



Universidad Veracruzana

Seminario de semántica

Doctorado en estudios del
lenguaje y lingüística aplicada

César Aguilar

Martes 5 de marzo de 2024

Síntesis de la clase anterior (1)

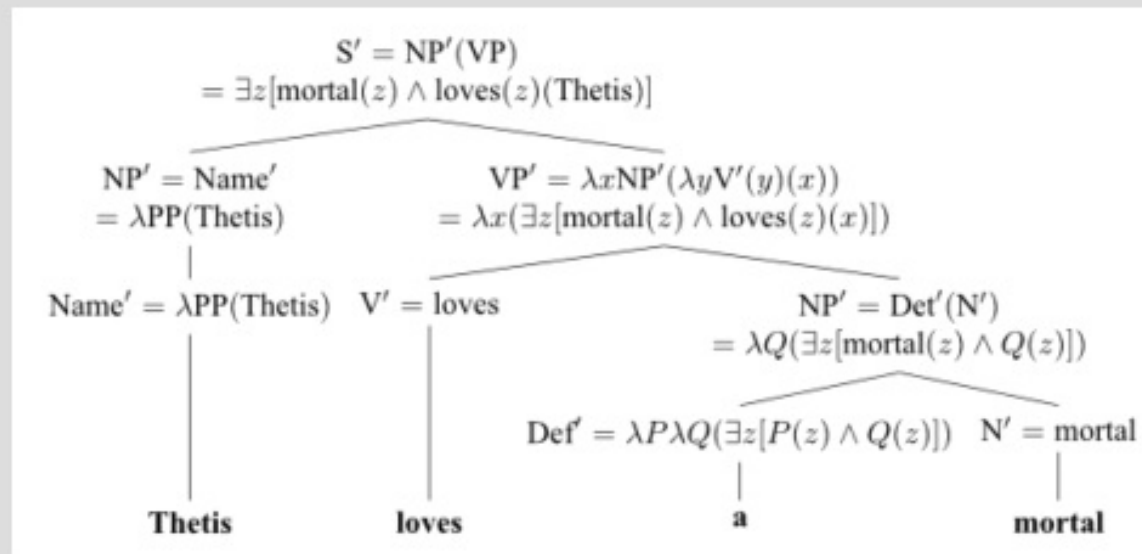
En la pasada unidad, hicimos un rápido recorrido por los orígenes de la semántica, revisando algunos de sus conceptos básicos, hasta concluir en la discusión que tuvieron Gottlob Frege y Bertrand Russell sobre las nociones de referencia y sentido.



Síntesis de la clase anterior (2)

La discusión y asimilación de las ideas de ambos autores ayudó a delinear —al menos— dos líneas de investigación en semántica:

1. La **semántica léxica**, que se enfoca en el estudio del significado de las palabras, así como las relaciones que se dan entre dichos significados.
2. La **semántica composicional**, que se concentra en el estudio de los fenómenos que subyacen en la construcción de significados insertos en frases y oraciones.



La semántica léxica (1)

Como ya dijimos, esta línea se enfoca en el significado de las palabras. Ésa sería la respuesta rápida. Empero, si atendemos a la explicación que nos ofrece la Dra. Escandell, la cuestión es más compleja:



Ma. Victoria
Escandell Vidal



La semántica léxica estudia el significado de las palabras (lexemas) que no se reduce a una frase o pequeña descripción, tampoco a un conjunto de rasgos que describen y componen algo, es una RED de significados. Cada significado se vincula con otros.

La semántica léxica (2)

Hasta aquí, todo bien... Sin embargo, esta idea que aparentemente sólo le interesa a los lingüistas en años recientes la han abordado gente que trabaja temas de inteligencia artificial, y se han dado cuenta de su potencial tecnológico.

Y para muestra, un botón: ¿quién sabe lo que es **IBM Watson**?



Todo está conectado (1)

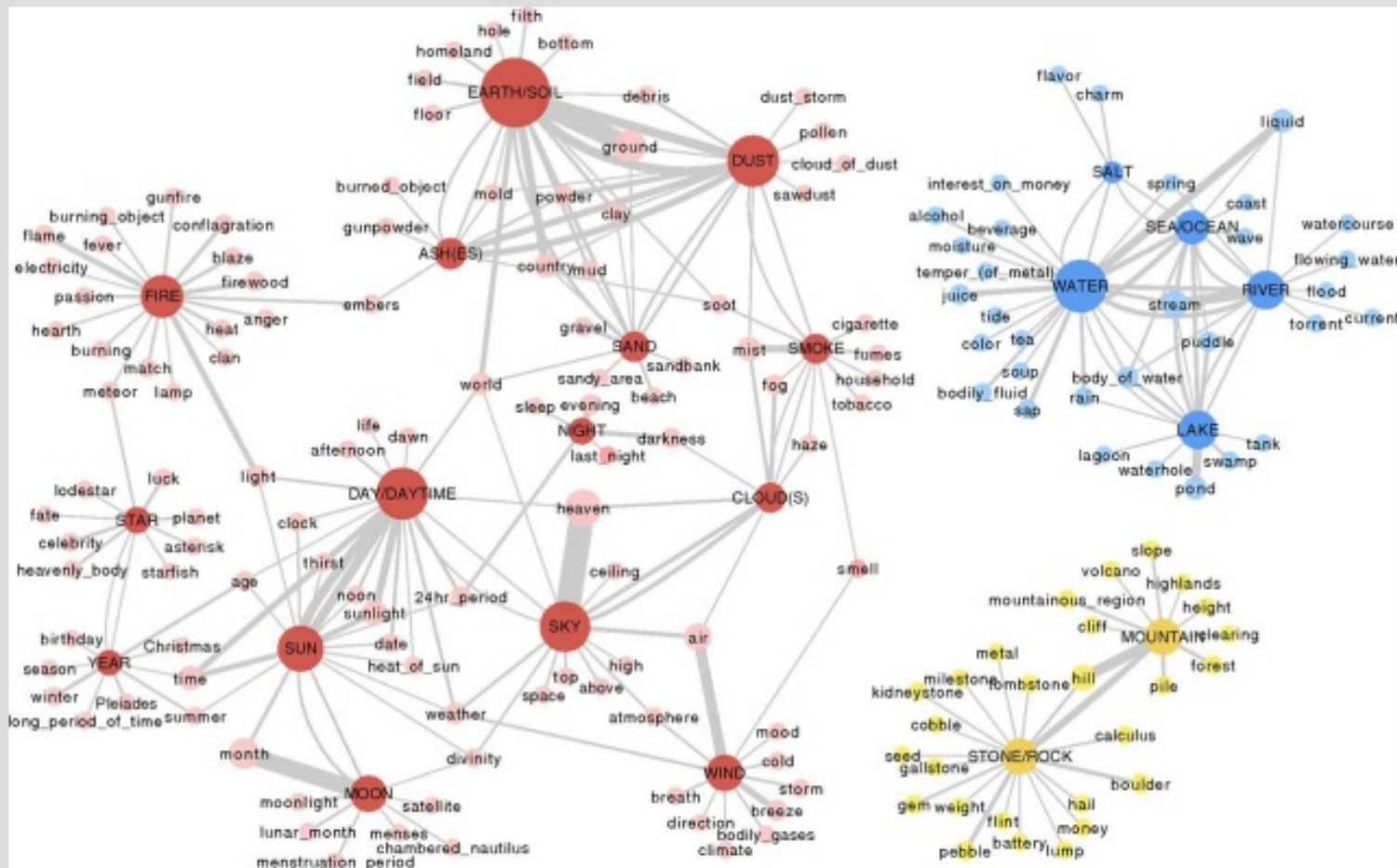
Seguro que muchos han pensado alguna vez que sus acciones están conectadas con otras, de tal suerte que si hacen algo, quizá eso tenga consecuencias a corto, mediano o largo plazo.

