

**El habla coloquial juvenil en el estado de Baja California.
Un análisis lexicológico y lexicográfico**
*Colloquial Speech of Young People in the State of Baja California.
A Lexicological and Lexicographic Analysis*

Jahiro Samar Andrade Preciado
Universidad Autónoma de Baja California, México
jahiro.andrade@uabc.edu.mx

Resumen

El habla coloquial es un fenómeno presente en la comunicación cotidiana, sobre todo en el contexto juvenil. El presente texto muestra elementos léxicos y fraseológicos propios del habla coloquial juvenil. Estos elementos han sido analizados con la finalidad de elaborar un diccionario de coloquialismos a partir de dos corpus lingüísticos. El proceso de análisis llevó, por un lado, al desarrollo de una ficha lexicográfica, integrando usos y frecuencias de cada término. Por otro lado, permitió la aplicación de un análisis lexicográfico similar al empleado para la elaboración del *Diccionario del Español de México* (2010). Como resultado de este conjunto de procesos analíticos, se elabora el *Diccionario de Coloquialismos de Baja California* con 441 términos coloquiales empleados por los jóvenes bajacalifornianos, con un rango de edad de entre 14 a 30 años. Algunos ejemplos de estos coloquialismos son, entre otros, *carrillero*, *perreado*, *puqui*, *chapulinear* y *masacuata*. Cabe mencionar que este *Diccionario de Coloquialismos de Baja California* tiene como propósito auxiliar las áreas de traducción, interpretación y estudios de lingüística aplicada.

Palabras clave: coloquialismos, lexicografía, lexicología, lingüística de corpus.

Abstract

Colloquial speech is a phenomenon that occurs in daily communication, especially among young people. This text illustrates lexical and phraseological elements of colloquial speech of young people. These elements were obtained from two linguistic corpora, as well as analyzed with the purpose of elaborating a dictionary of colloquialisms. The analysis process led to a development of a lexicographical record

integrating the uses and frequencies of each identified term. In addition, this analytical process permitted the implementation of a lexicographical method similar to the one implemented for the elaboration of the Dictionary of Mexican Spanish (Diccionario del Español de México, 2010). As result of these analytical processes, a Dictionary of Colloquialisms of Baja California was elaborated. It integrates 441 colloquial terms used by young people in a range of 14 to 30 years old. Some examples of this colloquial terminology are carrillero, perreado, puqui, chapulinear, and masacuata. It is worth mentioning that this dictionary aims to support translation, interpretation, and studies in applied linguistics.

Keywords: *colloquialism, corpus linguistics, lexicography, lexicology.*

Introducción

El objetivo general de esta propuesta es identificar elementos léxicos o fraseológicos propios del habla coloquial (HC) que se encuentran inmersos en la interacción social del habla, de manera particular, en la naturalidad y cotidianidad actual del discurso oral y escrito de los jóvenes en Baja California (B.C.). Esto se lleva a cabo con la finalidad de elaborar una herramienta lexicográfica, misma que consiste en un diccionario de coloquialismos empleados por los habitantes jóvenes de B.C. La identificación de dichos elementos se sitúa a través del Corpus del Habla de Baja California (CHBC)¹ y la elaboración de un corpus de texto tecleado que presenta rasgos propios de empleo del HC. Esta base de datos se denominó Corpus del Habla Coloquial de Baja California (CHCBJ). De manera particular, es necesario señalar que estos coloquialismos pueden referir a palabras o frases provenientes de otros estados.

La organización de este trabajo investigativo alude en la primera sección a un breve reporte del estudio del español en B.C. y a la noción de HC sobre la base de las investigaciones teóricas de quienes han desarrollado la temática, seguido de las postulaciones de las disciplinas lingüísticas que coadyuvan la compilación y análisis de los elementos identificados: la lexicología, lexicografía y lingüística de corpus. Posteriormente, se presenta la metodología utilizada para la compilación del corpus, su análisis y la elaboración de la matriz de datos donde fueron

¹ El CHBC es un proyecto llevado a cabo por el Cuerpo Académico: Lengua, Tecnología e Innovación de la Facultad de Idiomas de la Universidad Autónoma de Baja California, a nivel de estado. Este corpus es una selección de obras literarias regionales y entrevistas a habitantes del estado, clasificado en edades, nivel de estudio y géneros de tipo literario, periodístico, científico y técnico.

plasmados los hallazgos de este estudio, mismos que se presentan en la sección de resultados y, por último, se señala una breve conclusión sobre los mismos.

Rico (2014) postula que, en la actualidad, existen pocos trabajos lexicográficos enfocados en la confección de definiciones propias del léxico sub-estándar (coloquialismos). Por esta razón, este estudio representa una aportación significativa a la lexicología y lexicografía de dicha zona dialectal. La confección de un diccionario sobre coloquialismos a través de dos corpus lingüísticos, los cuales están concentrados exclusivamente en el habla de B.C., es un proyecto que no se ha desarrollado con anterioridad en dicho estado. Por lo anterior, desarrollar un proyecto de esta magnitud aporta al conocimiento de las variedades coloquiales presentes en el discurso cotidiano de los jóvenes bajacalifornianos.

