

Neología Semántica y Contexto: Un estudio de la detección de los neologismos semánticos tóxico, unicornio y hibernar usando dos corpus diacrónicos

Laura Moset Estruch
Universitat Pompeu Fabra
`laura.moset01@estudiant.upf.edu`

5 de abril de 2024

Resumen

La detección de los neologismos semánticos siempre se ha considerado compleja porque el término no sufre ningún cambio de forma, sino que solamente a nivel semántico. Se defiende entonces que la mejor manera de poder llegar a detectar este cambio sería observar el contexto que rodea al término. Si el término empieza a aparecer con cierta frecuencia en contextos en los que antes no ocurría, podemos tener fuertes sospechas de que este ha tenido un cambio de sentido neológico. Basándose en esto, este estudio realiza búsquedas de frecuencia en el contexto de tres neologismos semánticos (tóxico, unicornio e hibernar) en dos corpus diacrónicos, uno del 2011 y otro del 2021-22, para comprobar si sería posible observar este cambio de sentido. Se experimenta con dos tipos de contexto: el contexto directo, es decir, palabras que preceden directamente al neologismo; y un contexto más general de 100 palabras alrededor del término. Para esto se usa Sketch Engine con su herramienta de concordancia y frecuencia, y también un código escrito en Python para agilizar la búsqueda. Los resultados obtenidos son muy prometedores en cuanto se refiere a la detección de neologismos semánticos usando sus contextos.

Palabras Clave: corpus, neologismo semántico, detección, contexto

1. Introducción

Muchos de los que intentan realizar una clasificación de los neologismos se enfrentan a la complicación que presenta la subclasificación de esos categorizados como neologismos semánticos. Los neologismos de forma sufren un cambio observable y, en la mayoría de ocasiones, regularizado. Mientras que el neologismo semántico es aquella palabra ya existente en una lengua que sufre una nueva manifestación semántica y este cambio puede ser tanto “a un contenido conceptualmente nuevo o a un contenido previamente existente pero expresado ahora con este otro significante” (Díaz Hormigo, 2007, p.87 [5]). Cabré (2006 [4]) por ejemplo, propone una clasificación donde, dentro del tipo al que llama resematización, reconoce tres subtipos: la reducción, la ampliación y el cambio de significado. Aunque a primera vista parece bastante simple, se complica cuando uno intenta distinguir dónde está la línea entre cada uno de los subtipos.

La dificultad reside en que este cambio es solo a nivel semántico y observando solamente la palabra no se puede llegar a apreciar. Como menciona Lamaczová (2018 [6]), estas complicaciones provienen de la base de que el sentido léxico como tal es ya flexible y sufre cambios constantes para ajustarse a las necesidades comunicativas de los hablantes. Por lo tanto, estos cambios suelen ser rápidos, poco previsibles y poco regulares. Martínez Linares (2015 [7]) concluye entonces que, al contrario que los neologismos de forma, los neologismos semánticos se pueden identificar solamente usando el contexto en el que se encuentran, especialmente los cambios en cuanto a las palabras con las que se suelen asociar.

Este estudio tiene como objetivo analizar hasta qué punto es posible una detección de neologismos semánticos usando el contexto. Se emplean los neologismos: hibernar, tóxico y unicornio, como objeto de estudio. Y se usan dos corpus diacrónicos, del 2011 y del 2021-22, donde se observa la frecuencia absoluta de palabras en un contexto directo y un contexto general (100 palabras alrededor del término). Siguiendo como

base las conclusiones obtenidas por autores, como observaremos en la siguiente sección, se espera observar en el corpus más reciente (2021-22) la aparición de un contexto no habitual y diferente al del 2011 que permita confirmar el cambio.

El presente artículo se organiza en siete secciones principales. En la siguiente sección, empezamos revisando las dificultades de la detección de los neologismos semánticos y la literatura existente que defiende el uso del contexto para facilitar esta detección. En la sección 3, definimos los neologismos semánticos elegidos para este estudio (tóxico, unicornio e hibernar) para tener claro qué entenderemos como significado original o establecido y el significado neológico o novedoso, y así poder identificarlo en los resultados del corpus. La sección 4 recoge una explicación de la metodología usada y una descripción de los corpus. La sección 5, incluye los resultados y el análisis de estos en dos partes. En primer lugar, las observaciones obtenidas con el contexto directo y, a continuación, en el contexto general. Finalmente, la sección 6 resume las conclusiones obtenidas y la sección 7 incluye algunas de las limitaciones del estudio.

2. La Detección del Neologismo Semántico

La detección de los neologismos semánticos es una de las principales dificultades de este ámbito. A diferencia de los neologismos de forma que sufren un cambio observable, los neologismos de sentido solo sufren un cambio semántico que no es perceptible a simple vista. Por eso resultan más difíciles de identificar y describir para los hablantes y, como resultado, son menos presentes en los diccionarios y tienen una clasificación menos exhaustiva que otros neologismos (Adelstein, 2022 [1]; Díaz Hormigo, 2007 [5]).

A menudo, se destaca la importancia del contexto en cuanto se refiere a los neologismos semánticos. Bastuji (1974 [2]), en uno de los primeros artículos escritos sobre el tema, ya menciona como, la mayoría de veces, la variación en el contexto es lo bastante significativa como para provocar un cambio de significado en la interpretación de un término. Él pone como ejemplos, *réimporter* y *exportateur* los cuales solían ir acompañados de un sintagma nominal referente a objetos o cosas y cómo estos han pasado a incluir sustantivos con rasgos o características de persona. Podríamos clasificar este cambio como la ampliación de significado mencionada por Cabré (2006 [4]).

