

Traducción automática

Mateo Gonzalez Ocampo Juan Alejandro Uribe Ramirez

Junio 2020

Resumen

En el siguiente trabajo se realiza una revisión de algunos de los métodos que se han usado históricamente para resolver el problema de la traducción automática. Los métodos seleccionados fueron la traducción basada en reglas, basada en ejemplos y basada en redes neuronales. Para cada método se presenta una pequeña introducción histórica, un análisis de como funciona y aplicaciones destacadas. Finalmente se presentan algunas conclusiones acerca de lo encontrado en la literatura durante la investigación de estos métodos.

1. Introducción

La traducción automática se refiere al uso de computadores para realizar traducciones de un lenguaje SL (*Source Language*) a un lenguaje TL (*Target Language*).

primera parte de los 80. Estos sistemas usan un conjunto de reglas escritas a mano para la transformación de SL a TL, además de un diccionario para la traducción de palabras.

2. Métodos

2.1. Reglas

Los sistemas basados en reglas constituyen los primeros intentos usados para la creación de traductores automáticos y fueron ampliamente usados entre la década de las 60 y la

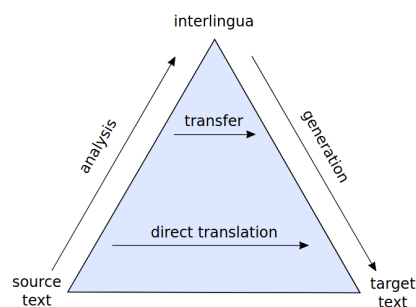


Figura 1: Triángulo de Vauquois.

Se pueden diferenciar los métodos usados en este tipo de sistemas en dos generaciones[1]: La primera, se caracteriza porque la transformación de SL a TL es directa, es decir, se traduce cada palabra por separado, usando un diccionario de SL a TL, y al resultado de esta transformación se le aplican un conjunto de reglas para hacer que la nueva oración sea coherente en TL, lo que puede involucrar una reorganización de las palabras del texto generado. A este tipo de métodos se les conoce como métodos de traducción directa.

La segunda generación de métodos, conocidos como de traducción indirecta, transforman inicialmente el texto en SL a una representación intermedia a partir de la cual se transforma a TL. Existen varios modelos para la generación de dicha representación, siendo los mas conocidos los métodos de interlingua y de transferencia [2][1][3]. En el primero se asume que la representación intermedia es unica para todos los lenguajes, por lo que solo es necesario definir reglas para transformar a este estado intermedio y desde este estado intermedio. Esto, basado en las ideas planteadas por Descartes en el siglo XVII acerca de la existencia de un lenguaje universal. En el segundo método cada lenguaje tiene su propia representación intermedia, lo que implica que además de las reglas relacionadas con la representación intermedia, es nece-

sario un conjunto de reglas adicional para transformar de una representación a otra. En la figura 1 se muestra el triangulo de Vauquois, el cual sirve para entender los diferentes niveles de análisis entre los métodos mencionados anteriormente.

SYSTRAN, APERTIUM y Gram-Trans son algunos ejemplos de aplicaciones que usan sistemas basados en reglas.

2.2. Ejemplos

Los sistemas basados en ejemplos surgieron durante la decada de los 80, debido principalmente al trabajo de Nagao[4], el cual planteaba una nueva forma de traducción automática basada en la idea de que el ser humano no traduce partiendo de análisis lingüísticos complejos, sino descomponiendo el texto a traducir en fragmentos los cuales son luego traducidos y reorganizados para formar el nuevo texto. De manera similar a los sistemas basados en reglas, los sistemas basados en ejemplos hacen uso de un diccionario de SL a TL junto con una base de datos de ejemplos, la cual se usa para como base para la traducción (En contraste con las reglas usadas en los sistemas basados en reglas). También suele usarse un Tesauro como parte de la base de datos. Un ejemplo esta constituido por dos textos en dos lenguajes, siendo ambos textos traducciones de si mismos[5].

Existen tres etapas en el proceso de traducción mediante ejemplos [4][6]:

1. Como primer paso, se realiza un análisis del texto, para identificar los fragmentos que mejor se relacionen con los textos de la base de datos.
2. Luego, usando los fragmentos obtenidos, se extraen de la base de datos los ejemplos que mas se asemejen a los fragmentos. Los ejemplos se seleccionan con base en alguna métrica que toma ambos textos y asigna un grado de similaridad entre ellos.
3. Finalmente, con base en los ejemplos obtenidos de la base de datos, se genera el texto traducido.

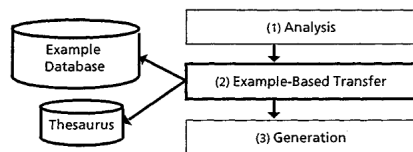


Figura 2: Configuración de un sistema basado en ejemplos Tomado de [6].

Los sistemas basados en ejemplos resuelven algunos de los problemas presentes en los sistemas basados en reglas[6], principalmente la dificultad y complejidad de añadir nuevas reglas a un sistema ya establecido, dado que

pueden haber conflictos con las reglas previamente establecidas. Cunei y CMU-EBMT son probablemente de los sistemas basados en ejemplos mas reconocidos.

2.3. Redes neuronales

Los fundamentos matematicos de las redes neuronales...

Referencias

- [1] W. John Hutchins. Machine translation: A brief history. In E.F.K. KOERNER and R.E. ASHER, editors, *Concise History of the Language Sciences*, pages 431 – 445. Pergamon, Amsterdam, 1995.
- [2] Musatafa Albadr, S. Tiun, and Fahad Al-Dhief. Evaluation of machine translation systems and related procedures. *ARPJ Journal of Engineering and Applied Sciences*, 13:3961–3972, 06 2018.
- [3] Cheragui Mohamed Amine. Theoretical overview of machine translation. *CEUR Workshop Proceedings*, 867:160–169, 01 2012.
- [4] Makoto Nagao. A framework of a mechanical translation between japanese and english by analogy

- principle. In *Proc. of the International NATO Symposium on Artificial and Human Intelligence*, page 173–180, USA, 1984. Elsevier North-Holland, Inc.
- [5] Haihua Pan. Example-based machine translation : A new paradigm. 2002.
- [6] Eiichiro Sumita, Hitoshi Iida, and Hideo Kohyama. Translating with examples: A new approach to machine translation. In *Proceedings Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, 1990.