Hay películas y series que abordan este tema, p.e.:



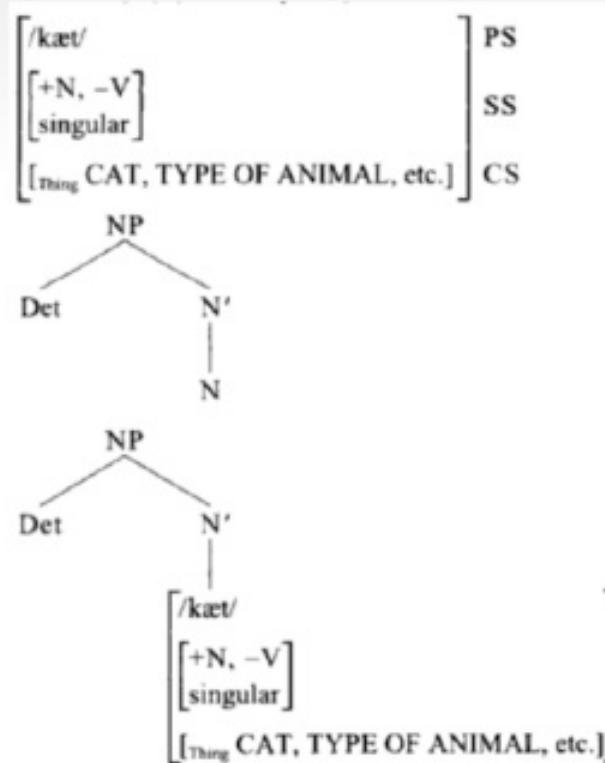
Todo está conectado (2)

Volviendo con el concepto de **red léxica**, esto no es una mera idea o una sensación, sino una realidad que la inteligencia artificial aprovecha todo lo más que puede:



Todo está conectado (3)

En lingüística, esta idea de las redes de palabras la hemos desarrollado mucho, sobre todo a partir de los años 60. De hecho, fue Noam Chomsky uno de los primeros en explicarnos cómo funciona:

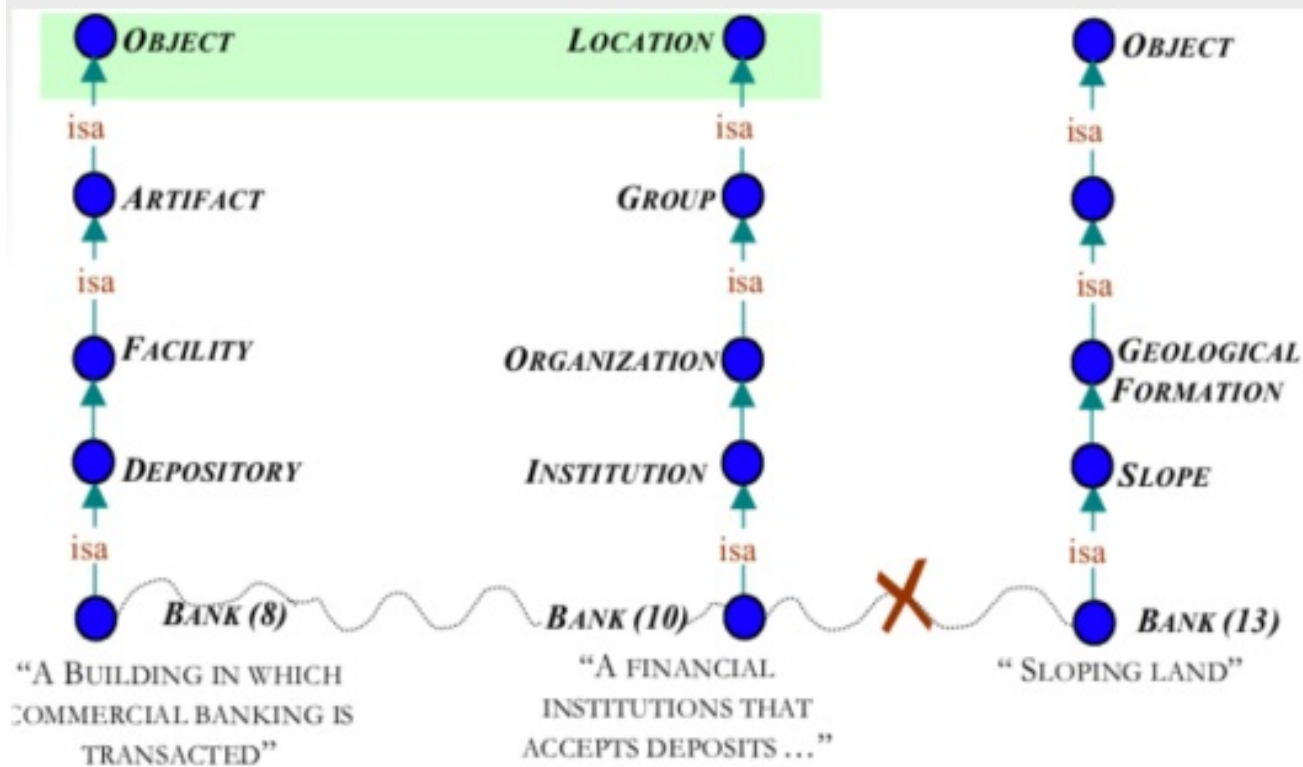


¿Qué es la palabra *cat*? Aparte de una secuencia de fonemas que asociamos con el inglés, aparte de ser un nombre, y aparte de ser el núcleo de frase nominal, es también la representación lingüística de **un tipo de animal**, del cual podemos decir que es *un mamífero de la familia de los felinos, el cual es domesticable*, etc.

