

Obtención del léxico para un diccionario incluyente en el área de sexualidad

Obtaining the lexicon for an inclusive dictionary in the area of sexuality

Jorge LÁZARO (UABC)
lazaroj@uabc.edu.mx

Gerardo SIERRA (UNAM)
gsierram@iingen.unam.mx

Mariana AGUILAR CONTRERAS (UNAM)
marianbano302@gmail.com

Recibido em: 30 de set. de 2020.
Aceito em: 28 de out. de 2020.

LÁZARO, Jorge; SIERRA, Gerardo; AGUILAR CONTRERAS, Mariana. Obtención del léxico para un diccionario incluyente en el área de sexualidad. *Entrepalavras*, Fortaleza, v. 11, n. esp., p. 236-263, mar. 2022. DOI: 10.22168/2237-6321-11esp2140.

Resumen: Con el fin de sistematizar la terminología de la sexualidad para facilitar la comunicación en el tema, se desarrolla actualmente el *Diccionario de las Sexualidades en México*. Para ello, y como base para conocer el léxico utilizado por los mexicanos para referirse a su sexualidad, se ha creado el *Corpus de las Sexualidades de México* (CSMX) en varias etapas. Actualmente, este corpus integral toma en cuenta los elementos sociales y culturales, así como la diversidad de registros y niveles del habla; con lo anterior, se asegura una obtención amplia del léxico en el área. Este artículo presenta la reconfiguración del corpus y la metodología para la obtención del léxico que conformará el lecionario del diccionario, ya que anteriormente sólo se contaba con textos especializados y semiespecializados. Después del procesamiento computacional de léxico, se discuten las ventajas y las aportaciones de incluir un mayor número de registros lingüísticos en trabajos de corte social y terminológico como este, lo que da pie a un diccionario incluyente.

Palabras clave: Lexicología. Lexicografía computacional. Terminología. Sexualidad.

Abstract: In order to systematize the terminology of sexuality to facilitate communication on the subject, the *Dictionary of Sexualities in Mexico* is currently being developed. To this end, and as a basis for knowing the lexicon used by Mexicans to refer to their sexuality, the *Corpus of Mexican Sexualities* (CSMX) has been created in several stages. Currently, this comprehensive corpus takes into account social and cultural elements, as well as the diversity of registers and levels of speech; this ensures a wide range of lexicon in the area. This article presents the reconfiguration of the corpus and the methodology for obtaining the vocabulary that will make up the dictionary lexicon, since previously only specialized and semi-specialized texts were available. After the computational processing of the lexicon, the advantages and contributions of including a greater number of linguistic records in studies of a social and terminological nature such as this one are discussed, giving way to an inclusive dictionary.

Keywords: Lexicology. Computational lexicography. Terminology. Sexuality

Introducción

Para aproximarse al tema de la sexualidad humana es necesario considerar que hay muchos elementos sociales y culturales que influyen en su estudio. Además, existen diversos registros y niveles del habla, por lo que, si queremos una muestra integral del léxico utilizado en el área de sexualidad, con base en la teoría sociocognitiva de la terminología (TEMMERMAN, 2000), es preciso tomar en cuenta muestras de los diferentes niveles de la lengua.

Hablar de sexualidad es un tema complicado, debido a que la sociedad lo ha convertido en tabú. Cuando la gente habla de sexo es natural la presencia de cierta incomodidad o dificultad para expresarse, lo que da pie a la aparición de eufemismos y coloquialismos. En el primer caso palabras elegidas con cautela tratando de suavizar lo que se está diciendo o darle un énfasis particular, en el segundo, palabras de uso más general para crear una sensación de cercanía social. Sin embargo, también sucede que se trata el tema con términos que pueden resultar ofensivos o despectivos.

Dada esta problemática se ha considerado la necesidad de crear un diccionario de la sexualidad que sea incluyente. Se propone que el *Diccionario de las Sexualidades en México* (MEDINA; SIERRA, 2004, SIERRA; MEDINA; LÁZARO, 2009), además de la terminología científica, contenga la información de un estrato menos culto y más coloquial para hacer de los productos derivados un material completo. Se trata, pues, de una optimización de los recursos diseñados para este proyecto, de manera que nos brinden un acercamiento global del léxico utilizado en el área de sexualidad, no sólo en el nivel culto, sino dando muestra de otros estratos de la lengua, consiguiendo así una aproximación integral.

El objetivo principal de este estudio es proporcionar una aproximación descriptiva del uso de palabras usadas para referirse a la sexualidad, sus variantes denominativas, coloquialismos y eufemismos, en los contextos donde se desarrollan atendiendo a su equivalencia dentro del discurso especializado.

Nos hemos dispuesto a abordar esta problemática desde un enfoque de la disponibilidad léxica y de la lingüística de corpus. Basados sobre todo en los estudios de López Morales (1999), Wotjack (2003), Butragueño (2005), Bartol (2006) y Paredes García (2014), hemos derivado una serie de pautas para, en primer lugar, extraer las palabras referidas a la sexualidad que son utilizadas por un amplio espectro de hablantes mexicanos y, en segundo lugar, para establecer las relaciones existentes entre estas unidades y sus variantes. La idea es reflejar el comportamiento del léxico en materia de sexualidad englobado en contextos especializados.

De López Morales retomamos, como varios estudios, la definición de léxico disponible como “el conjunto de palabras que los hablantes tienen en el lexicón mental y cuyo uso está condicionado por el tema concreto de la comunicación” (SARALEGUI; TABERNEIRO: 1). De igual manera abordaremos, como se podrá ver adelante, la noción de léxico básico como aquel registro que contiene las palabras contenidas en un inventario para un conjunto de hablantes determinados, gramaticales en su mayoría, y no gramaticales en un número menor, que van hacia un significado general.

De la recopilación de artículos coordinada por Wotjak (2003), tomamos que la investigación sobre la disponibilidad requiere de un análisis cuantitativo de los resultados obtenidos. Los autores presentan el método que utilizaron para el procesamiento de los datos, el cual se efectúa a través del cálculo del índice de la disponibilidad léxica. Nosotros utilizaremos herramientas léxicas que nos permitan realizar un análisis cuantitativo y, de esta manera, conocer cuáles son las palabras más relevantes o recurrentes del corpus.

Moreno Fernández (2006: 1), al presentar el *Proyecto Panhispánico de Léxico Disponible* y el *Proyecto para el Estudio Sociolingüístico del Español de España y de América* (PRESEEA), destaca la importancia de “reunir un gran corpus oral, técnicamente adecuado y sociolingüísticamente representativo de una amplia muestra de ciudades de todo el mundo hispánico”. En el caso de nuestro proyecto, la intención es crear un corpus disponible y representativo de México en el área de sexualidad.

Por otro lado, Paredes García (2014) propone renovar los centros de interés en los estudios de disponibilidad léxica; para una aproximación más acertada a las lenguas, plantea ampliar los horizontes y crear nuevas líneas de desarrollo, mantener los centros que se tienen y agregar otros que permitan tener un panorama más amplio y, por tanto, obtener estudios más enriquecidos sobre la disponibilidad léxica de manera que se recojan piezas léxicas de distintas categorías gramaticales y léxico abstracto. En nuestro proyecto podemos decir que estamos ampliando los horizontes metodológicos a un nuevo centro de interés, el cual sería el de la sexualidad (LÁZARO, 2011, 2015). Como antecedente inmediato, ya ha sido propuesto un acercamiento desde el léxico de los adolescentes mexicanos referido específicamente a la homosexualidad, contextualizado en una variante del norte (LÁZARO; RUBINO, 2020).

Metodología

Para el desarrollo de este proyecto nos valimos principalmente de GECO (Gestor de Corpus)¹, una herramienta desarrollada en el Grupo de Ingeniería Lingüística de la Universidad Nacional Autónoma de México, que ofrece aplicaciones diversas como etiquetado morfosintáctico, extracción de palabras clave y extracción de concordancias, entre otras (SIERRA; SOLÓRZANO; CURIEL, 2017). Para tareas que requerían un proceso más detallado, se usaron WordSmith Tools (LEE; SCOTT, 2018) y Sketch Engine², en sus versiones para el español.

Grosso modo, con base en los principios de la lingüística de corpus (2017), la metodología consta de cuatro etapas:

1. Recopilación de textos referentes al área de sexualidad. Son textos pertenecientes a diversas áreas en que se clasificó la temática de la sexualidad de acuerdo con el Instituto Kinsey³ y consultores privados, como la Federación Mexicana de Educación Sexual y Sexología A. C.⁴. Dichas áreas se detallan más adelante.
2. Limpiado y etiquetado de textos. Los nuevos documentos provenían de textos en internet, o de fuentes bibliográficas que se digitalizaron y se procesaron con un reconocedor óptico de caracteres (OCR por sus siglas en inglés). Todos ellos se pasaron a formato de texto plano y se recodificaron en UTF-8.