A través de dicho análisis, se busca identificar el significado, empleo y frecuencia de uso de los coloquialismos, con el fin de resolver problemáticas que puedan presentarse por el desconocimiento de esta terminología, las cuáles podrían suscitarse en disciplinas como la traducción audiovisual, la interpretación en cualquier modalidad o la enseñanza de español.

El español en Baja California

La presencia del HC en cada dialecto es indudable, por lo que B.C. no está exento de la existencia de coloquialismos en la interacción social de sus miembros jóvenes. Sin embargo, la entidad federativa carece de estudios científicos que traten la temática en específico. Algunos de los estudios más próximos en temática a esta propuesta, pero centrados en cuestiones del argot y el código delictivo, son los de Saldívar (2014) y Saldívar y Rodríguez (2015 y 2018).

Saldívar (2014), a partir de una recopilación de textos relacionados con el medio y el fenómeno del narco, entre ellos canciones (narcocorridos), notas periodísticas, entrevistas con gente del medio y encuestas, logró crear un corpus y hacer un análisis lexicológico de las palabras usadas en este contexto, y explica la razón de su popularización, quiénes usan estas palabras, el modo y sus diferentes usos.

Saldívar y Rodríguez (2015 y 2018) señalan que el empleo de estas formas de habla es adoptado en su mayoría por hombres (en menor medida por las mujeres) entre los 25 y 45 años de edad y, a su vez, la terminología más marginada es utilizada por los traficantes, consumidores y los integrantes de las clases bajas. Asimismo, se encontró que la cercanía del fenómeno narco en la sociedad actual está relacionada con el conocimiento del narcolenguaje, lo cual puede explicar que, aunque los individuos no escuchen narcocorridos o sean usuarios de drogas, conozcan palabras propias del fenómeno.

La posición geográfica de BC, al colindar al norte con el estado de California, en Estados Unidos, y debido al flujo migratorio que existe entre ambas fronteras, ha influenciado el empleo de distintos calcos o préstamos lingüísticos del idioma inglés en el discurso coloquial de los habitantes de la región (Rábago y Saldívar, 2008). A su vez, el flujo migratorio de connacionales de B.C. tiene un fuerte impacto en el léxico disponible de los hablantes, debido a que se han arraigado vocablos originarios de otros estados dentro del discurso cotidiano de los bajacalifornianos (Andrade y Cortez, 2020).

El habla coloquial

El HC es un fenómeno presente en cualquier dialecto y en todas las lenguas (Briz, 2019). Este tipo de vocabulario es considerado ajeno al registro estándar del español, por lo que este estudio se enfoca en el nivel sub-estándar del léxico, es decir, en el registro bajo y medio del habla natural y cotidiana. A diferencia de las manifestaciones lingüísticas conscientemente formuladas, se trata de un habla natural y espontánea, la cual forma parte de la conversación diaria (Beinhauer, 1996). A su vez, Blake (2010) indica que esta modalidad del habla normalmente está compuesta de vulgaridades, insultos, obscenidades y metáforas, de ahí que, en el HC, se empleen palabras para nombrar elementos, o hacer representaciones abstractas de situaciones en las que puede relacionarse algún objeto o animal con alguna acción.

Asimismo, el HC es un fenómeno muy complejo: se enraíza en la sociedad en distintos ámbitos y de diversos modos, añadiendo a su vez la complejidad que representa el estudio del lenguaje humano. Patridge (2015) afirma que el empleo de coloquialismos varía según el país y el grupo social. Asimismo, asevera que estos se emiten durante el discurso oral. Sin embargo, el auge de las redes sociales y la comunicación en línea han dado espacio para diversos elementos en la conversación teclada que figura en las plataformas sociales.

El HC está marcada por una región, un grupo social y otras características sociales que hacen que un grupo de individuos hablen de forma peculiar. Terrádez (2001) señala que, quienes usan un lenguaje extremadamente informal, en general, son aquellos individuos que se encuentran en los estratos sociales bajo y medio, integrando expresiones ligadas a comparaciones con objetos o situaciones cotidianas, vivencias y creencias mayormente metafóricas. A su vez, el autor sostiene que este tipo de habla carece de reglas aparentes, pero que generalmente sigue un patrón impuesto con el tiempo y con el pasar de persona a persona.