Adelstein (2022 [1]) recoge también diferentes subtipos de neologismo semántico, pero ella destaca la distancia semántica entre el significado novedoso y el establecido como categoría de clasificación de este tipo de neologismos. Concluye que “[los] tipos con mayor distancia semántica se asocian con procesos metafóricos y metonímicos, mientras que los de menor distancia, con la generalización y la especificación” (p.333). Adelstein crea cuatro categorías: los sentidos plenos homónimos, cuando no comparten ningún tipo de información y requieren información extralingüística para ser decodificados; los sentidos plenos polisémicos, la distancia también es grande y se necesita información extra para decodificar, pero existe una relación polisémica entre los dos sentidos; los microsensos, que tienen una mínima modificación semántica; y las modulaciones contextuales, son sentidos abstractos y surgen de considerar información del significado descriptivo original. Vemos entonces que la dificultad de clasificación de los neologismos semánticos es casi igual de conflictiva que su detección, y en la mayoría de casos van ligados.

Sánchez Manzanares (2009 [8]) afirma que “los neologismos de sentido se activan por el contexto, bien en un contexto sintagmático, bien en un contexto discursivo” (p.142). La autora concluye en su artículo que esta ruptura clasémica que se produce al insertar un neologismo semántico en un contexto no habitual puede ser esencial para su identificación. Lamaczová (2018 [6]) examina este cambio contextual en su artículo donde analiza los neologismos ‘viral’ y ‘bizarro’ usando un corpus del periódico ABC entre los años 1980 y 2015. En este confirma la existencia de un cambio de sentido en los términos, que este cambio es progresivo y que puede llegar a sustituir al sentido original. Este presente estudio está basado en estos resultados de Lamaczová en los cuales observa que cuando se produce el cambio de significado, los sustantivos y adjetivos con los que aparece el término también son diferentes. Ella realiza este análisis de manera manual, mientras que Torres Rivera (2019 [12]) realiza un intento similar automatizado usando la similitud o distancia léxica de los contextos y concordancias, a la vez que los diferentes campos semánticos de las palabras. Sus resultados son prometedores, pero están limitados a campos de conocimiento especializado.

3. Los Neologismos *tóxico*, *unicornio* y *hibernar*

Los neologismos semánticos usados para este estudio son: tóxico, unicornio e hibernar. Cuando se redacta este estudio, uno de los términos ya tiene recogido en el Diccionario de la lengua española [9] (en adelante DLE) su significado neológico, como indican Bernal, Freixa, y Torner (2020, [3]) esto implicaría que esta palabra “ya no [es neológica] desde un punto de vista lexicográfico” (p. 601). La cuestión de si se consideran aún neologismos o no, no resulta un problema para este estudio, ya que lo que se intenta comprobar es si la detección de este cambio es posible. Por lo tanto, lo que se busca es que en algún momento lo fueran. Para los otros dos hay una entrada en el Diccionario Panhispánico de Dudas [10] (en adelante DPD) donde se recoge su significado neológico. A continuación, tanto el significado original como novedoso de cada uno de los neologismos será definido.

Para el término ‘tóxico’, el DLE recoge tres acepciones, las primeras dos las atribuiremos al sentido original donde el adjetivo hace referencia a una sustancia: “1. adj. Que contiene veneno o produce envenenamiento. Apl. a una sustancia, u. t. c. s. m. / 2. adj. Perteneciente o relativo a una sustancia tóxica”. Luego la tercera acepción recoge el cambio de sentido neológico: “3. adj. Que tiene una influencia nociva o perniciosa sobre alguien”. Es decir, para encontrar este cambio neológico uno debe fijarse en el sustantivo al cual acompaña ‘tóxico’ que pasa de ser solamente sustancias a poder aplicarse también a personas, entre otros. Lo identificaríamos entonces como una amplificación de significado o una generalización.

La entrada del término ‘unicornio’ en el DLE incluye tres acepciones, las cuales hacen referencia al animal fabuloso, a un rinoceronte y al marfil que se cree que procede del animal. Muy lejos de estos significados está el que surgió del cambio semántico neológico que recoge el DPD como “[...] en el ámbito empresarial, especialmente cuando se habla de compañías emergentes, se emplea [unicornio] para aludir a un crecimiento tan rápido que es casi imposible de alcanzar”. Sería igual a un cambio total de sentido, resultado de un proceso metafórico o un sentido pleno homónimo.

En el caso del verbo ‘hibernar’, el DLE lo define como “1. intr. Pasar el invierno, especialmente en estado de hibernación”. Esta definición es un poco ambigua en cuanto a si incluye o no su significado neológico, ya que no especifica quién puede ser el sujeto del verbo. Por otra parte, el DPD en su entrada del término ofrece una definición más específica: “Dicho de ciertos mamíferos, ‘pasar el invierno en estado de hibernación o aletargamiento’ [...]; se emplea a veces, en sentido figurado, con sujeto de persona [...] o, referido a cosas, con el sentido de ‘dejar [algo] temporalmente inactivo’”. En este caso, entendemos como sentido original el que se emplea en referencia a mamíferos y neológico en referencia a personas o cosas. Muy similar al proceso por el que pasa ‘tóxico’ donde se generaliza el uso.

4. Metodología

En primer lugar, para este estudio se usó Sketch Engine porque se consideró una herramienta útil y fácil de usar para acceder a diferentes corpus y realizar búsquedas específicas. Los motivos de selección de los corpus eran que fueran en español, que estuvieran lematizados y también anotados con categoría gramatical. Finalmente, por un lado, se eligió el corpus ‘Spanish Web 2011 (esTenTen11, Eu + Am)’ para representar el periodo donde el cambio semántico aún no se ha producido. Por otro lado, el corpus ‘Timestamped JSI web corpus 2021-22 Spanish’ fue elegido como el corpus más actual y donde se espera poder observar en uso el neologismo semántico.

