



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY



Facultad de  
Psicología  
UNIVERSIDAD DE LA REPÚBLICA

# **Desempeño de modelos semánticos espaciales y redes semánticas asociativas en una tarea de analogías en español rioplatense**

Hernán Quintero Santos

Tutores: Álvaro Cabana y Juan Carlos Valle Lisboa

Revisor: Roberto Aguirre

Facultad de Psicología

Universidad de la República

2021

## Introducción

Entrenando modelos computacionales a partir de grandes corpus de textos o normas creadas a través de tareas relacionadas con el lenguaje es posible crear diferentes tipos de representaciones del significado de las palabras de los lenguajes naturales. Estas representaciones pueden utilizarse para realizar tareas lingüísticas y con el fin de compararlas con el desempeño humano (ver por ej.: Landauer et al., 2011).

En 1964 Osgood introdujo la idea de que el significado de una palabra se puede representar como una coordenada en un espacio semántico. Este espacio estaría formado de dimensiones cuantitativas que corresponden a la ubicación entre un par de calificadores opuestos, como *bueno-malo* o *fuerte-débil*. Al aumentar el número de dimensiones las ubicaciones posibles en el espacio semántico aumentan exponencialmente. Palabras con significados similares tendrían coordenadas cercanas, lo que nos permitiría agruparlas a partir de su cercanía en este espacio. Estas dimensiones pueden obtenerse a través de tareas de asignarle calificadores a las palabras llenando un espacio en blanco. Osgood (1964) utiliza este marco en un estudio translingüístico y transcultural y al realizar un análisis factorial encuentra evidencia de que las dimensiones se agrupan consistentemente en tres factores principales: un factor evaluativo (ej.: *bueno-malo*, *agradable-desagradable*), un factor de potencia (ej.: *fuerte-débil*, *pesado-ligero*) y un factor de actividad (ej.: *rápido-lento*, *activo-pasivo*).

Tversky (1977) propone y presenta evidencia a favor de la teoría del contraste (contrast theory) o *feature model* en la que la similitud o la distancia entre dos conceptos u objetos depende de la presencia o ausencia de numerosos rasgos (*features*) cualitativos (binarios o binarizables) en común o distintivos entre ellos, y se calcula utilizando operaciones de la teoría de conjuntos. Estos rasgos se obtienen a través de tareas de producción de rasgos (feature production tasks; ver p. ej. McRae et al., 2005) en las que se les pide a los sujetos que nombren la mayor cantidad de rasgos que puedan asignarle a cada palabra. El autor plantea esta teoría en respuesta a los modelos geométricos de similitud, que representan a los objetos como puntos en un espacio de coordenadas y asumen que la disimilitud es una especie de distancia geométrica en un espacio euclídeo. Una función  $\delta$  de distancia métrica en un espacio euclídeo le asigna un número no negativo a cada par de puntos cumpliendo los siguientes axiomas:

Minimalidad, que implica que la similitud entre un objeto y sí mismo es 0 para todos los objetos:

$$\delta(a, b) \geq \delta(a, a) = 0$$

Simetría, que implica que la similitud entre dos objetos es la misma en ambas direcciones:

$$\delta(a, b) = \delta(b, a)$$

La desigualdad triangular, que implica que, si los objetos  $a$  y  $b$  son similares, y  $b$  y  $c$  son similares, entonces  $a$  y  $c$  no pueden ser demasiado disímiles:

$$\delta(a, b) + \delta(b, c) \geq \delta(a, c)$$

Sin embargo, el autor presenta evidencia de que estos axiomas no siempre se cumplen para todas las medidas de similitud psicológica.

Asignar un valor manualmente a cada palabra tanto en el modelo de Osgood (1964) como en el de Tversky (1977) requiere una gran inversión de tiempo y recursos humanos. Si además consideramos que el esfuerzo debería replicarse para cada idioma y que dentro de una misma lengua el uso y significado de una palabra puede variar a lo largo del tiempo, las regiones, el registro e incluso el tema, se hace evidente la necesidad de contar con métodos automatizados aplicables a gran escala para lograr aplicaciones prácticas.

Por ejemplo, en el análisis semántico latente (LSA; Landauer y Dumais, 1997; Landauer et al., 2011) un corpus de gran tamaño es convertido en una enorme matriz de coocurrencias entre palabras y documentos en la que cada fila representa una palabra, cada columna representa un texto y en cada celda se pone el número de veces que dicha palabra aparece en dicho texto. Luego se aplican diversas transformaciones a la matriz, por ejemplo, para controlar el efecto de las palabras que aparecen con más frecuencia. Finalmente, se utiliza la descomposición en valores singulares (DVS) para reducir la cantidad de columnas de la matriz y se mantienen solamente las principales dimensiones. Esto da una representación en forma de vectores (las filas de la matriz) del significado de cada palabra. A este tipo de representaciones del significado en forma de vectores se les suele llamar *word embeddings*. También se puede representar un texto en forma de vector. Utilizando una medida de proximidad entre estos vectores se puede cuantificar la similitud semántica entre dos palabras, textos o entre una palabra y un texto. Una de las medidas más utilizadas es el coseno del ángulo que forman dos vectores en el espacio semántico. La reducción en la cantidad correcta de dimensiones (normalmente manteniendo alrededor de 300 se obtiene el mejor desempeño en diversas tareas y aplicaciones prácticas) es lo que permite el aprendizaje inductivo que lleva a la generalización de lo aprendido a situaciones nuevas y la

utilización de más o menos dimensiones lleva a un rendimiento menor (para una explicación de cómo funciona este procedimiento, ver p. ej. Valle Lisboa y Mizraji, 2007). Es importante destacar que estas dimensiones no tienen un significado interpretable (al menos *a priori*), a diferencia de las del modelo de Osgood (1964).

El algoritmo de aprendizaje del LSA (Landauer y Dumais, 1997) tiene la particularidad de que no solo aprende el significado de las palabras a partir de las ocasiones en las que aparecen, sino también a partir de los ejemplos en los que no aparecen. Esto tiene especial relevancia si tomamos en cuenta que el LSA aprende a un ritmo similar a un humano ("leyendo" una cantidad similar de texto), ya que, aunque no se puede afirmar que el cerebro realice una descomposición en valores singulares, aporta una posible respuesta al debate sobre el llamado problema de Platón o problema de la pobreza del estímulo, de cómo aprendemos tanto a partir de tan poca información.

Griffiths, Steyvers y Tenenbaum (2007) comparan la representación espacial creada utilizando LSA con utilizar una representación no espacial utilizando el *topic model* (Steyvers y Griffiths, 2007), que genera una representación del significado de las palabras basada en temas (*topics*), similar a la representación basada en características del modelo de Tversky (1977), pero generada a partir de una matriz de coocurrencias entre palabras y documentos, lo que permite crearlas de manera automatizada. Cada tema es una distribución de probabilidad de aparición en dicho tema para cada palabra del corpus. A diferencia del LSA, en el que a cada palabra le corresponde un único vector, independientemente de la polisemia, en el *topic model* el significado de cada palabra está distribuido a lo largo de todos los tópicos, lo que permite representar los diferentes significados de cada palabra por separado. Los autores comparan el desempeño de utilizar el coseno y el producto interno de los vectores de LSA como métrica de distancia con las probabilidades generadas por el *topic model* para predecir normas de asociación generadas por humanos. El *topic model* se desempeña mejor que el producto interno, que a su vez supera al coseno del LSA. Atribuyen esta diferencia a que ambas métricas respetan los axiomas de simetría y desigualdad triangular introducidos por Tversky (1977), mientras que las normas de asociación no lo hacen. La diferencia entre el rendimiento del producto interno y el coseno la atribuyen a que el primero es sensitivo a la frecuencia de las palabras mientras que el segundo no.