Lexicología, lexicografía y lingüística de corpus

La lexicología es la ciencia que tiene como propósito estudiar las palabras que han ido surgiendo o añadiéndose al habla de distintas regiones, ciudades o países. Asimismo, da a conocer de dónde provienen, cómo pueden usarse, en qué contexto se aplican, y la evolución de las mismas a lo largo de los años, en una visión general y completa (Ruano, 2011). Grzega y Schöner (2007) señalan ciertos criterios claves durante un análisis léxico. De manera inicial, las unidades léxicas comprenden diferentes formas flexivas, lo que permite adoptar una variación morfológica para expresar género, número, persona, modo y tiempo gramatical. De ahí que un análisis lexicológico debe estar enfocado exclusivamente a los lexemas que poseen unidades invariables con significados que presentan cierta autonomía y se combinan entre sí, o en su caso, con morfemas para formar unidades léxicas.

Además de una exclusividad léxica, otro punto que debe destacarse del trabajo de Grzega y Schöner (2007) es el análisis a nivel léxico-semántico, el cual está enfocado tanto en la interacción del lexema como en su significado, incluyendo las relaciones de sentido que las palabras establecen entre sí. Debido a la complejidad del lenguaje humano, es difícil lograr que un programa analizador de léxico, como AntConc (Anthony, 2017) identifique cuestiones semánticas, por lo que es importante analizar minuciosamente dichas cuestiones de manera manual una vez identificados los términos.

Bolinger (2008) señala que la lexicografía es una ocupación poco natural, ya que en esta se trata de arrancar las palabras de su contexto original y alinearlas en un orden no determinado por la naturaleza. Es decir, la definición de estas puede variar según el contexto del hablante que haga empleo de ella. Por su parte, Lara (2016) define la lexicografía como la “técnica de elaborar o confeccionar diccionarios” (p. 23).

Durante la confección de un diccionario, la lexicología, la semántica y la lexicografía trabajan en conjunto para la creación de definiciones precisas y coherentes (Lara, 2016). Lara (2006) asevera que la lexicología estudia las palabras desde su origen hasta su modo de empleo, en cambio, menciona que la lexicografía ayuda a crear una definición de diccionario para facilitar la comprensión del término. Por otro lado, otro aspecto que se toma en cuenta para la elaboración de un diccionario es saber cuáles son las fuentes de las que se debe partir para confeccionar un diccionario. Atkins y Rundell (2008) y Durkin (2015) sugieren el uso de corpus lingüísticos como una herramienta eficaz para identificar, obtener y crear nuevas definiciones con precisión lin-

güística para que sean incorporadas a nuevos diccionarios o añadidos como acepciones.

Los lexicógrafos recomiendan el uso de los corpus lingüísticos para confeccionar un diccionario. Estos corpus deben estar compuestos por novelas, periódicos y diversos textos (según la temática del diccionario), y deben apoyarse de otros diccionarios con la finalidad de obtener ejemplos variados y contextos diferentes. Por esta razón, la propuesta de este proyecto indaga en la formulación de entradas léxicas por medio de la identificación y extracción de elementos léxicos y fraseológicos en dos corpus. Para alcanzar dicho objetivo, es pertinente seguir los lineamientos lexicográficos de obras especializadas en la disciplina como las de Atkins y Rundell (2008) y Durkin (2015).

Una vez obtenidas las fuentes y composición del corpus, Durkin (2015) sugiere analizar los aspectos semánticos de la unidad léxica o fraseológica, según sea el caso, para detectar la categoría gramatical y redactar la definición. Posterior a ello, se procede a analizar las características pragmáticas que presenta la palabra para obtener acepciones de la unidad léxica y una definición completa que cubra todos los campos de empleo y uso.

De acuerdo con Berrissoul (2012), para la elaboración de un diccionario, el lexicógrafo debe tomar algunos aspectos preliminares antes de comenzar con el desarrollo de definiciones. Entre ellos destacan las necesidades que presentarán los usuarios de este y las funciones que el diccionario deberá desempeñar. Asimismo, dicho autor recomienda atender a tres cuestiones sustanciales: el tipo de diccionario, el tipo de usuario y el objetivo que tenemos para desarrollar el diccionario.

En el caso de este estudio, el tipo de diccionario propuesto es monolingüe y especializado, debido a que presenta las características de uno, al compararlo con la clasificación de Anglada (2005) pues se aportan definiciones para un campo específico. Con respecto al usuario, está enfocado en los especialistas dentro de la lingüística aplicada y su fin es establecer definiciones de palabras empleadas por individuos con registro bajo de habla en el estado de B.C.

La lingüística de corpus es una herramienta que permite distinguir aspectos semánticos y pragmáticos de datos lingüísticos de una manera óptima y sencilla, así como conocer el número de frecuencia que tienen las palabras, el contexto y el registro empleado por los hablantes. Estas bases de datos, como lo señala Ruano (2011), están fundamentadas en textos reales, las cuales permiten analizar de qué manera va evolucionando la lengua, cómo es que las nuevas generaciones emplean el vocabulario y el léxico, incluso qué alteraciones o nuevos significados

adquieren las palabras. En otras palabras, es una manera práctica de estudiar fenómenos desde la óptica de lexicología y lexicografía, por esta razón estas tres disciplinas son pieza clave para la creación de diccionarios monolingües.