¹ <http://www.corpus.unam.mx/geco>

² <https://www.sketchengine.eu/>

³ <http://institutokinsey.com/>

⁴ <https://femess.org.mx/>

3. Integración del corpus coloquial de las sexualidades al CSMX. Esta fase consistió en la creación de una base de datos con la nueva información correspondiente a cada uno de los textos que conforman el nuevo corpus. Es decir, se reestructuró el contenido del original CSMX, la indexación y los metadatos.

4. Procesamiento del corpus por diversas herramientas. Esta fase buscó obtener los términos más representativos del corpus y posteriormente la visualización de los contextos en los que se presentan. Aquí se hizo uso, sobre todo, de la herramienta de concordancias y el extractor de términos Termext, ambos incluidos en GECO, así como de WordSmith y Sketch Engine para las colocaciones, las cuales proporcionan un valioso método para la obtención de unidades léxicas multipalabra (FUERTES-OLIVERA; BERGENHOLTZ; NIELSEN; *et al.*, 2012).

Recopilación

La idea de esta investigación surge a partir del análisis de la primera versión del *Corpus de las sexualidades en México* (CSMX), publicado en 2010 en el portal de los corpus lingüísticos de la UNAM⁵ (SIERRA; MEDINA; LÁZARO, 2009). Para llevar a cabo la recopilación de textos sobre la temática de sexualidad, fue necesaria la ayuda de alumnos pertenecientes a las carreras de Ingeniería en Computación, licenciatura en Lengua y Literatura Hispánicas, maestría en Lingüística y licenciatura en Informática, quienes cursaban la asignatura de Lingüística de corpus en la UNAM. Se partió de este proyecto, pero con la idea de ampliar el nivel más bajo y cotidiano debido a que este corpus estaba centrado principalmente en documentos científicos o de divulgación de especialistas hacia hablantes legos. Con este corpus, se extrajo el léxico básico de las sexualidades en México (LÁZARO, 2010), que cubre la terminología más general del área. Después de ello, surgió el interés en acrecentar las muestras de habla coloquial con el fin de acercarnos a las palabras de uso común en México que son utilizadas para referirse a cuestiones de sexualidad y así constituir un material realmente fiable sobre el léxico sobre el que versa. El objetivo de esta fase ha sido, por tanto, mostrar el uso de las unidades léxicas que se presentan tanto en un discurso especializado como en el habla no especializada.

Se considera importante el acercamiento a este nivel, ya que nos da muestra de un lenguaje de uso cotidiano para la sociedad, un lenguaje que surge a partir de los conocimientos empíricos y que se presenta en un contexto especializado cuando los hablantes generales o

⁵ <http://www.iling.unam.mx/corpus/>

legos no encuentran unidades especializadas en, por ejemplo, consultas médicas y recurren a su experiencia para poder comunicar un problema o dolencias en el ámbito sexual. Este fue el punto de partida del presente proyecto de escalamiento del CSMX, pero hay que considerar también otras variables como que, por lo menos en el español de México, el mayor número de lexías referentes a sexualidad no se encuentran en el discurso especializado, sino en el nivel coloquial. Para la ampliación del corpus en el ámbito coloquial se decidió realizar una búsqueda exhaustiva en foros de discusión y redes sociales como Twitter y Facebook, así como Youtube.

El CSMX estaba organizado en cinco niveles, clasificados de acuerdo con el dominio o formato en el que habían sido localizados. Para dividirlos, se consideró el tipo de texto al que pertenecían y el lugar de la web en que habían sido obtenidos, así como su formato (SIERRA; MEDINA; LÁZARO, 2009).

A partir de estos datos, y como un trabajo de actualización, se decidió llevar a cabo una reestructuración de la clasificación anteriormente mostrada y se optó por reducir de 5 a 3 niveles la organización del corpus. Dicha reestructuración tuvo la finalidad de mostrar los textos de acuerdo con su nivel de especialización, es decir, de los textos más especializados hasta los más coloquiales. Posteriormente se agregó un cuarto nivel, el cual está compuesto por textos de ficción referentes a la literatura erótica mexicana.

De esta manera, actualmente el corpus se encuentra clasificado en cuatro niveles:

1. Científico. Como su nombre lo dice, en esta sección se encuentran documentos de carácter científico, investigaciones y artículos con bibliografías y referencias. Suelen ser documentos que tienen como finalidad informar a un sector especializado.
2. Informativo o de divulgación. En este nivel hay textos que brindan información, sin embargo, no alcanzan un carácter científico. Son documentos que tratan de informar al público y, por tanto, el lenguaje utilizado es general.
3. Coloquial. Es el nivel más bajo en cuanto a los registros lingüísticos. En él se encuentran textos obtenidos de diversos foros de internet y redes sociales como Facebook y Twitter. Estos documentos son relatos de experiencias, dudas y

opiniones entre hablantes comunes, no especialistas. En este nivel se encuentran los documentos de interés para nuestra investigación.

4. Ficción. Es un nivel que no se centra específicamente en lo especializado ni en lo coloquial. La literatura, gracias a su riqueza expresiva, nos ofrece una amplia muestra léxica de piezas que se usan para referirse al tema de la sexualidad en muy diversos contextos.

Limpiado

La fase de preprocesamiento (limpiado y etiquetado) consistió en eliminar imágenes, emoticones y algunos caracteres que entorpecieran la interpretación del texto. Además, para evitar problemas al momento de la lectura, los textos fueron convertidos de su formato original a texto plano (.txt) y a una codificación UTF-8, debido a la amplia gama de caracteres que nos ofrece. Esto se hace de manera automática a través del módulo “Convertidor de texto plano” incluido también en GECO. El etiquetado GECO, para esta fase, envía los archivos a un servidor para ser procesados con FreeLing (PADRÓ; STANILOVSKY, 2012).

La fase de diseño e integración consiste en la construcción del corpus referente al habla coloquial y se creó a partir de los textos recopilados por los colaboradores. Se basa, en esencia, en los parámetros propuestos por el Instituto Kinsey y adaptados al español de México de acuerdo con Sierra, Medina y Lázaro (2009). Se tomaron en cuenta las mismas áreas en las que está dividido el CSMX y, a partir de éstas, se realizó la clasificación de los documentos encontrados. Las áreas que podemos encontrar en el nuevo corpus son las siguientes:

1. Fundamentos biológicos de la sexualidad. Es el área donde se insertan los documentos relacionados con la anatomía humana, el cuerpo desde un punto de vista anatómico o como objeto de estudio. En el corpus coloquial, de esta área, encontramos principalmente textos sobre dudas ante métodos anticonceptivos, eyaculación precoz, menopausia y andropausia.

2. La respuesta y la expresión sexual. Esta área tiene que ver con asuntos metabiológicos, en la mayoría de los casos, que influyen en el desarrollo de la sexualidad. En el corpus coloquial,

esta se compone por textos en los que los participantes de los foros comentaban sobre sus diversas fantasías sexuales y consejos para mejorar en este ámbito.

3. Comportamiento sexual. Se trata de documentos relacionados con la actividad sexual en general: la masturbación, la estimulación oral y anal, el coito, la sexualidad postural y el aumento o disminución del placer en las relaciones sexuales. Para el corpus coloquial, la mayoría de los textos de nuestro corpus corresponden a discusiones sobre las mejores posiciones sexuales, juegos de cama y relatos sobre experiencias eróticas.

4. La identidad sexual. Son textos referentes a la identidad del sujeto, los roles y la orientación sexual; a los fundamentos de la identidad sexual; las fases del desarrollo de la identidad sexual; los roles y estereotipos masculinos y femeninos; la diversidad de maneras en que la identificación sexual y la sexualidad se desarrollan, y temas referentes a género. En el corpus coloquial, los textos que se encuentran en esta área son relatos, experiencias y dificultades sobre la homosexualidad, consejos sobre cómo manejar la homosexualidad en la familia, nuevas masculinidades, el rol del ser sociosexual, etc.

5. Las enfermedades de transmisión sexual (ETS). En esta sección hay documentos dedicados a la amplia gama de enfermedades transmisibles a través del ejercicio de la sexualidad. Para el caso del corpus coloquial, aquí encontramos textos en los que se habla sobre la experiencia del VIH principalmente, la manera en que fue contraído, los síntomas y la vida con esta enfermedad. Adicionalmente, se encuentran comentarios sobre las diversas infecciones de transmisión sexual y cómo afrontan esta situación las personas que las padecen.