El Spanish Web 2011 (esTenTen11, Eu + Am), el cual a partir de ahora me referiré solo como Spanish Web 2011, consta de 22.287.384 documentos que suponen un total de 10.985.547.573 tokens y fueron recopilados durante el 2011 por Heritrix y SpiderLing. Es un corpus de textos web que contiene español tanto de Latinoamérica como europeo. Está anotado con el conjunto de etiquetas FreeLing. El ‘Timestamped JSI web corpus 2021-22 Spanish’, en adelante corpus 2021-22, es construido por 18.274.752 artículos de prensa en línea y contiene en total 6.717.057.340 tokens recogidos entre el 2021 y 2022. En este caso, los textos pueden originar de cualquier parte del mundo, el único criterio es que estuvieran escritos en español. El sistema de etiquetas que usa es también FreeLing.

Para realizar búsquedas en estos corpus se usaron las herramientas de concordancia y frecuencia de Sketch Engine. En esta primera, se realizaron búsquedas avanzadas usando fórmulas CQL. Como los dos corpus utilizan el mismo sistema de etiquetas, las fórmulas se pueden utilizar en ambos sin necesidad de ningún cam-

bio. En total, se usaron tres fórmulas distintas. Para encontrar todas las ocurrencias del término/neologismo en el corpus se usó la fórmula [lemma = “neologismo”], donde el uso de lemma nos permite encontrar todas las variaciones morfológicas y conjugaciones de la palabra. En segundo lugar, la fórmula [tag=“N.*”] [lemma=“neologismo”] muestra todas las ocurrencias donde el término va precedido directamente por un sustantivo. Por último, [tag=“V.*”][lemma=“neologismo”], encuentra todas las ocurrencias del neologismo directamente detrás de un verbo. La primera fórmula se usó para conseguir un contexto del término más general. Mientras que las dos últimas se aplican junto a la herramienta de frecuencia de Sketch Engine y muestran en orden de más a menos frecuencia las colocaciones directas.

Para un estudio más exhaustivo y usando los avances tecnológicos del momento, de los resultados de la primera fórmula se obtuvieron diferentes archivos CSV ¹ con un contexto de 100 palabras, 50 delante y 50 después del neologismo. Se diseñó un código escrito en Python en la herramienta Colab², en el cual, usando la librería Spacy se cuenta la frecuencia total de sustantivos con mayor aparición en concordancia con el término. Esto nos dará acceso a un contexto mucho más general que el directo. Lo que hace el código es lematizar cada una de las palabras por cada los contextos y añadirla (excluyendo palabras vacías como ‘de’, ‘el’, ‘la’, etc.) en una lista donde se suma el número total de apariciones en el corpus. Usando la misma librería Spacy, se observan también las relaciones de dependencia sintáctica de los neologismos, por ejemplo, ver qué sustantivo aparece como sujeto del verbo.

5. Resultados y Análisis

En el primer apartado 5.1 observaremos los resultados obtenidos solamente usando la herramienta Sketch Engine, donde se puede observar la frecuencia de verbos y sustantivos en concordancia directa que precede al término. En la siguiente sección 5.2, se muestran los resultados obtenidos usando los contextos extraídos con Sketch Engine, pero el análisis de frecuencia está realizado usando Python, donde se observa un contexto total de 100 palabras que aparecen alrededor del neologismo.

5.1. El Contexto Directo

Para empezar, examinamos los resultados obtenidos usando el CQL [tag=“N.*”][lemma=“neologismo”], es decir, el término precedido por un sustantivo. Los resultados se muestran en cuadros donde solamente aparecen los diez primeros resultados que serían los más frecuentes.

Spanish Web 2011		JSI web corpus 2021-22	
Contexto Directo	Frecuencia	Contexto Directo	Frecuencia
sustancia	15.287	sustancia	6.900
efecto	6.088	relación	5.353
gas	5.645	gas	3.385
residuo	4.976	masculinidad	1.360
producto	3.162	residuo	1.320
desecho	2.787	persona	1.156
activo	1.710	efecto	905
elemento	1.654	nube	884
material	1.517	producto	830
compuesto	1.356	ambiente	688

Cuadro 1: Resultados de colocaciones directas con ‘tóxico’.

En primer lugar, los resultados obtenidos usando el término ‘tóxico’ se pueden observar en el Cuadro 1. De acuerdo con lo esperado, en el Spanish corpus 2011 todos los términos corresponden con sustancias o similares. Esto correspondería a las primeras dos acepciones del DLE. Por contraste, en el corpus 2021-22, aunque aún se repiten algunos sustantivos (sustancia, gas, residuo, efecto y producto), también aparecen y con frecuencias altas sustantivos que no tienen nada que ver con el sentido original. De esta manera,

¹En el caso de los neologismos que tienen muchas ocurrencias, Sketch Engine tiene un límite donde uno solo puede descargarse los 10.000 primeros contextos.

²Link para acceder al código: https://github.com/momolauris/NEO_SEM_corpus_di-cronicos

en el 2022 también pueden ser tóxicas una relación, la masculinidad, una persona o un ambiente. En este caso, podemos ver con claridad que ha habido un cambio de significado en el adjetivo. Esto concuerda con las observaciones de Bastuji (1974 [2]), en este caso un adjetivo que antes solamente se relacionaba con sustancias pasa a poder aplicarse también a personas entre otros nuevos ámbitos.