Metodología

De acuerdo con las características particulares mencionadas por Hernández, Fernández y Baptista (2014), para la realización de esta propuesta se desarrolló un estudio cuantitativo de alcance descriptivo. La terminología identificada fue extraída del corpus *ad-hoc* del habla estatal de B.C. y de un corpus diseñado exclusivamente para esta investigación, para el cual fueron seleccionados textos tecleados propios del HC. Estos textos se componen por diversos formatos de texto escrito en línea, como transcripciones de video blogs, música popular en la región, conversaciones en redes sociales, además de páginas de internet, así como interacciones y retroalimentación tecleada de los internautas que hacen uso de estas. Posterior a esta identificación de lemas,² se creó un instrumento que permitió el análisis léxico y posteriormente se desarrolló una confección lexicográfica para la elaboración del *Diccionario de Coloquialismos de Baja California*.

La muestra utilizada para este estudio la integra un conjunto de conversaciones realizadas por 420 jóvenes de ambos géneros, en un rango de edad de 14 a 30 años. Se seleccionó dicho rango de acuerdo con el estudio desarrollado por Islas (2015), mismo que analiza a los jóvenes de 14 a 30 años con respecto al uso de las redes sociales, por lo que, para efectos de representatividad, se decidió sumar 6 años más a la muestra, con la finalidad de ampliar la obtención de elementos léxicos. Cabe señalar que el ámbito temporal en el que se llevó a cabo la recolección de datos y el análisis de esta propuesta tuvo inicio en verano de 2016 y culminó en invierno de 2019.

Este estudio considera un acercamiento multidisciplinario desde distintos campos de conocimiento, tales como la lexicología, la lexicografía y la lingüística de corpus. De la lexicología se formula el análisis de los términos identificados a través de los corpus lingüísticos y la interacción social. De la lexicografía se toma el método y los fundamentos semánticos para la confección de nuevas definiciones. En cuanto a la lingüística de corpus, de ella se obtiene la metodología para la recopilación de textos y la identificación de terminología desconocida.

A continuación, se describe la metodología empleada para este estudio, a partir de la explicación acerca de la composición del corpus del HC, el proceso

² Según Hualde (2010), lema es la unidad autónoma constituyente de léxico en un idioma.

de obtención de elementos léxicos para su análisis, la aplicación del instrumento diseñado y la confección del diccionario. La figura 1 resume gráficamente esta metodología:

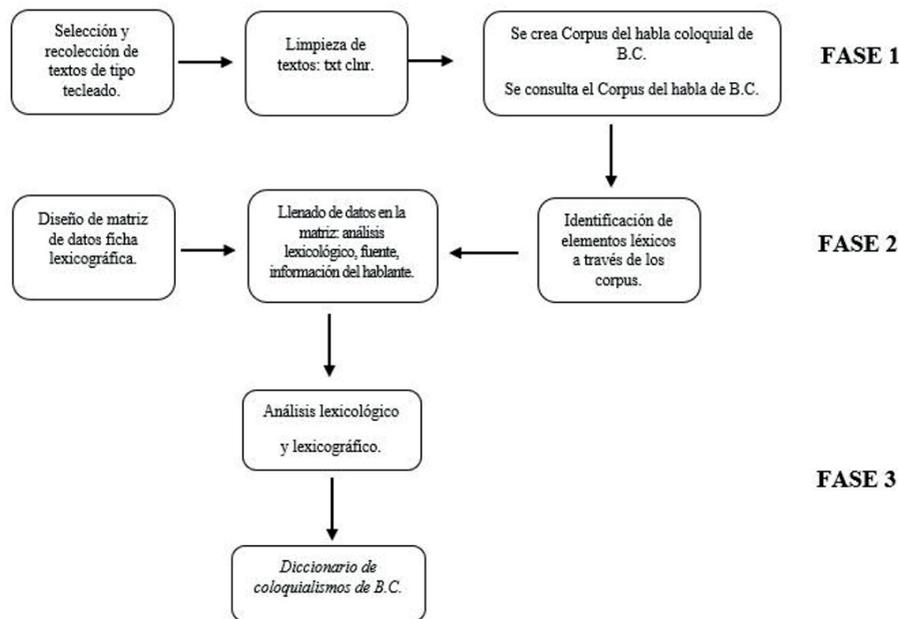


Figura 1. Organizador avanzado de metodología (elaboración propia).

De acuerdo con la figura 1, la metodología se encuentra dividida en tres fases. En la primera fase se seleccionaron aquellos textos de tipo teclado que conforman el CHCBJ (redes sociales, blogs periódicos, blogs y páginas populares, mensajes de texto, mensajes de chat y canción popular). Una vez obtenidos estos textos se introdujeron al formato *txt* y fueron limpiados y corregidos en la aplicación *txt clnr*. De este modo, se logró compilar el CHCBJ y se solicitó el CHBC.