6. La sexualidad variante. Aquí se agrupan los documentos referentes a las diferencias psicosociales en el área de la sexualidad: travestismo, heterosexualidad, bisexualidad, fetichismo, masoquismo, sadomasoquismo, etc. Los textos coloquiales encontrados en esta área son principalmente relatos de experiencias de los usuarios sobre prácticas poco comunes de la sexualidad, así como discusiones sobre la zoofilia, la lactofilia, entre otras.

7. Atracción sexual y relaciones entre individuos. Aquí se clasificaron documentos que tratan sobre el origen y el desarrollo de las relaciones de pareja, las habilidades sociales en la conquista de otros individuos, temas sobre fidelidad e infidelidad, conceptos culturales de las relaciones de pareja, función de las relaciones entre individuos, las terapias de pareja, etc. Aborda, también, temas sobre el abuso u ofensa sexual. El corpus coloquial contiene en esta área textos sobre infidelidades, el primer beso, consejos sobre la relación de pareja y cómo mantener la armonía del hogar, además de textos con sus respectivos comentarios sobre la visión feminista de la sexualidad.

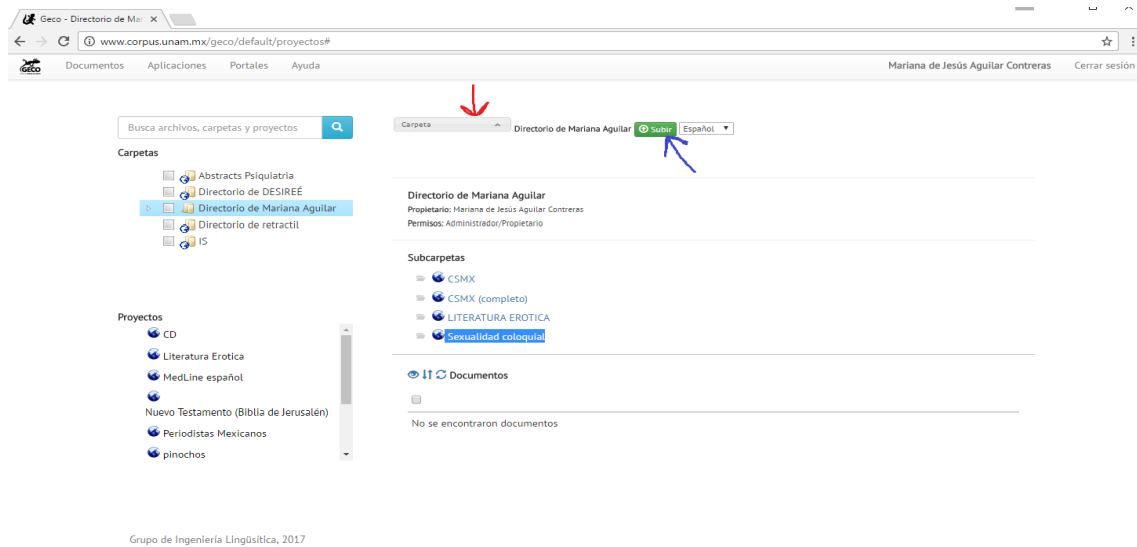
8. La educación sexual, cultura y sexualidad social. Documentos sobre los procesos de la educación sexual y sus limitaciones, la relación sexual en el tiempo y en la cultura occidental, las habilidades sociales en el comportamiento sexual y la sexualidad en diferentes culturas. En esta sección del corpus coloquial encontramos textos en los que se exponen dudas sobre cuál edad es la adecuada para hablar con los hijos sobre cuestiones de sexualidad, cómo hacerlo y cuáles son los límites entre lo que debe y no debe enseñarse.

Integración

Para que el CSMX estuviera disponible en el sistema GECO y la posterior integración del subcorpus de sexualidad coloquial, elegimos la opción “Gestión de documentos y proyectos” que se puede ver en la página principal del gestor. Seleccionando esta opción, es muy fácil integrar los datos. A continuación, describiremos los pasos más importantes del proceso:

Carga de documentos. Inicialmente se creó una carpeta que se pudiera administrar para poder tener acceso al total de los archivos. Para realizar esto, basta con dar clic en el botón despegable señalado con la flecha roja, seleccionar la opción “Nueva carpeta”, nombrarla y aceptar.

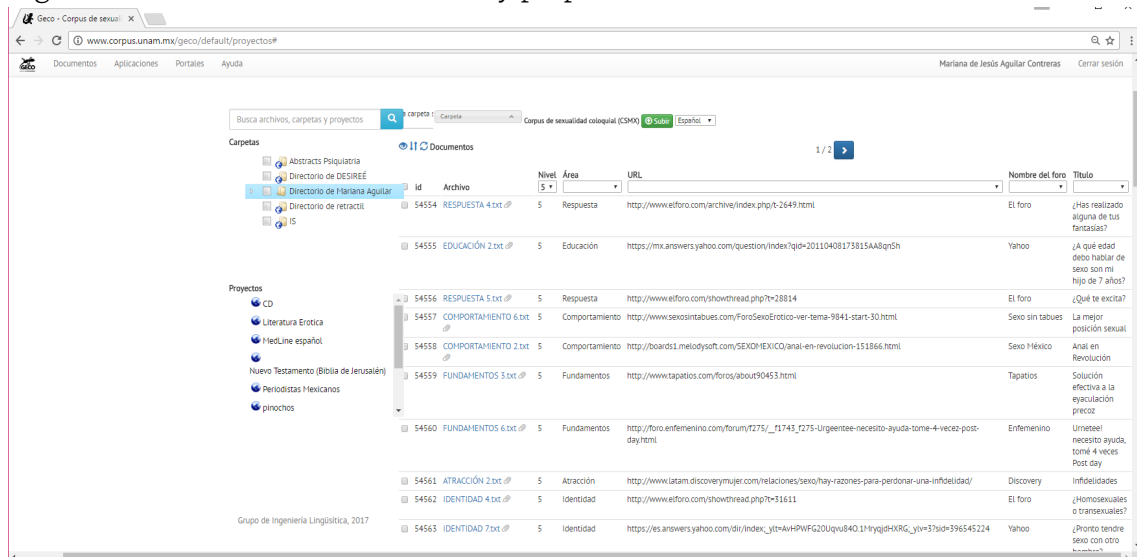
Figura 1- Creación del directorio principal



Fuente: <http://www.corpus.unam.mx/geco>.

Como puede observarse en la figura 1, en el directorio elegido se encuentran disponibles los corpus ya mencionados. Teniendo una carpeta limpia y habiéndola nombrado, hay que elegir el botón “Subir”; de ahí emergerá una ventana en donde aparecen nuestros documentos guardados en la PC, en dicha ventana podemos seleccionar el archivo o archivos que se desean agregar al GECO y damos aceptar (Figura 2).

Figura 2 – Vista de archivos subidos y preprocesados

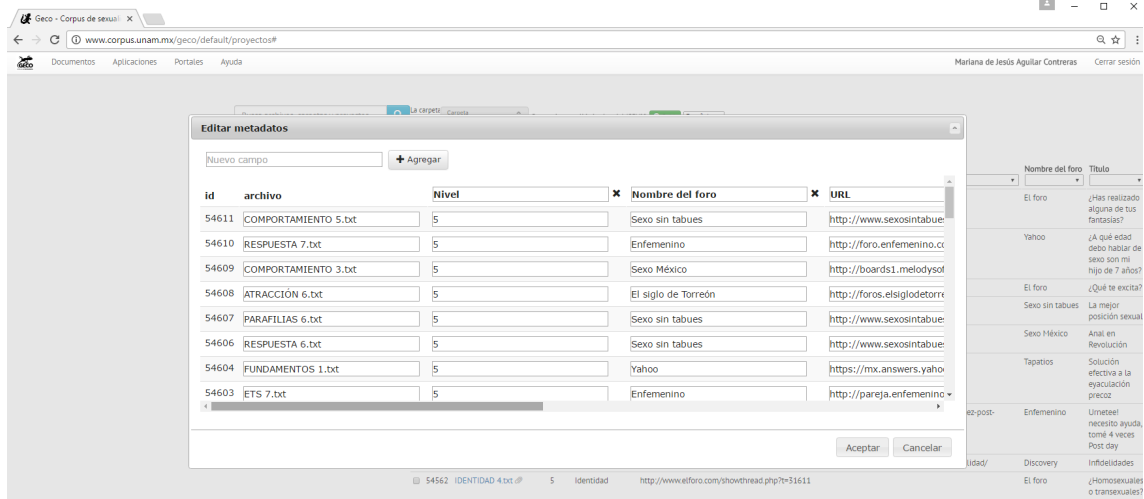


Fuente: <http://www.corpus.unam.mx/geco>.

