



Universidad Veracruzana

Seminario de semántica

Doctorado en estudios del lenguaje y lingüística aplicada

César Aguilar

Martes 30 de mayo de 2023

Lo que sabemos de semántica (1)

Tras haber hecho un recorrido en este seminario sobre algunas de las teorías que han ido configurando lo que es la semántica lingüística, podemos concluir algunas cosas:

- I. Existen diferentes niveles en los sistemas lingüísticos en donde se observan fenómenos semánticos. Aquí hemos observado dos: el nivel léxico y el sintáctico.
- II. Existen varios paradigmas teóricos que han modelado la visión de la semántica. En nuestro caso, hemos discutido al menos 2: el lógico-formal, y el funcional-cognitiva.
- III. Empero, estos paradigmas no son excluyentes, sino que tienden a coincidir y complementarse el uno a otro, apuntando hacia algunos fenómenos (p. e., la generación de relaciones léxicas o la composicionalidad) que son universales a cualquier lengua humana.

Lo que sabemos de semántica (2)

Con todo, existen problemas siempre por comprender y explicar, como es el caso de la ambigüedad, p. e.:



El aporte de Saussure

Como hemos dicho anteriormente, uno de los primeros que identificó particularidades relevantes respecto a la semántica fue **Ferdinand de Saussure**, quien (como recordarán) sentó las bases de la lingüística actual.



Ferdinand de Saussure
(1857-1913)



El aporte de Frege

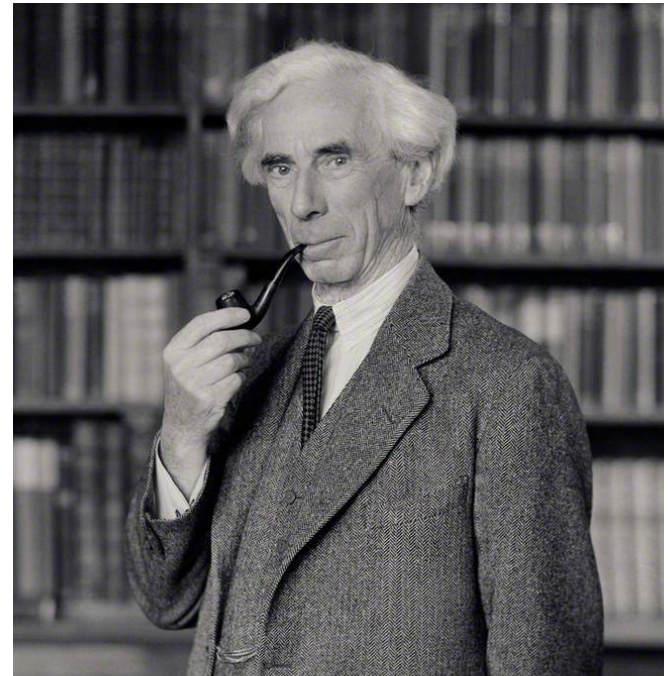


Gottlob Frege (1848-1925)

De forma paralela al trabajo de Saussure, y situado desde el marco de la lógica, el alemán **Gottlob Frege** desarrollaba un modelo que era un avance sustancial respecto a la propuesta de Aristóteles, y a la larga abrirá líneas de investigación relevantes en el terreno de la semántica.

El aporte de Russell

Igualmente, cabe mencionar aquí los aportes de **Sir Bertrand Russell**, quien ayudó a definir mejor el modelo semántico propuesto por Frege, a partir de varios ajustes que realizó, dando lugar al descubrimiento de un fenómeno esencial en la generación de significados: **la composicionalidad.**



Bertrand Russell
(1872-1970)

Composición y distribución (1)

Saussure y Frege sentaron las bases de dos hipótesis que configuran la semántica de los lenguajes naturales:

1. Los significados de las unidades y estructuras lingüísticas son arbitrarios. Una vez fijados, mantienen tanto un referente específico, como variantes de sentido.
2. Los significados pueden combinarse entre sí de gracias a la **composicionalidad**, la cual es un proceso universal a todas las lenguas.

Composición y distribución (2)


Otra manera de decirlo es de la siguiente manera:

Compositional Distributional Semantics

- **Compositional** - Based on the product of combining elementary word representations
- **Distributional** - Based on the co-occurrence statistics of words in a corpus
- **Semantics** - The study of the meaning of words and phrases in a language

Composición y distribución (3)

Partamos de un primer supuesto: cuando vemos que una serie de palabras co-ocurren en un contexto determinado (p. e., una búsqueda de colocaciones), ¿podemos decir que comparten siempre algún rasgo de significado? Veamos:



he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
they love the sun , the moon and the stars . None of
the light of an enormous moon . The splash of flowing w
man 's first step on the moon ; various exhibits , aer
the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind

Composición y distribución (4)

Con base en los resultados de la consulta anterior, ¿qué relación mantienen el sustantivo *star* con los verbos *to see*, *to bright* y *to shine*. Veamos:

Variations in the type of context features

	Doc1	Doc2	Doc3
stars	38	45	2
	\xleftarrow{dobj} see	\xrightarrow{mod} bright	\xrightarrow{mod} shiny
stars	38	45	44
	The nearest • to Earth	stories of • and their	
stars	12		10

Composición y distribución (5)

De primera instancia, sabemos que las estrellas son brillantes, ya que podemos decir que este rasgo es propio de ellas (este supuesto lo refuerza la noción de **referencia** de Frege).