La fase dos consistió en el trabajo de identificación de lemas y el diseño de la matriz de datos en el que se introdujeron las mismas con el fin de ser estudiadas. Una vez registrados los términos identificados, se realizó un análisis léxico y posteriormente un análisis lexicográfico, donde se generó una definición (con acepción según la palabra) a cada palabra y locución.

La matriz de datos, cuya función es realizar el análisis lexicológico y lexicográfico, se encuentra dividida en cinco secciones que combinan ambos tipos de análisis, los cuales conducen al *Diccionario de Coloquialismos de B.C.* En el cuadro 1 se aprecia un breve ejemplo del llenado de la misma y en breve la explicación de cada campo.

CUADRO I. EJEMPLIFICACIÓN DE MATRIZ DE DATOS (ELABORACIÓN PROPIA).

LEMA	CATEGORÍA GRAMATICAL	DEFINICIÓN	CAMPO SEMÁNTICO	TIPO DE PALABRA
Alucín	Nombre	Usuario de narcóticos	Atributos	Primitiva
Bronca	Adjetivo	Persona de comportamiento conflictivo	Atributos	Primitiva

Legenda:

- Lema: palabra o frase identificada.
- Categoría gramatical: en el caso de las frases, estas fueron clasificadas como locuciones.
- Definición: de acuerdo con el análisis de frecuencias en corpus y la lectura de ejemplos del lema y la asignación de categoría gramatical se asignó una definición a las palabras (y acepciones, según el caso).
- Campo semántico: se categorizaron las palabras reconocidas que comparten rasgos y elementos significantes o comparten cierto ambiente, estos campos son *partes del cuerpo, actividades, atributos, objetos y otros*.
- Tipo de palabra: de acuerdo con la clasificación de tipo de palabras de Lara (2006) como: primitiva, compuesta o expresión formulaica e idiomática.

Identificación de elementos coloquiales

La identificación de elementos léxicos y fraseológicos se realizó a través de ciertos procesos. De manera inicial, la recolección de un corpus es un trabajo que conlleva dedicación y una indagación minuciosa según el fenómeno lingüístico que se busque analizar (Zanettin, 2013), por lo que, después de limpiar los textos que componen el CHCBJ y verificar que estos no presentaran codificaciones y errores tipográficos, se procedió a la identificación de elementos coloquiales.

Con el apoyo del software *AntConc 3.4.4w* se analizaron todos los documentos que conforman los corpus de este estudio. Al utilizar la función de *Wordlist* se identificaron todos los elementos coloquiales de cualquier género disponible en los corpus. Una vez identificados dichos elementos, se empleó una función de *stoplist*. Esta función consiste en el empleo de un filtro que elimina los listados de cualquier palabra que el operador del software desee. En este caso, se utilizó el *Corpus de Referencia del Español Actual* (CREA) lematizado como lista de filtro y se aplicó a ambos corpus.

El resultado fue una lista de palabras que no aparecen en el español actual, las cuales fueron clasificadas como coloquialismos, pues estas no fueron reconocidas por la Academia Española en el discurso estandarizado. La mayoría de estas palabras identificadas provienen del CHCBJ, lo que da como resultado un

corpus viable para los objetivos de esta investigación. Después, se realizó una evaluación minuciosa de estas listas de palabras, de modo que se eliminó de manera manual aquellos lemas que no presentaron rasgos propios del HC y que el *stoplist* pasó por alto.

La siguiente etapa en este proceso de identificación fue la revisión del listado que se obtuvo de la restricción del *stoplist* y la eliminación manual. Los elementos identificados en este listado presentan rasgos coloquiales, por lo que quedó concluida la identificación de elementos léxicos. En cuanto a la identificación de elementos fraseológicos, se recolectó información de informantes de internet por medio de sondeos en redes sociales y foros populares. Una vez recabados los elementos fraseológicos, se procedió a buscarlos dentro de la lista de concordancias con el fin de comprobar su uso y empleo dentro del léxico disponible de los bajacalifornianos.

Una vez concluido el análisis, se procedió a la elaboración de definiciones de los coloquialismos identificados con el método lexicográfico de Lara (2016), utilizado en la actualidad en el *Diccionario del Español de México* (DEM),³ el cual consiste en la creación de nuevas definiciones a través del análisis léxico-semántico dentro de la compilación de nuevos archivos de texto para la obtención de nuevos mexicanismos.

Resultados

Tal como fue explicado en la metodología desarrollada para este estudio, los términos identificados se concentraron en una matriz de datos con el fin de complementar el perfil lexicológico de los mismos. Además, esta matriz funge como ficha lexicográfica para facilitar la confección del *Diccionario de Coloquialismos de Baja California*, en el que se abordan los diferentes rasgos que fueron encontrados en el análisis lexicológico, así como algunos aspectos relevantes de los coloquialismos estudiados.

