

Traducción Automática: Dimensiones y Estrategias

Luciana BENOTTI Sandra ROGER
lubenotti@infovia.com.ar sroger@uncoma.edu.ar

Departamento de Informática y Estadística.
UNIVERSIDAD NACIONAL DEL COMAHUE.
Buenos Aires 1400, CP 8300, Neuquén, Argentina.
FAX: (54)(0299)4490313.

Palabras Claves: Procesamiento en Lenguaje Natural, Traducción Automática,
Arquitectura por Transferencia.

1. Introducción

Entre las dificultades históricas de la comunicación podemos mencionar a la barrera del lenguaje. Es por ello que la traducción ha cobrado gran interés aún antes del nacimiento de la informática. Los avances tecnológicos, han propiciado el desarrollo de herramientas para la automatización de este proceso.

La traducción automática¹ (en adelante: TA) es sin duda un desafío científico. Sin embargo la razón por la cual despierta tanto interés no es de índole científica, sino por una necesidad netamente práctica.

No existe en la bibliografía una definición formal de lo que significa TA, podemos encontrar definiciones recursivas y/o aproximaciones de lo que debería entenderse: *“es el proceso por el cual una máquina traduce un texto de una lengua a otra, subdividiendo la sintaxis, identificando las partes del discurso, intentando resolver eventuales ambigüedades y, por último, traduciendo los componentes y la estructura en la lengua de destino”* [Mer02].

El propósito de este trabajo es presentar la motivación de nuestra línea de investigación, metas y desarrollos futuros. Fundamentalmente, nuestro interés estará en poder capturar el potencial de las técnicas y paradigmas actualmente desarrollados en el área traducción automática, para luego realizar una comparación y evaluación de los mismos.

2. Traducción Automática

La comunicación existe cuando el receptor logra reconocer la intención del emisor y no sólo el significado literal del mensaje. En este sentido, la traducción automática posee un punto

¹Se ha optado por denominar Traducción Automática al concepto en inglés *Machine Translation*(MT), el cual se refiere a los sistemas informáticos que realizan traducciones con o sin ayuda humana.

débil, y no es una tarea fácil el lograr poder transmitir la intención o la doble intención al receptor.

Un traductor automático no tiene en cuenta las condiciones extralingüísticas (la intención, el contexto situacional, histórico o cultural), y, por lo tanto, no puede hacerse cargo de las ambigüedades del lenguaje. De la misma manera, carece de sentido común, no distingue matices, posee una flexibilidad limitada y no puede apoyarse en un nivel pragmático ni incorporar su experiencia ni su conocimiento del mundo para solucionar problemas, cosas que sí puede hacer el traductor humano.

2.1. Arquitecturas del sistema de traducción automática

En la traducción automática podemos contar con cinco tipos de conocimiento[AS94]: lenguaje origen(LO), lenguaje meta(LM), correspondencia entre términos del LO y el LM (diccionario), conocimiento sobre el dominio de lo que se traduce(sentido común) y conocimiento del entorno cultural, convenciones sociales, costumbres, etc. del LO y del LM.

1. *Arquitectura directa o transformer:*

La idea principal en esta arquitectura consiste en la traducción directa palabra por palabra del lenguaje origen al lenguaje meta usando para ello, un diccionario bilingüe. Luego de ésto, las palabras, son reordenadas según reglas específicas de la lengua meta en cuestión.

Ejemplo de este tipo de TA son: SYSTRAN (grupo GAT de la Universidad de Georgetown y CEE), SPANAM (utilizado por la Pan-American Health Organization), GTS, PC-TRANSLATOR, etc.

2. *Arquitectura de traducción por transferencia y el uso de patrones.*

En este tipo de traducción contamos con 3 fases: análisis, transferencia y síntesis. La primera fase consiste en procesar la lengua origen a una representación dependiente de ésta (entrada de la siguiente fase). La transferencia (dependiente del LO y del LM) se puede producir en varios niveles: léxico, sintáctico y semántico,

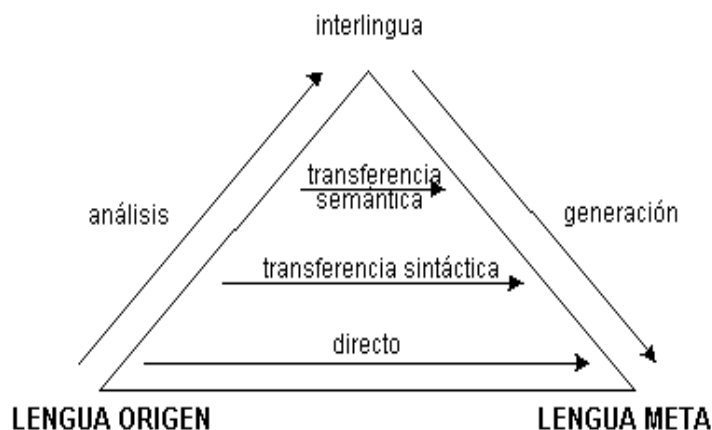
- Transferencia léxica: la búsqueda del término equivalente en el LM se realiza a partir de información contenida en el diccionario.
- Transferencia sintáctica: el árbol de análisis de la oración origen se transforma en un árbol de generación "equivalente" para la oración meta.
- Transferencia semántica: se transforman representaciones profundas (patrones semánticos).

Podemos mencionar los siguientes ejemplos en este tipo de traducción: METAL (Universidad de Texas, Siemens-Nixdorf), GETA (Universidad de Grenoble, Vauquois, 1985), TAUM-METEO/AVIATION (Universidad de Montreal, Isabelle, 1985), EUROTRA (CEE, Johnson, 1985), etc.