Captura de metadatos. El GECO brinda la opción de capturar los metadatos de los archivos, previamente agregados al sistema dando

click en la pestaña despegable “Carpeta > Captura metadatos”. Al seleccionar esa opción aparecerá un recuadro como el que aparece en la figura 3, en éste podemos añadir los metadatos que deseamos considerar y comenzar el listado. Al finalizar la captura de nuestros metadatos se debe aceptar y esperar la confirmación de que han sido agregados de manera exitosa.

Figura 3- Reorganización de los metadatos



Fuente: <http://www.corpus.unam.mx/geco>.

Para el CSMX consideramos los siguientes metadatos:

- Título (o tema tratado en el foro);
- Autor (en la mayoría de los casos del tercer nivel no aparece, ya que son foros y suelen ser conversaciones, sólo en algunos casos hay un autor específico);
- Área (de acuerdo con las ocho áreas de la sexualidad descritas previamente)
- Nivel (científico / informativo / foro / literatura);
- Nombre del sitio web en que se encontró;
- URL.

Procesamiento, extracción y selección de términos

La extracción terminológica es el proceso mediante el que se seleccionan unidades lingüísticas de un texto o corpus determinado, las cuales pueden construir términos. Las herramientas utilizadas para realizar esta tarea identifican las unidades que aparecieron con mayor frecuencia en el corpus.

La extracción de términos es sólo una de las aplicaciones de la extracción de información y puede ser utilizada en el desarrollo de herramientas lexicográficas en la delimitación de glosarios y en el análisis diacrónico de la lengua para mejorar la clasificación automática de documentos, entre otras aplicaciones. La extracción de información tiene como objetivo hallar en su interior la información útil para el usuario. Lo que se espera obtener de un sistema de extracción no es un conjunto de documentos, sino un conjunto de definiciones de una disciplina determinada.

Debe considerarse que la explotación de esta herramienta tiene un fuerte perfil lingüístico y requiere la creación de reglas para la extracción de candidatos. Si bien también se puede utilizar como un proceso previo de extracción de términos para su uso en tareas como *clustering* u otro tipo de clasificaciones de textos, existen otras herramientas que son independientes tanto de la lengua tratada como del área de los documentos analizados.

GECO nos brinda una herramienta que nos permite realizar la extracción de los términos de nuestro corpus. Esta herramienta, denominada TermExt (de la composición de *Term Extractor*), es un extractor de términos producto del Grupo de Ingeniería Lingüística que está basado en el algoritmo *NC-Value* (BARRÓN-CEDEÑO; SIERRA; DROUIN *et al.*, 2009). Este algoritmo fue diseñado originalmente para obtener candidatos a término del área de biomedicina en inglés.

La etapa inicial consiste en la generación de una lista de candidatos a término por medio de un filtro lingüístico basado en la búsqueda de patrones sintácticos y la aplicación de una lista de paro, que permita eliminar candidatos que contengan palabras que no se espera que formen parte de los términos del área tratada. La segunda parte consiste en el cálculo del potencial de cada candidato a ser un verdadero término por medio de su longitud y frecuencia de aparición en el corpus, este potencial es llamado *C-value*. La etapa final, llamada *NC-value*, considera que el contexto en el que ocurren los términos es significativo para su discriminación, por lo que se considera que se debe tomar en cuenta la vecindad de cada uno de ellos para favorecer la calidad de los resultados.

El objetivo del *C-value* y el *NC-value* es hacer que los verdaderos términos se localicen en la parte superior de una lista ordenada con base en sus valores calculados, lo que no forzosamente implica que un candidato en la parte inferior de la lista no sea un término.

Resultados de la extracción de términos

Al procesar nuestro corpus, Termext arrojó los términos extraídos como se muestra en la figura 4. En ella es posible observar la frecuencia con que apareció cada uno de los términos del CSMX, así como su evaluación según el algoritmo *NC-Value*. También nos permite ver la descripción del corpus y descargar los resultados.

Figura 4 - Vista de resultados del extractor terminológico

Extractor Terminológico (TermExt)
Extractor de términos.

Conectado como Mariana de Jesús Aguilar Contreras Cerrar sesión Ayuda Termext

Inicio Acerca de Aplicación Participantes Publicaciones

Resultados del proyecto - "Foros sexualidad"

Mostrar 50 términos por página Buscar:

#	Término	Frecuencia	NC-Value
20	parte	184	153.4523941
21	orgasmo	185	151.906841612
22	relación sexual	106	150.955951265
23	asi	177	148.552470605
24	momento	177	142.6
25	amor	172	142.5995338
26	verdad	170	142.365174945
27	placer	173	142.362747218
28	día	175	141.23731334
29	verga	170	136.420998246
30	caso	167	135.445110903
31	amigo	160	133.274856676

Mostrando 1 al 50 de 4,999 términos

Descripción

Tokens: - 254,279.
Types: - 29,759.

Total: 5,000 términos. [Descargar Resultados](#)

Fuente: <http://www.corpus.unam.mx/geco>.

Después de procesar el corpus y extraer la lista de términos, se decidió hacer una sola lista de 40 términos que aparecieron de manera recurrente a partir de una revisión manual. Se eliminaron elementos repetidos o formas no lematizadas, así como expresiones que tenían que ver con metadatos como fechas, enlaces o palabras funcionales. En la Tabla 1 se muestran los términos seleccionados para analizar y, a un costado, entre paréntesis, la frecuencia con que se presentaron a lo largo del corpus.

Tabla 1 - Lista general de términos extraídos

Términos extraídos	
1. mujer (4117)	21. pene (427)
2. hombre (2905)	22. coger (426)
3. madre (1675)	23. erotismo (425)
4. persona (1595)	24. pierna (424)
5. niño (1299)	25. condón (416)

6. <i>pareja</i> (1157)	26. <i>sida</i> (412)
7. <i>amor</i> (1139)	27. <i>médico</i> (394)
8. <i>sexo</i> (978)	28. <i>puta</i> (370)
9. <i>hijo</i> (858)	29. <i>pecho</i> (251)
10. <i>familia</i> (850)	30. <i>vagina</i> (248)
11. <i>placer</i> (236)	31. <i>homosexual</i> (236)
12. <i>joven</i> (830)	32. <i>nalgas</i> (222)
13. <i>placer</i> (810)	33. <i>masturbación</i> (157)
14. <i>semen</i> (734)	34. <i>caricia</i> (143)
15. <i>adolescente</i> (676)	35. <i>pasión</i> (141)
16. <i>enfermedad</i> (593)	36. <i>ano</i> (114)
17. <i>boca</i> (551)	37. <i>testículos</i> (83)
18. <i>orgasmo</i> (545)	38. <i>sexo oral</i> (79)
19. <i>embarazo</i> (451)	39. <i>menstruación</i> (75)
20. <i>excitación</i> (434)	40. <i>lesbiana</i> (22)

Fuente: propia.

Organización de los términos extraídos

Para poder llevar a cabo el análisis de los términos extraídos, se tomó la decisión de agruparlos en espacios donde el léxico tuviese relación conceptual con el fin de facilitar su distribución y posterior análisis, para ello se tomó en cuenta la teoría de los campos léxicos.

La teoría de los campos léxicos apareció en Alemania y Suiza en la década de 1920. Sus orígenes se encuentran en la filosofía y la lingüística. J. Trier fue el verdadero creador de la teoría de campos en 1931 con su monografía sobre los campos (conceptuales) semánticos (MARTÍNEZ, 2003a). Coseriu (1977, p.135) define un campo léxico como “un conjunto de lexemas unidos por un valor léxico común (zona de significación) que esos lexemas subdividen en valores determinados, oponiéndose entre sí por diferencias mínimas de contenido léxico (semas)”. Todos estos referentes están interrelacionados entre sí, por eso las palabras que los nombran en la lengua forman un campo (ROJAS PORRAS, 2006; CORRALES, 1991).

Otra definición que se le adjudica es la de considerarlo un conjunto paradigmático que comparte un área de significación y en el que la selección de uno de sus elementos dentro de la cadena hablada excluye directamente a los otros, es decir, es una asociación paradigmática que busca relaciones conmutables (MARTÍNEZ, 2003a, p. 267).

Siguiendo con Coseriu (1977, p.170), un campo léxico es “una estructura paradigmática constituida por unidades léxicas que se reparten una zona de significación común y que se encuentran en oposición inmediata las unas con las otras”. Al analizar detalladamente las palabras es posible observar que hay puntos de contacto entre ellas y, de igual manera, elementos discrepantes que sirven para oponerlos entre sí. El significado de cada palabra lo da el conjunto de rasgos semánticos. Cuando las palabras poseen los semas primarios, los más generales en común pertenecen al mismo campo léxico-semántico (MARTÍNEZ, 2003b).