Siguiendo la advertencia de Griffiths et al. (2007) de que sus resultados pueden deberse al uso del coseno como medida de distancia y no a las representaciones espaciales per se, Jones et al. (2018) argumentan que el espacio semántico es una representación de la

estructura de la memoria semántica, pero que el coseno no representa el proceso de recuperación de la memoria, por lo que no debe considerarse que es lo que las personas hacen al realizar una tarea. Los autores utilizan el mismo corpus y las mismas normas de asociación que Griffiths et al. (2007) para comparar sus resultados utilizando no solo LSA, sino también combinando la regla de elección de Luce (1959), que simula como los humanos eligen entre varias opciones, con LSA y otros dos espacios semánticos creados a partir del mismo corpus, uno utilizando el índice de Jaccard y utilizando el modelo BEAGLE. Utilizaron una versión simplificada de la regla de elección de Luce y no incluyeron el escalamiento exponencial basado en la ley universal de distancia y similitud percibida de Shepard (1987) que se suele utilizar en combinación con la regla de Luce. Aunque la utilización de la regla de Luce no generó grandes cambios en el desempeño del LSA, el desempeño del índice de Jaccard y BEAGLE al imitar las propiedades de las normas de asociación fue suficiente para que los autores concluyan que no se debe descartar a las representaciones espaciales de la semántica como modelos viables.

Mikolov et al. (2013) demostraron que en los modelos espaciales hay relaciones que están presentes en forma de offsets vectoriales cuando se da de manera consistente que la diferencia entre dos palabras  $a$  y  $b$  que comparten una cierta relación semántica o sintáctica es similar a la diferencia en las representaciones entre un par de palabras  $c$  y  $d$  que compartan la misma relación, por ejemplo, que la diferencia entre las representaciones de *hombre* y *rey* es similar a la diferencia entre las representaciones de *mujer* y *reina*. A esto lo llamaron regularidades lingüísticas. Estas regularidades permiten que ante una analogía del tipo  $a:b::c:d$ , que se expresaría “ $a$  es a  $b$  lo que  $c$  es a  $d$ ”, se pueda utilizar aritmética de vectores en las representaciones de tres de las palabras para predecir la representación de la cuarta, como en el conocido ejemplo de  $rey - hombre + mujer = reina$ .

Los modelos espaciales semánticos (junto con el *topic model*) son solo una de las formas de representar el significado de las palabras o conceptos. Otra de ellas es utilizando redes semánticas, como en la *spreading-activation theory* de Quillian (1962, 1967), que Collins y Loftus (1975) extienden con asunciones adicionales. En estas redes cada concepto es un nodo que está ligado a otros nodos (conceptos) por enlaces dirigidos, que a su vez son conceptos y tienen un valor numérico. Mediante un proceso en el que al activar uno o más nodos la activación se propaga hacia el resto de la gráfica a través de los enlaces entre los nodos se obtiene un modelo del funcionamiento de fenómenos como el priming y la memoria semántica. Una de las asunciones añadidas por Collins y Loftus (1975) es que la similitud semántica (*similarity* o *relatedness*) entre dos nodos está dada por el total de enlaces que comparten.

Más recientemente, De Deyne et al. (2019) recolectaron las normas de asociación para más de 12.000 claves (*cues*) en inglés, el proyecto Small World of Words en inglés (SWOW-EN). En la tarea de asociación utilizada para crear estas normas se les presenta a los sujetos una palabra (la clave) y se les pide que respondan con la o las primeras palabras que se le vengan a la mente (en este caso las tres primeras palabras). Estas normas son útiles para el estudio de las representaciones internas y el significado de las palabras. También es posible utilizarlas para construir redes semánticas. Los autores hacen esto y comparan el desempeño de las redes creadas a partir del SWOW-EN y de otras dos normas más antiguas y pequeñas para predecir los datos de similitud semántica utilizando tres medidas diferentes, es decir tres formas distintas de calcular la similitud a partir de las redes. Encuentran que el SWOW-EN tiene un buen desempeño y que es mayor al de las otras normas, pero que la medida utilizada influye considerablemente en el rendimiento.

En este estudio compararemos el desempeño de modelos espaciales semánticos entrenados a partir de corpus en español y un modelo espacial creado a partir del SWOW para el español rioplatense (SWOW-ES-RP: Cabana et al., en preparación) con las respuestas dadas por los participantes a una tarea de analogías del tipo “*a* es a *b* lo que *c* es a *d*”, en la que *c* es un espacio a completar con la palabra que crean más adecuada.

## Métodos

Se realizó una encuesta web en la que se les pidió a los participantes que dieran información acerca de su género, su máximo nivel educativo alcanzado y su lengua materna.

Luego se les presentaron progresivamente 40 ítems seleccionados al azar de un total de 79. Cada uno de estos ítems contiene una palabra clave de ejemplo y una palabra objetivo de ejemplo. Estas palabras guardan una cierta relación entre sí, que es diferente en cada ítem y que los participantes debían deducir. También se les presentó otra palabra clave y los participantes debían llenar un espacio en blanco con la palabra objetivo que les parecía conveniente, de modo que la relación entre estas dos palabras sea equivalente a la relación entre las palabras de ejemplo. Por ejemplo, al usuario puede presentarse el ítem “Buenos Aires es a Argentina lo que \_\_\_ es a Brasil”, en donde Buenos Aires es el objetivo de ejemplo, Argentina es la clave de ejemplo y Brasil es la clave. De esos ítems, 22 fueron tomados del test SVAT (Gallassi et al., 2014) y traducidos del italiano al español rioplatense, mientras que el resto fueron creados para este estudio. Los ítems tomados del SVAT tienen una respuesta correcta predefinida. El resto de los ítems varían desde ítems en los que se

consideró que había una única respuesta obvia hasta ítems pensados para tener múltiples respuestas posibles o no tener ninguna respuesta obvia predefinida. Los ítems fueron pensados para expresar relaciones diversas sin seguir *a priori* ningún sistema de clasificación de los tipos de relación.

## Participantes

La encuesta estuvo disponible en una web hosteada en el servidor del Centro de Investigación Básica en Psicología (CIBPsi) para cualquier persona que quisiera realizarla ([http://digital.psico.edu.uy/hernan\\_1/experiment.html](http://digital.psico.edu.uy/hernan_1/experiment.html)). La búsqueda de participantes se realizó a través de los contactos personales de los investigadores asociados y de publicidad en redes sociales. En total participaron 198 hablantes nativos de español rioplatense.

## Análisis de datos

Se compararon los resultados comportamentales de los participantes con las predicciones obtenidas de extrapolar la relación de las palabras ejemplo a la palabra clave utilizando los modelos espaciales semánticos y el SWOW. Para esto se utilizó el paquete Gensim en Python (Rehurek y Sojka, 2011).

Los embeddings utilizados para los análisis fueron: FastText-SUC (Cañete, 2019b), creados utilizando el algoritmo FastText (Bojanowski et al., 2017) en el Spanish Unannotated Corpora (Cañete, 2019a); FastText-SBWC (Universidad de Chile, 2019), creados utilizando FastText en el Spanish Billion Word Corpus (Cardellino, 2016); GloVe-SBWC (Universidad de Chile, 2019), creados utilizando el algoritmo GloVe (Pennington et al., 2014) en el Spanish Billion Word Corpus; FastText-wiki-ES (Bojanowski et al., 2017), creados utilizando el algoritmo FastText en el dump de la Wikipedia en español, y Word2Vec-SBWC (Cardellino, 2016), creados utilizando el algoritmo word2vec (en su versión Skip-gram) del paquete Gensim en el Spanish Billion Word Corpus.

Como red semántica se utilizó la versión en español rioplatense del proyecto Small World of Words, que incluye normas de asociación para más de 13.000 palabras recolectadas de más de 70.000 participantes de Uruguay y Argentina (Cabana et al., 2020). En concreto, se emplearon vectores derivados de la matriz de adyacencia, obtenidos mediante el procedimiento descrito en Steyvers et al. (2004). Estos vectores forman parte de un manuscrito en preparación (Cabana et al., 2020).

Seguendo a Mikolov et al. (2013), para las predicciones obtenidas de los modelos espaciales se realizó la operación:

$$\text{COS}(v_{\text{objetivo}E_j} - v_{\text{clave}E_j} + v_{\text{clave}}, v_{\text{objetivo}})$$

En la que  $v_{\text{objetivo}E_j}$  es el vector correspondiente a la palabra objetivo de ejemplo,  $v_{\text{clave}E_j}$ , el vector de la palabra clave de ejemplo,  $v_{\text{clave}}$ , el vector de la palabra clave, y  $v_{\text{objetivo}}$ , el vector de la palabra objetivo. Por cada ítem se realizó esta operación usando cada palabra presente en el embedding como palabra objetivo y tomando la palabra (o palabras) que diera un mayor coseno como la predicción.