Vemos los datos del mismo CQL aplicado al término ‘unicornio’ en el Cuadro 2. En este caso, los resultados del Spanish Web 2011 son más confusos, ya que parece ser que ‘unicornio’ se usa en nombres propios como es ‘sello unicornio’, ‘editorial unicornio’, ‘premio unicornio’, etc. y no únicamente como su significado original. Aun así, el cambio es más claro en el corpus 2021-22, donde empiezan a aparecer, y con frecuencias muy altas, términos relacionados con el ámbito empresarial como empresa, operación, compañía, startups y startup. Estas representarían este significado neológico que recoge el DPD. En el caso de ‘unicornio’ este cambio es mucho más claro que ‘tóxico’, ya que como sugiere Díaz Hormigo (2007 [5]) es un cambio a un contenido conceptualmente nuevo y que no tiene nada que ver con el significado establecido anterior. Se puede observar también en 2021-22 que aparece ‘el’ como sustantivo. Esto es resultado del programa de etiquetado de categoría gramatical que en ocasiones se equivoca con las etiquetas.

Spanish Web 2011		JSI web corpus 2021-22	
Contexto Directo	Frecuencia	Contexto Directo	Frecuencia
sello	98	empresa	642
editorial	20	operación	228
sistema	16	compañía	62
premio	16	el	52
teatro	11	startups	47
dios	11	sello	40
chico	9	startup	33
disco	9	firma	24
barbo	9	futuro	16
lavandería	8	término	15

Cuadro 2: Resultados de colocaciones directas con ‘unicornio’.

Hemos observado un adjetivo, un sustantivo, y ahora el verbo: ‘hibernar’, los resultados del cual se muestran en el Cuadro 3. Primero de todo, observamos que la frecuencia de este término es inferior a la de los otros en los dos corpus analizados. No obstante, esto no afecta a la posible observación de un cambio de sentido. Por un lado, casi todos los sustantivos del 2011 hacen referencia a animales o al tiempo. En 2021-22 podemos ver también animales y tiempo, pero en último lugar consigue colarse en la lista el ‘humano’, y así podemos ver este sentido figurado del verbo que menciona el DPD cuando este se aplica a personas. El proceso de ‘hibernar’ es relativamente similar al de ‘tóxico’, donde pasa de poder aplicarse solo a una cosa en concreto a sufrir una ampliación de significado, como indica Cabré (2007 [4]), para incluir personas y otras cosas. Por otro lado, podemos ver también en el corpus de 2011 referencias del ámbito de cardiología con ‘miocardio’ y ‘corazón’. Parece que este significado desaparece o se deja de usar en 2021-22. Es posible que como Rosa, et. al (2011 [11]) explican, este desuso sea a raíz de que la colocación ‘miocardio hibernado’ es una traducción directa del inglés pero que no acaba de funcionar en español y, como estos autores recomiendan en 2011, se debería dejar de usar.

En cuanto a los resultados obtenidos usando la fórmula [tag=“V.*”][lemma=“neologismo”], donde el contexto directo que precede a la palabra es un verbo, la información obtenida es muy abstracta y relativa. Por ejemplo, las listas de los diez verbos más frecuentes junto al término ‘tóxico’ incluyen tanto en 2011 como en 2021-22 los verbos: ser, resultar, considerar, volver, contener y evitar. Los verbos en 2021-22 que no aparecen en 2011 son: pesar, volver+se, llamar y tornar. Solo con la información del verbo más término como guía, concluimos que de manera general se podrían usar estos verbos tanto en contexto del significado original como también del neológico; por lo tanto, no nos da mucha información de si realmente está presente este cambio de sentido. De la misma manera, para el término ‘unicornio’ se repiten los verbos: ser, llamar, haber, ver, tener y buscar. En 2021-22 aparecen nuevos: considerar, crear, denominar y generar, de los cuales podríamos decir que ‘crear’, ‘denominar’ o ‘generar’ se están refiriendo exclusivamente al hecho de crear empresas unicornio, es decir, solamente al significado neológico, pero eso sería suponer. Una vez más, el cambio de sentido no está tan claro como cuando veíamos el sustantivo. Para el término ‘hibernar’, se repiten los verbos: estar, poder, haber, seguir, dejar y permanecer. En 2021-22 se añade: quedar, encontrar, mantener y decir, pero otra vez estos verbos se podrían aplicar tanto a un animal, como a una persona o una cosa que está hibernando. Podríamos concluir que hasta cierto punto estos neologismos semánticos,

Spanish Web 2011		JSI web corpus 2021-22	
Contexto Directo	Frecuencia	Contexto Directo	Frecuencia
miocardio	97	oso	23
oso	28	año	14
tejido	19	animal	9
corazón	17	modo	7
invierno	16	tiempo	7
animal	14	tortuga	6
tortuga	13	mariposa	6
estado	12	invierno	5
año	11	mes	4
modo	9	humano	3

Cuadro 3: Resultados de colocaciones directas con ‘hibernar’.

especialmente si son creados por generalización como es la ampliación de sentido de ‘tóxico’ e ‘hibernar’, tienen un uso muy similar al de su sentido original y por ello el verbo con el cual se relacionan no suele variar. Mientras que el proceso metafórico de ‘unicornio’ puede ser el que más distancia semántica tiene y el que nos puede llegar a dar más información.