Dicho lo anterior, una pregunta: dada esta relación entre *estrella* y *brillar*, ¿qué otras palabras pueden co-ocurrir con este par? Una respuesta posible:

E.g.: Co-occurrence with words, window of size 2, scaling by distance to target:

... two [*intensely bright stars in the*] night sky ...

	intensely	bright	in	the
stars	0.5	1	1	0.5

Composición y distribución (6)

Veamos el total de co-ocurrencias posibles entre estas palabras:



	bright	in	sky
stars	8	10	6
sun	10	15	4
dog	2	20	1

Composición y distribución (7)

Otra manera de establecer estas co-relaciones es determinar qué tan significativas son, asignándole un peso específico con base en su frecuencia de aparición. Para hacer esto, tenemos la medida PMI (**Pointwise Mutual Information**):

Point-wise Mutual Information (PMI)

$$\text{PMI}(\text{target}, \text{ctxt}) = \log \frac{P(\text{target}, \text{ctxt})}{P(\text{target})P(\text{ctxt})}$$

Composición y distribución (8)

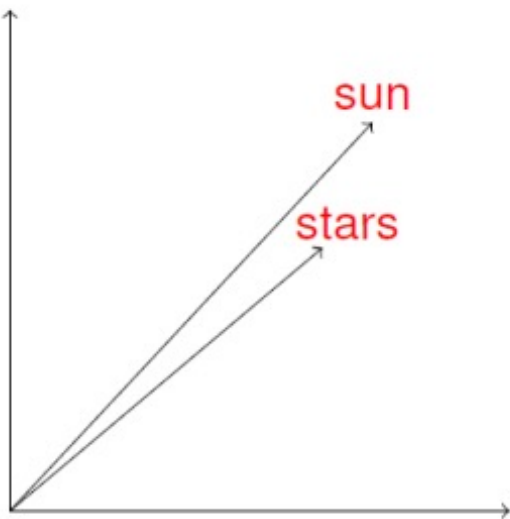
La medida de PMI nos ayuda a discernir y a hacer predecibles cuáles son las palabras que comparten mayor contenido semántico en un corpus o un vocabulario determinado, p. e.:

Adjusting raw co-occurrence counts:

	bright	in	...	
stars	385	10788	...	← Counts
stars	43.6	5.3	...	← PMI

Composición y distribución (9)

Ahora bien, si vamos a representar esta relación en un plano cartesiano, una pregunta que podemos hacernos sería: **¿qué clase de objeto, similar a una estrella, es capaz de brillar en el cielo?**
Respuesta:



Vectors

stars	2.5	2.1
sun	2.9	3.1

Cosine similarity

$$\begin{aligned}\cos(x, y) &= \frac{\langle x, y \rangle}{\|x\| \|y\|} \\ &= \frac{\sum_{i=1}^{i=n} x_i \times y_i}{\sqrt{\sum_{i=1}^{i=n} x_i^2} \times \sqrt{\sum_{i=1}^{i=n} y_i^2}}\end{aligned}$$

Composición y distribución (10)

Y como todas las cosas interesantes, volvemos al mismo punto de partida:

Principle of compositionality:

the “meaning (**vector**) of a complex expression (**sentence**) is determined by:

- the meanings of its constituent expressions (**words**) and
- the rules (**grammar**) used to combine them”