En este punto cabe destacar la diferencia entre un campo semántico y un campo léxico. Dicha diferencia consiste en que el primero está compuesto por un conjunto de palabras pertenecientes a la misma categoría gramatical que comparten algún rasgo de significado; mientras que en un campo léxico las palabras pueden ser de distintas categorías gramaticales, pero siempre compartiendo un elemento o idea de significado común.

Así, a partir de los términos extraídos se eligieron cuarenta piezas léxicas, las cuales fueron agrupadas en seis grandes campos léxicos distintos a partir de una selección manual teniendo en cuenta los preceptos teóricos ya mencionados. Hablamos de campos léxicos ya que no siempre las palabras que los componen pertenecen a la misma categoría gramatical. Entre los términos seleccionados hay sustantivos, principalmente, pero también verbos.

El grupo de campos semánticos se muestra a continuación, en la Tabla 2.

Tabla 2 - Campos semánticos y términos asociados

Partes del cuerpo	Identidad	Parentesco	Sexo	Sensaciones	Medicina
<i>ano</i>	<i>adolescente</i>	<i>familia</i>	<i>coger</i>	<i>amor</i>	<i>condón</i>
<i>boca</i>	<i>hombre</i>	<i>hijo</i>	<i>mastur- bación</i>	<i>caricia</i>	<i>embarazo</i>
<i>pecho</i>	<i>homosexual</i>	<i>madre</i>	<i>orgasmo</i>	<i>erotismo</i>	<i>enfermedad</i>
<i>pene</i>	<i>lesbiana</i>	<i>padre</i>	<i>semen</i>	<i>excitación</i>	<i>médico</i>
<i>pierna</i>	<i>mujer</i>	<i>pareja⁶</i>	<i>sexo</i>	<i>pasión</i>	<i>sida</i>
<i>nalgas</i>	<i>niño</i>		<i>sexo oral</i>	<i>placer</i>	
<i>testículos</i>	<i>persona</i>				
<i>vagina</i>	<i>puta</i>				

Fuente: propia.

⁶ Para este lexema se consideraron dos acepciones; como sinónimos consideramos por un lado *espos(o)* y por el otro, *novi(o)*, cada uno con sus respectivas variantes.

Se tomó cada una de las palabras mostradas en la tabla anterior para buscar sus equivalentes léxicos en diversos diccionarios de sinónimos, así como en la web, con el fin de obtener una mayor cantidad de piezas léxicas que compartieran la misma significación y fueran intercambiables en contextos determinados.

Por ejemplo, hay palabras como *beso*, que sólo presentó el sinónimo *ósculo*, el cual no arrojaba resultados en el CSMX. Por lo anterior, y en varios casos similares, decidimos quitarla. En otros casos como *sida* —y su sinónimo *VIH*— se presentaron en proporción de un 60% y 40%, respectivamente.

Con esta idea en mente, y para poder presentar resultados más precisos, así como explicar su aparición u omisión en la lista final, se tomó en cuenta la noción de *variación denominativa*.

Una variante denominativa es cada una de las alternativas denominativas que presenta un concepto determinado. Es una unidad terminológica, simple o compleja, que refleja en su estructura características del concepto que denomina.

De acuerdo con Freixa (2002), la variación denominativa obedece a causas de diversa naturaleza y no siempre se produce de manera consciente por el emisor. Sin embargo, otros autores afirman que existe una motivación cognitiva: la proyección de un concepto especializado en diferentes términos motivados se explica por la presencia de una motivación múltiple durante el acto denominativo (FERNÁNDEZ SILVA, 2010).

Los argumentos que justifican la motivación cognitiva de la variación denominativa se apoyan en tres principios fundamentales:

- a) La variabilidad de la categorización y estructuración conceptuales: Debido a que las estructuras lingüísticas son un reflejo de las estructuras conceptuales, es lógico que la variabilidad conceptual se manifieste a través de la variación denominativa.
- b) La diversidad de motivaciones y su carácter contextual: Cuando un emisor elige una denominación sobre otra, está adoptando un punto de vista determinado sobre el concepto, en función de las características que considera esenciales en una determinada situación.
- c) La complejidad de la relación concepto-término: Condiciones cognitivas que posibilitan y condicionan la variación.

La denominación terminológica se rige por factores de orden cognitivo o pragmático, como la intención y la perspectiva, alejada de ideales comunicativos. El emisor selecciona la denominación que considera más adecuada y comprensible en un contexto específico, de manera que diferentes contextos pueden requerir diferentes motivaciones. Cabré (1999) destaca que, si la comunicación especializada se produce en una variedad de escenarios, el lenguaje especializado debe necesariamente reflejar dicha variedad. A propósito de esto, Fernández Silva (2010) afirma, siguiendo a Tercedor Sánchez (2004, p. 199), que “La variación denominativa no se percibe como una falta de conocimiento experto o terminológico, sino que *pone de relieve matices cognitivos en la forma de ver una misma realidad*”. Esta afirmación resulta interesante, ya que podemos afirmar que la diversidad de lexemas utilizados para referirse a un mismo concepto nos da información sobre los hablantes y los contextos en que se prefiere el uso de uno u otro sinónimo.

Como resultado, Freixa (2002) propone una tipología de las causas de la variación denominativa en los textos especializados a partir de diversa información recopilada. Su propuesta pretende agrupar y articular todas o la mayoría de las causas observadas. La tabla 3 es la siguiente:

Tabla 3 - Resumen de las causas de la variación denominativa

Tipo	Subtipos
1. Causas previas	La redundancia lingüística La arbitrariedad del signo lingüístico Las posibilidades de variación de la lengua
2. Causas dialectales	Variación geográfica Variación cronológica Variación social
3. Causas funcionales	Adecuación al nivel de la lengua Adecuación al nivel de especialización
4. Causas discursivas	Evitar la repetición Economía lingüística Creatividad, énfasis y expresividad
5. Causas interlingüísticas	Convivencia del término “local” con el préstamo Diversidad de propuestas alternativas
6. Causas cognitivas	Imprecisión conceptual Distanciación ideológica Diferencias en la conceptualización

Fuente: propia.

Análisis de los resultados

Teniendo ya los campos semánticos con las piezas léxicas que los conforman y sus respectivos sinónimos, nos dimos a la tarea de buscar la frecuencia con que aparecían en el corpus, así como el número de documentos diferentes en los que se presentaron. Después de esta exhaustiva búsqueda obtuvimos los resultados que se muestran en la Tabla 3⁷.

En la primera columna se observa resaltado el lexema seleccionado y, en las filas consecuentes, los sinónimos obtenidos. Posteriormente, en las cuatro columnas sucesivas encontramos los niveles en que clasificamos el corpus con las frecuencias de cada lexema y sus variantes. El valor de la izquierda muestra el número de documentos distintos en los que se presentó, mientras que el número entre paréntesis nos indica su frecuencia por nivel en el CSMX. Al final, la última columna nos muestra el valor de la sumatoria de los cuatro niveles, el número de documentos y frecuencia total en el corpus.

Esta tabla nos expone de manera gráfica el uso de determinados lexemas de acuerdo con el nivel en que se presenta.

Tabla 4 - Términos asociados al campo semántico *Identidad*

IDENTIDAD					
Palabra	Científico	Informativo	Foro	Ficción	Total
adolescente	36 (467)	26 (121)	21 (41)	13 (47)	96 (676)
joven	41 (359)	26 (124)	38 (80)	24 (267)	129 (830)
muchacho	12 (24)	4 (9)	13 (28)	23 (222)	52 (283)
chavo		1 (1)	18 (32) ⁸	5 (97)	24 (130)
mozo				7 (27)	7 (27)
púber	3 (7)		1 (1)		4 (8)
efebo	1 (1)			1 (1)	2 (2)
					314 (1956)
hombre	60 (972)	61 (429)	106 (795)	29 (709)	256 (2905)
varón	33 (312)	16 (65)	15 (43)	10 (23)	74 (443)

⁷ Cabe mencionar que, como esta, existe una para cada campo semántico y que por cuestiones de espacio no incluimos. Tómese la presente como un ejemplo representativo de la presentación de los datos. Las demás tablas, así como su explicación exhaustiva, se presentan en la tesis de Aguilar Contreras (en preparación).

