

La sinonimia desde una perspectiva lingüístico-cognitiva. Medición de la distancia semántica

*Synonymy from a linguistic-cognitive perspective.
Measuring semantic distance*

Martin Zapico

Universidad Nacional de Mar del Plata
Argentina

Jorge Vivas

Universidad Nacional de Mar del Plata
Argentina

ONOMÁZEIN 32 (diciembre de 2015): 198-211

DOI: 10.7764/onomazein.32.11



Martin Zapico: Centro de Procesos Básicos, Metodología y Educación, Facultad de Psicología, Universidad Nacional de Mar del Plata, Argentina. | Correo electrónico: mgzapico@mdp.edu.ar

Jorge Vivas: Centro de Procesos Básicos, Metodología y Educación, Facultad de Psicología, Universidad Nacional de Mar del Plata, Argentina. | Correo electrónico: jvivas@mdp.edu.ar

Fecha de recepción: diciembre de 2014

Fecha de aceptación: abril de 2015

Resumen

La sinonimia, entendida como la capacidad de intercambiar dos palabras en un contexto determinado sin que se pierda el significado original, ha presentado una serie de problemas teóricos y técnicos en lingüística a lo largo del siglo XX. Ante estas controversias, la medición de la distancia semántica entre palabras se perfila como una vía sugerente para la elucidación de este problema. La perspectiva propuesta coloca la sinonimia como un fenómeno gradual y progresivo. La hipótesis que subyace a este trabajo es que el grado de sinonimia se constituye como un caso extremo de la mayor cantidad de atributos compartidos por el concepto al que hace referencia la palabra. De este modo se estableció la correspondencia entre la sinonimia tal como es presentada en diccionarios y tesauros y la sinonimia obtenida experimentalmente a partir de las distancias. Se utilizaron las Normas de

Producción de Atributos Semánticos en Castellano obtenidas a partir de 900 jóvenes universitarios. Se obtuvieron las distancias semánticas para 60 palabras pertenecientes a las normas y a los diccionarios por medio del cálculo del coseno entre vectores en un espacio n-dimensional para determinar los grados de coincidencia o sinonimia. Las palabras consideradas sinónimos por los diccionarios fluctúan en su distancia semántica entre 0.6 y 0.95. Estos resultados sugieren que las distancias entre los pares considerados sinónimos por los diccionarios resultó superior a 0.9 cuando pertenecen a una misma subcategoría específica. En las palabras con una distancia próxima a 0.8 se advierte la pertenencia a una misma categoría de base y cuando puntuaron entre 0.6 y 0.8 pertenecen a una misma categoría tomada en un sentido amplio.

Palabras clave: sinonimia; distancia semántica; atributos semánticos.

Abstract

The synonymy, understood as the ability to exchange two words in a given context without loss original meaning, has presented a number of theoretical and technical issues in linguistics throughout the twentieth century. Given these controversies, measuring the semantic distance between words is emerging as a good way to solve this problem. These proposal puts perspective synonymy as a gradual and progressive phenomenon. The hypothesis underlying this work is that the degree of synonymy is established as an extreme case of as many shared the concept referenced word attributes. Thus the correspondence between synonymy as it is presented in dictionaries and thesaurus and synonymy experimentally obtained from the distances established. For this work "Normas de Producción de

Atributos Semánticos en Castellano" obtained from 900 university students were used. Semantic distances to 60 words belonging to the rules and dictionaries by calculating the cosine between vectors in a n-dimensional space were obtained. Considered synonyms for words fluctuate dictionaries semantic distance between 0.6 and 0.95. These results suggest that the distances between the pairs considered synonymous for dictionaries 0.9 was higher than when they belong to one specific subcategory. In the words of a close distance 0.8, warns belonging to the same category of base and when they scored between 0.6 and 0.8 belong to the same category taken in a broad sense.

Keywords: synonymy; semantic distance; semantic features.

1. Introducción

Desde los siglos XVIII-XIX la sinonimia ha sido trabajada bajo la forma de grandes diccionarios de sinónimos donde se ha problematizado el alcance de la noción misma de la sinonimia y su posibilidad (González Pérez, 1994). Entre estas podemos destacar *Ensayo de los Synonimos* de Manuel Dendo y Ávila (1756), *Ensayo de la posibilidad de fixar la significación de los sinónimos de la lengua castellana* de José López Huerta (1789) y *Colección de los sinónimos de la lengua castellana* de José Joaquín de Mora (1855), entre otros. No obstante, tal como señala el mismo González Pérez, esta gran proliferación de obras marcaba una tendencia bien definida desde los mismos títulos de sus obras. Se afirma la existencia de sinónimos y se los clasifica o presenta sin dar fuertes argumentos para confirmar la sinonimia, asentándose la idea en la buena comprensión del lector al darla por sentada.

Pese a esta diversidad, llegando a mediados del siglo XX, Bernard Levy (1942) marcaba que “la historia de la sinonimia española es, en gran parte, un terreno relativamente poco explorado”. Pérez, en el artículo ya citado, explicita que la sinonimia, dentro de la semántica, ha sido dejada de lado bajo premisas axiomáticas que se quedan en la idea de que nunca se podrá lograr la igualdad total de significados. González Martínez (1989) señala que, dentro de la poca atención que ha recibido el tema, cada una de las posturas que se han ido presentado, si bien resultan impecables en sí mismas, no llegan nunca a resultar viables si se busca su aplicación al análisis de situaciones comunicativas reales.