3. *Arquitectura de traducción por interlingua:*

En este tipo de arquitectura es más antiguo que el anterior y consiste en traducir el lenguaje origen a una representación independiente (interlingüa o pivot). Dicha representación es utilizada como entrada en la generación de la traducción al lenguaje meta.

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



En la figura anterior se distinguen los tres enfoques principales enunciados anteriormente, es decir, enfoques directos, los de interlingua y los de transferencia (normalmente sintáctica y en contados casos también semántica). Esta pirámide se basa en las diferencias de “longitudes relativas” de los tres componentes de la traducción: análisis, transferencia y síntesis o generación.

2.2. Dimensiones de la TA

Una dimensión es un factor que permite clasificar el ámbito de aplicación de la TA, acotando así su diseño con el objetivo de mejorar su efectividad. Se distingue entre seis dimensiones [Aba02]:

- Tipología del texto: Para afrontar las peculiaridades de los distintos textos se suele recurrir a los siguientes conceptos:
 - Registro (o jerga): Se relaciona con la variedad del lenguaje elegido para la comunicación. Se puede distinguir entre distintos lenguajes de especialidad definidos por una terminología o fraseología particular, en contraste con la lengua común cuyas expresiones y vocabulario son de uso genérico.
 - Estilo: Dentro de un mismo registro se pueden encontrar diferentes estilos (directo o indirecto, formal o informal, etc.). La noción de estilo distingue los textos por sus propiedades lingüísticas: longitud de oraciones, uso de conectores, etc.
 - Género: Distingue los textos según su pragmática (novela, poesía, teatro, ensayo, etc.).
 - Tipo: Diferencia textos dentro de un mismo género. Por ejemplo, dentro del género periodístico se puede distinguir entre: noticias, reportajes, columnas, anuncios, etc.
- Densidad y granularidad: La densidad léxica de un texto es inversamente proporcional a su granularidad. La granularidad de un texto se define en términos del tamaño de las unidades léxicas en las que se puede segmentar un texto. Los textos de granularidad más gruesa son aquellos que tienen una mayor combinación de palabras: fórmulas, frases hechas y términos compuestos. Un sistema de TA debe ser capaz de reconocer las unidades mayores primero, antes de tratar cada palabra por separado. Una lengua con una densidad mayor (como un lenguaje de especialidad) es más apropiado para la TA.

- Distancia lingüística y cultural: Cuando el LO y el LM son lenguas afines (por ejemplo, un par de lenguas latinas), es posible obtener resultados aceptables con un sistema simple de traducción palabra por palabra. Los problemas típicos de la traducción son minimizados por la proximidad lingüística. Por el contrario, la estrategia de traducción entre lenguajes más distantes debe ser necesariamente más elaborada. Al problema de la lingüística, hay que sumarle el de la distancia cultural; principalmente cuando alguno de los lenguajes intervinientes pertenece a una cultura con convenciones exóticas.
- Desarrollo lingüístico: El grado de disponibilidad de recursos lingüísticos como diccionarios, gramáticas y, sobre todo, corpus (tanto para el LO como para el LM), afectará la facilidad de desarrollo de una herramienta de TA.
- Medio y modo: Podemos distinguir tres modos de traducción según el medio involucrado:
 - Interpretación: Cuando el medio es oral el lenguaje es espontáneo, contextualizado, discontinuo y, muchas veces, agramatical.
 - Traducción: Un traductor de textos escritos generalmente se enfrenta a textos cuidadosamente redactados.
 - Localización: Tratamiento automático de textos en soporte electrónico.
- Fin: El factor más importante a tener en cuenta cuando se diseña un programa es definir claramente su objetivo. Las herramientas de TA no son la excepción a esta regla. Algunos propósitos comunes de los sistemas de TA son: dar una idea del contenido, informar y publicar.

3. Conclusiones y Trabajos Futuros

Se ha presentado una de las líneas de investigación que se está desarrollando dentro del marco de un proyecto de investigación de la Universidad Nacional del Comahue.

Los objetivos específicos de esta línea están encuadrados dentro de la TA, definidos dentro de las arquitecturas de transferencias expuestas en el punto 2 de la sección 2.1. Aunque aún estamos en sus primeras etapas, se ha comenzado a desarrollar un análisis de las herramientas a fin de poder establecer el grado de complejidad de los algoritmos, orientando nuestro desarrollo trabajos sobre textos informáticos, de estilo formal(en inglés) para poder realizar una traducción al idioma español.

Referencias

- [Aba02] Joseba Abaitua. Introducción a la traducción automática. http://sirio.deusto.es/abaitua/konzeptu/ta/mt10_es/ta10h-1es.htm. *Material preparado para los cursos de postgrado en Traducción por la UD y en Tecnología Lingüística por la UEU/EHU*, 2002.
- [AS94] D. Arnold, L. B. S. M. R. L. H. and L. Sadler. Machine translation: An introductory guide. *London: Blackwel I-NCC*, 1994.
- [Mer02] Pilar Hernández Mercedes. En torno a la traducción automática. <http://internet.cervantes.es/internetcentros/cultura/pdf/traducionautomatica.pdf>. *Cervantes N° 2.*, 2002.

[Tru00] Arturo Trujillo. Estrategias de traducción automática. *Quark. Ciencia, Medicina, Comunicación y Cultura*, 19:53 – 57, 2000.