5.2. Contexto General

En esta sección vemos dos tipos de cuadros de frecuencia. En primer lugar, el número total de apariciones de cada sustantivo en un contexto de 100 palabras alrededor del término, 50 delante y 50 después. Por otro lugar, el siguiente cuadro tendrá en cuenta solo los sustantivos que tienen una relación de dependencia sintáctica de algún tipo con el neologismo, es decir, es el sujeto de la oración donde aparece, lo modifica, es el objeto, etc.³. Por motivos de tiempo y observando los resultados de la sección anterior, en este caso solo se examinarán sustantivos, ya que se considera que con estos se obtiene más información para determinar la presencia del neologismo. En este caso, los cuadros también muestran solamente los diez sustantivos más frecuentes, pero se recomienda emplear el código uno mismo para ver la totalidad de los resultados.

Una vez más empezamos viendo los resultados de frecuencia absoluta del término ‘tóxico’, estos se encuentran en el Cuadro 4. Podemos ver que la mayoría de sustantivos que aparecen en esta tabla también aparecían en los resultados obtenidos solamente mirando el contexto directo con Sketch Engine. Esto no es del todo sorprendente, ya que se ha usado el mismo corpus para realizar la búsqueda, pero considerando el enorme aumento que se ha realizado al contexto, merece la pena ser mencionado. A su misma vez, los resultados del corpus de 2011 también hacen en su mayoría referencia a productos y sustancias, mientras que en 2021-22 surgen entre estos nuevos sustantivos relativos a persona, como sería relación, persona y ambiente.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
sustancia	1695	relación	1259
producto	1116	sustancia	1203
efecto	1001	persona	989
residuo	966	año	670
agua	682	producto	627
caso	551	agua	604
gas	503	gas	545
planta	468	vida	478
consumo	466	ambiente	446
año	458	salud	445

Cuadro 4: Frecuencia de sustantivos en un contexto de 100 palabras alrededor de ‘tóxico’.

³Esta dependencia sintáctica siempre es en relación con la frase completa.

En los datos obtenidos por dependencia sintáctica, podemos encontrar en el Cuadro 5 tres etiquetas de función diferentes *nsubj*, *conj* y *nmod*. La etiqueta *nsubj* identifica un sustantivo que es sujeto nominal, es decir, que forma parte de un sintagma nominal el cual es también parte del sujeto. Lo que nos dice esto es que en 2011 ‘tóxico’ forma parte de oraciones donde los sustantivos planta, seta, sustancia, producto y humo forman parte del sujeto. Pero en 2021-22 este sujeto también pasa a ser relación, persona, alcohol y ambiente. Por otro lado, *nmod* es un sustantivo que modifica a otro. En el caso del adjetivo ‘tóxico’, estos resultados no nos dan mucha información, ya que los sustantivos como ‘humano’ con etiqueta *nmod* también podrían hacer referencia al sentido original si forman parte de formaciones como ‘tóxico para los humanos’. Por último, la etiqueta *conj* marca que el sustantivo está unido a otro con la conjunción ‘y’ o ‘o’. En este caso no se puede saber con exactitud la función de este sustantivo, es decir, podría ser parte de un sujeto, un objeto, etc. Pero esta información no es necesaria para el análisis y podemos identificar igualmente que son sustantivos a los que hace referencia el adjetivo. Podemos ver que en 2011 son colocaciones con cosas tóxicas como estupefaciente y sustancia.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
planta	nsubj	relación	nsubj
seta	nsubj	adolescente	nmod
sustancia	nsubj	humano	nmod
estupefaciente	conj	organismo	nmod
célula	nmod	humo	nsubj
producto	nsubj	persona	nsubj
sustancia	conj	alcohol	nsubj
humo	nsubj	ser	nmod
organismo	nmod	ambiente	nsubj
ser	nmod	salud	nmod

Cuadro 5: Frecuencia de dependencias sintacticas de ‘tóxico’.

El Cuadro 6 muestra los resultados de frecuencia absoluta del término ‘unicornio’. En este caso, los resultados del 2011 son mucho más claros que los obtenidos con Sketch Engine en cuanto al hecho de identificar si hacen referencia o no al sentido original. La aparición de cuerno, secreto, mundo, animal, libro, aventura, canción y obra, nos hace pensar en este animal fabuloso o de fantasía que se identifica por su cuerno y que seguramente solo aparece en historias. Una vez más, los resultados del 2021-22 son muy claros en relación a este cambio de sentido, ya que siete de los diez sustantivos podrían ser relacionados con el mundo empresarial y ninguno de ellos con el sentido original. También se puede apreciar el gran aumento de frecuencia como resultado de la aparición del sentido neológico.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
cuerno	485	empresa	3075
año	419	año	1610
secreto	354	dólares	1433
mundo	343	startup	1100
animal	342	compañía	1027
libro	321	país	802
aventura	316	inversión	708
canción	283	valoración	697
tiempo	250	mercado	654
obra	241	mundo	587

Cuadro 6: Frecuencia de sustantivos en un contexto de 100 palabras alrededor de ‘unicornio’.