Asimismo, diversos autores (Hernández García, 1995) señalan que, salvo contando excepciones, el panorama no ha variado demasiado. Entre estas excepciones, se destacan los trabajos de Salvador (1985) y Lyons (1979) que plantean posiciones definidas en torno al tema y las defienden a partir del análisis de grandes diccionarios o corpus particulares. Otra gran variedad de estudios que no se tienen en cuenta para este

trabajo hacen énfasis en aspectos históricos sobre el tema o se limitan a proponer grandes corpus de sinónimos. Martínez López (1997) explica que la divergencia en las posturas está dada por puntos de partida diferentes (no se considera la misma definición de sinonimia, algunos analizan la lengua, otros analizan el habla, etc.) y en su propio análisis toma en cuenta que al actualizarse la lengua en el habla se pierde cualquier neutralidad pretendida en los diccionarios. No hay que dejar nunca de lado que el diccionario en sí es una forma de definir el significado de una palabra fuera de su uso que, si bien está elaborada a partir de estudios sobre el uso, nunca podrá ser totalmente fiel al mismo, puesto que en cada acto de habla la lengua cambia lentamente. Porque un diccionario y una gramática son una abstracción; ilusión de que hay un sistema único y homogéneo. Sucede lo mismo con trabajos propuestos desde el ámbito de la Lingüística Computacional, que trabajan la proximidad de conceptos a partir del análisis y procesamiento de la co-ocurrencia lexical en grandes corpus de datos, como WordNet (Maki, McKinley y Thompson, 2004; Lin Zhao, 2003). Estos trabajos, a partir del estudio de conceptos entendidos como conjuntos de atributos, apuntan a la estructuración de categorías de análisis para la comprensión del funcionamiento del sistema semántico en el proceso de interpretación. Es por esto que acuden a bases de datos que, por su naturaleza misma, distan de proponer la sinonimia en grupos de habla como una problemática, salvo en los casos de desarrollo de buscadores semánticos en la *web*. No obstante, cabe hacer una aclaración. La cuestión del diccionario es en función de la validez ecológica, pero como señala Vergara (2011) nunca hay que dejar de lado el uso del diccionario de sinónimos cuando se encare cualquier tarea de redacción dado que “es de suma importancia revisar los textos y sustituir elementos reiterados, que provocan monotonía y falta de fluidez, por otros términos más precisos, los diccionarios de sinónimos pueden ser una herramienta perfecta para tal objetivo”.

Quedan definidas así tres grandes posturas en el asunto de la sinonimia. La de García Cordero que afirma que no es posible que dos conceptos sean intercambiables, caso contrario no existirían dos conceptos diferentes y, en todo caso, que los hablantes elijan usar como sinónimos pares de palabras es accidental. La de Gregorio Salvador que afirma que existen pares de palabras que son intercambiables y que en determinados conceptos no se pierde en ningún aspecto el significado de la frase. Es probable que estas tres posturas no conciliables (sinonimia no —García Cordero: 2004—, sinonimia si —Gregorio Salvador—, sinonimia solo parcial —RAE—) estén dadas por la falta de validez ecológica de los análisis. Esta apreciación, si bien puede parecer evidente, se realizó sobre el filo del fin de siglo XX. Hay que destacar que no ha sido un problema que ocupe la agenda de la disciplina, salvando a la lexicografía que se preocupa por el diseño de los diccionarios, entre los cuales se hallan los diccionarios de sinónimos.

Dado que ninguna de las corrientes de estudio del siglo XX se ha ocupado en superar esta problemática, se trabajará a partir de la Teoría Neurocognitiva de Sidney Lamb (1999) que, se cree, es capaz de dar una respuesta tanto descriptiva como explicativa y hasta predictiva de las relaciones de proximidad semántica que pueden darse entre pares de conceptos.

Los aportes de la Teoría Neurocognitiva para el estudio de los sinónimos serían los siguientes:

1. El fenómeno sirve para fundamentar que el sistema lingüístico es una red de relaciones.
2. Muestra de qué forma se relaciona la información léxica con la información semántica.
3. Permite representar a través de redes relacionales los aspectos de la comunicación humana mencionados al comienzo.
4. Pueden establecerse consideraciones en términos de la plausibilidad operativa, la plausibilidad de desarrollo y la plausibilidad neurológica.

Como complemento de la teoría lingüística, se acudirá, además, a planteos propios de la perspectiva cognitiva de la cognición situada, particularmente de Barsalou (2008, 2009) y McRae (2005, 2013), en lo que se refiere a la definición de lexema, concepto, y las relaciones que se establecen entre ambos, además de cómo se integran en la propuesta de Lamb. Lexema en la teoría de redes relacionales son “aquellas unidades que se aprenden (es decir, que se manifiestan en el sistema cognitivo) como unidades, lo que implica que no se constituyen en el momento en base a sus morfemas constituyentes” (Lamb, 1999: 66). Es decir, unidades como “cortadora de pasto”, “diccionario de bolsillo”, “bolso de mano” en el sistema de conocimiento de una persona tienen representación conceptual particular, lo mismo en el sistema léxico-gramatical, ya que al hablar de “cortadora de pasto” no se piensa en un elemento que corta, y en pasto, sino que se piensa en la máquina empleada para cortar pasto como una unidad. Lo mismo sucede con “pájaro carpintero”, que en la representación mental de un individuo remite a una unidad, un pájaro con determinadas características, no a un pájaro que tiene por oficio la carpintería. No se construye el significado de los lexemas a partir de la combinación de lexemas, por eso mismo los casos antedichos son unidades como tales. Además, la idea de concepto que se utilizará será la propia de la propuesta de Barsalou que, como señala Ramos (2010: 126-127), se puede entender un concepto como un constructo aglomeración de atributos, siendo los atributos otros conceptos. Y los conceptos, de cuya relación se constituye el significado, emergen como producto de un sofisticado proceso que comienza con la percepción, es decir, la construcción activa a partir de los estímulos del medio. Así, si ubicamos la formación de conceptos y su definición en el sistema de conocimiento de un individuo, son compatibles los planteos de Barsalou para trabajar en el marco del sistema conceptual de Lamb. Guiados por esta perspectiva reticular podemos sugerir que el significado de un objeto no emer-

ge directamente de las propiedades intrínsecas del mismo. No se trata de algo que se debe descubrir como si fuese una propiedad inherente ni la labor del agente consiste en desvelar dicho significado. En lugar de ello, el significado, como diría Wittgenstein, emerge de las interacciones lingüísticas en que ese agente se ha participado en la comunidad, es decir, de su experiencia como hablante.