Supongamos que encapsulamos toda esta información dentro de la palabra *cat*. ¿Qué obtenemos? Una unidad léxica, que junto otras similares conforman un conjunto que llamamos **lexicón**.

Todo está conectado (4)

Este problema de inferir conceptos es muy interesante para la semántica léxica computacional: más allá de la cuestión de “encapsular” toda esta información dentro del lexicón, el reto es: ¿cómo activar esta información latente cuando es necesaria?



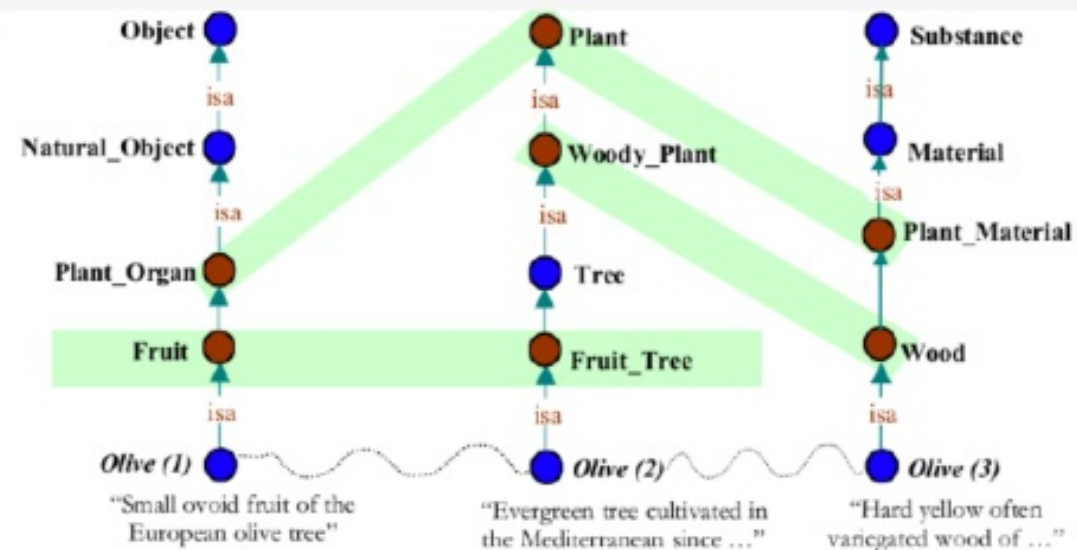
Un nuevo ejemplo: el término *banco*, aparentemente tiene un referente claro. Empero, ¿por qué sabemos que son cosas distintas el *banco de la cocina*, el *Bando del Bajío*, el *banco de arena*, y el *banco de peces*?

Todo está conectado (5)

Lo anterior nos lleva entonces a lo siguiente: suponiendo que nuestro cebrero, al momento de procesar palabras y conceptos, no tiene problemas para almacenar estas relaciones (para eso tenemos el lexicón), lo que queremos es entender cómo se dan esas relaciones.

En otras palabras: nuestro problema no es saber cuánta capacidad tiene la mente humana para guardar datos, sino explicar cómo se organizan esos datos.

Si hablamos de *organización*, ello implica pensar en cosas como: *taxonomías, vocabularios, terminologías, redes, ontologías, etc.*



WordNet (1)

El problema de saber cómo es que el cerebro asocia y organiza conceptos fue el punto de partida para un proyecto de investigación, realizado en Princeton University en 1985, el cual fue coordinado por un psicolingüista llamado George Armitage Miller (1920).



A grandes rasgos, Miller estaba interesado en modelar y reproducir a través de sistemas computacionales la forma en como la mente procesa y organiza conceptos referidos por palabras. Retomando la idea de Chomsky, podemos decir que Miller desarrolló un lexicón electrónico, organizado a partir de relaciones léxicas.

Con el paso del tiempo, lo que Miller y sus colegas lograron concretar fue una valiosa herramienta muy usada en PLN para realizar muchas tareas. Claro está, entre ellas destaca la búsqueda y clasificación de información semántica y conceptual en documentos.

PRINCETON UNIVERSITY

WordNet
A lexical database for English

WordNet (2)

Como todo buen lexicón, WordNet es una enorme base de datos que agrupa y clasifica palabras, con base en 4 categorías sintácticas: nombres, verbos, adjetivos y adverbios.