En el caso de la dependencia sintáctica de ‘unicornio’, observable en el Cuadro 7, vemos resultados muy similares en cuanto a la claridad del cambio semántico. En 2011, vemos con la etiqueta *conj* palabras que seguramente tendrán el mismo nivel de importancia que ‘unicornio’ uniéndolas usando la conjunción ‘y’ o ‘o’. Así se encuentran otros animales fabulosos como dragón, hada, centauro y pegaso. Con esta etiqueta, en 2021-22 aparecen, por otra parte, solo sustantivos con el significado neológico: empresa y startup. En 2011 casi no aparece pero en 2021-22 se encuentran varias instancias de *nmod*, ‘unicornio’ pasa de ser simplemente

modificado, por ejemplo, por ‘unicornio de peluche’ a poder ser modificado dependiendo del lugar donde origina una empresa etiquetada como unicornio. Aparece otra nueva etiqueta *appos* que significa modificador aposicional y se refiere al sintagma nominal el cual está situado justo después de otro sintagma nominal para definir o modificar este último. Las palabras con esta etiqueta servirán para describir el término. En 2011, estas descripciones hacen referencia al significado original, exponiendo que el ‘unicornio’ es un animal, es como un dragón o es un símbolo de alguna cosa. Mientras que en 2021-22, ‘unicornio’ es descrito como una empresa, una startup, o una compañía, cosa que nos indica muy explícitamente la presencia de este sentido neológico.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
dragón	conj	empresa	appos
hada	conj	región	nmod
centauro	conj	país	nmod
animal	appos	startup	appos
pegaso	conj	mundo	nmod
color	nmod	origen	nmod
unicornio	appos	compañía	appos
peluche	nmod	empresa	conj
dragón	appos	startup	conj
símbolo	appos	sector	nmod

Cuadro 7: Frecuencia de dependencias sintácticas de ‘unicornio’.

La frecuencia absoluta de sustantivos en el contexto general del verbo ‘hibernar’ se encuentra en el Cuadro 8. Como pasaba con los resultados de Sketch Engine, de los tres neologismos estudiados en este artículo, este verbo es en el cual resulta más difícil observar el cambio de sentido y la presencia del sentido neológico. Usando un contexto general, en los diez sustantivos más frecuentes de 2021-22 no aparece ninguna referencia a una persona que pueda insinuar este cambio de sentido. Este es uno de los casos donde es necesario mirar todos los resultados obtenidos por el código, ya que usar los diez primeros no es suficiente. En el código donde aparecen hasta los primeros 50 resultados se puede encontrar ‘persona’ en la posición 41 con una frecuencia absoluta de 37, por otra parte, no aparece en ningún momento en 2011. Los sustantivos como por ejemplo año, invierno y tiempo podían ser resultado del uso neológico del verbo como cosas que están un tiempo inactivas, pero como hablantes de la lengua podemos inferir rápidamente que la razón por la cual están aquí es porque son comúnmente usados como complemento circunstancial de tiempo en conjunto con el verbo ‘hibernar’. Pero podemos confirmar esto con los resultados de dependencia sintáctica.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
año	428	año	300
invierno	350	invierno	179
tiempo	251	oso	171
animal	250	mes	154
oso	230	animal	138
mes	229	tiempo	127
temperatura	203	vida	90
especie	167	especie	87
tortuga	152	primavera	85
lugar	151	lugar	85

Cuadro 8: Frecuencia de sustantivos en un contexto de 100 palabras alrededor de ‘hibernar’.

Desafortunadamente, los resultados mostrados en el Cuadro 9 sobre la dependencia sintáctica no presentan mejoras significativas para la detección. En este caso, como ya sabemos que con el significado neológico de ‘hibernar’ lo que ha cambiado es que el sujeto que solía ser solo animales ahora puede ser ocupado también por personas o cosas, podemos fijarnos especialmente en las palabras etiquetadas con *nsubj*. Como se ve en el cuadro tanto en 2011 como en 2021-22, de los diez sustantivos más frecuentes y que forman parte del sujeto, todos hacen referencia a animales, es decir, al sentido original. De la misma manera que con la frecuencia absoluta total, observando la totalidad de los resultados del código Python con 50 sustantivos, sí

que se llega a ver que palabras como individuo, humano, adulto, grupo e información, también se encuentran como sujetos de ‘hibernar’ y por consiguiente confirmando la existencia de este cambio semántico.

Spanish Web 2011		JSI web corpus 2021-22	
Sustantivo	Función	Sustantivo	Función
invierno	obl	invierno	obl
año	obl	año	obl
mes	obl	mes	obl
tortuga	nsubj	oso	nsubj
animal	nsubj	animal	nsubj
oso	nsubj	mariposa	nsubj
equipo	obj	especie	nsubj
año	obj	casa	obl
mes	obj	bosque	obl
mariposa	nsubj	economía	obj

Cuadro 9: Frecuencia de dependencias sintácticas de ‘hibernar’.

A pesar de esto, fue sorprendente descubrir que la palabra ‘persona’ estaba ausente de esta lista, por lo tanto, se llevó a cabo otro test buscando solo oraciones donde una palabra ocurre con el neologismo. Por ejemplo, mostrar cada oración en la que el lema ‘persona’ ocurre en la misma oración que el neologismo ‘hibernar’ y luego verificar manualmente las oraciones para ver si la palabra ‘persona’ actuaba como sujeto del neologismo. Y ese fue el caso, al menos en tres de las oraciones, p. ej. “De esta manera, tendrán que repararla a tiempo para poder salvar a las 5.000 personas que permanecen hibernando ... y también a ellos mismos”, donde el programa debería detectar que ‘personas’ forma parte del sujeto de ‘permanecen hibernando’. Encontrar estas limitaciones es muy importante. Como se observa aquí, en funciones como la detección de dependencias sintácticas, los programas están aún lejos de ser perfectos, así que los resultados obtenidos a través de estos se tienen que coger con pinzas. En este caso, hemos visto que en neologismos con frecuencias de aparición muy altas esto no es tan relevante, ya que el número de ocurrencias compensará para aquellas que el programa pase por alto o se equivoque etiquetando, pero esto no será así con neologismos de más baja frecuencia y donde este sentido neológico sigue siendo aún una parte tan baja del uso del término.