En otro paradigma y a los fines de modelar la memoria semántica, el concepto de distancia semántica fue acuñado por Smith, Shoben y Rips (1976), y ha recibido diferentes definiciones, que repasaremos brevemente a continuación. Distintas definiciones darán como resultado distintos modelos de comprensión de la distancia entre dos o más palabras.

En una red semántica clásica (Collins y Loftus, 1975) dos conceptos se hallan semánticamente relacionados si se encuentran juntos próximos en la red. Podemos medir la “proximidad” como la distancia literal entre ambos, esto es, la longitud del camino que ambos comparten. El modelo se presenta adecuado para abordar el problema planteado.

En el modelo de Clave Compuesta de Ratcliff y McKoon (1988), la distancia semántica se explica por la probabilidad con que dos palabras coexistan relacionadas más frecuentemente que palabras no relacionadas. Esta circunstancia explica la existencia del efecto de *priming* significativo para relaciones asociativas no necesariamente semánticas. Este proceso es, sin embargo, transparente cuando se solicita a las personas que estimen la proximidad semántica entre dos términos relacionados semánticamente por relaciones de distinta naturaleza y con diferente peso en su asociación.

Los modelos Conexionistas Distribuidos (Plaut, 1997; McRae y Boisvert, 1988) proponen que los conceptos relacionados presentan un patrón de activación similar. Si bien el método de comparación de distancias aquí propuesto no

se muestra, en principio, incompatible con este modelo, sugerimos que la mensura de la distancia semántica dentro de esta perspectiva probablemente encuentre un instrumento adecuado y coherente para su ejecución por medio de la utilización de los algoritmos Self-Organizing Map (SOM), introducidos por T. Kohonen (1988, 1997).

Los modelos basados en la co-ocurrencia de ítems lexicales (Lund, Burgess y Atchley, 1995; Burgess y Lund, 2000) se basan en la coexistencia de términos lexicales en un corpus de texto. La distancia semántica, en este caso, se calcula de acuerdo a la similitud entre los vectores semánticos calculados para cada término. Este modelo está orientado a obtener la organización semántica de grandes corpus de términos lexicales y facilitan el reconocimiento de lo que tienen en común ciertas comunidades. Poco nos aportan sobre la organización particular de la memoria semántica de una comunidad lingüística particular. Más recientemente, Maki, McKinley y Thompson (2004) han generado una monumental base de datos a partir de Word Reference que contiene actualmente cerca de 50.000 pares de palabras y que contiene los valores para la distancia semántica, fuerza asociativa, y la similitud sobre la base de co-ocurrencia.

Ahora bien, cuando una persona estima la similitud semántica entre dos o más ideas puede establecer entre ellas diferentes tipos de relaciones semánticas. Su proximidad puede estar dada porque ambos conceptos presentan una relación inferencial entre sí, de modo que evocar un concepto supone la propagación de la activación hacia otro concepto con el que se halla vinculado lógicamente. Pero también ambos conceptos pueden compartir numerosos atributos por medio de los cuales se establezcan relaciones no necesariamente lógicas. Las semejanzas en los atributos compartidos entre dos conceptos pueden promover el establecimiento de relaciones analógicas —identificaciones por el predicado— que se hallan facilitadas por la presencia de activación en las etiquetas respectivas.

Del mismo modo, y tal como lo han señalado algunos estudios sobre relaciones semánticas (Bejar, Chaffin y Embretson, 1991; Mayor y López, 1995), distintos procesos cognitivos permiten elicitar relaciones parte-todo, contraste, causa-propósito, etc. De hecho, estos estudios recuperan trabajos que proponen taxonomías de, al menos, trece sistemas de clasificación diferentes, que varían, en el marco de la teoría de la propagación de la activación, del control ejecutivo que la persona pueda ejercer cuando produce un *prime* de la red semántica.

De hecho, la idea de ‘vitalidad’ propuesta por San Martín Nuñez (2011) avala lo antedicho. La existencia de palabras sinónimas está dada por la vigencia, el uso y las distintas apropiaciones que distintos sectores de una comunidad lingüística hacen de las palabras en función de sus propias necesidades comunicativas.

Así planteado el problema, la sinonimia no puede definirse de manera clara y concisa dada la diversidad de definiciones y perspectivas en lingüística. El estudio de la distancia semántica presenta varias metodologías de análisis que podrían contribuir a su elucidación. Entonces, en este estudio exploratorio, a partir de estas divergencias, proponemos entender que la sinonimia puede ser interpretada y calculada como una forma particular de distancia semántica cuando la proximidad es del orden de .9.

Este trabajo es la continuación de un estudio de medición de distancia realizado en conceptos correspondientes a sustantivos y emociones. Para una descripción más detallada véase “Synonymy as a particular case of semantic distance” (Zapico y Vivas, 2014).

2. Metodología

2.1. Materiales

Diccionario de sinónimos de la Real Academia Española (2014).

WordReference diccionario de sinónimos (2014).

Normas de Producción de Atributos Semánticos en Castellano Rioplatense (Vivas y otros, 2013).

2.2. Procedimiento

Se utilizaron dos procedimientos.

Por un lado se conformó el corpus de palabras consideradas sinónimos a partir de la intersección entre las 400 palabras de las Normas de Atributos y los diccionarios de referencia. Se obtuvieron 30 pares de palabras consideradas sinónimos (15 sustantivos y 15 adjetivos) y se calculó la distancia semántica para cada par de palabras.

Por otro lado, se realizó el cálculo de la distancia semántica para todos los pares de sustantivos y todos los pares de adjetivos. Los datos se colocaron en una matriz de modo uno, y se realizó análisis de cluster.