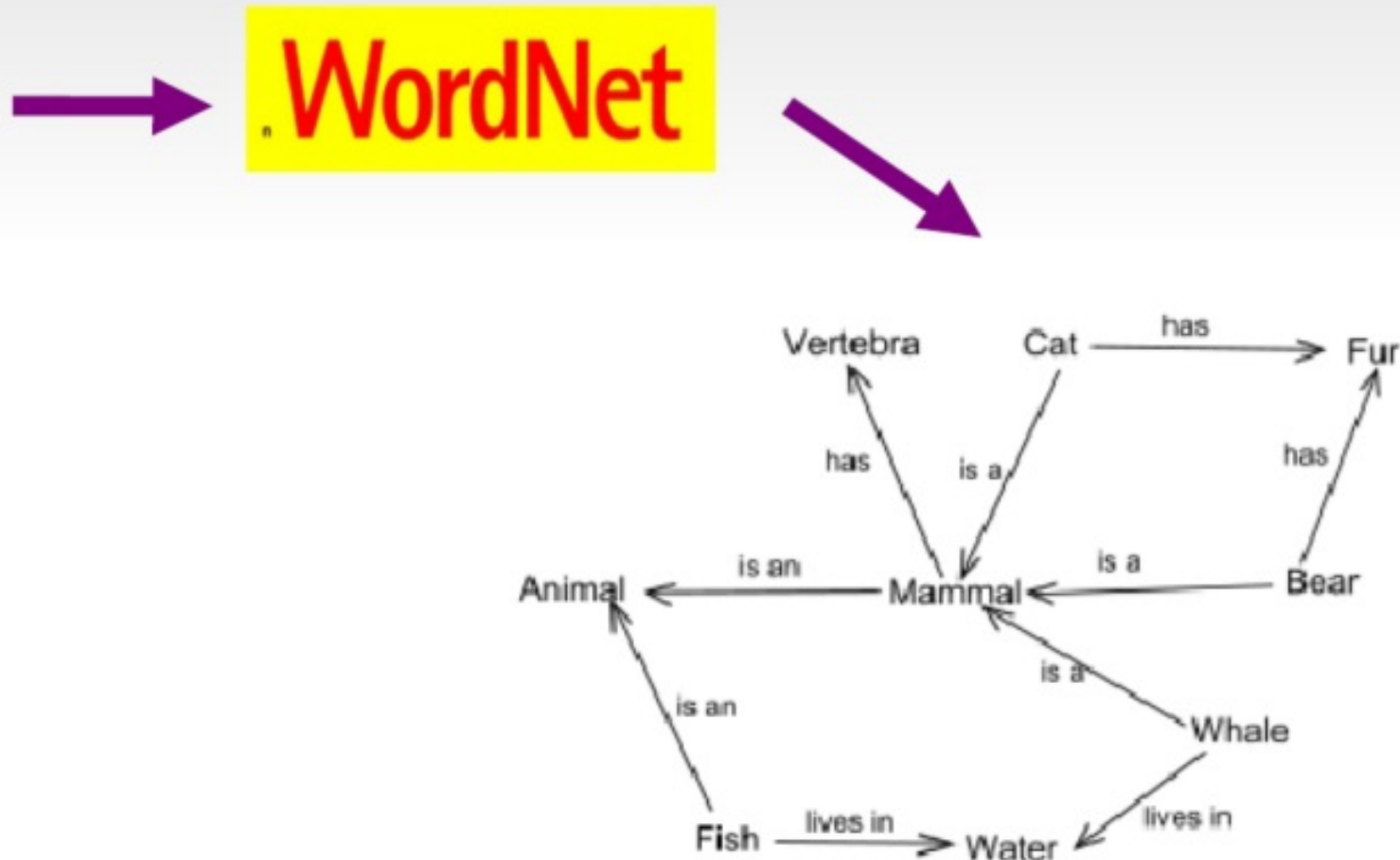
Con base en estas 4 categorías, WordNet determina una serie de relaciones entre palabras a partir de su contenido léxico. A este contenido que WordNet liga a cada palabra se le denomina *synset*, y equivale a un descriptor *s* que permite identificar y jerarquizar a una palabra dentro de una red léxica.

```
dog, domestic dog, Canis familiaris
=> canine, canid
    => carnivore
        => placental, placental mammal, eutherian, eutherian mammal
            => mammal
                => vertebrate, craniate
                    => chordate
                        => animal, animate being, beast, brute, creature, fauna
                            => ...
```

WordNet (3)

En síntesis, a partir del uso de relaciones léxicas WordNet logra crear una red de conceptos, organizada de manera jerárquica.

- Holonimia
- Polisemia
- Sinonimia
- Antonimia
- Hipernimia
- Hiponimia
- Meronimia



WordNet (4)

Veamos ahora qué podemos hacer con **WordNet**, accediendo a su sitio electrónico:



WordNet Research
@ PRINCETON

People Current Projects Publications Downloads Related Research

The WordNet group at Princeton is a multidisciplinary group composed of linguists, computer scientists, and psychologists. Our current focus is on leveraging human expertise via machine learning to improve WordNet as a linguistic resource.

Downloads of WordNet and more information about WordNet is available [at the Princeton WordNet site](#).

research sponsored by

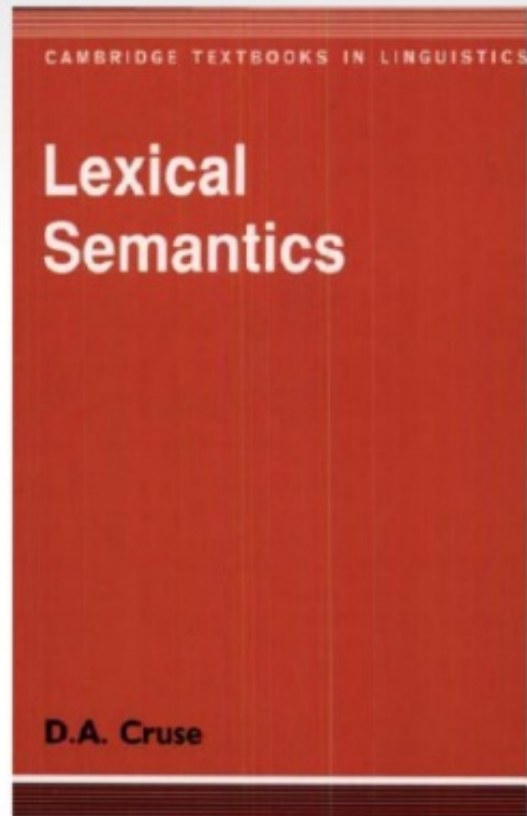
 

Questions or comments? E-mail the [reluctant webmaster](#).

<https://wordnet.princeton.edu/>

Relaciones léxicas (1)

Cuando hablamos de relaciones entre conceptos, estamos aludiendo a asociaciones que han sido analizadas y explicadas desde la lexicografía y la semántica. En concreto, hablamos de **relaciones léxicas**.