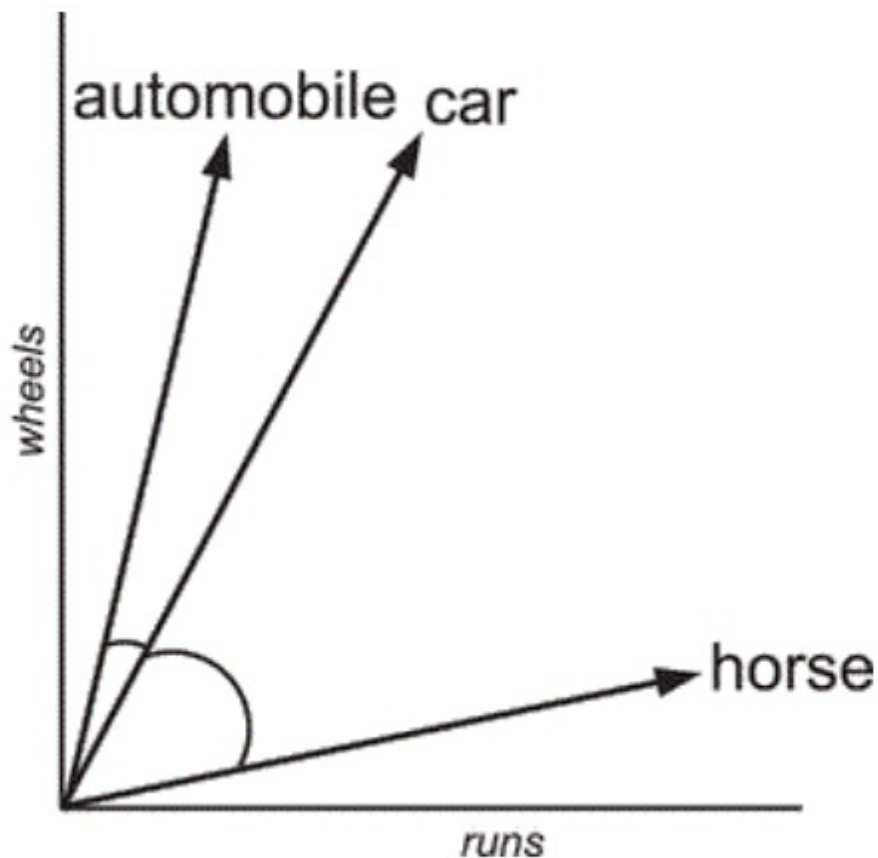


— Gottlob Frege
(1848 - 1925)

Composición y distribución (11)

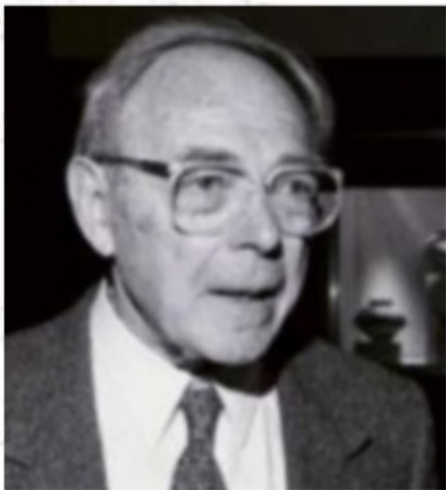
Para terminar, una forma muy simple de representar la diferencia entre autos y caballos según la semántica distributiva:

	<i>runs</i>	<i>wheels</i>
automobile	1	4
car	2	4
horse	4	1



La visión distributiva del lenguaje (1)

En los años 50, la lingüística norteamericana estaba asumiendo una visión matemática sobre el funcionamiento del lenguaje humano. El principal representante de esta corriente en ese momento era **Zellig Harris**.



Zellig Harris
(1909-1992)

Harris fue uno de los primeros lingüistas que puso atención en cómo las lenguas operaban de manera similar a un sistema matemático, de tal suerte que:

1. Sus unidades de construcción se pueden agrupar en dos conjuntos: simples y complejas
2. Sus procesos recursivos, los cuales dan lugar a su capacidad combinatoria
3. Esta combinatoria da lugar a la identificación de patrones regulares, los cuales muestran frecuencias de uso claras y medibles

La visión distributiva del lenguaje (2)

Veamos un ejemplo sobre cómo entendía Harris todo esto. Pensemos en la co-ocurrencia de palabras como **cat** y **dog**, según la siguiente tabla:

	dog	cat	bread	pasta	meat	mouse
dog	40	7	1	0	1	5
cat	7	32	0	1	0	8
bread	1	0	22	15	8	0
pasta	0	1	15	24	10	1
meat	1	0	8	10	30	2
mouse	5	8	0	1	2	31

La visión distributiva del lenguaje (3)

Dados estos valores, ¿cuáles son para ustedes las palabras que pueden asociarse mejor? Veamos:

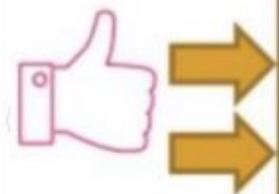


	dog	cat	bread	pasta	meat	mouse
dog	40	7	1	0	1	5
cat	7	32	0	1	0	8
bread	1	0	22	15	8	0
pasta	0	1	15	24	10	1
meat	1	0	8	10	30	2
mouse	5	8	0	1	2	31

2.000.000

La visión distributiva del lenguaje (4)

Ahora bien, ¿qué otro grupo de palabras se pueden relacionar de manera cercana, dada su distribución?:

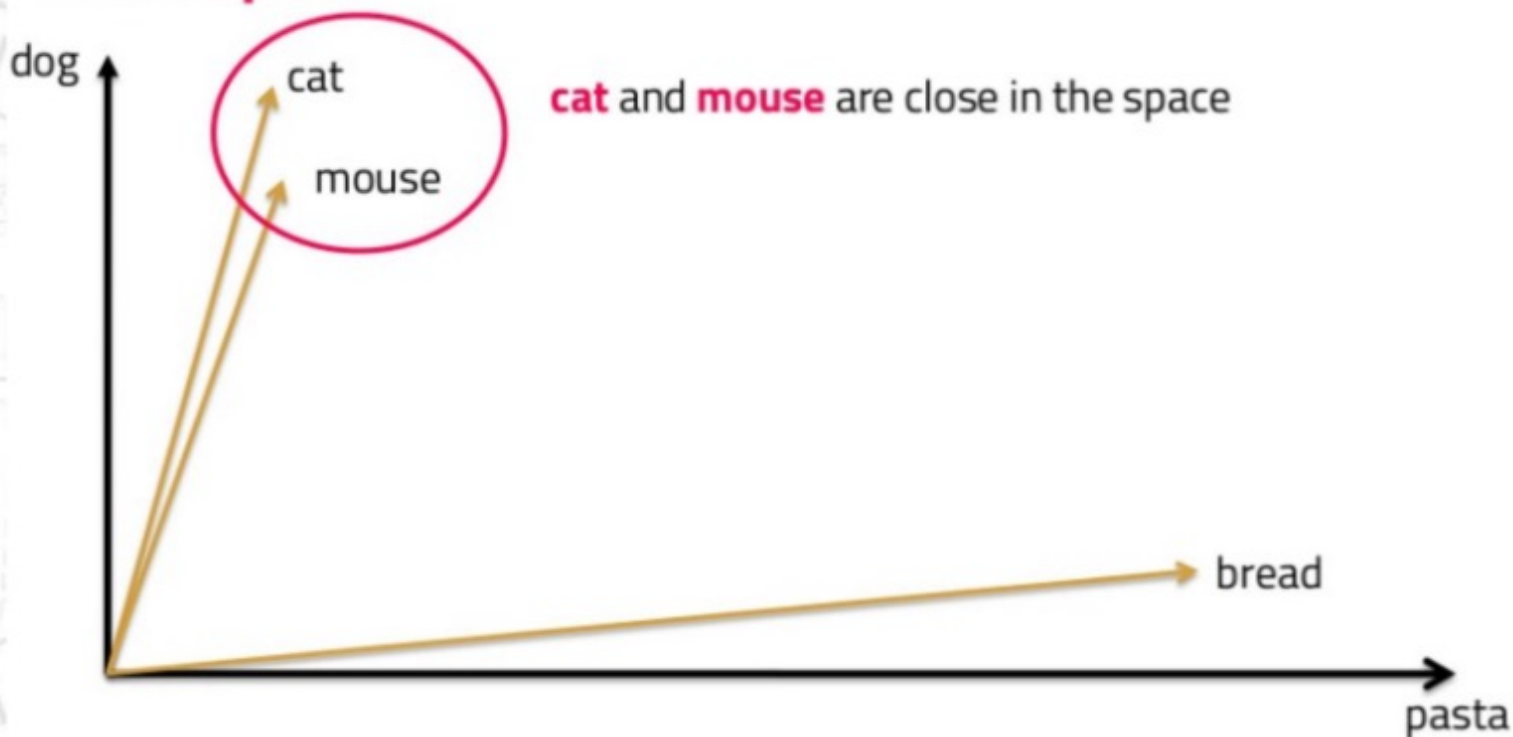


	dog	cat	bread	pasta	meat	mouse
dog	40	7	1	0	1	5
cat	7	32	0	1	0	8
bread	1	0	22	15	8	0
pasta	0	1	15	24	10	1
meat	1	0	8	10	30	2
mouse	5	8	0	1	2	31

La visión distributiva del lenguaje (5)

Si trasladamos las relaciones entre estos pares de palabras (**cat/ dog** y **bread/pasta**) a un plano espacial, podemos modelar dos vectores, siguiendo la explicación que nos dio Ana Bertha en clases anteriores:

Geometric space
WordSpace

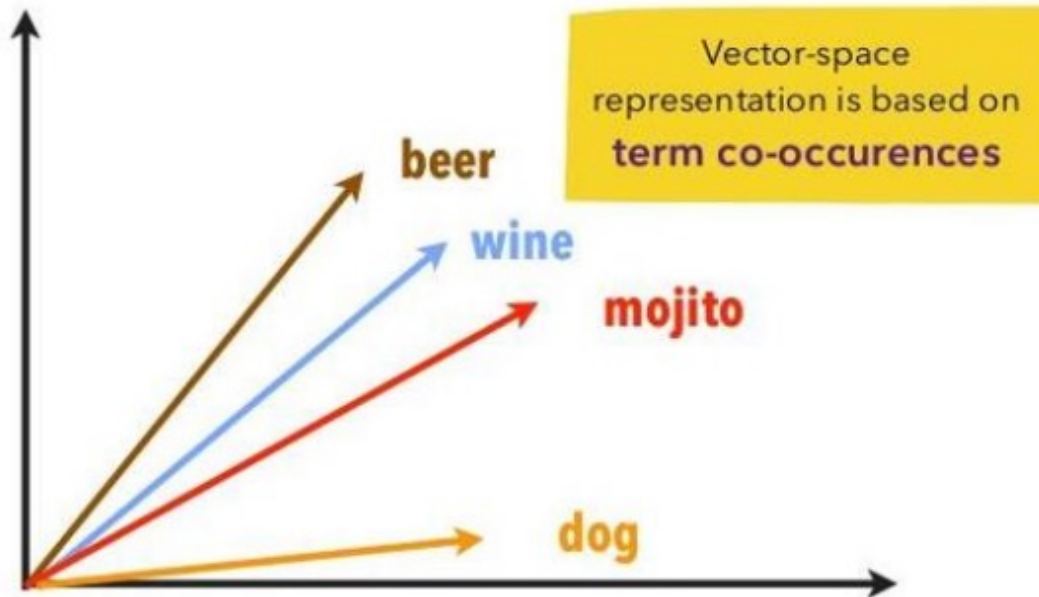


La visión distributiva del lenguaje (6)

Harris lo que nos diría al observar estos vectores es lo siguiente:

Distributional Hypothesis

"Words occurring in similar (linguistic) contexts tend to be semantically similar"



Equivalencias entre semántica formal y distributiva (1)

El análisis de pares de palabras como las anteriores nos da pistas sobre la información semántica que pueden compartir entre sí. Ahora, ¿qué pasa cuando analizamos frases y oraciones?