Para evaluar el desempeño de los embeddings se utilizaron tres métricas: en la primera se tomó en cuenta solamente la respuesta más común dada por los participantes para cada ítem y se comparó con la primera respuesta obtenida al utilizar los embeddings. Al dividir la cantidad de aciertos sobre el total de ítems se obtuvo la precisión de las predicciones.

Para la segunda métrica se convirtió la lista de predicciones de cada ítem en una distribución de probabilidades utilizando la regla de elección introducida por Luce (1959), que define que la probabilidad de responder al estímulo  $S_i$  con la respuesta  $R_j$  es:

$$p(R_j|S_i) = \frac{\beta_j \eta_{i,j}}{\sum_{k \in M} \beta_k \eta_{i,k}}$$

En donde  $\beta_j$  es el sesgo de la respuesta para el ítem  $j$  y  $\eta_{ij}$  es la similitud entre los estímulos  $i$  y  $j$ .

Esta fórmula se usó combinada con la ley de distancia y similitud percibida de Shepard (1987), que permite simular el hecho de que la probabilidad empírica de seleccionar los ítems toma la forma de un descenso exponencial. A esta combinación generalmente se la conoce como axioma de elección de Shepard-Luce (ver Jones et al., 2017):

$$p(R_j|S_i) = \frac{\beta_j e^{-\lambda d(S_i, R_j)}}{\sum_{k \in M} \beta_k e^{-\lambda d(S_i, R_k)}}$$



En donde  $d$  es una función de distancia psicológica y  $\lambda$  es un parámetro libre para la pendiente de la exponencial, un mayor valor de  $\lambda$  implica una pendiente más pronunciada, y por lo tanto un mayor aumento en la diferencia entre un valor y el siguiente. Para el  $\beta$  se utilizó inicialmente el logaritmo de la frecuencia de la palabra, tomado del SUBTLEX-ESP (Cuetos et al., 2011), en una versión con correcciones posteriores a la publicación del artículo original tomada de <http://crr.ugent.be/archives/679>. Sin embargo, varias de las respuestas no se encontraban en el SUBTLEX-ESP y en los análisis el valor del  $\beta$  tenía mayor influencia que la medida de distancia, lo que llevó a descartar el uso de este valor.

Asimismo, se utilizó el coseno entre los vectores de los estímulos  $S_i$  y  $R_j$ , que es una medida de similitud y no de distancia, por lo que la fórmula utilizada finalmente fue:

$$p(R_j|S_i) = \frac{e^{\lambda \cos(S_i, R_j)}}{\sum_{k \in M} e^{\lambda \cos(S_i, R_k)}}$$

Por cada ítem se utilizó la divergencia de Kullback-Leibler (KL, también llamada entropía relativa; introducida por Kullback y Leibler, 1951) entre las probabilidades empíricas de las respuestas dadas por los sujetos y las distribuciones de probabilidad obtenidas a partir de aplicar el axioma de Shepard-Luce a las predicciones obtenidas a partir de los embeddings. KL representa la cantidad de información necesaria para transformar una distribución teórica  $Q$  en su distribución real  $P$ . En este estudio se utilizó la siguiente implementación (ver Brownlee, 2019):

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}$$

Donde el operador  $||$  representa “divergencia”,  $P$  es la distribución de probabilidad empírica de las respuestas de los sujetos y  $Q$  la distribución de probabilidad de las predicciones obtenidas de los embeddings.

Esta medida se utiliza para cuantificar la divergencia entre dos distribuciones de probabilidad, lo que nos permitirá evaluar mejor los ítems en los que no haya una única respuesta predominante.

Por último, se promedió la mediana de la posición que las 5 primeras respuestas de cada ítem tenían dentro de las 100 primeras predicciones correspondientes a dicho ítem.

## Resultados

En primer lugar, se calculó la precisión, entendida como la proporción de veces que la primera predicción dada por el embedding coincidía con la primera respuesta dada por los sujetos. Como se puede observar en la Tabla 1, el SWOW fue el tercero mejor posicionado de los seis embeddings y obtuvo un valor (0,126761) casi en el centro entre el embedding con mejor desempeño (el FastText-SUC, con una precisión de 0,189873) y el peor (el FastText-wiki-ES, con 0,072464).

Embedding	Precisión
FastText-SUC	0,189873
FastText-SBWC	0,151899
GloVe-SBWC	0,075949
FastText-wiki-ES	0,072464
Word2Vec-SBWC	0,101266
SWOW-ES-RP	0,126761

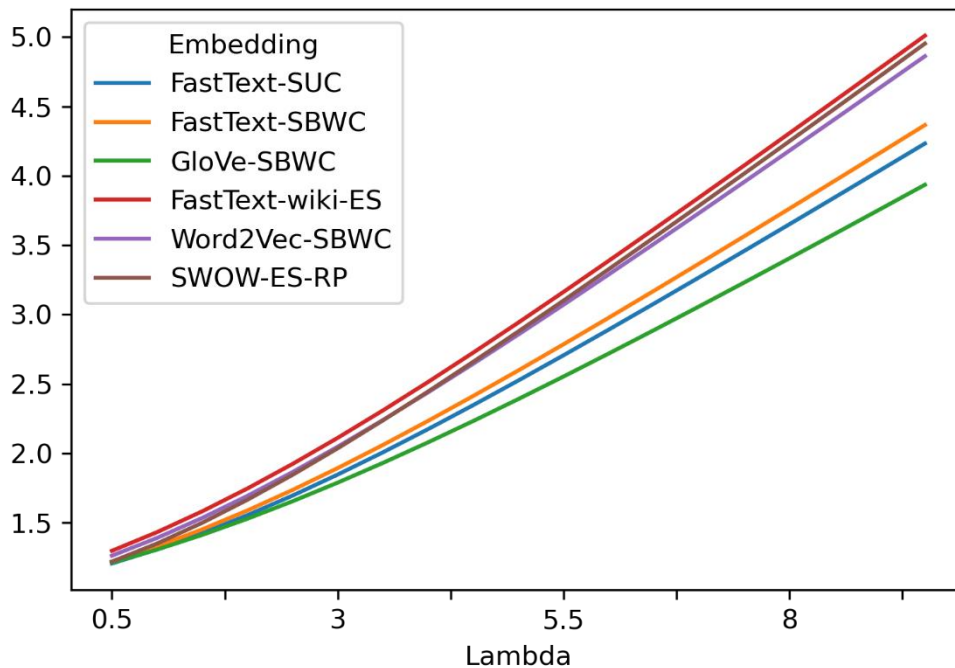
**Tabla 1:** Precisión de cada embedding

Consideramos que la precisión es una medida muy limitada, ya que solo toma en cuenta la primera de las respuestas y la primera de las predicciones, por lo que decidimos utilizar la divergencia de Kullback-Leibler para comparar la distribución de las respuestas con la distribución obtenida de aplicar el axioma de Shepard-Luce a las distancias de las predicciones. Con esto esperábamos obtener un reflejo más fiel de la verdadera capacidad de los embeddings para captar las relaciones semánticas.

El axioma de Shepard-Luce tiene un parámetro libre, el lambda, que permite aumentar la importancia relativa de la primera opción (la primera predicción en este caso) al mismo tiempo que se reduce el peso del resto de las opciones, lo que le aproxima forma de la distribución a un descenso exponencial. Calculamos la divergencia KL para lambdas de 0,5 a 9,5, en intervalos de 0,5.

En los valores más bajos de lambda todos los embeddings dieron resultados similares. A medida que el lambda aumenta los embeddings empiezan a tener valores más altos de KL, lo que implica un peor desempeño, pero también aumenta la diferencia entre los embeddings, con el GloVe-SBWC posicionándose como el mejor y el FastText-wiki-ES, el SWOW-ES-RP y el Word2Vec-SBWC teniendo el peor desempeño, con valores similares entre ellos.

Un lambda más alto favorece a los casos en lo que haya una respuesta predominante y la primera predicción coincida con dicha respuesta, y penaliza el resto de los casos, por lo que el aumento del valor de KL a medida que aumenta el lambda puede explicarse porque para la mayoría de los ítems no hubo una respuesta claramente predominante, además del hecho de que los casos en los que la primera respuesta y la primera predicción coincidían fueron minoritarios.



