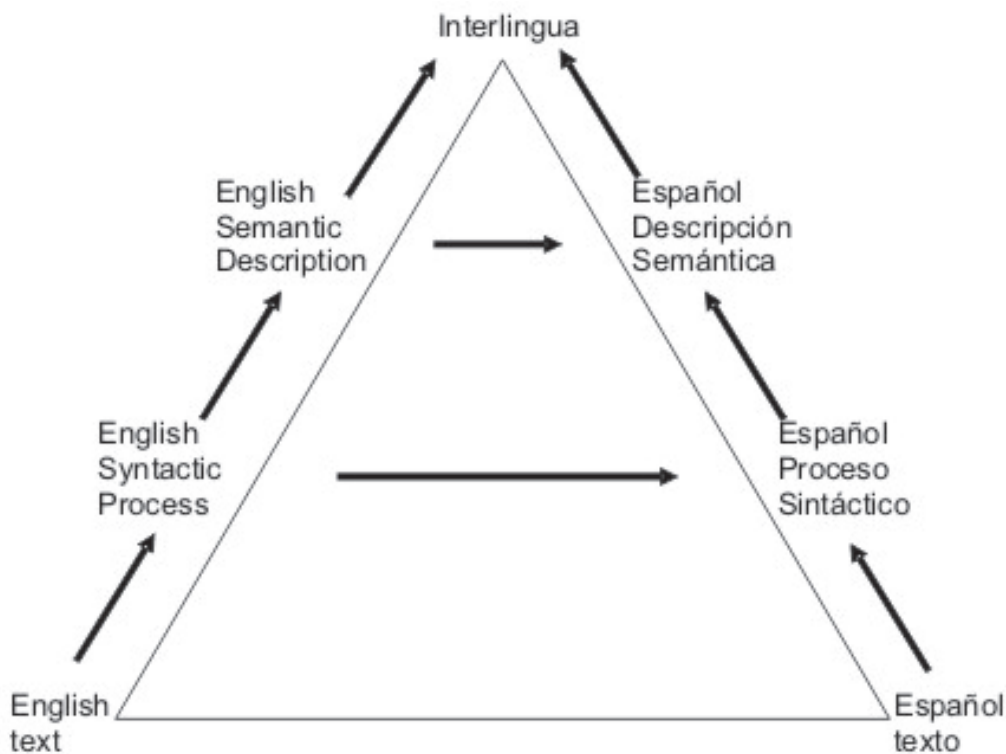


Figura 1 Estructura Piramidal para la Automatización de la Traducción del Lenguaje Natural.

Un problema adicional en este enfoque léxico es el orden de las palabras. Por ejemplo, la traducción léxica –basada únicamente en palabras– de una frase simple como *the blue car* y su equivalente *el azul auto* es inadecuada en el lenguaje castellano. Este problema hace evidente el hecho de que cada lenguaje natural es representado por un conjunto de reglas –conocido como gramática– que deben ser observadas para una comunicación eficiente.

El propósito de la segunda estrategia en el proceso de automatización es el análisis de este tipo de problemas. En este nuevo enfoque, las palabras como unidades básicas de análisis son reemplazadas por las estructuras sintácticas. Es así como de esta manera, el texto original debe ser compilado con el fin de obtener la estructura sintáctica (conocida como árbol sintáctico) la cual será usada para crear el equivalente árbol sintáctico en términos del lenguaje deseado para posteriormente generar la traducción final a partir de éste último árbol producido. Descrito este proceso de otra manera, podemos asumir que en este enfoque hacemos uso de diccionarios que contienen reglas gramaticales, en vez de palabras, para producir árboles sintácticos equivalentes. Podemos por ejemplo asumir, para el caso de las frases mencionadas líneas arriba, que hacemos uso de las siguientes reglas sintácticas¹:

Inglés	Castellano
NP DT JJ NN	NP DT NN JJ

Aun cuando este enfoque representa un progreso significativo comparado con la traducción léxica, la ambigüedad sigue siendo el problema a abordar. En este enfoque, el problema de ambigüedad se hace evidente cuando el proceso de compilación para una oración genera más de un árbol sintáctico. El texto clásico que representa este

problema de ambigüedad es la oración *the girl saw the boy with a telescope* (*la niña vio al niño con un telescopio*). Los árboles sintácticos generados por el proceso de compilación de esta oración son los siguientes:

En base a los árboles sintácticos mostrados en la Figura 2, se puede observar que el problema en este caso radica en determinar quien es la persona que hace uso del telescopio: la niña o el niño. Mientras que el enfoque previo hace evidente la confusión en cuanto a la determinación del significado correcto para una palabra en particular, en este caso cualquier estructura sintáctica representa una interpretación plausible. Por lo tanto, la disyuntiva que se presenta en este caso se conoce como ambigüedad sintáctica. Resumiendo, en esta segunda perspectiva del proceso de automatización, el problema de ambigüedad sintáctica es el obstáculo a superar con el fin de elegir la interpretación correcta.

La tercera alternativa en el proceso de automatización tiene por objeto abordar la ambigüedad que se presenta en el caso previo. El propósito en este enfoque es elaborar una representación semántica del texto que se pretende traducir para entonces transformar dicha representación en una representación semántica equivalente descrita en términos del lenguaje deseado de manera tal que la traducción final se genere a partir de esta última representación semántica. Existen varias técnicas para crear dichas representaciones semánticas (que posiblemente se basan en el preprocesamiento sintáctico del enfoque previo), las cuales hacen uso de diversas notaciones para la descripción del significado que transmite la oración. El uso de la Lógica Formal, Redes Semánticas y Gramáticas Semánticas son solo algunas de las diversas estrategias creadas con el fin de describir el

