



---

# WORDNET

---

De Luna Ocampo Yanina, 4CDM1



16 DE FEBRERO DE 2022  
INSTITUTO POLITÉCNICO NACIONAL  
ESCUELA SUPERIOR DE CÓMPUTO (ESCOM)

## WordNet

Es una base de datos léxico-conceptual del inglés estructurada en forma de red semántica. El proyecto, desde el principio estuvo respaldada por instituciones gubernamentales como privadas norteamericanas. En 1985 se propuso crear una herramienta que permitiera moverse por la estructura de un diccionario conceptualmente y no alfabéticamente.

Este, divide el léxico en 4 categorías: nombres, adjetivos, verbos y adverbios, estos se agrupan en conjuntos de sinónimos cognitivos denominados “synsets” y cada uno expresa un concepto distinto. Las preposiciones y conjunciones no se presentan porque no forman parte del léxico-semántico, solo del semántico. La estructura básica de esta, es de forma que las relaciones se establecen fundamentalmente entre conceptos, no entre palabras, esto hace que la sinonimia junto con la taxonomía, sea la relación léxica primordial, más no la única.

Se parece superficialmente a un diccionario de sinónimos, ya que agrupa palabras en función de su significado, pero existen distinciones importantes:

1. No solo interconecta formas de palabras, sino sentidos específicos de palabras.
2. Etiqueta las relaciones semánticas entre palabras, mientras que en un diccionario de sinónimos no se sigue ningún patrón explícito más que la similitud del significado.

La matriz daría acceso a la información de dos maneras, se podría acceder a una columna e ir bajando hasta el final, de esta forma obtendríamos todos los sentidos que una palabra puede tener en diferentes contextos. También podríamos acceder por una fila y seguirla hasta el final, de este modo obtendríamos todas las maneras posibles de expresar un determinado concepto. Así, la matriz de vocabulario contempla dos de los principales problemas de la semántica léxica, la polisemia y la sinonimia, como dos aspectos complementarios de una única estructura abstracta.

Significados Léxicos	Formas Léxicas				
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	.....	F <sub>n</sub>
M <sub>1</sub>	E <sub>1,1</sub>	E <sub>1,2</sub>			
M <sub>2</sub>		E <sub>2,2</sub>			
M <sub>3</sub>			E <sub>3,3</sub>		
:				..	
:				.....	E <sub>m,n</sub>
M <sub>n</sub>					

Tabla 2.1 Matriz de vocabulario de WordNet

En los sustantivos, este contiene aproximadamente 57,000 formas nominales organizadas en unos 48,000 synsets. Las definiciones están organizadas en jerarquías semánticas construidas con base a los términos superordinados que aparecen en las definiciones de los sustantivos, junto con los rasgos distintivos que diferencian un sustantivo de su hiperónimo. Se duplica por medio de punteros entre los “synsets”, la mayoría de los casos no contiene más de doce niveles de organización. Los rasgos distintivos se introducen de manera que se crea un sistema de herencia léxica en el que cada palabra hereda los rasgos distintivos de su término superordinado, creándose una jerarquía que va

## WordNet

desde los términos inferiores o subordinados de los niveles inferiores hacia un número de términos genéricos en la parte superior de la jerarquía.

Un ejemplo de lo explicado anteriormente sería:

oak @ -> tree @ -> plant @ -> organism

El @-> representa una relación semántica transitiva y asimétrica de los tipos “es un” o “es un tipo de”. En el sistema de herencia léxica los hipónimos están conectados a sus hiperónimos y viceversa. Poniendo un ejemplo, la entrada de tree contiene una referencia hacia la entrada de la forma plant, de este modo el “synset” de tree sería algo parecido a:

{tree, plant, @ conifer, ~...}

Donde los puntos suspensivos representan los demás posibles punteros a otros hipónimos. Los sustantivos no están estructurados en torno a una jerarquía única que contenga un término superordinado general del tipo {entity} que englobe todos los demás. Los consideran componentes semánticos primitivos porque a los que adscribe un campo léxico y los lexemas que en él se contiene.