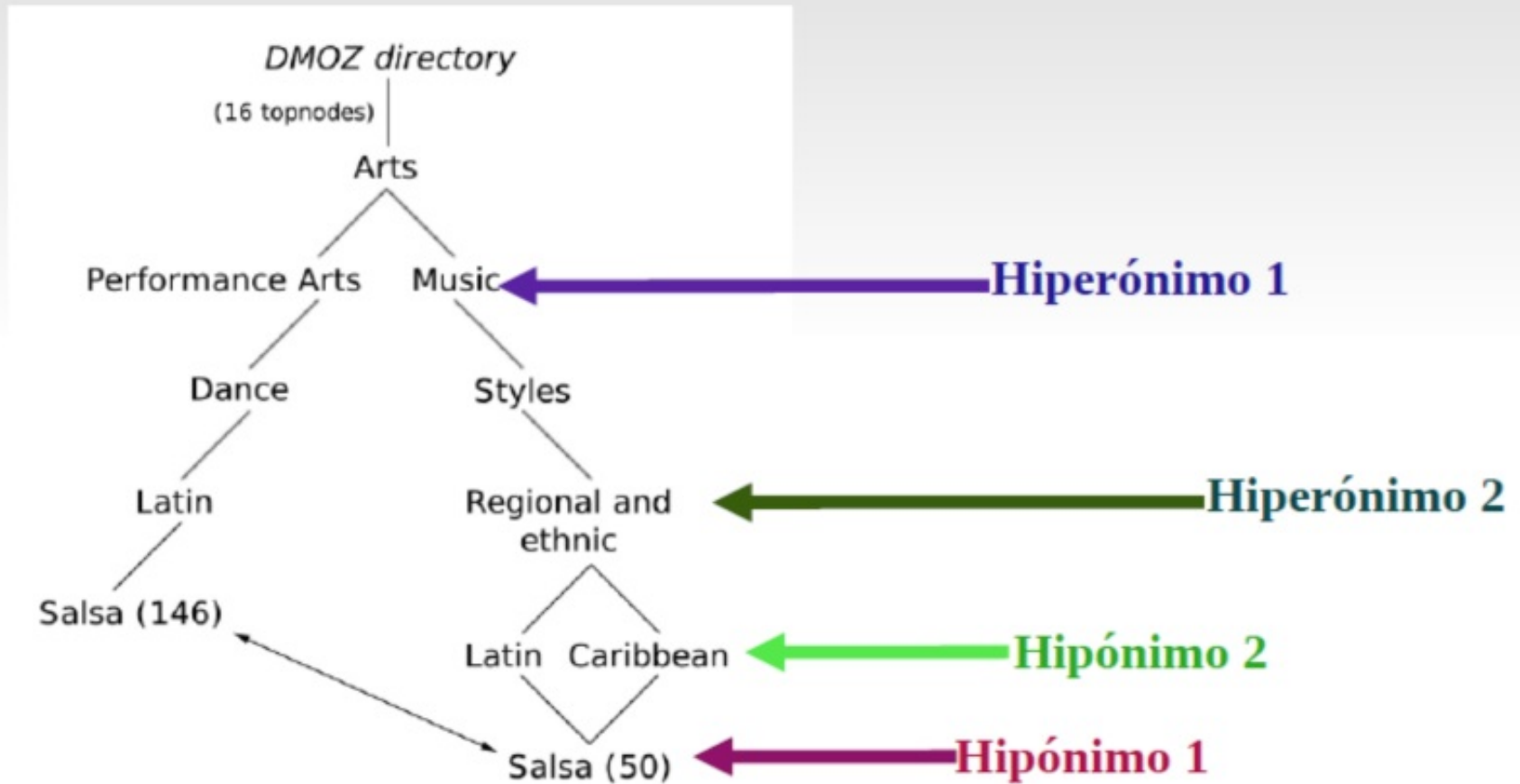
Sigamos manejando el supuesto de que las palabras de una lengua natural (por lo menos, la mayoría de ellas) se refieren a conceptos que describen objetos del mundo.

Considerando que estos conceptos están organizados de algún modo, podemos deducir que un tipo de relación es justo pensar que algunos son “superiores” a otros: p. e., una *maquina* conceptualmente es más genérica que una *computadora*, un *scanner*, una *cámara fotográfica*, etc.

A esta relación entre superiores e inferiores se le conoce como *hiperonimia* e *hiponimia*.

Relaciones léxicas (2)

Veamos un ejemplo simple: ¿qué es la *salsa*?



Relaciones léxicas (3)

La relación léxica hiponimia/hiperonimia ha sido una de las más recurridas para generar esta clase de tesauros, ya que conceptualmente permiten establecer jerarquías claras entre los elementos que la conforman.

Un patrón típico en esta clase de relaciones es el operador IS-A.

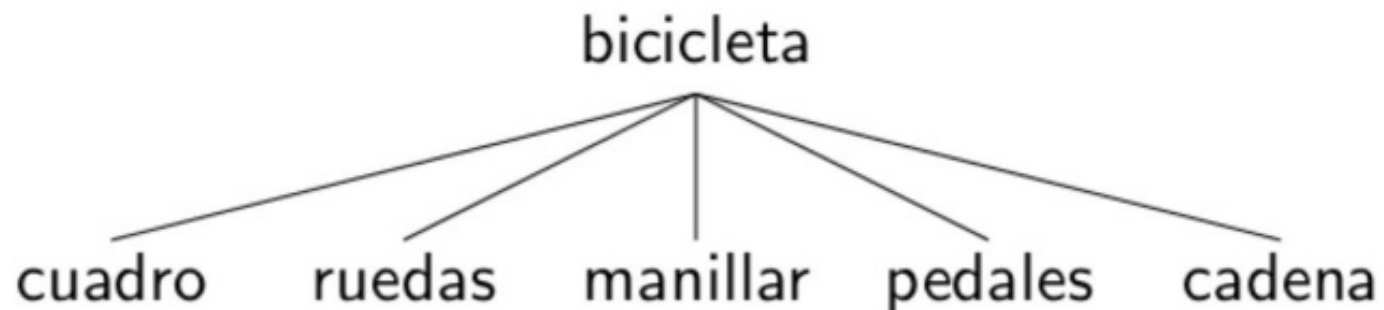
IS-A lo único que indica es que un objeto es un tipo/una clase/una categoría de otro.