Lemas y locuciones

El número total de términos identificados y catalogados como coloquialismos en el CHBC y el CHCBC está conformado por 441 coloquialismos. A estos términos se les consignó como lemas ya que se relacionan a una sola palabra las variaciones de la misma (Sebastián-Gallés, 2000), particularmente tratándose de

³ El DEM es el resultado de varias investigaciones del vocabulario utilizado en México desde 1921 a la fecha.

conjugaciones verbales. Cabe mencionar que se tomó dicha decisión con el propósito de reducir a categorías gramaticales los coloquialismos que conforman los hallazgos de esta investigación. En cuanto a las expresiones o frases, estas fueron categorizadas en un solo grupo como locuciones.

Categoría gramatical, locuciones y préstamos

La identificación y la asignación de categorías gramaticales de los coloquialismos es parte fundamental, pues la confección lexicográfica exige la identificación de una categoría gramatical como punto de partida de un nuevo concepto del significado (Bolinger, 2008). Asimismo, un léxico no podría estar completo sin la información gramatical. A cada coloquialismo le fue asignado una categoría, no obstante, es substancial señalar que fueron encontradas palabras que funcionan como sustantivo y adjetivo, esto según el uso de empleo.

En la tabla 1 se muestran los coloquialismos identificados divididos en categorías gramaticales. La mayoría de los términos identificados se centraron en verbos, adjetivos y nombres respectivamente. Sin embargo, dentro de estos términos se encontraron adverbios.

TABLA 1. FRECUENCIAS POR CATEGORÍA GRAMATICAL

CATEGORÍA	FRECUENCIAS	EJEMPLO
1) Verbo	100	<i>chacalear, enganchar, paniquear, etc.</i>
2) Adjetivo	150	<i>grifo, malaleche, rajón, etc.</i>
3) Nombre	121	<i>cuacha, lima, meco, etc.</i>
4) Adverbio	13	<i>ferro, simón, sincho, etc.</i>

En cuanto a los coloquialismos identificados como adverbios, se encontró que estos son de tipo afirmativo, es decir, palabras como *ferro, simón, sincho* son innovaciones coloquiales para hacer afirmaciones a cuestionamientos sustituyendo al adverbio *sí*. De igual forma palabras como *nel, nones* son adverbios de negación los cuales sustituyen al adverbio *no* como una forma graciosa de hacer una negación ante un cuestionamiento o la indisposición de realizar alguna actividad.

La identificación de locuciones en el HC de los jóvenes bajacalifornianos se encuentra dividido entre locuciones de tipo adjetival, nominal y verbal, pues durante el análisis en *Antconc* es evidente que los núcleos de estas construcciones oracionales son adjetivos, verbos y nombres en menor medida. En la tabla 2 se aprecian algunos ejemplos y el número de frecuencias de cada tipo de locución de un total de 47 identificadas en ambos corpus.

TABLA 2. LOCUCIONES

Locuciones	Frecuencias	Ejemplo
1) Locución nominal	6	<i>ni al caso, qué pedo, cuál es tu pedo</i> , etc.
2) Locución adjetival	11	<i>bien lamido, bien descremado</i> , etc.
3) Locución verbal	30	<i>hacer el paro, le sacas, echar hueva</i> , etc.

Por otra parte, los préstamos no representan un número significativo de frecuencias como se estimaba dentro de los corpus lingüísticos, pues se consideraba que la frecuencia de préstamos en el discurso coloquial sería mayor, en especial en una zona fronteriza como lo es B.C. En total se encontraron 10 préstamos provenientes del idioma inglés como *chechar* (*to check*), *raite* (*to ride*), *guachar* (*to watch*), *baica* (*bike*), *chilo* (*to chill*) entre otros.

Otro dato substancial de este reporte es la función de algunos adjetivos que pueden funcionar como dos categorías gramaticales. Este es el caso de adjetivos con función de nombre, términos denominados adjetivos sustantivados. En la tabla 3 se presenta el número total de adjetivos sustantivados coloquiales identificados.

TABLA 3. FRECUENCIAS DE ADJETIVOS SUSTANTIVADOS

	FRECUENCIAS	EJEMPLOS
Adjetivos sustantivados	39	<i>periquero, morro, mandilón</i> , etc.

Martínez *et al.* (2011) postulan que el verbo es la categoría léxica que expresa acción, movimiento, existencia, consecución, condición o estado del nombre. Tal como fue mostrado en la tabla 1, se encontraron un total de 88 verbos, los cuales funcionan como transitivos, intransitivos, ditransitivos y preposicionales. En la tabla 4 se muestran las frecuencias de estos cuatro tipos de verbos.

TABLA 4. FRECUENCIAS DE TIPOS DE VERBOS

TIPOS DE VERBOS	FRECUENCIAS	EJEMPLOS
Transitivos	41	<i>chorrear, desafanar, enciclar</i> , etc.
Intransitivos	37	<i>rifar, rolar, tirar</i> , etc.
Ditransitivos	8	<i>zumbar, quemar, manchosear</i> , etc.
Preposicionales	14	<i>ábranse, atorarle, bofearse</i> , etc.