Como he mencionado antes, ahora aparece también la nueva etiqueta *obl* que nos marca esos sustantivos que forman parte de complementos circunstanciales, en este caso, tanto de lugar como de tiempo. De esta manera, podemos fácilmente descartar que cualquiera de estas cosas formará parte del significado neológico.

6. Conclusiones

En general, podemos concluir una cosa clara: el análisis del contexto de un neologismo semántico resulta muy útil cuando se trata de identificar este cambio de significado. Ambas pruebas realizadas en este estudio, usando el contexto directo y el contexto general, generalmente nos proporcionan evidencia sólida para afirmar que ha tenido lugar un cambio de significado. Podríamos argumentar que fue el neologismo ‘unicornio’ el que resultó más fácil de identificar con ambos métodos, ya que experimentó un cambio completo de significado, por lo que los resultados obtenidos pertenecían a un campo semántico completamente diferente y no había dudas para decir si provenían del significado original o del neológico. Mientras tanto, para ‘tóxico’ y ‘hibernar’, que experimentaron una ampliación del significado, si bien había sustantivos que eran claros para uno u otro, también había otros que podían generar dudas. Por lo tanto, esto podría indicar que el proceso de creación del neologismo podría ser un factor determinante en cuanto a si será más fácil o no la detección un neologismo semántico usando su contexto.

En cuanto a las diferencias observadas del análisis de diferentes contextos, por un lado, el contexto directo fue suficiente en la mayoría de los casos para proporcionar indicios de este cambio. Se revisaron principalmente dos elementos que precedían al término: los sustantivos y los verbos. Los sustantivos resultaron ser muy útiles en su mayoría, ya que proporcionan mucha información semántica como, por ejemplo, si se trata de un ser humano, un animal o un objeto, y también información sobre su campo semántico. Esto fue relevante para distinguir si formaban parte del significado original o del neológico. Sin embargo, la búsqueda utilizando verbos resultó ser muy ambigua, ya que no contenían toda esta información y muchos de ellos se repetían en ambos años. Esto fue especialmente confuso para los neologismos semánticos como ‘tóxico’

e ‘hibernar’, que solo sufren una amplificación del significado. Podemos entonces concluir que el significado neológico se usa de manera muy similar al original, especialmente si son resultado de generalización, por lo que la construcción verbal a su alrededor no cambia y por eso los verbos no resultan útiles para realizar esta distinción. Mientras que los sustantivos ligados a este término sí que cambian considerablemente y su observación nos sirve de gran ayuda.

No obstante, no podemos olvidar las limitaciones que surgen al observar solo la palabra que precede al neologismo. Aunque funcionó especialmente bien para adjetivos como ‘tóxico’, porque en español suelen modificar al sustantivo que les precede, los resultados de los sustantivos y verbos, aunque eran suficientes para observar el cambio, no eran tan buenos y claros. En el caso del sustantivo ‘unicornio’, en 2011 principalmente se encontraron nombres propios, y para el verbo ‘hibernar’ debemos tener en cuenta que el sujeto no siempre aparecerá antes del verbo. Es por eso que se optó por analizar también un contexto más general.

Los resultados obtenidos a partir de un contexto más general también fueron muy útiles para observar el cambio del neologismo semántico, pero, por supuesto, una observación más amplia también proporcionó resultados más ambiguos. Por ejemplo, un recuento más general de las ocurrencias alrededor del término ‘tóxico’ en 2011 también podría llegar a incluir el sustantivo ‘persona’ debido a construcciones como ‘sustancias tóxicas para personas’, que aunque hace referencia a personas, sigue siendo parte del sentido original de la palabra porque lo que está modificando ‘tóxico’ aquí es una sustancia.

El contexto más amplio brindó mejores resultados para el problema con ‘unicornio’ y los nombres propios que surgían del corpus de 2011, ya que los sustantivos que recoge están mucho más claramente relacionados con el significado original. El significado neológico del verbo ‘hibernar’ es el que trajo más problemas para ser identificado con ambos métodos, pero sospecho que esto puede deberse a que la frecuencia del uso neológico es mucho menor que en los otros dos casos. A pesar de esto, en los resultados obtenidos con el código Python, que proporciona una lista de los 50 sustantivos más frecuentes (y que se puede cambiar para obtener aún más resultados), aún es posible observar algunos de los usos neológicos del término. De esta manera, se encontró otro factor que puede afectar los resultados y la dificultad de identificación: la frecuencia de uso del significado neológico. El caso más claro fue ‘hibernar’, ya que para empezar el verbo en su totalidad tenía una frecuencia absoluta mucho menor a los demás, pero luego en el corpus de 2021-22, a diferencia de ‘unicornio’ cuyo uso neológico superaba con creces al del significado original, y ‘tóxico’ cuyos dos sentidos parecían tener una cantidad igual de usos, el uso de ‘hibernar’ en su sentido neológico era muy reducido en comparación con el del significado establecido. Por lo tanto, mientras que fue suficiente mirar solamente los diez primeros resultados obtenidos de los primeros dos términos para descubrir que había habido un cambio semántico, necesitamos al menos revisar 50 resultados para hacer lo mismo con ‘hibernar’.