2.3. Tratamientos numéricos

Para la obtención de la distancia semántica se utilizó la técnica geométrica de comparación de vectores en el espacio euclidiano n-dimensional usual a partir del ángulo formado entre los mismos, representando el paralelismo el caso de mayor semejanza y la ortogonalidad el de mayor diferencia. El cálculo de dicho ángulo se realizó mediante la fórmula estándar de cociente entre el producto interno (usual, componente a componente) de los vectores y el producto de sus normas. Cabe aclarar que la idea de medir la distancia semántica mediante la conformación de dos vectores a partir del conjunto de atributos que definirán un determinado concepto fue propuesta originalmente por Kintsch (2001).

Con el resultado de la distancia entre cada par de concepto se generó una matriz cuadrada modo-1. En Análisis de Redes Sociales se denominan matrices modo-2 (2-dimensionales) a aquellas en las cuales sus filas y columnas corresponden a diferentes entidades. Por el contrario, una ma-

triz modo-1 es aquella en la cual filas y columnas refieren al mismo conjunto de entidades (Borgatti y Everett, 1997). Luego se aplicó el método de Johnson (1967) para analizar los cluster emergentes. Este método, que tiene como objetivo agrupar por similitudes y diferencias en un conjunto de n elementos, se aplica en una matriz simétrica $n \times n$. Las diferentes particiones de la matriz se ordenan de acuerdo al aumento (o disminución) de los niveles de similitud (o disimilitud). El algoritmo comienza con la partición de identidad (en el que todos los ítems están en agrupamientos diferentes). Luego junta los ítems que se hallen más similares (o menos diferentes) para luego considerarlos una sola entidad. El algoritmo continúa recursivamente hasta que todos los ítems se han unido en un solo grupo (partición completa).

2.4. Análisis

Para cada concepto se obtendrá la siguiente información: cada atributo será registrado con su frecuencia de aparición, que es el número de participantes que enumeraron ese atributo. Los atributos que son sinónimos se registrarán de un único modo, dentro del mismo y en distintos conceptos. Por ejemplo, el atributo “usado para el transporte”, “utilizado para transportar”, “se usa para el transporte” y “la gente lo usa para transporte” tendrá una sola codificación estándar, por ejemplo: usado_como_transporte. Es igualmente importante asegurarse de que los atributos que diferencian el significado tengan etiquetas verbales distintas. Las respuestas deben ser interpretadas conservadoramente y la validez de todas las interpretaciones, aún las más obvias, ser verificadas por múltiples colegas. Siguiendo las sugerencias de McRae y otros (2005), no alteraremos nombres de los atributos para evitar ambigüedad potencial. Por eso, la lógica conservadora indica que trataremos de mantener la mayor fidelidad a lo expresado por los participantes. De este modo, la matriz reflejará esta realidad y los usuarios podrán clasificar los archivos de conceptos por nombre de

atributo y hacer sus propios juicios, si les resulta conveniente. Se utilizará el método propuesto por McRae y otros (2005) para interpretar y organizar el registro de atributos. En todos los casos, cuantificadores como “generalmente” o “usualmente” serán sacados de las respuestas de los participantes. Consideramos que la información proporcionada por estos cuantificadores será emergente de la frecuencia de la producción. Esta circunstancia es especialmente válida para el tratamiento del Analizador de Definidoras: Cuando un participante presente un atributo adjetivo-sustantivo (e. g. <tiene cuatro ruedas>), su respuesta será dividida (<tiene ruedas> y <tiene 4 ruedas>) bajo el supuesto de que el participante proporcionó dos bits de información. Los atributos disyuntivos (e. g., <es verde o rojo>) serán divididos en <es verde> y <es rojo>. Si es necesario se usará un número de palabras clave para organizar y codificar los atributos, como lo han sugerido los autores. Este será un tratamiento guiado por los datos y ajustado a las categorías propuestas por Wu y Barsalou (2009, 2010).

3. Resultados

El primer procedimiento que se realizó fue obtener el corpus de palabras consideradas sinónimos, a partir de la búsqueda entre las 400 palabras de las Normas de Atributos aquellas que se encontraban definidas como sinónimos en el Diccionario de la Real Academia Española y en Word Reference Diccionario de sinónimos. Como resultado solo se obtuvieron 30 pares de palabras consideradas sinónimos que, a su vez, también tuvieran disponibles sus listas de atributos definidores de acuerdo a las Normas de Producción de Atributos Semánticos en Castellano Rioplatense (Vivas y otros, 2013). Con estas 60 palabras se calculó la distancia semántica en los pares considerados sinónimos por los diccionarios.

La tabla 1, a continuación, muestra los resultados de obtener la distancia semántica para 15

TABLA 1

Distancia semántica entre palabras

ADJETIVOS			SUSTANTIVOS		
PALABRA	SINÓNIMO	DISTANCIA	PALABRA	SINÓNIMO	DISTANCIA
Lindo	Atractivo	0,92	Baúl	Jaula	0,81
Bueno	Noble	0,83	Foco	Lámpara	0,81
Estudioso	Aplicado	0,81	Mesa	Escritorio	0,84
Divertido	Entretenido	0,86	Cigarrillo	Habano	0,85
Ingenioso	Perspicaz	0,74	Lápiz	Lapicera	0,86
Audaz	Valiente	0,79	Silla	Sillón	0,89
Inteligente	Instruido	0,82	Taza	Vaso	0,83
Irracional	Impulsivo	0,83	Tornillo	Tuerca	0,91
Feo	Repulsivo	0,91	Peine	Cepillo	0,93
Malo	Perverso	0,85	Vaso	Copa	0,89
Vago	Dejado	0,81	Caja fuerte	Candado	0,81
Aburrido	Desganado	0,75	Armario	Baúl	0,83
Lento	Gradual	0,78	Cómoda	Escritorio	0,79
Tonto	Estúpido	0,93	Hamaca	Cama	0,77
Tímido	Cobarde	0,82	Bota	Zapato	0,81

pares de conceptos considerados sinónimos por el diccionario de sinónimos Word Reference. En el sistema de medición que arroja esta forma de calcular la distancia semántica el número 1 se corresponde con la identidad y el 0 con la total ausencia de sinonimia.