Puede tener variantes, por ejemplo, la **meronimia**, que indica que un objeto pertenece a un conjunto, o a otro objeto mayor: *el mouse es una parte de una computadora.*

```
book is_formal phys-object,  
book is_telic literature,  
book is_agent literature,  
dictionary is_formal book,  
dictionary is_telic reference,  
dictionary is_agent compiled-material,  
play is_agent literature,  
play is_telic book.
```

Relaciones léxicas (4)

Veamos otro ejemplo para entender cómo funciona la meronimia, respondiendo a la pregunta: *¿qué es una bicicleta?*



Relaciones léxicas (5)

Otra relación a considerar es la **sinonimia**, la cual consiste en el vínculo que establecen dos o más palabras con significados muy relacionados que (a menudo, aunque no siempre) pueden intercambiarse en una oración, p. e.:



Relaciones léxicas (6)

Las palabras sinónimas suelen diferenciarse en términos de registro, como es el caso de los formales e informales, p. e.:

Mi padre adquirió un vehículo espacioso.

Mi papá compró un coche grande.

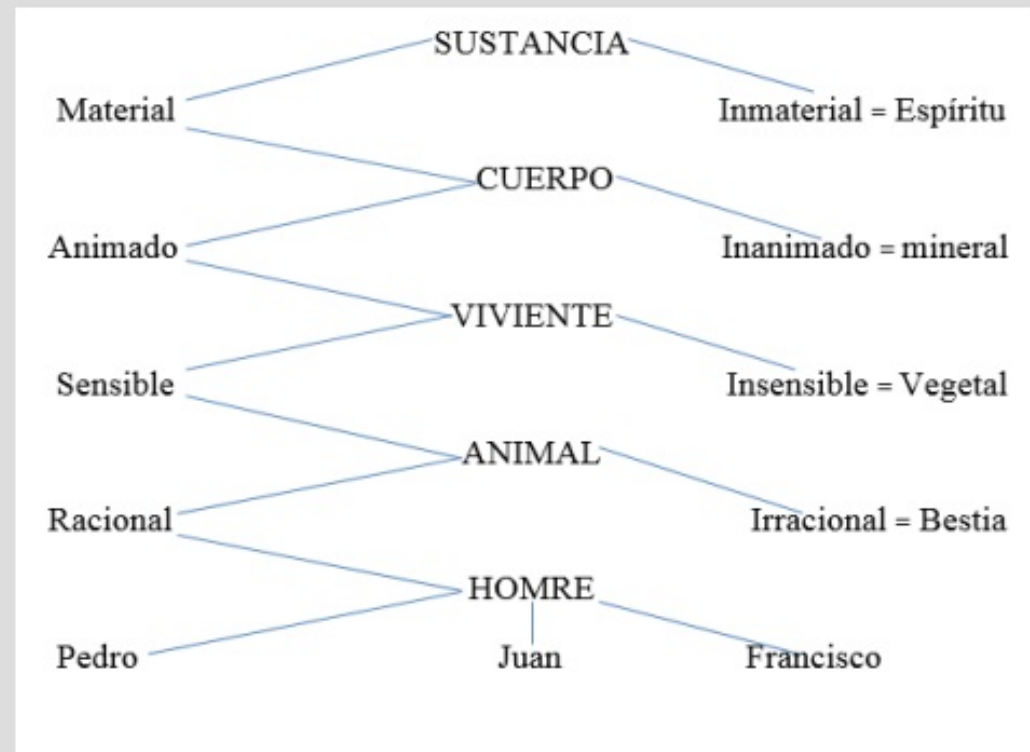


¿La semántica nos ayuda a organizar el mundo? (2)

Uno de los primeros que vislumbró esta relación fue el filósofo griego **Porfirio**, quien desarrolló un modelo de clasificación basado en Aristóteles.



**Porfirio
(233-304)**

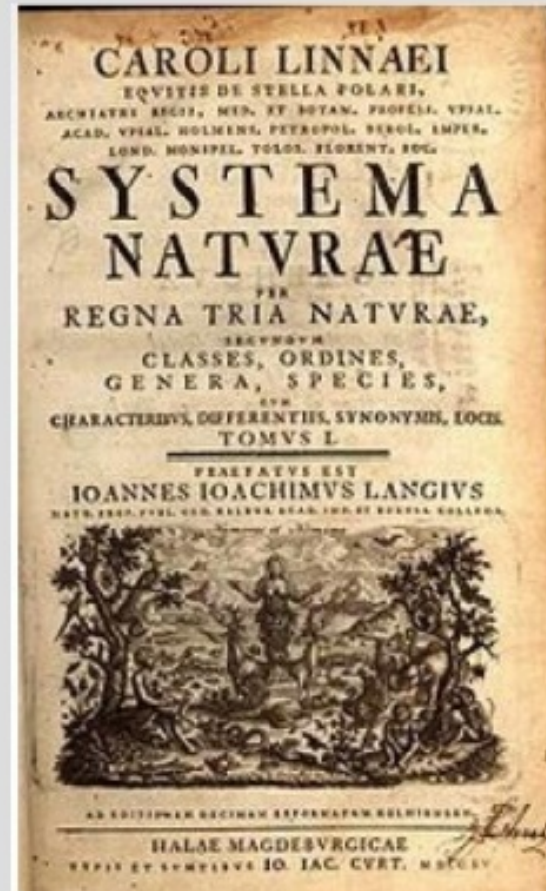


¿La semántica nos ayuda a organizar el mundo? (3)

Otro ejemplo similar a la propuesta de Porfirio es la que planteó en el S. XVIII **Carl Linnæus**, desarrollando una taxonomía para asignar una denominación a todas las especies de seres vivos existentes.



Carl Linnæus
(1707-1778)



¿La semántica nos ayuda a organizar el mundo? (4)

Ahora bien, quizá el ejemplo más clásico y conocido por todos de lo que es un sistema de clasificación de conceptos es el **diccionario**.