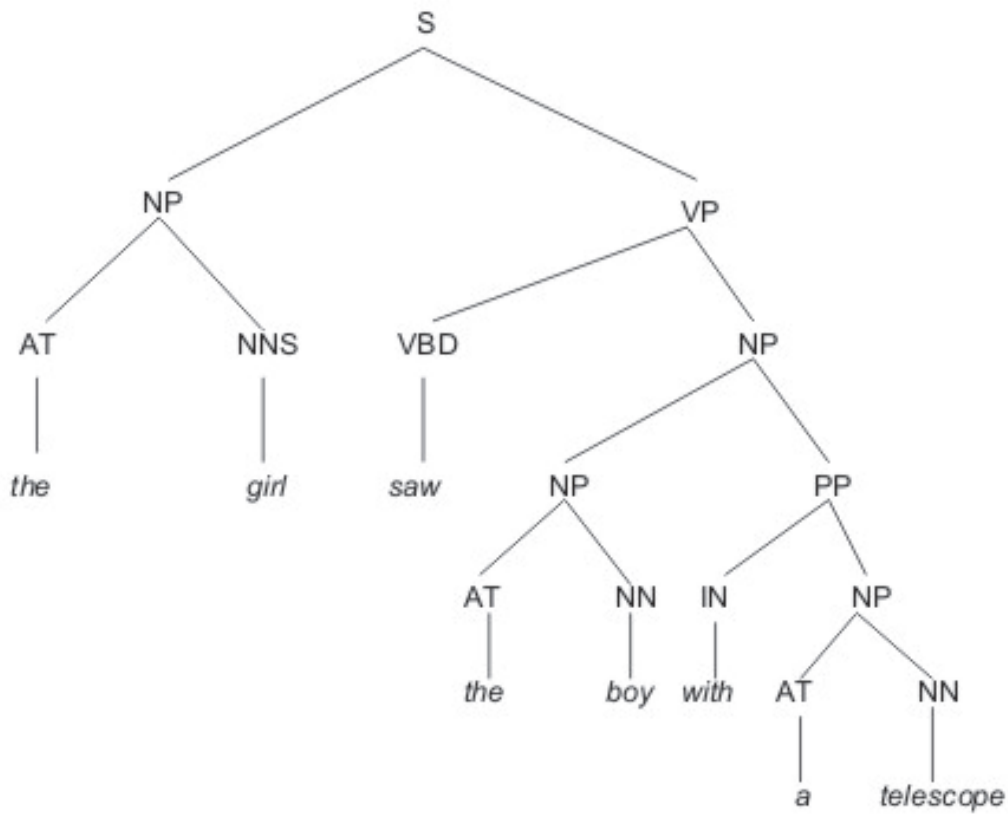
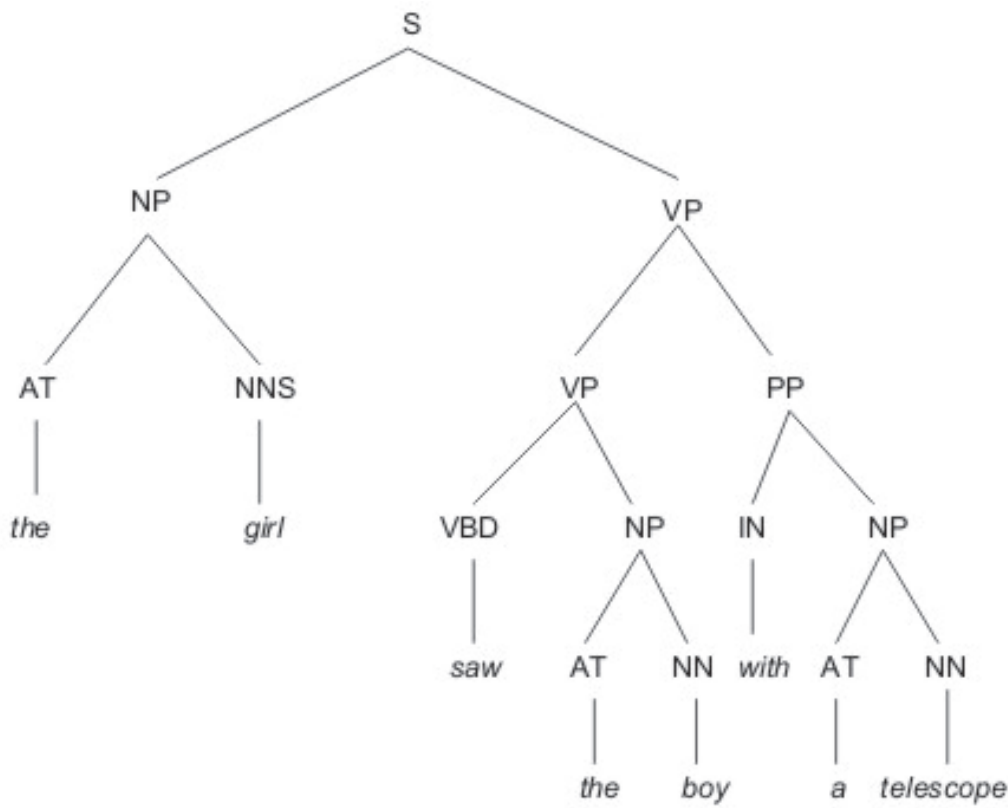


Figura 2. Árboles sintácticos generados para *the girl saw the boy with a telescope*. En el primer árbol sintáctico, es la niña quien hace uso del telescopio; mientras que en el segundo árbol sintáctico es el niño quien usa dicho artefacto.

significado de nuestras cláusulas o enunciados. A continuación mostramos el ejemplo de Jurafsky (2000) que ilustra

el uso de la Lógica Formal para representar el significado de una simple frase:

I have a car	Tengo un auto
? _{x,y} Having(x) ? Haver(Speaker, x) ?	? _{x,y} Poseer(x) ? Poseedor(Hablante, x) ?
HadThing(y, x) ? Car(y)	ObjetoPoseido(y, x) ? Auto(y)

Existen casos en los cuales la traducción literal obtenida a través de representaciones semánticas es correcta, sin embargo, la traducción no es del todo adecuada. Talmy (1985) ilustra este caso al mostrar la diferencia en la manera en la cual lenguajes tales como inglés y castellano describen relaciones de movimiento o de espacio entre los objetos. El texto clásico que ejemplifica esta diferencia es la siguiente oración:

La botella entró a la cueva flotando.