Los 25 componentes semánticos primitivos:

{act, action, activity}	{natural object}
{animal, fauna}	{natural phenomenon}
{artifact}	{person, human being}
{attribute, property}	{plant, flora}
{body, corpus}	{possession}
{cognition, knowledge}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation}
{feeling, emotion}	{shape}
{food}	{state, condition}
{group, collection}	{substance}
{location, place}	{time}
{motive}	

Con los adjetivos, es diferente, se dividen en dos clases principales: descriptivos y relacionales, contiene aproximadamente 19,000 formas adjetivales organizadas en unos 10,000 “synsets”. Además de estos, contiene un grupo cerrado de adjetivos que se consideran de modificación referencia. Los investigadores consideran como adjetivos descriptivos aquellos que adscriben a los sustantivos valores de atributos bipolares, por tanto, están organizados en base binaria o sinonimia. Los relacionales son aquellos que significan algo parecido a “relativo con” el sustantivo al que modifican.

La antonimia es la relación semántica básica de los adjetivos descriptivos, la relación esta estructurada de forma que la función de estos adjetivos es expresar los valores opuestos de atributos que en la mayoría de los casos son bipolares. Los problemas se plantean, por ejemplo, en el caso de dos adjetivos que poseen un significado muy parecido pero que tienen antónimos diferentes o en aquellos casos en los que el adjetivo no tiene ningún antónimo, y no se le puede asignar el de otro adjetivo similar.

## WordNet

En cambio, los verbos, contiene más de 21,000 verbos y aproximadamente significados léxicos o “synsets”. Estos están divididos en 15 archivos diferentes, con base a criterios semánticos, todos denotan acciones o eventos, a excepción de los que no forman un dominio semántico ni comparten otra propiedad semántica que no sea la de referirse a estados.

Las relaciones que organizan los verbos, se pegan en torno al principio de implicación léxica, esto se refiere a la relación que existe entre dos verbos cuando la oración *AlguienV1* implica lógicamente la oración *AlguienV2*. Por ejemplo, *snore* implica *sleeping* porque la oración “He is snoring” implica “he is sleeping”. Esta relación debe ser solamente unilateral, si no se convierten en sinónimos. En el caso de los verbos se denomina “troponimia” ya que la relación semántica de hiponimia es considerada entre sustantivos, ya que se considera que las distinciones de “modo” son las más importantes a la hora de diferenciar un hipónimo verbal de su hiperónimo.

Las otras dos relaciones de implicación considerada son las de oposición y de relación causal. La relación de oposición entre verbos es bastante compleja, ya que está basada en muchos casos en un proceso morfológico que se aplica a uno de los miembros de la oposición. Por otro lado, la causativa se compone de dos conceptos, uno causativo y otro resultativo, incluye aquellos pares causativo-resultativos que están lexicalizados, de modo que los sinónimos heredan la relación causativa, indicando que la relación se sostiene entre el concepto y no entre las palabras.

Su finalidad principal era convertirse en un reflejo computacional de la memoria léxica y no la representación del conocimiento léxico, por lo que incluye muchas de la información que un hablante nativo posee acerca de las propiedades semánticas y sintácticas de los verbos. El resultado de este es impresionante debido a la cantidad de información que contiene, toda esta fue incluida manualmente por el grupo de lexicógrafos del proyecto.

La información que contienen los diferentes ficheros es ciertamente valiosa, lo que hace que futuros proyectos que tomen WordNet como base, puedan realmente sacar partido a ella si se integra en un sistema computacional apropiado. Este, también es un recurso de traducción muy importante, ya que representa uno de los diccionarios multilingües más grandes que existen en la actualidad por la cantidad de idiomas y la cantidad de palabras y conceptos cubiertos. Incluso lo utiliza Google Translate como parte del proceso de traducción entre idiomas que poseen WordNets.

En la actualidad existen versiones en varias etapas de desarrollo para una enorme variedad de idiomas, esto incluye que no solo lo más extendidos y conocidos como el francés, alemán, español chino, sino también los menos conocidos y minoritarios, como el croata, catalán, gallego y galés.

### Referencias:

- Moreno Ortiz, A. 2.4.2 *WordNet*. WordNet. <http://elies.rediris.es/elies9/2-4-2.htm>
- *WordNet: una taxonomía léxica de palabras en inglés*. (2021, 7 febrero). ICHI.PRO. <https://ichi.pro/es/wordnet-una-taxonomia-lexica-de-palabras-en-ingles-74164241289215>
- <http://www.it.uc3m.es/jvillena/irc/practicas/estudios/WordNet.pdf>