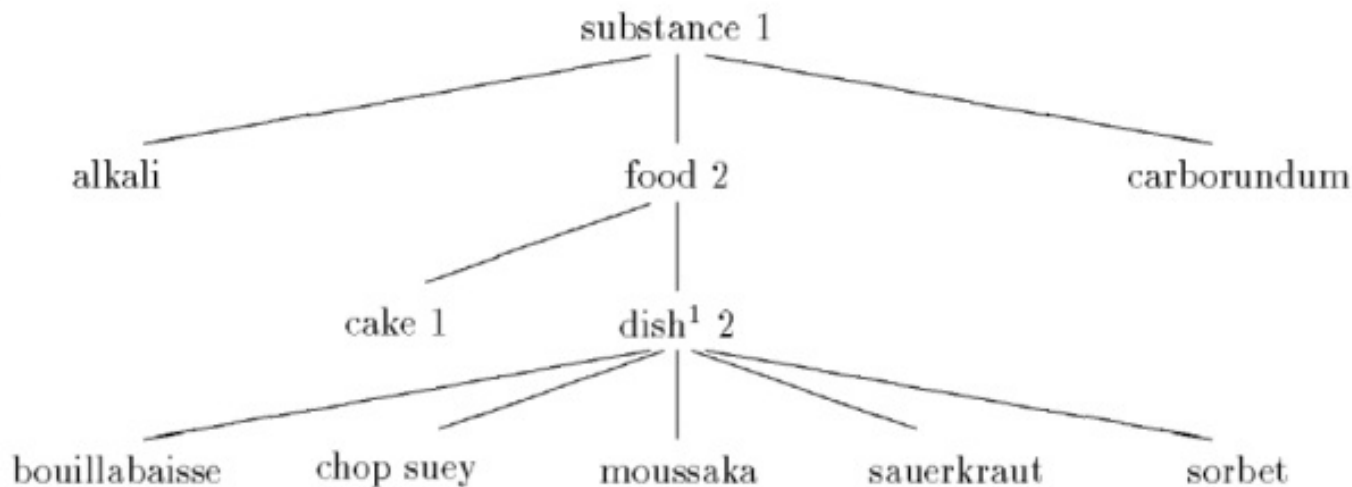
Pongamos por caso los diccionarios, los cuales podemos ver justo como objetos lingüísticos que integran y organizan de manera sistemática el vocabulario de una lengua, asociando a cada palabra una definición.

El primer diccionario del que se tenga registro en la historia data del año 2300 A. C., Se trata de una estela hallada en Siria, escrita en caldeo antiguo, y se le conoce como *Urra=hubullu*.



¿La semántica nos ayuda a organizar el mundo? (5)

Los diccionarios son objetos lingüísticos interesantes, porque se estructuran en torno a un *esqueleto* conceptual que se conoce como **taxonomía**, una evolución del árbol conceptual de Porfirio.



En este caso, Anne Copestake elabora una simple taxonomía de alimentos basada en sus ingredientes. El método para hacerlo es la relación género próximo/diferencia específica: un *alimento* es **un tipo de substancia** que cuenta con **alcanos** y **carbonos** (así podríamos definir el concepto *alimento*, de acuerdo con su taxonomía)



Una mirada al futuro (1)

La semántica léxica, conforme lo que hemos visto, es una línea de investigación importante para la lingüística, la cual ha tenido un impacto considerable, y hoy en día particularmente en el ámbito de las ciencias computacionales.

Un ejemplo muy sencillo: si ustedes tienen en su celular un servicio como *iTunes* o *Spotify*, ¿qué clase de información aparece asociada con la palabra *Interpol*?



Una mirada al futuro (2)

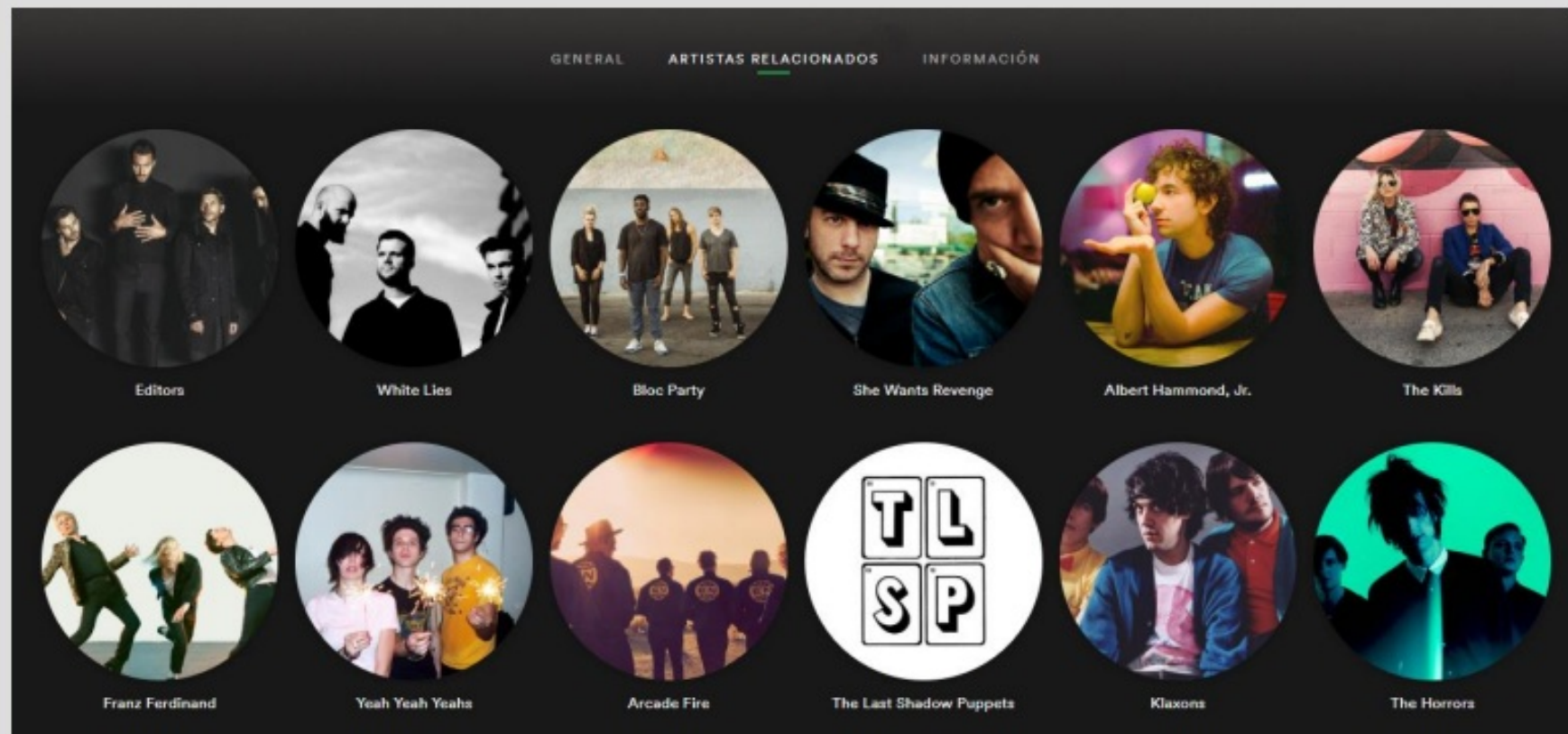
Si consideramos este ejemplo, por supuesto, uno podría pensar que no hay nada semántico de por medio: trabajan con un motor de búsqueda que sencillamente localiza lo que le pides, en este caso, alguna de estas dos entidades:



Ahora bien, dado que estamos hablando de música, es más probable que hablemos del grupo de rock, y no de la Policía Internacional (a menos que estemos buscando temas legales relacionados con música).