En esta oración, el verbo *entró* denota la dirección de la acción mientras que la manera en que se realiza la acción es expresada a través de la frase *flotando*. La traducción literal en inglés sería la siguiente oración¹:

The bottle entered the cave floating.

Aunque dicha traducción no representa impedimento alguno para que toda persona cuya lengua materna es el inglés pueda captar el significado correcto de la oración, un sutil error es también percibido por el interlocutor anglosajón. La traducción correcta es la siguiente:

The bottle floated into the cave.

Finalmente, podemos observar en la punta de la pirámide de la Figura 2 el enfoque basado en *interlingua*. Dicha estrategia consiste en transformar el

texto a ser traducido en una representación independiente tanto del lenguaje original como del lenguaje final deseado; siendo la principal ventaja de este enfoque la drástica reducción en el número de traducciones requeridas para que un lenguaje pueda ser reproducido en múltiples lenguajes². Sin embargo, el principal reto en este formalismo de representación independiente sigue siendo la resolución de ambigüedad.

De esta manera, he tratado de describir en forma breve algunas de las dificultades que implica automatizar el proceso de traducción del lenguaje natural. Como se ha mencionado, diversas estrategias han sido definidas para la optimización del proceso de manera tal que hoy en día es evidente el gran avance que esta área de la tecnología ha experimentado si comparamos las traducciones de hoy con las de 20 o 15 años atrás. En mi opinión, hoy es posible hacer uso de herramientas que asisten (apoyan) el proceso de traducción de manera satisfactoria en tareas o actividades no críticas. Sin embargo, a pesar de los grandes avances, hay mucho camino por recorrer. Por ejemplo, la evolución del lenguaje implica retos, ya que constantemente se crean nuevos términos, nuevos nombres propios, y nuevos modismos. De esta manera, es difícil no coincidir con Kuhn [Geer 2005] cuando nos dice que el manejo de textos sofisticados es algo que todavía compete sólo al quehacer humano. Y como muestra de ello aquí dejamos para

concluir la traducción de los versos de Shakespeare en las siguientes líneas:

Y puesto que usted sabe usted no puede verse, tan pozo como por la reflexion, I, su cristal, la voluntad descubre modesto a se, el de se cuál usted con todo sabe no.

- William Shakespeare 

Notas

¹ NP denota un sustantivo, DT es un artículo, JJ es un adjetivo y finalmente NN representa un sustantivo singular.

² esta traducción se puede observar en Systran (<http://www.systransoft.com/>)

³ por ejemplo, en lugar de traducir el castellano al inglés, alemán, francés, etc. , se trabaja solamente en la transcripción del lenguaje en términos de la notación o formalismo independiente (interlingua) para de esta manera poder ser reproducido en cualesquier otro lenguaje.

Bibliografía

[Manning and Schutze 1999] Manning, C.D., and Schutze, H. Foundations of Statistical Natural Language Processing. The MIT Press, 1999.

[Jurafsky and Martin 2000] Jurafsky, D., and Martin, J.H. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall, Inc., 2000.

[Talmy 1985] Talmy, L. Lexicalization patterns: Semantic structure in lexical form. In Timothy Shopen (ed.), Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon, pp. 57-149. Cambridge, MA: Cambridge University Press.

[Geer 2005] Geer D. Statistical Machine Translation Gains Respect. In: IEEE Computer, volume 38, number 1, October 2005, pp. 18-21.



Universidad Iberoamericana, A.C.

Ciudad de México

Dirección de Servicios para la Formación Integral

**Prolongación Paseo de la Reforma 880
Lomas de Santa Fe, Deleg. Álvaro Obregón
México, D.F., C.P. 01210**

**Tel: 5950-4000 exts. 4919 o 7600. FAX: 5950-4331
didac@uia.mx**