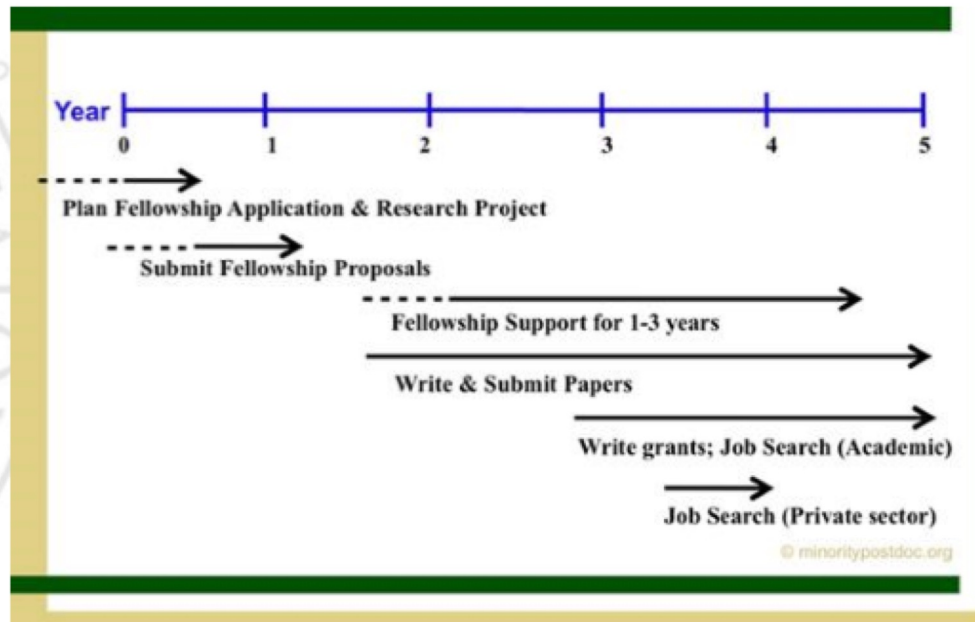
En breve, se puede ver que la **composicionalidad** puede ser vista también como un fenómeno de distribución. P. e., consideren la siguiente proposición:

$$\exists x, y [new(postdoc(x)) \wedge article^*(y) \wedge write(x, y)]$$

Equivalencias entre semántica formal y distributiva (2)

A partir de una proposición como la anterior, somos capaces de sintetizar una buena parte del contenido de información semántica que tiene el nombre **postdoc**, e incluso podemos inferir que puede aparecer en el contexto de oraciones como:

The article was written by the postdoctoral student yesterday



Equivalencias entre semántica formal y distributiva (3)

Contrastemos lo anterior con la siguiente idea: ¿qué pasa si encuentro las mismas relaciones, pero esta vez en oraciones contenidas, p. e., en los reportes de actividades que los investigadores universitarios llenan para justificar la labor de sus postdoctorales, p.e.:

The **postdoc** shook her head.

Any grad student or **postdoc** he'd have would be a clonal copy of himself.

That tattoo is actually a custom-made cartoon generated by the **postdoc** who led this work.

During that **postdoc**, I didn't publish anything.

(a)

\implies 0.43679 1.93841 ...

(b)

(c)

Figure 1

Distributional Semantics: The linguistic contexts in which an expression appears, for example, the words in the *postdoc* sentences in (a), are mapped to an algebraic representation (see the vector in (c)) through a function, represented by the arrow in (b). The resulting representations (in the example, the vector for *postdoc*) are thus abstractions over contexts of use. Examples adapted from the COCA corpus.

Equivalencias entre semántica formal y distributiva (4)

Ahora bien, proyectemos en un plano cartesiano las relaciones que establecen las palabras **postdoc**, **student** y **wealth**, y generemos sus vectores, esto es:

	dim1	dim2
postdoc	0.43679	1.93841
student	0.71038	1.76058
wealth	1.77337	0.00012

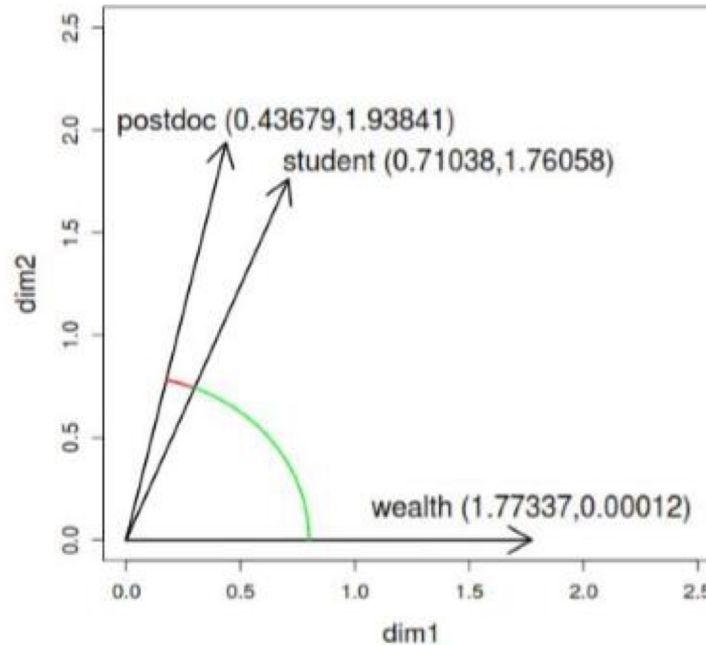


Figure 2

Semantic distance as geometric distance: A toy lexicon (left) and its representation in semantic space (right). The distance (see arcs) between the vectors for *postdoc* and *student* is smaller than the distance between the vectors for *postdoc* and *wealth*. The converse is true for similarity as

Parametrizando el sentido común (1)

Veamos otro ejemplo para comprender el sentido de la pregunta. Supónganse que analizamos un corpus, y hacemos una identificación de nombres y verbos como:

“The **dog** **barked** in the **park**. The **owner** of the **dog** put him on the **leash** since he **barked**.”

contexts = nouns and verbs in the same sentence

Si lo pensamos un rato, todos estos nombres y verbos describen **atributos y datos** que validan (o invalidan, también es posible) nuestro conocimiento del mundo.

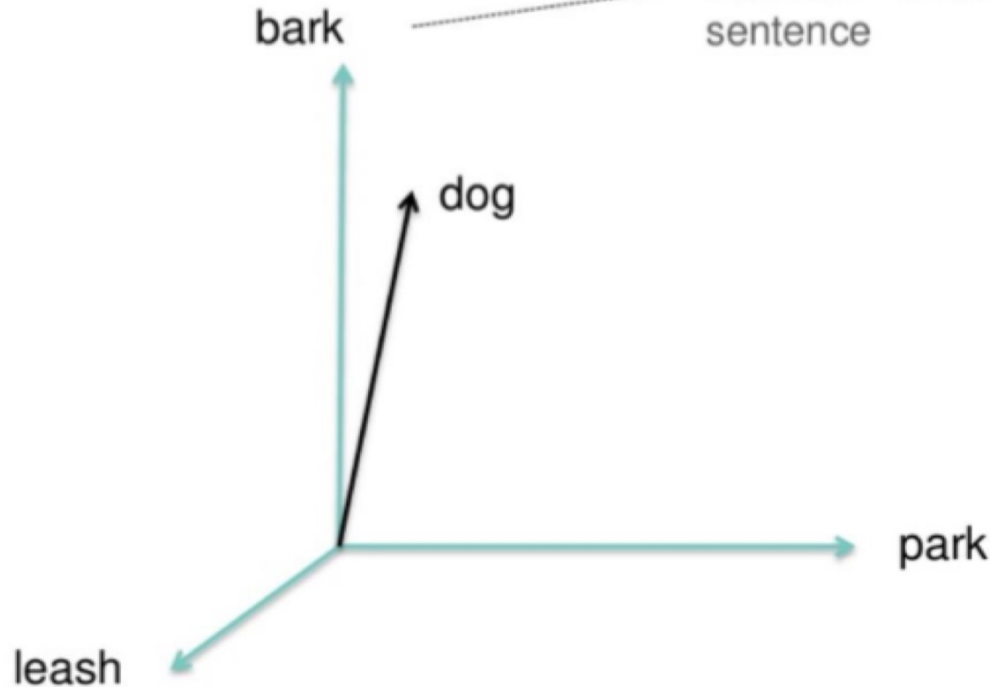


Parametrizando el sentido común (2)

Ahora bien, generemos de nuevo nuestro plano cartesiano y tracemos nuestros vectores, asignando un valor a cada relación para determinar qué palabras están más cercanas o alejadas:

“The **dog** **barked** in the **park**. The **owner** of the **dog** put him on the **leash** since he **barked**.”

contexts = nouns and verbs in the same sentence



bark : 2
park : 1
leash : 1
owner : 1

Distributional Semantic Models (DSMs)