Una mirada al futuro (3)

Decidimos entonces la opción de la banda de rock. Pregunta, en el caso de *Spotify*, si eligen una opción como *artistas relacionados*, una pregunta: ¿bajo qué criterios se establecen estas relaciones?



Una mirada al futuro (4)

Otro ejemplo con otro sistema que ustedes también manejan: *Linkendin*, una red que les permite hacer contactos profesionales, tener información sobre sus temas laborales e incluso buscar trabajo, entre otras cosas.



Considerando esta última opción, la de buscar trabajo, ¿cómo resuelve Linkendin esta tarea?

Una mirada al futuro (5)

Lo primero que se les pide es crear un perfil laboral (esto es, un CV), el cual resuma toda la información que ustedes consideren relevante, p. e.:



The image shows a LinkedIn profile for César Aguilar. The profile picture is a circular image of a man in a blue jacket standing outdoors. The background of the profile header is a blue network diagram. The profile information includes the name "César Aguilar", his current position "Profesor asistente en Pontificia Universidad Católica de Chile", and his location "Chile". There are two buttons: "Añadir sección" (Add section) and "Más..." (More...). To the right, there are three icons with corresponding text: a university icon for "Pontificia Universidad Católica de Chile", another university icon for "Universidad Nacional Autónoma de México", and a contact icon for "Ver información de contacto". Below these, there is a contact icon for "Ver contactos (308)". At the bottom, there is a bio: "Profesor asistente, afiliado a la Facultad de Letras de la Pontificia Universidad Católica de Chile. Áreas de investigación: Procesamiento del lenguaje natural, ingeniería lingüística, extracción de información, sintaxis/semántica formal, taxonomías y ontologías."

César Aguilar
Profesor asistente en Pontificia Universidad Católica de Chile
Chile

[Añadir sección](#) [Más...](#)

-  Pontificia Universidad Católica de Chile
-  Universidad Nacional Autónoma de México
-  Ver información de contacto
-  Ver contactos (308)

Profesor asistente, afiliado a la Facultad de Letras de la Pontificia Universidad Católica de Chile. Áreas de investigación: Procesamiento del lenguaje natural, ingeniería lingüística, extracción de información, sintaxis/semántica formal, taxonomías y ontologías.

Una mirada al futuro (6)

Luego, es cierto, necesitan darle al sistema algunas pistas más específicas para que *Linkendin* pueda delimitar qué les interesa, digamos:

The screenshot displays the LinkedIn job search interface. At the top, there are navigation links: "Hacer un seguimiento de mis empleos", "Intereses de empleo", and "¿Estás buscando candidatos?". A button labeled "Anunciar empleo" is also visible. Below this is a search bar with the text "El empleo de tus sueños está a una búsqueda...". The search bar contains two input fields: "Buscar empleos" and "Buscar ubicación", followed by a "Búsqueda" button. Underneath the search bar, the section "Tus búsquedas de empleo" is shown, with a "Gestionar" link and navigation arrows. Three job search filters are listed, each with a bell icon for notifications:

- Spanish Computational Linguist (Todo el mundo)
- Natural Language Processing (Nuevos: 446) (Todo el mundo)
- Computational Linguistics (Nuevos: 12) (Todo el mundo)

¿Qué tan buenas opciones me puede ofrecer *Linkendin*, considerando mi perfil? Resolvamos esta duda a continuación.

Una mirada al futuro (7)

Veamos esta opción: hay un puesto en **Amazon Lab 126**, en donde se requiere un *Applied Scientist*, especializado en *Natural Language Processing*. Y da la casualidad de que alguien que estudió en mi Universidad también trabaja con ellos:



The screenshot shows a job listing for Amazon Lab 126. The job title is "Applied Scientist, Natural Language Processing". The location is "Amazon Lab126 · Sunnyvale, CA, US". The listing was published "hace 1 semana" and has "251 visualizaciones". There are two buttons: "Guardar" (Save) and "Solicitar" (Apply). Below the buttons, it says "Ve una comparación con los otros 27 solicitantes." and "Probar Premium gratis durante 1 mes". At the bottom, there are three columns: "Empleo" (27 applicants, No corresponde), "Empresa" (1,001-5,000 employees, Electrónica de consumo), and "Contactos" (1 antiguo alumno).

amazon
Lab126

Applied Scientist, Natural Language Processing

Amazon Lab126 · Sunnyvale, CA, US

Publicado: hace 1 semana · 251 visualizaciones

[Guardar](#) [Solicitar](#)

Ve una comparación con los otros 27 solicitantes.

[Probar Premium gratis durante 1 mes](#)

Empleo	Empresa	Contactos
<ul style="list-style-type: none">• 27 solicitantes• No corresponde	<ul style="list-style-type: none">• 1.001-5.000 empleados• Electrónica de consumo	 1 antiguo alumno

Una mirada al futuro (8)

Quizá no me guste esa oferta porque no me sienta capaz de hacer lingüística computacional en otros idiomas..., pero puedo trabajar quizá haciendo procesamiento para el español. ¿Alguna posibilidad? Veamos:



Spanish Computational Linguist - Alexa AI



Amazon · Turin, IT

Publicado: hace 3 semanas · 704 visualizaciones

Guardar

Solicitar

Ve una comparación con los otros 75
solicitantes.

[Probar Premium gratis durante 1 mes](#)



Empleo

- 75 solicitantes
- No corresponde

Empresa

- 10.001 empleados
- Internet

Contactos

-  8 antiguos empleados
-  62 antiguos alumnos

Muchas gracias