

Una contribución al procesamiento automático de la sinonimia utilizando Prolog

Santiago Fernández Lanza

Universidad de Santiago de Compostela

Dpto. de Lógica y Filosofía Moral, Facultad de Filosofía, Campus Sur s/n
sflanza@usc.es

Resumen: Se presenta un resumen de la tesis doctoral “Una contribución al procesamiento automático de la sinonimia utilizando Prolog”, defendida el 16 de Julio de 2001 en la Universidad de Santiago de Compostela y dirigida por el doctor Alejandro Sobrino Cerdeiriña.

Palabras clave: Sinonimia, Procesamiento del lenguaje natural, Diccionario de sinónimos.

Abstract: We present a summary of the thesis “A contribution to automatic processing of synonymy using Prolog”, put forward in 16th July, 2001 by Santiago Fernández Lanza at University of Santiago de Compostela and directed by doctor Alejandro Sobrino Cerdeiriña.

Keywords: Synonymy, Natural Language Processing, Dictionary of Synonyms

Se parte de una presentación del lenguaje de programación Prolog, que constituye la herramienta utilizada en toda la tesis para el tratamiento computacional de la sinonimia. En dicha presentación se describen las estrategias comunes de la programación en dicho lenguaje, entre las que se puede destacar el procesamiento de listas y gramáticas, aspectos que ponen de relieve su adecuación para el procesamiento del lenguaje natural.

El paso siguiente fue realizar un recorrido por distintos modos en cómo fue tratada históricamente la sinonimia. Entre las diversas concepciones de la sinonimia podían vislumbrarse, de modo general, dos tendencias. La que persigue su definición y la que trata de describirla tal y como es usada en la práctica. El tratamiento de la sinonimia de este trabajo está más próximo a la segunda tendencia. Las ideas que subyacen son, por una parte, que un diccionario de sinónimos puede constituir una muestra razonablemente fiel del uso de la sinonimia en la práctica y que dos expresiones son sinónimas si su significado es el mismo o parecido. Se ha perseguido la descripción del comportamiento de la sinonimia dentro un diccionario de sinónimos concreto (Blecua Perdices, J. M. (1997) (dir.) *Diccionario avanzado de sinónimos y antónimos de la lengua española*, Vox.). Esto ha tenido como consecuencia la elaboración de un diccionario electrónico de sinónimos. El comportamiento

de la sinonimia en el diccionario refleja su carácter gradual, lo que hace que la similaridad sea un modelo más adecuado para tratarla que la equivalencia, alternativa que históricamente se había pensado como adecuada para este propósito en algunos ámbitos.

Posteriormente se estudian, mediante la utilización de las herramientas de la programación lógica, algunos problemas importantes dentro del marco del procesamiento del lenguaje natural como la traducción automática, la sinonimia entre palabras, la sinonimia entre oraciones, la desambiguación, etc. En la última parte, se realiza una descripción de un diccionario de sinónimos impreso que servirá de base de datos para el diccionario electrónico de sinónimos cuyo diseño es la aportación aplicada fundamental de la tesis. Finalmente, se analiza la utilidad de dicho diccionario electrónico en el procesamiento del lenguaje natural, en lexicografía y en recuperación de información, así como se compara su aportación con respecto a WordNet.

Partiendo de una concepción gradual de la sinonimia y de la utilización de un diccionario de sinónimos (impreso) como base de datos empírica entendida como una muestra razonablemente fiel del comportamiento de la sinonimia en la práctica, en esta tesis, se propone la elaboración de un diccionario electrónico de sinónimos que calcula el grado

de sinonimia de dos de sus entradas. El cálculo del grado permite mejorar las prestaciones del diccionario electrónico frente al diccionario impreso. Las mejoras obtenidas son las siguientes:

1. El diccionario electrónico detecta, de forma automática, las acepciones y homógrafas de los sinónimos que proporciona como respuesta.
2. Ordena los sinónimos según su grado de sinonimia.
3. Obtiene más sinónimos que los que aparecen en el diccionario impreso bajo el criterio de que dos palabras son sinónimas si su grado de sinonimia es distinto de 0.
4. Permite el establecimiento de umbrales más abajo de los cuales dos palabras no serán consideradas sinónimas.

Todo ello se ha realizado utilizando herramientas y técnicas de distintos ámbitos. Por una parte, se han analizado distintas propuestas teóricas de estudio de la sinonimia y se ha elegido aquella que la concibe como una relación aproximada o de grado por responder de un modo más adecuado a nuestras intuiciones de cómo debería comportarse la sinonimia en la práctica. Se ha utilizado el lenguaje de programación Prolog para implementar toda la propuesta (Visual Prolog para el interfaz gráfico de usuario) y se han usado medidas de similaridad para calcular el grado de sinonimia de dos entradas del diccionario.

La tesis constituye una contribución, por una parte, al procesamiento del lenguaje natural, ya que consiste en la automatización de un aspecto importante del lenguaje, y por otra, a la lexicografía, ya sea ésta computacional o no, puesto que el resultado es un diccionario electrónico de sinónimos que podría ser utilizado por cualquier usuario de la lengua y que además puede servir de herramienta para conseguir mejores diccionarios de sinónimos impresos. También puede ser útil como campo de pruebas para comprobar la validez y potencia explicativa de muchas de las propuestas proporcionadas por los estudios teóricos sobre sinonimia.

A todo este trabajo subyace la idea de que el tratamiento computacional de cuestiones relativas al lenguaje natural es una tarea de gran complejidad en la que no se pueden establecer reglas excesivamente estrictas y de la que no cabe esperar la ausencia de problemas ante cualquier resultado. En ningún momento se ha

pretendido afirmar que el diccionario electrónico de sinónimos presentado realice una descripción infalible del fenómeno de la sinonimia, simplemente constituye una herramienta relativamente fiable que podría ser útil en campos como el procesamiento del lenguaje natural, la lexicografía o la recuperación de información y que entre sus usuarios podrían verse beneficiados los lexicógrafos, informáticos, teóricos de la sinonimia y usuarios de la lengua en general. Se ha procurado ser lo más cuidadoso posible respecto a la fiabilidad de los resultados siempre que la complejidad del lenguaje lo permitiera.