**Figura 1:** Kullback-Leibler en función del valor de lambda utilizado en la fórmula de Shepard-Luce

Al comparar las predicciones y respuestas para cada ítem notamos que una limitación de la divergencia KL es que no hace distinción entre palabras que no coinciden y no están relacionadas semánticamente y palabras que no coinciden pero que guardan una cierta relación semántica. Esto implica que un ítem sin coincidencias exactas, pero con predicciones razonables, y un ítem con predicciones completamente erradas podrían tener la misma divergencia KL. Para tratar de subsanar esta limitación, tomamos las 5 primeras respuestas de cada ítem y para cada embedding les asignamos el valor que ocupaba dicha palabra en el ranking de predicciones (de modo que, si la respuesta era *hormiga*, y *hormiga* era la predicción número 8, le asignamos el valor 8, si la respuesta no estaba en las 100 primeras predicciones le asignamos el valor 101), tomamos la mediana de esos valores dentro de cada ítem, y luego tomamos la media de las medianas dentro de cada embedding. Con esta métrica esperábamos tener un mejor reflejo de qué tan pertinentes eran las predicciones de los embeddings, aunque no hubieran acertado exactamente.

FastText-SBWC	FastText-SUC	FastText-wiki-ES	GloVe-SBWC	SWOW-ES-RP	Word2Vec-SBWC
93,500000	89,275641	98,139706	88,525641	67,100000	94,153846

**Tabla 2:** Promedio de la mediana de la posición de las 5 primeras respuestas dentro de las predicciones de cada ítem.

El SWOW fue el embedding que tuvo una mejor puntuación, con respuestas en promedio de 67,1 en el ranking de predicciones. Esto corroboró nuestra observación de que en ocasiones el SWOW daba respuestas más pertinentes que el resto de los embeddings, aunque no tuviera una puntuación sustancialmente diferente en la divergencia KL ya que no daba una coincidencia exacta.

## Discusión

Basados en la precisión de las predicciones, el embedding basado en el SWOW tiene un desempeño intermedio entre los embeddings que mejor y que peor desempeño tuvieron, mientras que utilizando en la divergencia de Kullback-Leibler puntuó sistemáticamente peor que el resto y en la posición de las respuestas dentro de las predicciones obtuvo la mejor puntuación.

Una posible razón por la cual el desempeño del SWOW es menor al de otros embeddings es que el porcentaje de las palabras encontradas en las respuestas que no se encontraban en el SWOW es significativamente mayor al porcentaje de respuestas que no se encontraban en los otros embeddings, a causa de la diferencia de tamaño entre sus vocabularios (más de 12.000 palabras en SWOW contra alrededor de 1.000.000 en los otros embeddings), por lo que si se extendiera el vocabulario del SWOW es de esperar que su desempeño mejore significativamente.

En cambio, hay tres razones que podrían explicar el mejor desempeño del SWOW frente a algunos embeddings y que permiten teorizar que el SWOW podría alcanzar o superar el mejor desempeño si incluyera todas las respuestas en su vocabulario.

En primer lugar, hay un efecto de corpus, mientras que el SWOW se basa en respuestas orales del español rioplatense al igual que las respuestas a la tarea del presente estudio, el

resto de los embeddings se basa en corpus escritos que incluyen otras variedades del español y otras situaciones de comunicación.

Un ejemplo de esto es que en el ítem abogado:corte::doctor:? el SWOW predice palabras relacionadas a la medicina, de manera similar a los participantes de la tarea. En cambio, el embedding FastText-wiki-ES (basado en el dump de la Wikipedia en español) predice palabras como *Gallifrey*, *Tardis*, *daleks* y *cybermen*. Esto probablemente se deba a que la representación de la palabra *doctor* está muy asociada al nombre del protagonista de la serie *Doctor Who*, algo poco común en el español rioplatense hablado, donde se usa *doctor* como sinónimo de *médico*, pero más esperable de una enciclopedia que evitaría este uso técnicamente impreciso.

En segundo lugar, tanto la tarea de analogía realizada para este estudio como la tarea de asociación libre en la que se basa el SWOW son tareas de recuperar de la memoria palabras individuales a partir de un estímulo o estímulos. Esta similitud entre tareas, que es una diferencia entre el embedding SWOW-ES-RP y el resto de los embeddings, basados en corpus escritos, podría ser uno de los factores que influyen en el desempeño del SWOW-ES-RP frente a los otros embeddings.

Por último, un vocabulario de menor tamaño significa menos opciones a la hora de decidir entre todas las predicciones posibles, por lo que si las respuestas se encuentran en el vocabulario del embedding, un vocabulario menor implica menor probabilidad de introducir ruido en las predicciones. Como se mencionó anteriormente, la tarea de asociación libre, de la que se obtienen los vectores del SWOW, guarda ciertas similitudes con la tarea realizada en este estudio, tanto en corpus como en tipo de tarea, por lo que es de esperarse una mayor probabilidad de coincidencia entre el vocabulario del embedding y las respuestas dadas por los sujetos en comparación a los otros embeddings.

El que hayamos encontrado diferentes desempeños relativos dependiendo de la métrica utilizada y que hayamos notado diferencias cualitativas al observar las predicciones y compararlas con las respuestas lleva a pensar que los modelos espaciales pueden representar diferentes aspectos del significado dependiendo de cómo sean construidos.

En futuras investigaciones sería conveniente extender el SWOW-ES-RP para utilizar una versión del embedding comparable al resto de los embeddings en términos del vocabulario relevante para el estudio o bien restringir el análisis a las palabras objetivo que están actualmente en todos los embeddings. Se necesita una investigación más exhaustiva para determinar las condiciones que afectan el desempeño relativo de los tipos de embedding.

## Referencias bibliográficas

- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.
- Brownlee, J. (2019). *How to Calculate the KL Divergence for Machine Learning*. Recuperado el 12 de noviembre del 2021 de <https://machinelearningmastery.com/divergence-between-probability-distributions/>
- Cabana, Á., Zugarramurdi, C., & De Deyne, S. (2020). The “Small World of Words” Free Association Norms for Rioplatense Spanish. En Molina Mejía, J. M., Valdivia Martin, P., y Venegas Velásquez, R. A. (eds.), *Actas III Congreso Internacional de Lingüística Computacional y de Corpus - CILCC 2020 y V Workshop en Procesamiento Automatizado de Textos y Corpus - WoPATeC 2020* (pp. 191-192). Universidad de Antioquia, Medellín.
- Cabana, Á., Zugarramurdi, C., Valle-Lisboa, J.C., & De Deyne, S. (en preparación). The “Small World of Words” Free Association Norms for Rioplatense Spanish.
- Cañete, J. (2019a). *Spanish Unannotated Corpora*. Recuperado el 12 de noviembre del 2021 de <https://doi.org/10.5281/zenodo.3247731>
- Cañete, J. (2019b). *Spanish Word Embeddings*. Recuperado el 12 de noviembre del 2021 de <https://github.com/BotCenter/spanishWordEmbeddings>
- Cardellino, C. (2016). *Spanish Billion Words Corpus and Embeddings*. Recuperado el 12 de noviembre del 2021 de <https://crscardellino.github.io/SBWCE/>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.
- Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica*, 32, 133-143.
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior research methods*, 51(3), 987-1006.

- Gallassi, R., Sambati, L., Stanzani Maserati, M., Poda, R., Oppi, F., De Matteis, M. & Marano, G. (2014). Simple verbal analogies test: normative data on a short task exploring abstract thinking. *Aging Clinical and Experimental Research*, 26, 67–71.
- Griffiths, T.L., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, 114, 211-244.
- Jones, M. N., Gruenenfelder, T. M., & Recchia, G. (2018). In defense of spatial models of semantic representation. *New Ideas in Psychology*, 50, 54-60.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1), 79-86.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2011). *Handbook of latent semantic analysis*. Routledge.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, 37(4), 547-559.
- Mikolov, T., Yih, W. T., & Zweig, G. (2013, junio). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 746-751).
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- Quillian, M. R. (1962). A revised design for an understanding machine. *Mech. Transl. Comput. Linguistics*, 7(1), 17-29.
- Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science*, 12(5), 410-430.

- Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317-1323.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In *Handbook of latent semantic analysis* (pp. 439-460). Psychology Press.
- Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 237-249). Washington, DC: American Psychological Association.
- Universidad de Chile. (2019). *Spanish Word Embeddings*. Recuperado el 12 de noviembre del 2021 de <https://github.com/dccuchile/spanish-word-embeddings>
- Valle-Lisboa, J. C., & Mizraji, E. (2007). The uncovering of hidden structures by latent semantic analysis. *Information sciences*, 177(19), 4122-4147.



# Anexo 1

## Lista de ítems

<b>Clave ejemplo</b>	<b>Objetivo ejemplo</b>	<b>Clave</b>	<b>Origen</b>
Tigre	rayas	mulita	propio
Cristianismo	iglesia	judaismo	propio
Ave	ala	delfín	propio
Arbol	hojas	flor	propio
Picnic	parque	asado	propio
Desayuno	cocina	cena	propio
Selva	león	desierto	propio
Paz	paloma	guerra	propio
Francia	París	Estados Unidos	propio
Argentina	Buenos Aires	Brasil	propio
Perro	gato	gato	propio
Tiburón	pez	pez	propio
Martillo	clavo	sierra	propio
Tornillo	destornillador	tuerca	propio
Caballo	jinete	moto	propio
Avión	piloto	tren	propio
Caballero	espada	vaquero	propio
Peluquera	tijera	cocinero	propio
Blanco	negro	rojo	propio
Frío	caliente	seco	propio
China	Asia	Rusia	propio
Sushi	Japón	pizza	propio
Hamburguesa	Estados Unidos	tallarines	propio
Vodka	Rusia	cerveza	propio
Tereré	Paraguay	vino	propio
Gato	felino	perro	propio
Gallina	ave	vaca	propio
Fruta	árbol	verdura	propio
Bebé	cigüeña	regalo	propio
Perro	ladrar	ave	propio
Bailarín	bailar	albañil	propio
Lento	tortuga	rápido	propio
Plato	mesa	libro	propio
Paraguas	lluvia	gorro	propio
Lluvia	tristeza	sol	propio
Amigos	felicidad	jefe	propio
Escuela	niño	universidad	propio
Memoria	pasado	imaginación	propio
Cerebro	pensamiento	corazón	propio
Fruta	manzana	carne	propio
Tigre	rayas	mulita	propio

Campera	buzo	buzo	propio
Uruguay	fútbol	Estados Unidos	propio
Brasil	samba	Cuba	propio
Rock	guitarra	tango	propio
Pirámide	momia	ataúd	propio
Ropa	pantalones	mueble	propio
Tela	algodón	metal	propio
Cubierto	tenedor	herramienta	propio
Bebida	leche	comida	propio
Ave	paloma	insecto	propio
Ave	pingüino	mamifero	propio
Ciudad	Montevideo	país	propio
Fútbol	estadio	estudio	propio
Juicio	juzgado	recital	propio
Puerta	pestillo	caja	propio
Cirujano	bisturí	locutor	propio
Rojo	ira	azul	propio
Oveja	cordero	vaca	SVAT
Volante	auto	manubrio	SVAT
Perro	perrera	ave	SVAT
Red	pesca	rifle	SVAT
destornillador	tornillo	martillo	SVAT
Pie	zapato	mano	SVAT
Cuadro	pintor	estatua	SVAT
París	Francia	Roma	SVAT
Mujer	niño	gallina	SVAT
Esmalte	uñas	lápiz labial	SVAT
Tornillo	ferretería	carne	SVAT
Cuerda	guitarra	tecla	SVAT
Cordón	zapato	cinturón	SVAT
Garage	auto	puerto	SVAT
Libro	papel	mesa	SVAT
Sábana	cama	mantel	SVAT
Pistola	bala	arco	SVAT
Abogado	corte	doctor	SVAT
Tren	vías	auto	SVAT
Mermelada	fruta	queso	SVAT
Estufa	gas	chimenea	SVAT
Pez	submarino	ave	SVAT

## Anexo 2

Divergencia de Kullback-Leibler en función del valor utilizado en el parámetro lambda del axioma de Shepard-Luce.

<b>Lambda</b>	<b>FastText-SUC</b>	<b>FastText-SBWC</b>	<b>GloVe-SBWC</b>	<b>FastText-wiki-ES</b>	<b>Word2Vec-SBWC</b>	<b>SWOW-ES-RP</b>
<b>0,5</b>	1,206834	1,221154	1,212515	1,298174	1,262932	1,220180
<b>1</b>	1,308591	1,329673	1,306964	1,432576	1,390830	1,350683
<b>1,5</b>	1,424998	1,452623	1,412813	1,582655	1,534972	1,499962
<b>2</b>	1,554681	1,588715	1,529200	1,747042	1,693803	1,665728
<b>2,5</b>	1,696080	1,736460	1,655127	1,924118	1,865531	1,845494
<b>3</b>	1,847580	1,894296	1,789534	2,112175	2,048299	2,036841
<b>3,5</b>	2,007624	2,060702	1,931365	2,309537	2,240320	2,237577
<b>4</b>	2,174779	2,234266	2,079611	2,514654	2,439962	2,445827
<b>4,5</b>	2,347779	2,413733	2,233344	2,726144	2,645796	2,660038
<b>5</b>	2,525537	2,598019	2,391731	2,942821	2,856608	2,878958
<b>5,5</b>	2,707140	2,786207	2,554042	3,163686	3,071389	3,101598
<b>6</b>	2,891837	2,977541	2,719648	3,387916	3,289319	3,327184
<b>6,5</b>	3,079014	3,171401	2,888015	3,614843	3,509737	3,555118
<b>7</b>	3,268179	3,367288	3,058690	3,843930	3,732119	3,784938
<b>7,5</b>	3,458938	3,564801	3,231298	4,074747	3,956051	4,016290
<b>8</b>	3,650979	3,763621	3,405527	4,306957	4,181207	4,248904
<b>8,5</b>	3,844055	3,963498	3,581120	4,540292	4,407333	4,482572
<b>9</b>	4,037972	4,164231	3,757865	4,774542	4,634233	4,717135
<b>9,5</b>	4,232576	4,365664	3,935590	5,009544	4,861754	4,952470

## Anexo 3

Tres primeras predicciones obtenidas utilizando los embeddings y tres primeras respuestas dadas por los participantes.

Estímulo	Posición	FastText-SUC	FastText-SBWC	GloVe-SBWC	FastText-wiki-ES	Word2Vec-SBWC	SWOW-ES-RP	Respuestas	Conteo de respuestas
abogado : corte :: doctor : ?	1	suprema	suprema	cual	tardis	máxilofacial	bisturí	hospital	58
abogado : corte :: doctor : ?	2	dr	necrocirugía	mismo	dalek	torácicas	curación	consultorio	16
abogado : corte :: doctor : ?	3	majestad	cortes	finalmente	acompañantes	cortes	médico	clínica	4
amigos : felicidad :: jefe : ?	1	subjefe	subjefe	subjefe	subjefatura	jefatura	mandamas	trabajo	15
amigos : felicidad :: jefe : ?	2	sufelicidad	exjefe	jefatura	jefatura	Jefe	gerente	autoridad	7
amigos : felicidad :: jefe : ?	3	telicidad	jefatura	comandante	'jefe	jefa	manda	responsabilidad	6
ave : ala :: delfín : ?	1	delfines	delfin	alero	delfines	derecha	aleta	aleta	79
ave : ala :: delfín : ?	2	chateamorand	delfines	pivot	delfín»	timón	ballena	aletas	7
ave : ala :: delfín : ?	3	timón	alero	escolta	delfí	Martignac	foca	cola	3
ave : paloma :: insecto : ?	1	abeja	himenóptero	chrysomelidae	lipaleyrodes	Chrysomelidae	cucaracha	mosca	25