Se puede observar que para la comunidad lingüística en estudio (jóvenes universitarios entre 20-40 años) presentan tres grupos correspondientes a los percentiles .90 y 1, .80 y .89 y .70 y .79. En lo que respecta a los adjetivos en el primer grupo aparecen los pares LINDO-ATRACTIVO y TONTO-ESTUPIDO (puntuaciones superiores a .9). En el primer par la comunidad lingüística los reconoce y utiliza recurrentemente como sinónimos, considerando que ambos son matices muy similares en torno a la idea de belleza. Similar el segundo par, donde prácticamente no hay dis-

tinción de significado, sino apenas una gradualidad para referirse a las cualidades negativas de personas.

En el otro extremo de la similitud, el segundo grupo, que corresponde a los pares de puntuaciones inferiores a .8, encontramos un total de cuatro casos. Sin realizar un análisis detallado de cada par, se puede observar que todos tienen en común el tratarse de palabras que, si bien comparten algún matiz de significado derivado a partir del tipo de objeto al cual suelen describir (ABURRIDO-DESGANADO suele usarse para describir una actitud, LENTO-GRADUAL para describir procesos), cada uno de los integrantes del par tiene un significado que le es propio y bien definido. El empleo de este tipo de palabras como sinónimos sería solo en casos muy específicos y en contextos lingüísticos particulares.

Finalmente, el tercer grupo (las puntuaciones fluctuantes entre .8 y .9) comprende pares que podrían ser considerados intercontextualmente. Pares como INTELIGENTE-INSTRUIDO refieren al mismo tipo de objeto al cual describen (refieren a las características de una persona) y además en algunos contextos no admiten distinción semántica. Es el uso particular y específico dado por los hablantes el que impone el matiz en este caso. No se trata de palabras definitivamente sinónimos pero sí pueden ser utilizadas como tal con relativa frecuencia en diversos contextos.

En lo que respecta a los sustantivos en el primer grupo (TORNILLO-TUERCA, PEINE-CEPILLO) encontramos los pares con mayor proximidad semántica. TORNILLO-TUERCA se hayan asociados por una función complementaria, sin embargo, para esta población poseen casi los mismos atributos, lo que, si bien en sentido estricto es un error, permite caracterizar este par de palabras como sinónimos. Ambas palabras pertenecen a una subcategoría muy específica. En el otro extremo COMODA-ESCRITORIO y HAMACA-CAMA, aunque guardan una alta similitud, poseen una menor proximidad semántica, por ende, la comunidad no los puede utilizar de manera intercambiable en diferentes contextos sin que se pierda el significado. La sinonimia parece estar dada, en este caso, por pertenecer a una misma categoría amplia (MUEBLES, LUGARES_DONDE_ACOSTARSE).

En el resto de los pares de palabras la distancia semántica fluctúa entre .8 y .9. Tienen como particularidad el hecho de pertenecer a una misma categoría (muebles, útiles, indumentaria, etcétera) y aún de pertenecer a una subcategoría específica dentro de estas (dentro de indumentaria BOTAS-ZAPATOS son calzado, VASO-COPA son cristalería, etcétera). Sin embargo, resulta dudoso que puedan ser utilizadas como sinónimos sin pérdida de sentido.

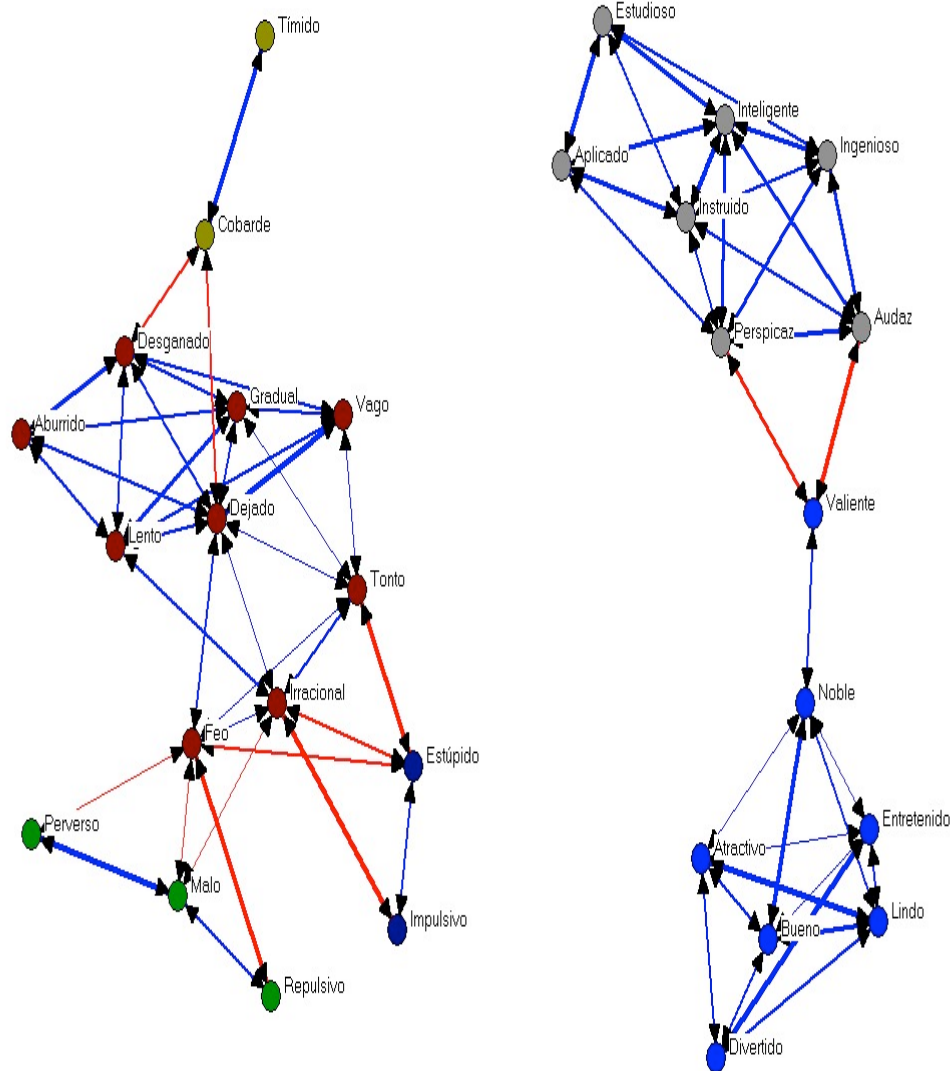
A partir de estos resultados se procedió a realizar otro análisis que, en función de la proximidad de las distancias, estableciese los agrupa-