⁸ Los datos que se encuentran subrayados de color azul son aquellos que se localizaron con mayor frecuencia en el corpus; en algunos casos hay datos resaltados de color rosa, éstos son casos en los que las piezas léxicas se presentaron en un número mayor de documentos distintos, sin embargo, la frecuencia total de apariciones fue menor que el de color azul.

señor	1 (1)	1 (1)	18 (21)	23 (206)	43 (229)
muchacho	6 (16)	4 (6)	9 (22)	18 (76)	37 (120)
chavo		1 (1)	16 (29)	5 (92)	22 (122)
macho	6 (13)	2 (2)	12 (38)	11 (35)	31 (88)
mozo				6 (24)	6 (24)
señorito			2 (2)	1 (1)	3 (3)
					472 (3934)
homosexual	22 (78)	11 (62)	19 (64)	6 (32)	58 (236)
gay	7 (16)	7 (8)	30 (105)	7 (19)	51 (148)
puto			4 (5)	6 (47)	10 (52)
maricón	2 (2)		2 (2)	7 (34)	11 (38)
marica			1(2)	3 (14)	4 (16)
joto			1(1)	2 (2)	3 (3)
mariquita				2 (2)	2 (2)
puñal				1 (1)	1 (1)
puñetas				1 (1)	1 (1)
					141 (497)
lesbiana	1 (1)	3 (13)	4 (4)	2 (4)	10 (22)
marimacho	1 (8)		1 (1)		2 (9)
tortillera				1(1)	1 (1)
					13 (32)
mujer	63 (1765)	60 (504)	103 (1018)	42 (830)	268 (4117)
señora	1 (1)		15 (19)	19 (267)	35 (287)
muchacha	7 (12)	1 (2)	4 (5)	16 (109)	28 (128)
dama			6 (8)	17 (66)	23 (74)
señorita			5 (13)	9 (44)	14 (57)
chava	1 (1)	3 (3)	18 (24)	5 (27)	27 (55)
hembra	9 (17)	3 (3)	6 (12)	8 (19)	26 (51)
moza				2 (2)	2 (2)
					423 (4771)
niño	33 (386)	11 (24)	46 (190)	27 (699)	117 (1299)
nene		1 (1)	17 (27)	5 (18)	23 (46)
infante	4 (11)	4 (4)	3 (3)	1(1)	12 (19)
escuincle				4 (14)	4 (14)
mocosos			2 (2)	3 (3)	5 (5)
					161 (1383)
persona	56 (464)	57 (452)	101 (480)	25 (199)	239 (1595)

<i>individuo</i>	35 (189)	18 (69)	6 (30)	4 (4)	63 (292)
<i>ser humano</i>	24 (96)	21 (40)	18 (65)	10 (15)	73 (216)
<i>sujeto</i>	30 (139)	10 (19)	2 (6)	5 (8)	47 (172)
					422 (2275)
<i>puta</i>		2 (2)	33 (85)	17 (283)	52 (370)
<i>prostituta</i>	12 (33)	6 (9)	12 (23)	5 (10)	35 (75)
<i>piruja</i>			1 (1)	4 (25)	5 (26)
<i>perra</i>			4 (14)	4 (9)	8 (23)
<i>zorra</i>			4 (4)	5 (13)	9 (17)
<i>ramera</i>				4 (5)	4 (5)
<i>de la vida galante</i>				3 (3)	3 (3)
<i>cortesana</i>				2 (2)	2 (2)
<i>mujer de la calle</i>				1 (1)	1 (1)
					119 (522)

Fuente: propia.

Para referirse a una persona de poca edad, *joven* es la palabra que se presentó en un mayor número de ocasiones, seguido de *adolescente*. Ambos vocablos se observan sobre todo en textos de carácter científico, al igual que *púber*, por lo que, de acuerdo con Ullman (1972), son nuestros sinónimos más profesionales. En foros, *chavo* fue más usado que en las demás áreas; mientras que *muchacho* y *mozo* son las variantes literarias que se localizaron en documentos de ficción. También se encontró la palabra *efebo* una ocasión en el área científica y una ocasión en el área de ficción.

Hombre es el término más común para referirnos a este concepto y se observó con mayor frecuencia en documentos científicos; sin embargo, esta pieza léxica es recurrente en más foros distintos e incluso un documento informativo (61) más que en científico (60). Con mayor frecuencia también en textos de carácter científico se presentaron los vocablos *varón*, *señor* y *muchacho*, *mozo* (en ficción) y, finalmente, *señorito* y *macho* en foros. En *chavo*, tenemos la misma situación que en *hombre*: se presentan varios casos en el área de ficción, pero en mayor número de textos sobre foros. En esta palabra encontramos formaciones interesantes como ‘tapar el ojo al macho’, refiriéndose a ocultar algo, y ‘es muy macho’, haciendo alusión a alguien valiente y fuerte.

Homosexual es más abundante en textos de carácter científico, aunque también se encuentra en foros y documentos informativos de

manera significativa. En el nivel en que es casi nulo es en ficción, sin embargo, en éste encontramos diversos sinónimos de dicho lexema, tales como *joto*, *puto* (presentes en foros también), *maricón* (foros y científico), *mariquita*, *puñal* y *puñetas*. En foros el más abundante es *gay*. El caso de *puto* es curioso ya que no se usa exclusivamente para referirse a una persona varón homosexual, sino que también es utilizado como un adjetivo de carácter despectivo o molesto como en el caso de “el puto problema”. Normalmente este adjetivo, cuando se refiere a una persona homosexual, suele estar cerca de los verbos *volverse* y *ser*, por ejemplo: “Juan se volvió gay cuando conoció a Gilberto” o “Todos sabemos que Mario es gay desde joven”.

Cabe destacar que en la palabra *homosexual* decidimos hacer distinción entre hombres y mujeres, ya que las palabras que suele utilizarse para denominar a personas con gustos por alguien del mismo sexo varían de acuerdo con el género. Para mujeres se suele usar el lexema *lesbiana*. En este caso, únicamente aparecieron 2 casos de sinonimia: el primero, es la palabra *marimacho*, la cual se encontró 8 ocasiones en un texto científico. El segundo caso es *tortillera*, localizado en ficción. *Lesbiana* se presentó en 3 documentos informativos y 4 de ficción, sin embargo, es más frecuente (13 ocasiones) a nivel informativo.

El lexema *mujer*, lo encontramos principalmente en documentos científicos y foros, mientras que sus distintos sinónimos aparecen sobre todo en ficción, exceptuando *chava* en foros y *hembra* en documentos científicos (usado casi en todo momento para referirse a la especie animal).

En los términos utilizados para referirse a ‘niño’ encontramos mayor frecuencia en foros, sin embargo, es contenido en más documentos de ficción y científicos. *Nene* es utilizado en el estrato coloquial (foros), mientras que su vocablo similar, *infante*, es más usado en documentos científicos; así como *escuincla* y *mocoso* se presentan en documentos relativos a la literatura.

El lexema *persona* y sus sinónimos no presentan gran variedad en cuanto al estrato. *Persona* es más común en foros, mientras que todas sus variantes se presentaron en contextos más académicos. Cabe destacar que en su variante *sujeto* tuvimos que hacer distinción, mediante el contexto, entre lexemas que se referían a un individuo y las conjugaciones del verbo *sujetar*.

Para referirse a *puta*, encontramos que se presenta más veces en ficción que en foros, pero el número de textos en los que se

observa es mayor en foros, por lo que podría decirse que tiene mayor representatividad. *Prostituta* se manifiesta en más textos de carácter científico; mientras que términos como *zorra*, *ramera*, *piruja*, *mujer de la calle*, *de la vida galante* y *cortesana* únicamente aparecieron en ficción (*perra* se presenta en ficción y en foros). Se observaron expresiones como ‘posición de perra’, ‘perra en celo’, ‘hijo de perra’, ‘perra despiadada’ y ‘ser la perra de alguien’.

Finalmente nótese el conjunto de resultados que arroja la herramienta *concordancias* de GECO cuando se hacen dos búsquedas con elementos coloquiales del español de México referidos a la sexualidad.