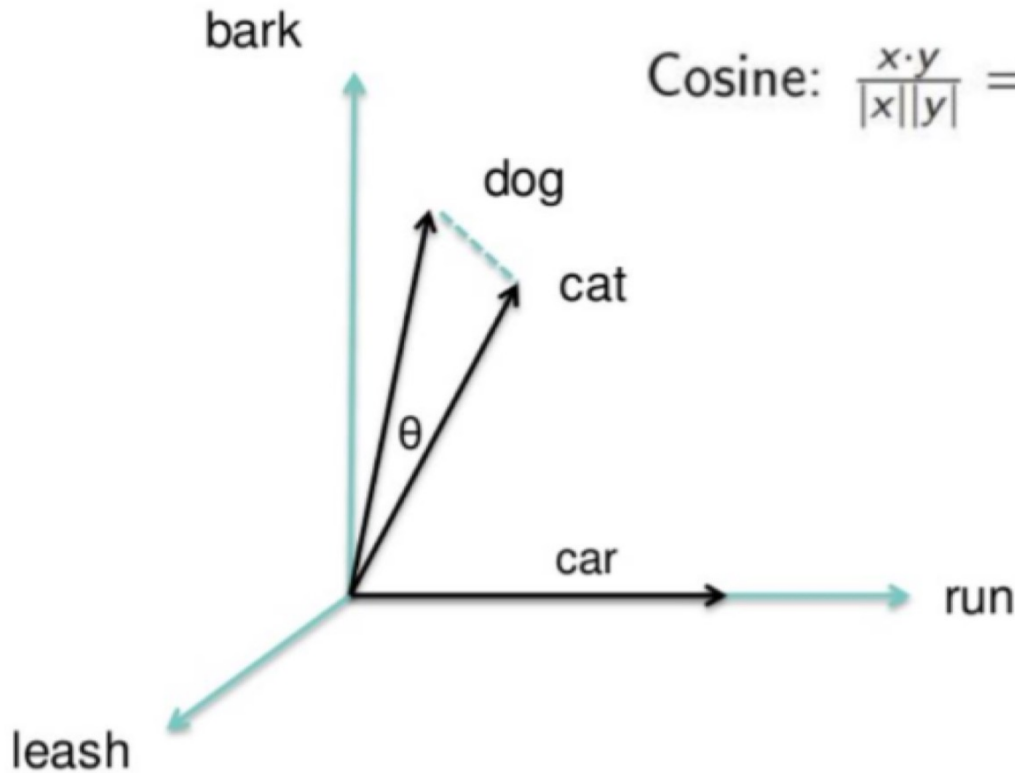
Vector Space Model (VSM)

distributional matrix = **targets** x **contexts**

	contexts						
	leash	walk	run	owner	leg	bark	
targets							
dog	3	5	1	5	4	2	
cat	0	3	3	1	5	0	
lion	0	3	2	0	1	0	
light	0	0	0	0	0	0	
bark	1	0	0	2	1	0	
car	0	0	4	3	0	0	

Parametrizando el sentido común (4)

Ahora bien, una forma de establecer una similitud semántica entre estas palabras es empleando una fórmula geométrica como el cálculo de coseno, esto es:



$$\text{Cosine: } \frac{x \cdot y}{|x||y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Parametrizando el sentido común (5)

Con estos datos ya podemos inferir varias relaciones semánticas, p. e.:

- **Semantic similarity - two words sharing a high number of salient**
 - features (attributes)
 - synonymy (car/automobile)
 - hyperonymy (car/vehicle)
 - co-hyponymy (car/van/truck)
- **Semantic relatedness (Budanitsky & Hirst 2006) - two words semantically associated without being necessarily similar**
 - function (car/drive)
 - meronymy (car/tyre)
 - location (car/road)
 - attribute (car/fast)

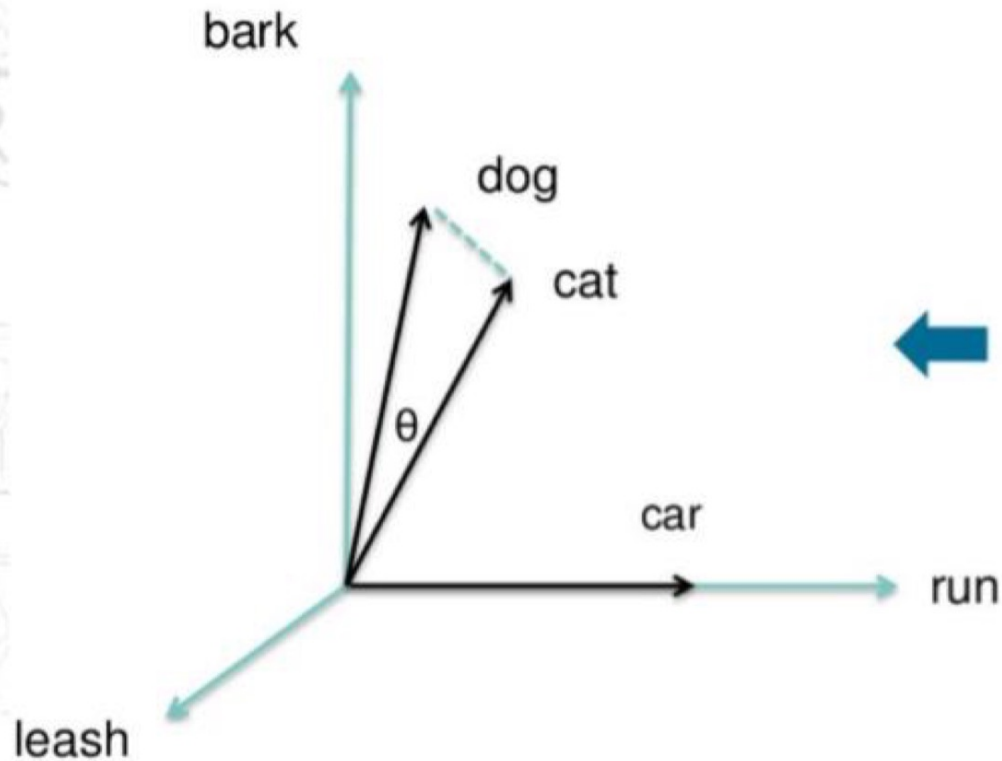
Parametrizando el sentido común (6)