Coloquialismos frecuentes

Dentro de los distintos contextos y textos estudiados, se encontraron distintos coloquialismos empleados en la cotidianidad del discurso. De acuerdo con el análisis desarrollado a partir de la tabla 1, los nombres y adjetivos fueron las categorías gramaticales de mayor frecuencia, por lo que surge la siguiente cuestión: ¿cuáles son las palabras de mayor frecuencia en los corpus? En la tabla número 5 se presentan las 10 palabras más frecuentes, sus números de frecuencia y ejemplos de uso.

TABLA 5. PALABRAS MÁS FRECUENTES

#	PALABRA	FRECUENCIAS	EJEMPLO:
1	Güey	35	Estás bien <i>güey</i> .
2	Equis	34	Está muy <i>equis</i> tu ropa, ve y cámbiatela.
3	Pinche	32	<i>Pinche</i> carro feo que traes.
4	<i>Morrola</i>	29	Estoy muy <i>morro</i> para pensar en eso.
5	Culero	28	No seas <i>culero</i> y hazlo tú.
6	Machín	28	Me duele bien <i>machín</i> las piernas.
7	Cotorrear	25	¿Puedes salir a <i>cotorrear</i> ?
8	Curada	24	Está <i>curada</i> la película nueva de...
9	Joto	24	Es medio <i>joto</i> , llegó y me abrazó.
10	Chamaco	23	¿Cuántos <i>chamacos</i> tiene?

Campo semántico

Cada lengua tiene su propia manera de percibir la realidad al considerar qué objetos comparten rasgos semánticos para ser considerados en diferentes ámbitos (Faber y Mairal, 1999). La estructuración de los campos semánticos y los rasgos más relevantes dependen, en gran medida, de factores extralingüísticos y tiene que ver con aspectos culturales y sociales. Un campo semántico está constituido por un grupo de palabras relacionadas por su significado y que comparten ciertas características en común. Estos campos pueden representarse como un conjunto de formas léxicas (Martínez, 2003).

Para la clasificación de campos semánticos de los coloquialismos identificados dentro de la matriz, se establecieron cinco campos, mismos que fueron clasificados según las propiedades semánticas compartidas. Estas categorías están representadas como *partes del cuerpo*, *actividades*, *objetos*, *atributos* y *otros*.

TABLA 6. FRECUENCIAS POR CAMPO SEMÁNTICO

CAMPO SEMÁNTICO	FRECUENCIAS	EJEMPLO
Partes del cuerpo	14	<i>nepe, pancha, riata</i> , etc.
Actividades	125	<i>tabanear, displayar, chacalea</i> .
Objetos	52	<i>güiri, lima, pinga</i> , etc.
Atributos	218	<i>ruca, pelada, mariposón</i> , etc.
Otros	32	<i>al cien, sincho, simón</i> , etc.

La tabla 6 permite visualizar que el campo semántico con mayor frecuencia es el de los *atributos* seguido del de *actividades*. Dentro del corpus, se observó que el léxico juvenil bajacaliforniano resulta coloquial gracias a que estas palabras se encuentran estrechamente ligadas a la similitud que los objetos o actividades con las que son sustituidos poseen con relación de sus características o propiedades físicas. Así, pueden ser interpretadas en el acto comunicativo por mecanismos metafóricos, o en su caso, metonímicos cuando los hablantes comparten las

representaciones mentales necesarias. En relación con los campos de *partes del cuerpo y objetos*, estos representan una menor frecuencia, sin embargo, requieren del mismo conocimiento enciclopédico para su empleo (emisor) y decodificación (receptor) durante el discurso.

Tipo de palabra

La última categoría que comprende la matriz de datos refiere el tipo de palabras que se divide en primitivas, compuestas o derivadas en el sentido lexicológico de Lara (2006). En cuanto a las locuciones, estas fueron divididas en expresiones formulaicas o expresiones idiomáticas. Este tipo de análisis proporciona un perfil más completo acerca del análisis lexicológico. El número de frecuencias por tipo de palabra identificado en la muestra se aprecian en la tabla 7.

TABLA 7. FRECUENCIAS DE TIPO DE PALABRAS

TIPO DE PALABRA	FRECUENCIAS	EJEMPLOS
Primitivas	204	<i>filo, calote, chacal, chairo, etc.</i>
Compuestas	16	<i>malacopa, malaleche, chupafaros, etc.</i>
Derivadas	171	<i>bateado, cuernudo, forjado, etc.</i>
Expresiones formulaicas	21	<i>en caliente, en chinga, en corto, etc.</i>
Expresiones idiomáticas	29	<i>a la bestia, agarrar el rollo, etc.</i>

En el estudio realizado por Saldívar (2014), tanto en el narcolenguaje como en el habla coloquial, el tipo de palabra que predomina es de origen primitivo y por consecuente derivado. El fenómeno de derivación permite la designación de conceptos relacionados semánticamente en cierto sentido considerados como primitivos (Lara, 2006). Por su parte, Martínez, Muñoz, Mora y Ángel (2011) sustentan que la derivación en varias lenguas es la principal fuente de nuevas palabras además de la semejanza de rasgos físicos o característicos, generando la formación de nuevas palabras con nuevos significados en el discurso coloquial.