Figura 5 – Concordancias de la lexía *puto* en el Corpus de las Sexualidades de México



Acerca del CSMX	Proyecto ▾	Ayuda	Ligas de interés	Consulta ▾
	especialmente atractivo le parece el uso de la palabra "	puto	" para otros). Es un ser retardado en	
	muy rudo , me mojo mucho en esas situaciones	puto	Publicado : 12/08/2015 00:44 Yo prefiero ser violado por	
	lo que no sabe mi doctor es que porque soy	puto	y me meten la verga en el culo y por	
	es el día , ni la hora , ni el	puto	momento . Pero qué detestable que cuando empiezo a hablar	
	decía a su madre : " Sirva le a el	puto	" . Su madre , que era una dama muy	
	todas esas indiferencias juntas con trabajos vas a sacar un	puto	quarter . Claro que todavía a la hora de levantar	
	le que después de esas cuatro semanas de cargar con su	puto	plan de austeridad yo sufría cada vez menos con sus	
	suena tan feo . ¿ Sabes cómo me dijo el	puto	viejo de Miami cuando le llamé ? Youfuckin ' slut	
	el gramo . ¿ 0 a poco crees que el	puto	Nefastófeles me la iba a regalar ? Digo , ese	
	para que me dijera su talla y clic : el	puto	Nefastófeles en la línea . Quise tranquilizar lo con el	
	encargo . ¡ Él se encargaba ! Haz me el	puto	favor : Tío_Nefas lo iba a resolver todo . Y	
	Chivo_Viejo salía muy temprano con mi coche ? Era un	puto	tormento kampucheano , pero también servía de acicate . Vivía	
	reír se igual . Yo me estaba tirando a su	puto	cliente y el idiota pensaba que podía chalanear me delante de	
	veía que eras celoso , que te quebrabas de el	puto	berrínche cada vez que me hablaban a el celular ,	
	que te dicen que traes los pantalones pegados porque eres	puto	y cosas por el estilo aparte teníamos una abuela la	
	que algo le había ido a decir que había un	puto	por ahí o no sé y corrí como nunca había	
	que un tipo pudiera pagar por coger se a un	puto	o_sea lo que yo no entendía no sabía era que	
	no sabía era que el que se cogía a el	puto	también era homosexual ¿ mentientes ? pero dije " bueno	
	porque te mandan a el carajo te dicen " pinche	puto	" y en un descuido a lo mejor hasta te	
	le saliva gritando " ¡ es hombre ! ¡ es	puto	¡ ¡ tiene voz de mujer pero es puto !	

Fuente: <http://www.corpus.unam.mx/geco/portal/index/CSMX>.

Nótese que una vez ajustado el sistema y pulidas las listas de términos, el sistema ha sido capaz de extraer un conjunto de concordancias que abarcan las diferentes manifestaciones de la lexía *puto* en el español de México. Estos campos cubren tanto aquellas denominaciones despectivas como las que hacen referencia sólo a la condición de homosexualidad, pasando por el filtro gramatical, pues se ha discernido entre las veces que es utilizado como sustantivo y como adjetivo.

En el caso de otra lexía como *verga*, que también tiene una amplia sinonimia, sucede algo que nos llamó la atención. Obsérvese el conjunto de concordancias:

Figura 6- Concordancias de la lexía *verga* en el Corpus de las Sexualidades de México


The screenshot shows the CSMX Corpus de las Sexualidades de México interface. It features a navigation bar with links for 'Acercas del CSMX', 'Proyecto', 'Ayuda', 'Ligas de interés', and 'Consulta'. Below this is a concordance table for the word 'verga'.

Izquierda	Petición	Derecha
11-47 No mames pinche feroso , a ti no hay	verga	que te llene ni pito que te de gusto ,
sexí que ya me imagina montada arriba de el en su	verga	yo para provocar lo traia un shorts negro muy ajustado
ya estaba demaciado exitada , termine de chupar le la	verga	y el se puso encima de mi acaricio mis pezones con
en 4 esa posicion me encanto ps mientras metia su	verga	en mi vagina me nalgeaba y eso me gustaba tenia
a el hombro , me gusta meter la toda mi	verga	hasta si es posible mis testiculos que ella grite de
ver le la cara y ademas te entra toda la	verga	, así me encanta a mi , aunque si el
mujeres nos gusta tomar de la mamila o biberon (verga) .A mí me supo como entre saladito y
pasa , me duele mucho a el principio aguanto la	verga	, pero si me la sacan para cambiar de posición
doctor es que porque soy puto y me meten la	verga	en el culo y por eso estoy desflorado , porque
problema fisico o mental porque el que ha probado una	verga	, siempre quiere más ... ¿ o no ? Que
toda la razón karlitos a los que nos meten la	verga	siempre queremos mas aunque nos vuelva a doler es delicioso
el hombro , perrito , que se siente sobre tu	verga	etc ... o sera que tienea una verga cabezona !
sobre tu verga etc ... o sera que tienea una	verga	cabezona!!! , larga!!! o
sucesivamente , normalmente siempre les duele algo ya que tengo una	verga	de buen tamaño responsable y respetable pero el dolor que
si yo no los busco ellos vienen solo por mas	verga	y ademas a mí me gusta hacer les el sexo
foto de su perfil es de ellos mamando le la	verga	a un perro , Cada uno tiene sus fetiches y
. No le importaba que me quedara con ganas de	verga	... Después de el aquel día de la fiesta en que
no le importaba que yo me quedara con ganas de	verga	no me dio remordimiento de haber le puesto los

Fuente: <http://www.corpus.unam.mx/geco/portal/index/CSMX>.

Cualquier observador aguzado y conocedor del léxico mexicano puede notar enseguida, y echar en falta, las denotaciones de *verga* como un adjetivo. Es decir, en el 100% de los casos, por lo menos en las primeras cien filas, la palabra *verga* es utilizada únicamente como sustantivo refiriéndose al miembro viril. Pero esto no es casual, resulta que los datos que conforman el corpus cubren un periodo amplio, de 2007 a 2016, con algunos documentos previos, y lo que nos quiere decir esta imagen es que hasta ese año la palabra *verga* como adjetivo o atributo (del tipo “No muy vergas, cabrón” o “Juana es bien vergas”). De esta manera, en concordancia con los estudios de variación léxica, como los de Hiroto y Ruiz (2008) y de Seghezzi (2013), nuestro proyecto también nos muestra que el léxico referido a sexualidad, y sobre todo aquel de orden coloquial, evoluciona de una manera mucho más dinámica que el resto del léxico general, sin contar, acaso, el léxico referido a la tecnología, la computación o las redes sociales.

Conclusiones

La variación léxica ha sido de mucho interés para la lexicografía y ha sido considerada como una parte del lenguaje especializado (ZGUSTA, 1971; KOSTINA, 2009). En el caso del lenguaje utilizado para tratar los temas de sexualidad, se toman aspectos sociales y culturales

que conlleva una diversidad denominativa muy amplia. Este léxico resulta fundamental en la elaboración de diccionarios y terminologías (FREIXA, CABRÉ, KOSTINA, 2008).

Se han presentados los análisis de dos campos semánticos, identidad y medicina, para hacer notar al lector dos características capitales tanto de la distribución de los términos asociados a dichos conceptos como de su variación⁹.

En primer lugar, a partir de los análisis presentados podemos corroborar los estudios de Escobar Hernández (2002), en que es en el área coloquial donde se presenta mayor riqueza a la hora de hablar en el español de México, específicamente cuando se toca el tema de la sexualidad. Es en foros o redes sociales donde encontramos que los hablantes desarrollan más su creatividad, ya que no se sienten tan limitados al momento de expresarse al no haber mecanismos reguladores como la escritura académica. Caso similar pasa en la literatura, área en la que en ocasiones también se encuentran piezas léxicas en desuso o poco comunes en lo cotidiano, pero funcionan en lo literario como eufemismos o para dar un matiz distinto a lo que se está diciendo. En este caso, la creatividad lingüística para hablar de la sexualidad se basa en la revitalización.

En segundo lugar, el punto principal a destacar es la enorme cantidad de sinónimos que se encuentran para referirse a algunas palabras, sobre todo, relacionadas a partes del cuerpo o para hacer referencia al acto sexual. Se observa que, debido a que es un tema tabú (aunque cada vez se hable más abiertamente de él en nuestros días), los hablantes buscan maneras de suavizar esos lexemas en foros de preguntas o dudas médicas, sobre todo. Otra posibilidad, por el contrario, es que se intenta enfatizarlas si se trata de relatos o foros en los que la intención es obtener atracción o atención de quien está leyéndolos. En este caso hablamos de riqueza léxica basada en variación denominativa.

Finalmente es imperante hacer notar la dinamicidad de la variación tanto denominativa como conceptual en las áreas que cubre esta investigación (ÁLVAREZ, 2006). Se han mostrado dos búsquedas con las lexías *puto* y *verga*, pero la experiencia nos mostró que este mismo fenómeno sucede con muchas de las palabras que conforman nuestra lista de términos y vocablos referidos a la sexualidad en el español

⁹ El análisis completo de los campos semánticos, así como el estudio cuantitativo de todos los términos encontrados, se puede encontrar en Aguilar Contreras (2020).

de México¹⁰. Adicionalmente, cabe decir que las mismas concordancias hacen ver el uso especializado y no especializado de las lexías por su contexto de uso, como ha sido tradición en la terminología y lexicografía (CABRÉ, 1993; BERGENHOLTZ; TARP, 1995; ATKINS; RUNDELL, 2008). A manera de cierre, hemos mencionado que no es difícil suponer, por tanto, que las nociones de *variación* y *cambio* — en tanto términos lingüísticos — en el español de México, por lo menos, han tomado un matiz distinto en cuanto a la velocidad con que se llevan a cabo, pues se puede observar en este estudio que fenómenos como la sinonimia o la recategorización son una constante en apenas unos 20 años dentro de ciertos campos semánticos. Pero esto constituye un trabajo a futuro que merece, por lo menos, otro capítulo con respecto a esta línea de investigación.