mientos de palabras próximas entre sí a partir de un análisis de cluster. Se calculó la distancia semántica para todas las combinaciones de pares posibles $T = (N * N - 1 / 2)$. Donde las T combinaciones son el resultados del número de palabras consideradas por en número menos uno dividido 2 ya que las distancias son bidireccionales. Esta simetría no se verifica necesariamente para otro tipo de relaciones (e. g. cantidad de llamadas telefónicas entre dos personas). Luego, se colocaron los resultados de todas las combinaciones en una matriz de modo uno y se realizó un análisis de cluster.

El gráfico 1 pone de manifiesto la división en dos grandes sectores de significado. Al identificar los elementos que conforman estos dos grandes grupos, rápidamente se advierte que se puede hablar de la inclusión de adjetivos “positivos” por un lado, y “negativos” por el otro. El hecho de que no tengan conexión entre sí en ningún elemento (no poseen atributos en común ni siquiera con valencia opuesta) interdicta la creencia de que los antónimos, palabras de significado opuesto, se hallen próximas en una red semántica particular pudiendo estar asociados por su misma antonimia. Que LINDO y FEO (antónimos por excelencia) no tengan nada en común significa que probablemente activen significados asociados de manera diferencial, y los significados próximos a cada uno estén dados por proximidad semántica y no por asociación contextual o de oposición. Se podría explorar si el almacenamiento por oposición no es propio de la base reticular lexical aunque no de la semántica, ya que la mayoría de la evidencia colectada, incluso la experiencia personal subjetiva, no distingue entre la representación conceptual y la etiqueta lexical que la representa fonológicamente. En otras palabras, distintos grados de proximidad, desde la sinonimia hasta la distancia nula, determinarán la aparición de activaciones conjuntas en una red semántica particular. Para esta comunidad lingüística hay una fuerte asociación entre adjetivos positivos por un lado

GRÁFICO 1

Formación de cluster a partir de una matriz de distancias entre adjetivos



y negativos por el otro, no habiendo relación alguna entre ambos grupos.

En el grupo de la izquierda, el de los adjetivos negativos, se observa la formación de cuatro cluster distinguibles por el color de sus nodos. Uno de central importancia, constituido de 9 nodos, y otros tres periféricos y de tamaño considerablemente menor. El cluster central, que relaciona los conceptos asociados con la idea de la dejadez, el aburrimiento, el desgano, plantea fuertes asociaciones entre sus miembros, mostrando cómo muchos de estos términos pueden

ser considerados sinónimos en diversos contextos, o que al menos la activación de uno suele implicar la asociación de otro de ese grupo. Además, cada uno de los otros tres cluster se conforma a partir de al menos un vértice vinculado con el central. El que asocia el par de sinónimos TIMIDO-COBARDE lo hace con DESGANO y DEJADO; el que asocia ESTUPIDO-IMPULSIVO lo hace con FEO, IRRACIONAL y TONTO, y el que asocia PERVERSO-MALO-REPULSIVO con FEO e IRRACIONAL.

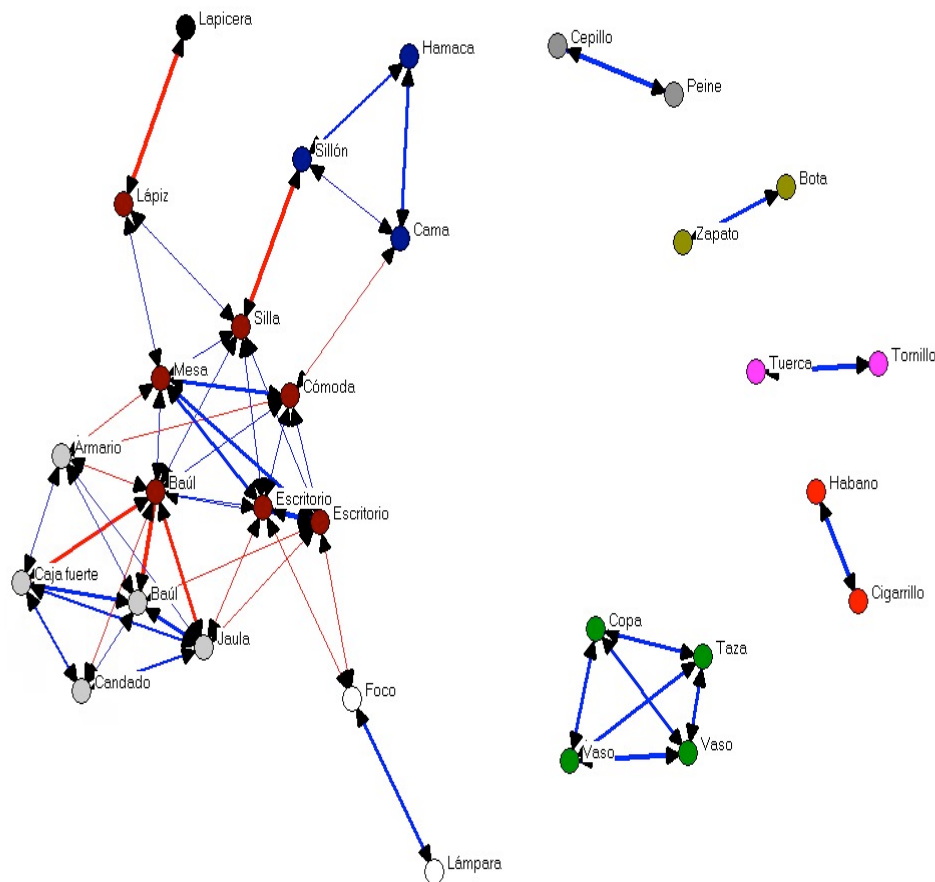
Hay que destacar, además, la importancia del nexo FEO-IRRACIONAL en este gráfico. Los