Luego, fue especialmente interesante observar las dependencias sintácticas más frecuentes vinculadas al neologismo. Aunque la herramienta aún está en sus primeras etapas y dista mucho de tener resultados perfectos, nos permite ver solo las palabras que tienen algún tipo de relación con nuestro término. Por eso, fue útil para distinguir los casos como, por ejemplo, donde ‘persona’ podría estar modificando ‘tóxico’ pero no ser el sujeto de este como con ‘tóxico para las personas’. Y para ‘hibernar’, confirmó que palabras como año, invierno o mes son solo complementos que funcionan como circunstanciales de tiempo. Sería interesante experimentar más con esta herramienta y hacer predicciones: por ejemplo, normalmente un neologismo semántico que es un verbo sufriría cambios en el sujeto o el objeto al que está vinculado, y observar otros elementos de la oración podría ser innecesario; o el adjetivo normalmente sufre cambios en el sustantivo que acompaña, etc.

Los resultados obtenidos en este estudio refuerzan aún más las afirmaciones mencionadas anteriormente de que el contexto es una excelente manera de identificar neologismos semánticos. Entre ambas opciones, el contexto general sería más deseable, ya que los resultados son más fieles a la realidad. Además, las herramientas computacionales pueden ser de gran ayuda para ahorrar esfuerzo y tiempo a la hora de obtener resultados. Sin embargo, estos resultados no siempre son correctos, como se vio antes con ‘el’ siendo incorrectamente etiquetado como un sustantivo. También se ha visto cómo otras herramientas computacionales podrían ser útiles en este proceso de identificación, como la herramienta de dependencia sintáctica utilizada; sin embargo, en el estado actual, las limitaciones de estas herramientas pueden ser observadas en los resultados. Con más trabajo y a medida que estas herramientas avancen, esta detección podría volverse aún más rápida y hasta cierto punto automática.

7. Limitaciones

Este estudio tuvo algunas limitaciones que deben tenerse en cuenta. En primer lugar, solo se analizaron tres neologismos, por lo que ninguna de las conclusiones puede generalizarse. Además, cada neologismo era una categoría gramatical diferente para observar si los resultados variarían dependiendo de ella, pero eso también significa que otros sustantivos, verbos y adjetivos pueden mostrar resultados completamente diferentes a los observados.

En cuanto al corpus, me gustaría mencionar que un corpus más antiguo habría sido ideal, aunque los resultados obtenidos no se han visto afectados. El corpus más antiguo, el de 2011, estaba aún muy cerca del inicio del uso neológico de la mayoría de los neologismos estudiados, lo que podría crear interferencias en los resultados. Luego, está el problema de que Sketch Engine solo permite la descarga de un máximo de 10.000 contextos, por lo que los tests con el código Python no se realizaron con la totalidad del corpus, por lo que los resultados podrían estar sesgados. Como ya se mencionó, la función de dependencia sintáctica de la librería Spacy aún tiene bastantes problemas de categorización, por lo que los resultados podrían verse afectados por ello, especialmente en casos como ‘hibernar’ donde había menos ocurrencias y algunos usos neológicos podrían perderse.

Finalmente, es importante mencionar que aunque se usó herramientas computacionales para el recuento de palabras, al final, el análisis de los resultados fue manual, por lo que los resultados no son completamente objetivos si nos atenemos a mi propio juicio sobre lo que contaba o no como parte del significado original o el neológico.

Referencias

- [1] Adelstein, A. (2022). Neología y semántica: grados de neologicidad en el ámbito nominal. *Neología y semántica: grados de neologicidad en el ámbito nominal*, 327-345.
- [2] Bastuji, J. (1974). Aspects de la néologie sémantique. *Langages*, (36), 6-19.
- [3] Bernal, E., Freixa, J., y Torner Castells, S. (2020). *Criterios para la diccionarización de neologismos: de la teoría a la práctica*.
- [4] Cabré, M. T. (2006). La clasificación de neologismos: una tarea compleja. *Alfa*, 50(2), 229-250.
- [5] Díaz Hormigo, M. T. (2007). Aproximación lingüística a la neología léxica. En *Morfología: investigación, docencia, aplicaciones: Actas del II Encuentro de Morfología: investigación, docencia* (pp. 33-54). Universidad de Extremadura.
- [6] Lamaczová, K. (2018). Análisis de los neologismos semánticos viral y bizarro: un estudio de corpus. *Language design: journal of theoretical and experimental linguistics*, 20, 0093-120.
- [7] Martínez Linares, M. A. (2015). Cuestiones en torno al tratamiento de los neologismos semánticos: entre la intuición, el análisis lingüístico y el diccionario. En Cuniță, A.; Lupu, C. (ed.). *Neologie, neologism. Conepte, analize*. Bucarest: Universidad de Bucarest, 69-100.
- [8] Sánchez Manzanares, M. C. (2009). Procedimientos trópicos en la neología semántica: sistematicidad y creatividad. *Revista de investigación lingüística*, 12, 123-146.
- [9] Real Academia Española (s.f.). *Diccionario de la lengua española*, 23.^a ed., [versión 23.7 en línea]. Recuperado de <https://dle.rae.es>. [Consulta: 26/02/2024].
- [10] Real Academia Española y Asociación de Academias de la Lengua Española. (s.f.). *Diccionario pan-hispánico de dudas (DPD) [en línea]*. Recuperado de <https://www.rae.es/dpd>, 2.^a edición (versión provisional). [Consulta: 26/02/2024].
- [11] Rosa, Y. H., Moreno-Martínez, F. L., y de Armas Castro, M. A. (2011). ¿Realmente existe un miocardio aturdido o hibernado? *Cardiacore*, 46(4), 167-168. <https://doi.org/10.1016/j.carcor.2011.06.001>
- [12] Torres Rivera, A. (2019). *Deteccion y extraccion de neologismos semanticos especializados: un acercamiento mediante clasificacion automatica de documentos y estrategias de aprendieaje profundo* (Doctoral dissertation). Université d’Avignon; Universitat Pompeu Fabra.