ave : paloma :: insecto : ?	2	hormiga	insectopía	coleoptero	xenaleyrodes	coleoptera	bicho	hormiga	20
ave : paloma :: insecto : ?	3	chinche	chrysomela	coleoptera	aleyrodes	escarabajos	mosca	mosquito	18
ave : pingüino :: mamífero : ?	1	gorila	pingüinos	saltarrocas	pingüino»	mamut	esquimal	ballena	20
ave : pingüino :: mamífero : ?	2	mamíferos	mamíferos	roedor	pingüinil	Spheniscus	mamíferos	vaca	17
ave : pingüino :: mamífero : ?	3	pinnípedo	micromamífero	spheniscus	pingüinos	adeliae	Antártida	perro	12
avión : piloto :: tren : ?	1	maquinista	conductor	conductor	expiloto	lastrero	locomotora	maquinista	34
avión : piloto :: tren : ?	2	delpiloto	motociclista	marcha	#piloto	ferrobús	tranvía	conductor	31
avión : piloto :: tren : ?	3	conductor	maquinista	parada	descarrilador	TAVe	ferrocarril	chofer	10
bailarín : bailar :: albañil : ?	1	albañiles	albañiles	gritar	albañiles	albañiles	construcción	construir	69
bailarín : bailar :: albañil : ?	2	labrar	albañilería	levantarse	albañilería	changas	constructor	revocar	6
bailarín : bailar :: albañil : ?	3	pelotear	albañilerías	llorar	albañilería	lustrador	ladrillos	construcción	5
bebida : leche :: comida : ?	1	lácteos	lácteos	alimentos	alimentos	alimentos	alimento	pan	40
bebida : leche :: comida : ?	2	alimentos	carne	carne	carne	lácteos	ordeñar	carne	15

bebida : leche :: comida : ?	3	mantequilla	alimentos	lácteos	verduras	carne	lácteo	queso	10
bebé : cigüeña :: regalo : ?	1	obsequio	obsequio	obsequio	cigüeñas	obsequio	obsequio	papá noel	32
bebé : cigüeña :: regalo : ?	2	cigüeñas	regalar	peregrino	cigüeña»	palomero	regalos	moño	6
bebé : cigüeña :: regalo : ?	3	obsequia	regalao	águila	cigüeñón	buho	obsequiar	reyes magos	5
blanco : negro :: rojo : ?	1	amarillo	amarillo	azul	rojo,	azul	rojizo	azul	52
blanco : negro :: rojo : ?	2	azul	carmesí	color	amarillo	amarillo	color	verde	36
blanco : negro :: rojo : ?	3	anaranjado	anaranjado	amarillo	rojo»	color	bordó	amarillo	3
caballero : espada :: vaquero : ?	1	cuchillo	chaqueta	hacha	vaquera	bandana	vaqueros	pistola	32
caballero : espada :: vaquero : ?	2	machete	hacha	látigo	vaqueros	cinto	cowboy	revólver	28
caballero : espada :: vaquero : ?	3	hacha	látigo	daga	vaqueta	látigo	jeans	lazo	9
caballo : jinete :: moto : ?	1	motorista	motociclista	ducati	motociclista	motocicleta	motocicleta	motociclista	44
caballo : jinete :: moto	2	motocicleta	motorista	motogp	motocicleta	motociclista	ciclomotor	conductor	13

: ?									
caballo : jinete :: moto : ?	2	motocicleta	motorista	motogp	motocicleta	motociclista	ciclomotor	motoquero	13
caballo : jinete :: moto : ?	3	motociclista	motocicleta	customizada	motos	motos	motos	piloto	8
campera : buzo :: buzo : ?	1	buceador	buceador	cualificado	buceador	buzos	suéter	remera	33
campera : buzo :: buzo : ?	2	submarinista	buzos	cuatrimotor	buzos	buceador	buceo	campera	22
campera : buzo :: buzo : ?	3	submarinero	submarinista	buceo	buceo	buceadores	buceador	camiseta	19
cerebro : pensamiento :: corazón : ?	1	librepensamiento	humanismo	espíritu	corazón,	humanismo	latido	sentimiento	48
cerebro : pensamiento :: corazón : ?	2	repensamiento	librepensamiento	amor	librepensamiento	providencialismo	latidos	latido	10
cerebro : pensamiento :: corazón : ?	3	ensamiento	tradicionalismo	alma	'pensamiento	filosófico	pulso	sangre	6
cirujano : bisturí :: locutor : ?	1	teleantillas	radiofónico	radiofónico	locutor/a	Mariskal	locución	micrófono	89
cirujano : bisturí :: locutor : ?	2	locutora	presentador	presentador	exlocutor	goldo	comunicador	cincel	2
cirujano : bisturí :: locutor : ?	2	locutora	presentador	presentador	exlocutor	goldo	comunicador	voz	2

cirujano : bisturí :: locutor : ?	3	locutores	esradio	radial	radiofónico	Fernandisco	radio		
cordón : zapato :: cinturón : ?	1	baticinturón	zapatos	zapatos	cinturón»	asteroide	cinto	pantalón	77
cordón : zapato :: cinturón : ?	2	elcinturón	zapatillo	asteroide	asteroides	asteroides	pantalón	hebilla	11
cordón : zapato :: cinturón : ?	3	asteroide	asteroide	pedilon	asteroide»	Luboš	calza	cintura	1
cordón : zapato :: cinturón : ?	3	asteroide	asteroide	pedilon	asteroide»	Luboš	calza	hebillas	1
cordón : zapato :: cinturón : ?	3	asteroide	asteroide	pedilon	asteroide»	Luboš	calza	jean	1
cordón : zapato :: cinturón : ?	3	asteroide	asteroide	pedilon	asteroide»	Luboš	calza	vaquero	1
cuadro : pintor :: estatua : ?	1	escultor	escultor	escultor	escultor	escultor	estatuas	escultor	93
cuadro : pintor :: estatua : ?	2	esculto	esculto	escultura	esculto	escultura	escultura	cuadro	1
cuadro : pintor :: estatua : ?	2	esculto	esculto	escultura	esculto	escultura	escultura	escultora	1
cuadro : pintor :: estatua : ?	2	esculto	esculto	escultura	esculto	escultura	escultura	leña	1
cuadro : pintor ::	2	esculto	esculto	escultura	esculto	escultura	escultura	mármol	1



estatua : ?									
cuadro : pintor :: estatua : ?	2	esculto	esculto	escultura	esculto	escultura	escultura	persona	1
cuadro : pintor :: estatua : ?	3	escultura	escultura	monumento	escultura	esculpida	escultor		
cubierto : tenedor :: herramienta : ?	1	herramienta	herramienta	azadon	herramienta»	criptomoneda	herramientas	martillo	56
cubierto : tenedor :: herramienta : ?	2	herrami	herramiento	herramientas	herramientas»	pharming	destornillador	destornillador	17
cubierto : tenedor :: herramienta : ?	3	herramiento	herramien	interfaz	herramienta	herramientas	llave_inglesa	pinza	6
cuerda : guitarra :: tecla : ?	1	teclado	teclado	teclado	teclado	teclado	teclas	piano	84
cuerda : guitarra :: tecla : ?	2	teclar	teclados	teclados	teclados	teclados	teclado	teclado	5
cuerda : guitarra :: tecla : ?	3	teclax	teclas	teclas	teclado/piano	teclas	piano	gamer	1
cuerda : guitarra :: tecla : ?	3	teclax	teclas	teclas	teclado/piano	teclas	piano	órgano	1
desayuno : cocina :: cena : ?	1	cocinería	culinaria	tradicional	cocinas	cocinas	chef	cocina	39

desayuno : cocina :: cena : ?	2	barbacoa	cocinería	tradición	culinaria	platos	cocinero	comedor	33
desayuno : cocina :: cena : ?	3	parrillada	platos	platos	cocinar	Cocina	gastronomía	living	8
destornillador : tornillo :: martillo : ?	1	martilla	poleas	espiral	martilla	biela	clavo	clavo	88
destornillador : tornillo :: martillo : ?	2	perno	martillado	resorte	martillos	perno	clavos	martillar	1
destornillador : tornillo :: martillo : ?	3	martillado	tornillos	percutor	martillito	roldana	tornillos		
escuela : niño :: universidad : ?	1	niño9	niña	infancia	bebé	niña	nene	joven	39
escuela : niño :: universidad : ?	2	bebé	niño0	hombre	vniversidad	bebé	universitario	adulto	30
escuela : niño :: universidad : ?	3	niño4	niño	derecho	stanforduniversidad	muchacho	doctorado	estudiante	9
estufa : gas :: chimenea : ?	1	comprimido	chimeneas	gases	gases	licuado	garrafa	leña	77
estufa : gas :: chimenea : ?	2	metano	licuado	humo	chimeneas	petróleo	gases	humo	13
estufa : gas :: chimenea : ?	3	gases	gases	aire	gaseo	Snøhvit	gaseoso	fuego	9
fruta : manzana ::	1	ternera	cerdo	cerdo	carnes	cerdo	churrasco	vaca	14