elementos que la comunidad les atribuye en el cluster son PERVERSO-MALO-REPULSIVO-ESTUPIDO-DEJADO. Es notable porque los adjetivos feo e irracional se utilizan para señalar como negativo las cualidades estéticas e intelectuales de un objeto, pero, en este caso, la palabra trae también aparejados adjetivos negativos provenientes de distintos ámbitos. Por ejemplo, PERVERSO, que se supone atribuye actitudes desviadas de lo socialmente esperado; MALO, que atribuye actitudes moralmente reprochables por lo social; ESTUPIDO, que atribuye un grado de incapacidad operativa. Se puede decir que el único adjetivo que era “lógicamente esperable” asociar con FEO sería REPULSIVO, no obstante, la comunidad atribuye otros adjetivos en alguna medida a aquello que es feo.

Por otro lado, en el grupo de la derecha, los adjetivos “positivos”, podemos observar la formación de dos grandes cluster donde se agrupan los significados de manera bien delimitada. Podemos hablar de que un grupo contiene los adjetivos que refieren a los procesos cognitivos de la persona: ESTUDIOSO-APLICADO-INTELIGENTE-INSTRUIDO-INGENIOSO-PERSPICAZ-AUDAZ. Todos estos adjetivos se caracterizan por describir la capacidad particular de un sujeto de enfrentar diversos tipos de situaciones en diversos contextos, pero están siempre orientados respecto a la voluntad y capacidad del sujeto como individuo. En el otro grupo, están los adjetivos que describen los procesos socio-emocionales de la persona: LINDO-BUENO-ATRACTIVO-NOBLE-ENTRETENIDO-DIVERTIDO son todos adjetivos con un fuerte color emocional y afectivo.

GRÁFICO 2

Formación de cluster a partir de una matriz de distancias en sustantivos



Al observar el gráfico 2, donde se observan los cluster que conforman los sustantivos, llama la atención la existencia de díadas que constituyen pares de palabras sinónimas que se relacionan exclusivamente entre sí. Un total de cuatro pares de palabras que solo se relacionan entre sí y la tétrada VASO-TAZA-COPA-BOL. Luego, hay un cluster central correspondiente a la categoría “muebles” que contiene dos sub-clusters que pueden ser identificados con una sub-categoría de muebles. El cluster que incluye los elementos MESA-SILLA-COMODA-ESCRITORIO-BAÚL identifica claramente una variedad de muebles. A su alrededor se puede observar la sub-categoría “muebles para descansar”, conformada por SILLÓN-HAMACA-CAMA, y la sub-categoría “muebles para guardar”, conformada por CAJA FUERTE-ARMARIO-JAULA-BAÚL. Cada uno de estos cluster contiene elementos que pueden ser considerados términos que las personas suelen utilizar como sinónimos, puesto que la especificidad dentro de la categoría base les da un conjunto de atributos comunes.

4. Discusión y conclusiones

A partir de los resultados expresados en la tabla 1, podemos sugerir que aquello que diversos autores han polemizado bajo el nombre de sinonimia (con todos sus matices, desde las posturas que plantean la inexistencia teórica de la sinonimia hasta los que afirman sin lugar a dudas que existe) puede comprenderse mejor desde una perspectiva más amplia y mensurable si se considera la sinonimia como un caso particular de la distancia semántica. Así pues, el problema establecido en la introducción conceptual de este trabajo se resuelve con una propuesta integradora. Los autores que afirman la existencia de la sinonimia lo realizan a partir de la formalización del diccionario y la reflexión tautológica que indica que, si las palabras se utilizan en algunos casos como sinónimos, deben tener la identidad de tales. Por otro lado, las propuestas negacionistas se amparan en un problema filosófico: ¿Puede

existir la coincidencia absoluta de significados entre dos palabras? Si es así, ¿por qué existen dos palabras diferentes? La respuesta a este interrogante apela al principio de la comunicación como rectora del lenguaje. Esta idea de Halliday de que el sistema lingüístico se basa en poner sonidos a significados. Si una palabra se utiliza es para comunicar algo. Si se utiliza otra diferente es para comunicar algo diferente. El sesgo teórico subyacente reside en no admitir un postulado esencial de la lingüística estructural: el significado que obtiene una palabra es a partir de su relación con las otras del sistema lingüístico. La propuesta que sostenemos es la gradualidad a partir de atributos compartidos. Compartir gran cantidad de atributos es a la vez ser parecido y diferente, pues solo compartiendo todos habría sinonimia total que, como caso extremo, refiere a la identidad. No obstante, compartir atributos pone en relación dos palabras, en mayor o menor medida. Esta posibilidad, más allá de su simpleza, no ha sido todavía debidamente considerada por los teóricos de la lingüística.

Al analizar los dos gráficos de agrupamientos de palabras de manera comparativa, se puede apreciar cómo la comunidad lingüística parece tener una mayor riqueza semántica a la hora de describir objetos concretos. Esta facilidad redundante en la formación de cluster bastante bien definidos por categorías bastante simples, además de pares de palabras sueltas y una tétrada. Por otro lado, a la hora de describir los adjetivos, los resultados parecen indicar que las dos grandes distinciones parecen solo estar dadas por los aspectos POSITIVO/NEGATIVO COGNITIVO/SOCIO EMOCIONAL. Es decir, una menor particularización de los conceptos otorga un gran matiz de significados donde ocurren asociaciones que, pese a estar naturalizadas, al ser analizadas a la luz del diagrama de relaciones no parecen lógicas, como los significados relacionados con FEO.