Lo que advierte aquí es lo siguiente:

- Computational models that build contextual semantic representations from corpus data
- **Semantic context** is represented by a **vector**
- Vectors are obtained through the statistical analysis of the linguistic contexts of a word
- Saliency of contexts (cf. context weighting scheme)
- Semantic similarity/relatedness as the core operation over the model

Parametrizando el sentido común (7)

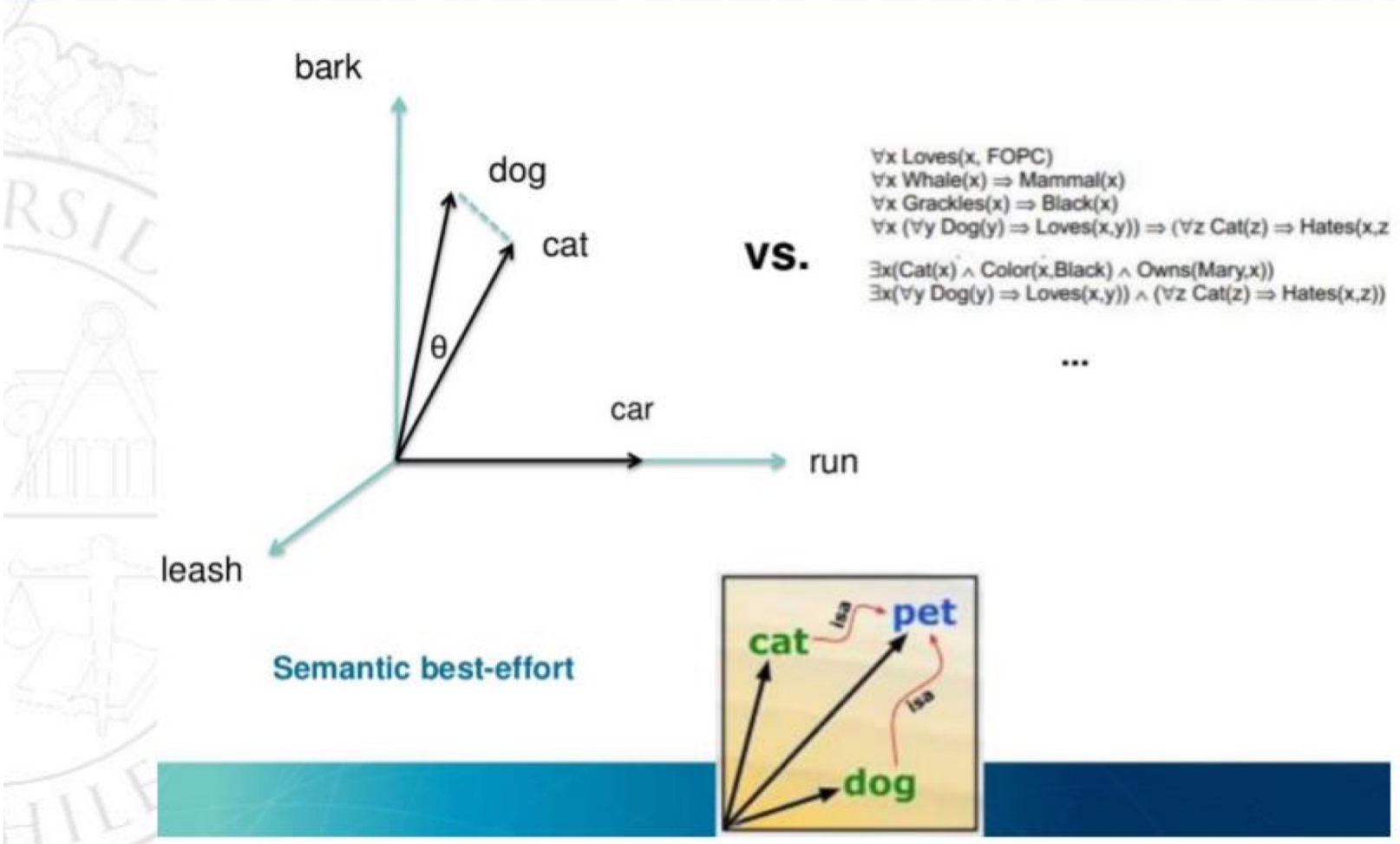
Y considerando lo anterior, digamos que nuestro sistema de análisis semántico computacional está haciendo inferencias usando un tipo de **sentido común artificial**, esto es:



Commonsense is here

Parametrizando el sentido común (8)

Tales inferencias no tienen por qué estar desligadas de los modelos tradicionales de semántica formal; de hecho, se pueden combinar y generar métodos híbridos, p. e.:



Muchas gracias