carne : ?									
fruta : manzana :: carne : ?	2	filete	pechuga	pollo	ternera	res	bife	asado	12
fruta : manzana :: carne : ?	3	chuleta	carnes	picada	missmanzana	ternera	carnicería	churrasco	11
fruta : árbol :: verdura : ?	1	árboles	árboles	árboles	'árbol	arbolito	vegetal	planta	43
fruta : árbol :: verdura : ?	2	arbolillo	tronco	Árbol	tronco	arbol	sauce	tierra	30
fruta : árbol :: verdura : ?	3	arbusto	árbolito	tronco	árbolito	siempreverde	ramas	huerta	4
frío : caliente :: seco : ?	1	seca	seca	seca	remojado	tocosh	sequedad	mojado	55
frío : caliente :: seco : ?	2	húmedo	salada	agua	caliendo	siru	hirviendo	húmedo	41
frío : caliente :: seco : ?	3	jabonoso	calentada	húmedo	remojada	tostarlo	sequía	frío	2
fútbol : estadio :: estudio : ?	1	elestudio	estudios	estudios	estudio,	Estudio	facultad	escuela	17
fútbol : estadio :: estudio : ?	2	unestudio	auditorio	análisis	estudio,	análisis	estudios	universidad	11
fútbol : estadio :: estudio : ?	3	prospectivo	comparativo	realizado	estudio	Phonalex	universidad	facultad	10
gallina : ave :: vaca : ?	1	paseriforme	caprimulgiforme	especie	paseriforme	pelecaniforme	vacuno	mamífero	66
gallina : ave :: vaca : ?	2	paloma	paseriforme	león	piciforme	anseriforme	bovino	bovino	8
gallina : ave	3	accipitriforme	estrutioniforme	núñez	bucerotiforme	falconiforme	ternero	rumiante	7

:: vaca : ?									
garage : auto :: puerto : ?	1	puertos	manzanillo	cual	—puerto	puertos	barcos	barco	93
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	avión	1
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	buque	1
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	embarcación	1
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	hangar	1
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	marina	1
garage : auto :: puerto : ?	2	portua	muelle	bordo	puerto,	Henecan	barco	yate	1
garage : auto :: puerto : ?	3	portuaria	cuadillo	rico	#puerto	zarpase	buque		
gato : felino :: perro : ?	1	canino	felinos	rebaño	felinos	felinos	fiera	canino	64
gato : felino :: perro : ?	2	félido	rebaño	auriazul	perros	felina	feroz	can	13
gato : felino :: perro : ?	3	cánido	perros	adiestrado	cánido	mastín	lobos	cánido	9
juicio : juzgado :: recital : ?	1	concierto	concierto	concierto	recitales	concierto	concierto	estadio	20
juicio : juzgado :: recital : ?	2	recitale	recitales	recitales	megarecital	recitales	cantantes	escenario	19
juicio : juzgado :: recital : ?	2	recitale	recitales	recitales	megarecital	recitales	cantantes	teatro	19
juicio :	3	recitales	conciertos	conciertos	concierto	conciertos	cantante	músico	5

juzgado :: recital : ?									
juicio : juzgado :: recital : ?	3	recitales	conciertos	conciertos	concierto	conciertos	cantante	recitado	5
lento : tortuga :: rápido : ?	1	latortuga	tortugas	tortugas	'tortuga	tortugas	veloz	liebre	60
lento : tortuga :: rápido : ?	2	tortugas	carettochelys	marina	#tortuga	golfina	rápidamente	conejo	8
lento : tortuga :: rápido : ?	3	carettochelys	golfina	golfina	tortuga»	Lepidochelys	rapidez	chita	7
libro : papel :: mesa : ?	1	mesas	desempeñar	desempeñar	rol	mesas	mesas	madera	75
libro : papel :: mesa : ?	2	desempeñana	mesas	desempeñando	mesas	rol	comedor	mantel	9
libro : papel :: mesa : ?	3	rol	desempeñas	mesas	papeles	tapete	sillas	pata	3
libro : papel :: mesa : ?	3	rol	desempeñas	mesas	papeles	tapete	sillas	patas	3
lluvia : tristeza :: sol : ?	1	amargura	añoranza	angustia	tristeza,	melancolía	infelicidad	alegría	71
lluvia : tristeza :: sol : ?	2	melancolía	nostalgia	alegría	tristeza»	nostalgia	angustia	felicidad	26
lluvia : tristeza :: sol : ?	3	angustia	melancólicamente	melancolía	tristezas	angustia	desconsuelo	calor	2
lluvia : tristeza :: sol : ?	3	angustia	melancólicamente	melancolía	tristezas	angustia	desconsuelo	tristeza	2

martillo : clavo :: sierra : ?	1	sierro	leona	leona	sierrra	serranía	sierras	madera	60
martillo : clavo :: sierra : ?	2	sierran	sierran	tomé	sierran	Chuacús	cerro	hoja	6
martillo : clavo :: sierra : ?	3	presierra	sierratahumara	liberia	sierrana	Xucaneb	cerros	serrucho	3
martillo : clavo :: sierra : ?	3	presierra	sierratahumara	liberia	sierrana	Xucaneb	cerros	tronco	3
memoria : pasado :: imaginación : ?	1	Imaginación	pasada	pasada	imaginado	despunte	imaginativo	futuro	83
memoria : pasado :: imaginación : ?	2	imaginació	desbordante	creatividad	imaginativo	cargadito	fantasioso	creatividad	2
memoria : pasado :: imaginación : ?	2	imaginació	desbordante	creatividad	imaginativo	cargadito	fantasioso	presente	2
memoria : pasado :: imaginación : ?	2	imaginació	desbordante	creatividad	imaginativo	cargadito	fantasioso	sueños	2
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	abstracción	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	capacidad generar ideas	1

?									
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	fantasía	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	ficción	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	ideas	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	inspiración	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	irrealidad	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	ligero	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	literatura	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	mundo	1
memoria : pasado ::	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	mundo contrafáctico	1

imaginación : ?									
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	niñez	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	nuevo	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	representación	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	soñar	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	tiempo	1
memoria : pasado :: imaginación : ?	3	Imaginaciones	insólito	hace	imaginabas	abrumante	creatividad	volar	1
mermelada : fruta :: queso : ?	1	quesos	quesos	carne	frutas	carne	quesos	leche	80
mermelada : fruta :: queso : ?	2	frutas	frutas	frutas	quesos	quesos	picada	vaca	3
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	bridge	1



mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	fermento	1
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	fondue	1
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	lechee	1
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	lácteos	1
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	pan	1
mermelada : fruta :: queso : ?	3	pescado	verduras	verduras	'queso	verduras	kiwi	requesón	1
mujer : niño :: gallina : ?	1	pollito	pollito	pavo	gallinaza	pollito	pollito	pollito	56
mujer : niño :: gallina : ?	2	cerdito	pollo	pollo	pollito	huevo	pollitos	pollo	27
mujer : niño :: gallina : ?	3	polluelo	huevo	huevo	gallinazus	pollo	gallinas	huevo	12
oveja : cordero :: vaca : ?	1	pollo	villalobos	pérez	febres	cerdo	ternero	ternero	78
oveja : cordero :: vaca : ?	2	carnero	vargas	mota	corderas	ternera	vacuno	toro	8
oveja : cordero :: vaca : ?	3	cerdo	solís	duarte	guisado	carne	bovino	ternera	3
paraguas :	1	elluvia	lluvias	lluvias	#lluvia	lluvias	gorra	frío	47