Bibliografía

AGUILAR CONTRERAS, M. **CSMX**: De lo especializado a lo coloquial, estudio del léxico sobre sexualidad en el español de México. (Tesis de licenciatura). Ciudad de México: Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, en preparación.

ÁLVAREZ, A. **La variación lingüística y el léxico**: conceptos fundamentales y problemas metodológicos. Hermosillo: Universidad de Sonora, 2006.

ATKINS, B. T.; RUNDELL, M. **The Oxford guide to Practical Lexicography**. Oxford: Oxford University Press, 2008

BARRÓN-CEDEÑO, A.; SIERRA, G.; DROUIN, P.; ANANIADOU, S. An improved automatic term recognition method for Spanish. In: GELBUKH, A. (Org.). **Computational Linguistics and Intelligent Text Processing**. Berlin Heidelberg: Springer-Verlag, 2009 p. 125-136.

BARTOL HERNÁNDEZ, J. A. La disponibilidad léxica. **Revista Española de Lingüística**, v. 36, 379-391, 2006.

BERGENHOLTZ, H.; TARP, S. **Manual of Specialised Lexicography**: The Preparation of Specialised Dictionaries. Amsterdam: John Benjamins Publishing. 1995.

BUTRAGUEÑO, P. M. Reseña sobre Pautas y pistas en el análisis del léxico hispano (americano). **Nueva Revista de Filología Hispánica**, Ciudad de México, v. 53, n. 1, p. 212-216, 2005.

CABRÉ, M.T. **La terminología**: Teoría, metodología, aplicaciones. Barcelona: Antártida, 1993.

CABRÉ, M.T. **La terminología**: representación y comunicación: elementos

¹⁰ Para todo aquel que desee hacer búsquedas más detalladas, puede ingresar al CSMX desde: <http://www.corpus.unam.mx/csmx>.

para una teoría de base comunicativa y otros artículos. Barcelona: IULA, 1999.

CORRALES, C. El estudio de los campos semánticos. **Revista de Filología**, Universidad de la Laguna, Tenerife, n. 10, p. 79-93, 1991.

COSERIU, E. **Principios de semántica Estructural**. Madrid: Gredos. 1977.

ESCOBAR HERNÁNDEZ, J. C. **Manual del español coloquial de México**. México: Trillas. 2002.

FERNÁNDEZ SILVA, S. **Variación terminológica y cognición**. Factores cognitivos en la denominación del concepto especializado. (Tesis doctoral). Barcelona: IULA, 2010.

FREIXA AYMERICH, J. **La variació terminològica**: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient. (Tesis doctoral). Barcelona: IULA, 2002.

FREIXA, J.; CABRÉ, M.T.; KOSTINA, I. La variación terminológica en las aplicaciones terminográficas. **Memorias de RITerm**, VIII Simposio Cartagena de Indias. 2008.

FUERTES-OLIVERA, P., BERGENHOLTZ, H., NIELSEN, S. & AMO, M. N. Classification in Lexicography: The Concept of Collocation in the Accounting Dictionaries. **Lexicographica**, Berlín, v. 28 n. 1, p. 293-308, 2012.

HIROTO, U.; RUÍZ, A. Investigaciones sobre la variación léxica del español: Proyectos y resultados de 1992 a 2007. **Actas de XV Congreso de la Asociación de Lingüística y Filología de América Latina**. Montevideo, p. 21-24. 2008.

KOSTINA, I. **La variación conceptual de los términos en el discurso especializado** (Tesis doctoral). Barcelona: IULA, 2009.

LÁZARO, J.; NÚÑEZ RUBINO, M. Disponibilidad léxica en informantes culiacanenses: homosexualidad como centro de interés. **Lingüística Mexicana**, Nueva Época, v. II, n. 2. p. 17-40, 2020.

LÁZARO, J. **El ejemplo en terminología**: caracterización y extracción automática (Tesis de doctorado). Barcelona: IULA, 2015.

LÁZARO, J. **Relevancia cuantitativa y cualitativa de ejemplos y contextos de uso de algunos términos referidos a sexualidad** (Trabajo de fin de máster). Barcelona: IULA, 2011.

LÁZARO, J. **Extracción de la terminología básica de las sexualidades en México a partir de un corpus lingüístico** (Tesis de licenciatura). Distrito Federal: UNAM, 2010.

LEE, S. M.; SCOTT, M.; SHIN, D. **A Practical Approach to Corpus-based Analysis**: Introducing WordSmith Tools, Range, and Other Packages. Seoul: Hankookmunhwas, 2018.

LÓPEZ MORALES, H. **El léxico disponible de Puerto Rico**. Madrid: Arco Libros, 1999.

MARTÍNEZ, M. Setenta años de teoría de los campos: balance provisional. **Revista Española de Lingüística**, v. 33, n. 2, p. 261-314. 2003a.

MARTÍNEZ, M. Definiciones del concepto Campo en semántica: antes y después de la lexemática de E. Coseriu. **Odisea: Revista de Estudios Ingleses**, n. 03 (número extraordinario), 2003b.

MEDINA, A.; SIERRA, G. Criteria for the Construction of a Corpus for a Mexican Spanish Dictionary of Sexuality. **Proceedings 11th Euralex International Congress**. Vol. 2. Francia: Université de Bretagne-Sud, 2004.

MORENO FERNÁNDEZ, F. Información básica sobre el Proyecto para el Estudio Sociolingüístico del Español de España y de América» - PRESEEA (1996 - 2010). **Revista Española de Lingüística**, n. 36, p. 385-391, 2006.

PADRÓ, L.; STANILOVSKY, E. FreeLing 3.0: Towards Wider Multilinguality. **International Conference on Language Resources and Evaluation**. Istanbul: LREC, 2012.

PAREDES GARCÍA, F. A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE. **Revista Nebrija de Lingüística Aplicada**, n.16, p. 54-59, 2014.

ROJAS PORRAS, M. Esbozo de lineamientos conceptuales para la enseñanza y aprendizaje del léxico. **Educación**, v. 30, n. 2. p. 9-29, 2006.

SARALEGUI, C.; TABERNERO, M. Aportación al proyecto panhispánico de léxico disponible: Navarra. In: OLZA MORENO, I.; CASADO VELARDE, M.; GONZÁLEZ RUIZ, R. **Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)**. España: Departamento de Lingüística hispánica y Lenguas modernas. Pamplona, Servicio de Publicaciones de la Universidad de Navarra. 2008.

SEGHEZZI, N. Variación terminológica: de la escritura a la oralidad. **Debate Terminológico**, n. 9. p. 62-80, 2013.

SIERRA, G. **Introducción a los corpus lingüísticos**. Ciudad de México: Universidad Nacional Autónoma de México, 2017.

SIERRA, G.; MEDINA, A.; LÁZARO, J. Determinación de la terminología básica en sexualidad a partir de la Web como corpus. In: CANTOS GÓMEZ, P; SÁNCHEZ PÉREZ, A. (Org.). **Panorama de investigaciones basadas en corpus**. Murcia: AELINCO, 2009, p. 374-385.

SIERRA, G.; SOLÓRZANO, J.; CURIEL, A. GECO, un Gestor de Corpus colaborativo basado en web. **Linguamática**, v. 9, n. 2, p. 57-72, 2017.

TEMMERMAN, R. **Towards new ways of terminology description: the sociocognitive-approach**. Amsterdam: John Benjamins Publishing, 2000.

TERCEDOR SÁNCHEZ, M. Esquemas metafóricos en el español de la ciencia y la tecnología. In: FABER, P.; JIMÉNEZ HURTADO, C.; WOTJAK, G. (Org.). **Léxico especializado y comunicación interlingüística**. Granada Lingüística, p. 233-242, 2004.

ULLMANN, S. Semántica: Introducción a la ciencia del significado. Madrid: Aguilar. 1972.

WOTJAK, G. (Org.). **Pautas y pistas en el análisis del léxico hispano(americano)**. Madrid: Iberoamericana/Vervuert, 2003.

ZGUSTA, L. **Manual of Lexicography**. Paris: Walter de Gruyter, 1971.