A la luz de la Teoría Neurocognitiva, la agrupación desigual de sustantivos y adjetivos permite reafirmar la hipótesis de que el sistema

lingüístico de un individuo se dispone como una red de significados donde las diferentes activaciones están dadas por distintos grados de asociación Lexema-Concepto.

La aplicación de este método de análisis sobre una matriz más poblada ($n \times n$ con $n=400$) que estamos desarrollando en la actualidad generará, sin duda, una configuración de agrupamientos con diferentes relaciones intra e inter cluster, que, en un sentido más amplio que el objetivo de este estudio, probablemente aporte a la comprensión de los deslices de significado entre diferentes palabras y en diferentes comunidades. Se espera realizar las pruebas ampliadas ya no solo en jóvenes, sino también en adultos, adultos mayores y niños.

5. Bibliografía citada

- BARSALOU, Lawrence W., 2008: "Grounded cognition", *Annual Review of Psychology* 59, 617-645.
- BARSALOU, Lawrence W., 2009: "Simulation, situated conceptualization and prediction", *Philosophy Transactions of the Royal Society B* 364, 1281-1289.
- DE MORA, Joaquín, 1855: *Colección de los sinónimos de la lengua castellana*, España: Imp. Nacional.
- DENDO Y ÁVILA, Manuel, 1756: *Ensayo de los sinónimos*, Madrid, España.
- GARCÍA, G., L. PAGNOTTA, E. PAZGON y J. VIVAS, 2013: "Poder de discriminación de los atributos semánticos. Mínima cantidad de descriptores requeridos para identificar conceptos emocionales" en V. JAICHENCO e Y. SEVILLA (coords.): *Psicolingüística en español. Homenaje a Juan Seguí*, Bs. As.: Editorial FFyL UBA.
- GONZÁLEZ MARTÍNEZ, Juan, 1989: "La sinonimia: Problema metalingüístico", *Anales de Filología Hispánica* 4, 193-210.
- GÓNZALEZ PÉREZ, Nieves, 1994: "Sinonimia y teoría semántica en diccionarios de sinónimos españoles de los siglos XVIII y XIX", *Revista española de lingüística* 24, 39-48.
- HERNÁNDEZ GARCÍA, Benjamín, 1997: "Sinonimia y Diferencia de significado", *Revista española de lingüística* 27, 1-31.
- JOHNSON, Samuel, 1967: "Hierarchical clustering schemes", *Psychometrika* 32, 241-253.
- KINTSCH, Walter, 2001: "Predication", *Cognitive Science* 25, 173-202.
- LAMB, Sidney, 2004: "Dimensions of the Territory of Neurolinguistics" en *Language and Reality*, London & New York: Continuum, 318-323.
- LAMB, Sidney, 2004b: "On the Perception of Speech" en *Language and Reality*, London & New York: Continuum, capítulo 18.
- LAMB, Sidney, 1999: *Senderos del cerebro*, traducción de José María Gil y Adolfo Martín García (2011), Argentina: Eudem.
- LEVY, Bernard, 1942: "Libros de sinonimia", *Hispanic Review* X/4, 285-313.
- LÓPEZ HUERTA, José, 1789: *Ensayo de la posibilidad de fixar la significación de los sinónimos de la lengua castellana*, Madrid.
- LYONS, John, 1979: *Introducción a la lingüística teórica*, Barcelona: Teide.
- MAKI, W. S., L. N. MCKINLEY y A. G. THOMPSON, 2004: "Semantic distance norms computed from an electronic dictionary (WordNet)", *BRMI&C* 36 (3), 421-431.
- MARTÍNEZ LÓPEZ, Juan, 1997: "Concreción y abstracción en el estudio de la sinonimia", *Moenia* 3, 157-169.
- MAYOR, R. y R. LÓPEZ, 1995: "Relaciones Semánticas" en *Anexos de la Revista de Psicología del Lenguaje*, anexo 2, Madrid: Departamento de Psicología Básica, U. C. M.
- MCRÆ, Kenny y Serge BOISVERT, 1988: "Automatic Semantic Similarity Priming", *Journal of Experimental Psychology: Learning, Memory and Cognition* 24, 3, 558-572.

MCRÆ, Ken y Michel JONES, 2013: "Semantic Memory" en D. REISBERG (ed.): *The Oxford Handbook of Cognitive Psychology*.

MCRÆ, K., G. CREE, M. SEIDENBERG y C. McNORGAN, 2005: "Semantic feature production norms for a large set of living and nonliving things", *Beh Research Methods* 37, 547-559.

RAE, 2014: *Diccionario de la lengua española* [www.rae.es].

SALVADOR, Gregorio, 1985: "Sí hay sinónimos" en Gregorio SALVADOR (ed.): *Semántica y lexicología del español*, Madrid: Thomson Parainfo, 51-66.

SAN MARTÍN NÚÑEZ, Abelardo, 2011: "Voces de origen lunfardo en el registro festivo del diario chileno La Cuarta", *Onomázein* 23, 105-147.

VERGARA GÓNZALEZ, Carlos, 2011: "Anexos a clases", anexos I-V [disponible en skemman.is/stream/get/1946/15186/36874/2/Anexos.pdf].

VIVAS, J. R., A. COMESAÑA, A. GARCÍA CONI, L. VIVAS, M. YERRO, 2011: "Distribución de los atributos semánticos en función del tipo de categoría y campo semántico. Resultados preliminares para la confección de normas de atributos" en M. C. RICHAUD y V. LEMOS (comp.): *Psicología y otras ciencias del comportamiento. Compendio de investigaciones actuales I*, Libertador San Martín: Editorial Universidad Adventista del Plata, 311-333.

WORDREFERENCE, 2014: *Diccionario de Sinónimos* [www.wordreference.com/es].

ZAPICO, Martín Gonzalo y Jorge Ricardo VIVAS, 2014: "Synonymy as a particular case of semantic distance", *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, vol. 19, n. 40, 253-266.