lluvia :: gorro : ?									
paraguas : lluvia :: gorro : ?	1	elluvia	lluvias	lluvias	#lluvia	lluvias	gorra	sol	47
paraguas : lluvia :: gorro : ?	2	torrencial	torrencial	temperaturas	'lluvia	neblina	sombrero	viento	6
paraguas : lluvia :: gorro : ?	3	lloviznado	llovizna	frío	lluvias	llovizna	boina	cabeza	5
paz : paloma :: guerra : ?	1	tórtola	gavilana	águila	paloma,	ave	pájaro	arma	17
paz : paloma :: guerra : ?	2	águila	torcacita	cucurrucucú	paloma»	chocha	bélico	águila	14
paz : paloma :: guerra : ?	3	guerrita	lagartijero	isabel	*guerra	torcaz	ave	bomba	11
perro : gato :: gato : ?	1	conejo	zorro	montés	gatos	montés	gata	perro	50
perro : gato :: gato : ?	2	montés	ardilla	zorro	ratón	Gato	felino	ratón	46
perro : gato :: gato : ?	3	gatito	mapache	cocolino	montés	zorro	garras	pájaro	2
pez : submarino :: ave : ?	1	minisubmarino	tren	tren	submarinos	avión	aves	avión	86
pez : submarino :: ave : ?	2	submarinos	portaaviones	túnel	submarino	submarinos	pájaro	aéreo	3
pez : submarino :: ave : ?	3	superportaaviones	superportaaviones	trenes	submariner	sumergible	plumas	aire	2
picnic : parque ::	1	ternera	cocido	monumento	#parque	lechal	asador	parrillero	32

asado : ?									
picnic : parque :: asado : ?	2	asador	chorizo	prado	parquesol	cochinillo	bife	parrilla	22
picnic : parque :: asado : ?	3	asada	asados	museo	cabrito	Parque	parrilla	barbacoa	17
pie : zapato :: mano : ?	1	guante	calzapato	zapatos	zapatos	guante	manos	guante	88
pie : zapato :: mano : ?	2	servilleta	guante	bolso	zapatones	zapatos	aprieta	guantes	5
pie : zapato :: mano : ?	3	servilletero	zapatos	pincel	guante	pañuelo	agarrar	anillo	3
pistola : bala :: arco : ?	1	iris	arcos	ángulo	arcos	arcos	arquero	flecha	103
pistola : bala :: arco : ?	2	arcos	tímpano	izquierdo	arcos—	escarzano	golero	pelota	5
pistola : bala :: arco : ?	3	tímpano	escarzano	arcos	arcos,	abocinado	gol		
plato : mesa :: libro : ?	1	librode	libros	libros	libros	libros	biblioteca	biblioteca	44
plato : mesa :: libro : ?	2	libros	prologado	publicación	libro,	coeditó	libros	escritorio	11
plato : mesa :: libro : ?	3	libroa	leyendo	publicó	\$libro	Martropía	lectura	estantería	7
plato : mesa :: libro : ?	3	libroa	leyendo	publicó	\$libro	Martropía	lectura	hoja	7
puerta : pestillo :: caja : ?	1	cajas	pestillos	redoblante	cajas	eyector	cajas	tapa	29
puerta : pestillo :: caja : ?	2	pestillos	desembrague	recopilatoria	pestillos	monodisco	paquete	manija	9

puerta : pestillo :: caja : ?	3	registradora	cajas	tóracica	ahorros	percutora	paquetes	cinta	8
red : pesca :: rifle : ?	1	escopeta	fusil	fusil	pistola	escopeta	escopeta	caza	77
red : pesca :: rifle : ?	2	fusil	escopeta	escopeta	escopeta	fusil	fusil	cacería	9
red : pesca :: rifle : ?	3	caza	carabina	calibre	fusil	rifles	bala	arma	2
red : pesca :: rifle : ?	3	caza	carabina	calibre	fusil	rifles	bala	bala	2
red : pesca :: rifle : ?	3	caza	carabina	calibre	fusil	rifles	bala	disparo	2
red : pesca :: rifle : ?	3	caza	carabina	calibre	fusil	rifles	bala	tiro	2
rock : guitarra :: tango : ?	1	bandoneón	bandoneón	violín	bandoneón	bandoneón	bandoneón	bandoneón	84
rock : guitarra :: tango : ?	2	violín	contrabajo	piano	-guitarra	contrabajo	Gardel	acordeón	12
rock : guitarra :: tango : ?	3	piano	tangos	bandoneón	guitarra...	tangos	acordeón	piano	8
rojo : ira :: azul : ?	1	enojo	furia	furia	tristeza	furia	rabia	tristeza	28
rojo : ira :: azul : ?	2	furia	enojo	enojo	desesperación	enojo	bronca	calma	17
rojo : ira :: azul : ?	3	frustración	frustración	frustración	resentimiento»	frustración	enojo	paz	13
ropa : pantalones :: mueble : ?	1	taburete	pantaloncillo	peto	mueblero	taburete	muebles	mesa	29

ropa : pantalones :: mueble : ?	2	mueblero	claveteado	accesorio	pantalonetas	vaqueta	de_madera	ropero	15
ropa : pantalones :: mueble : ?	2	mueblero	claveteado	accesorio	pantalonetas	vaqueta	de_madera	silla	15
ropa : pantalones :: mueble : ?	3	torneador	muebles	bombachos	muebles	perchero	mesas	sillón	10
selva : león :: desierto : ?	1	desiero	gto	durango	león»	dromedario	desértico	camello	63
selva : león :: desierto : ?	2	sahagún	guanajuato	valladolid	león,	leones	árido	escorpión	8
selva : león :: desierto : ?	3	chihuahuero	zacatecas	guanajuato	desierto»	leopardado	sequía	arena	3
sábana : cama :: mantel : ?	1	sofá	sofá	sofá	dormir	sofá	mesa	mesa	99
sábana : cama :: mantel : ?	2	comedor	manteles	mesa	manteles	cenar	comedor	pastito	1
sábana : cama :: mantel : ?	3	desayuno	cenar	baño	sofá	mesa	platos		
tela : algodón :: metal : ?	1	níquel	heavy	heavy	metalmetal	heavy	metálico	hierro	31
tela : algodón :: metal : ?	2	heavy	metales	metales	metalco	death	metales	acero	13
tela : algodón :: metal : ?	3	metales	thrash	rock	industrialmetal	grindcore	acero	aluminio	11
tiburón : pez :: pez : ?	1	peces	perciforme	singnatiforme	perciforme	ciprinodontiformes	pez_dorado	tiburón	19
tiburón : pez :: pez : ?	2	pececillo	percichthys	pleuronectiforme	ciprinodontiforme	rivulines	pescados	plancton	13

tiburón : pez :: pez : ?	3	pececito	larimichthys	actinopterio	syngnathidae	Lacustricola	pescado	animal	8
tigre : rayas :: mulita : ?	1	mulitas	ribeteadas	orlovi	mulitas	capucho	carpincho	caparazón	126
tigre : rayas :: mulita : ?	2	sombrerillos	mulitas	longitudinales	plastrón	resaque	líneas	coraza	7
tigre : rayas :: mulita : ?	3	hamulitas	blancas	dasyopus	botonaduras	motitas	hamster	anillos	6
tigre : rayas :: mulita : ?	3	hamulitas	blancas	dasyopus	botonaduras	motitas	hamster	armadura	6
tornillo : destornillador :: tuerca : ?	1	destornilla	destornillar	carpetazo	destornilladores	Widower	llave_inglesa	llave	65
tornillo : destornillador :: tuerca : ?	2	destornillado	destornillada	bofetada	tuerca»	Frontios	ferretería	pinza	15
tornillo : destornillador :: tuerca : ?	3	destornillar	destornilladores	empujón	tuercas	Vyon	tornillos	tornillo	5
tornillo : ferretería :: carne : ?	1	carnes	ferreterías	lácteos	ferreterías	charcutería	carnicería	carnicería	96
tornillo : ferretería :: carne : ?	2	ferreterías	carnicería	fresca	carnes	abarrotes	carnes	pulpería	1
tornillo : ferretería :: carne : ?	2	ferreterías	carnicería	fresca	carnes	abarrotes	carnes	rojo	1
tornillo : ferretería :: carne : ?	3	abarrotes	carnicerías	leche	carnicerías	carnicerías	carnicero		
tren : vías :: auto : ?	1	vias	autos	autos	theft	ridades	autos	carretera	36
tren : vías ::	2	autos	personadas	recurso	conducentes	autos	vehículos	calle	30

auto : ?									
tren : vías :: auto : ?	3	ridades	recurso	alternativas	defectivas	TSJR	automóvil	ruta	16
árbol : hojas :: flor : ?	1	pétalos	pétalos	flores	sinflorescencias	flores	flores	pétalos	67
árbol : hojas :: flor : ?	2	flores	flores	pétalos	sinflorescencia	inflorescencias	jazmín	pétalo	19
árbol : hojas :: flor : ?	3	floreillas	cabezuelas	florales	inflorescencias	cabezuelas	orquídea	tallo	7