Para el tratamiento de locuciones y su clasificación, se utilizaron los parámetros de expresiones formulaicas y expresiones idiomáticas, esto debido a que, durante el análisis en el nivel semántico, a las expresiones formulaicas se les atribuyó un solo uso o significado en el contexto, mientras que las expresiones idiomáticas poseían dos o tres significados y el empleo de las mismas varía en la situación y contexto.

Conclusiones

Esta investigación cumple con los objetivos establecidos, los cuales dan respuesta a la pregunta de investigación. En primer lugar, se compiló un corpus con distintos textos de tipo teclado, que permitieron la identificación de coloquialismos

empleados por los jóvenes de B.C. Además, se realizó un análisis lexicológico que permitió conocer los usos de estas palabras mediante ejemplos de uso real. Gracias a este análisis lexicológico se abordó el análisis lexicográfico que posibilitó la creación del producto final de esta propuesta.

El diccionario de coloquialismos del habla coloquial de B.C. es una herramienta lexicográfica que presenta distintas palabras y frases de carácter coloquial y con rasgos del nivel léxico sub-estándar en el dialecto de B.C. Este diccionario no refiere a un listado léxico exclusivo de palabras que tienen su origen en la entidad, sino que muestra 441 términos coloquiales que bien pueden tener su origen en distintos estados de México. Es importante mencionar que este diccionario, al no ser un diccionario etimológico, no se centra en los orígenes de los términos presentados, solo en su definición de acuerdo con las muestras de uso real en el habla coloquial.

Un diccionario de coloquialismos permite conocer un conjunto de fenómenos lingüísticos presentes en el discurso juvenil. Si bien existen trabajos lexicográficos que buscan difundir distintos coloquialismos (Fitch, 2006 y 2011), hasta este momento existen pocos trabajos de este tipo concentrados en una sola zona dialectal. El *Diccionario de Coloquialismos del Estado de Baja California*⁴ se encuentra disponible en línea, de igual forma existe un breve glosario español-inglés de coloquialismos utilizados en la ciudad de Mexicali, B.C.⁵

El CHCBJ, por su gran número de entradas identificadas, es considerado un corpus viable para dar seguimiento a este estudio u otros con temática similar. Con especial énfasis, se pretende trabajar en los aspectos tanto semánticos como estructurales de las locuciones encontradas, así como en la búsqueda de otras nuevas utilizando la metodología de recolección de datos aplicada en este estudio. De tal modo, los resultados obtenidos podrían dar inicio a nuevos proyectos de tipo lingüístico que brindarán aporte científico e innovador, ya sea desde un análisis de datos lingüísticos a través de la lingüística de corpus, como en el análisis de variables sociales en la sociolingüística variacionista o a nivel semántico-pragmático de los elementos léxicos identificados.

Las temáticas del HC son variadas y diversas por lo que la clasificación de las palabras encontradas en campos semánticos y dividir las expresiones en idiomáticas y formulaicas optimizó el proceso de designación de significado para la

⁴ Diccionario disponible en https://www.researchgate.net/publication/340363989_DICCIONARIO_DE_COLOQUIALISMOS_DEL_ESTADO_DE_BAJA_CALIFORNIA

⁵ Glosario disponible en https://docs.google.com/spreadsheets/u/1/d/1SI09pjQMBf8sk3_zV3MH0Y8iI_8MMY6jdJEPQNFvoc/edit?usp=drive_web&zouid=100641237077574516188

propuesta lexicográfica. Estas clasificaciones semánticas auxilian al acotamiento de acepciones dado el contexto y aplicación de los términos encontrados en los corpus.

Dividir y clasificar los verbos en sus distintos tipos ayudó a entender mejor su uso en el contexto, especialmente en relación con aquellos verbos preposicionales a los que la derivación morfológica les añade un nuevo significado de índole coloquial, ya que en su estado infinitivo no presentan rasgos coloquiales. Dicha asignación de significado y uso dependerá del contexto en el estado preposicional verbal, lo cual ocurre también en las locuciones verbales.

La clasificación de categorías gramaticales permitió conocer el funcionamiento de las palabras dentro del discurso juvenil. Como se apreció en la tabla 1, los nombres y adjetivos suelen ser más recurrentes en el acto comunicativo, debido a que estas categorías describen características y atributos para la construcción de un imaginario social compartido por los mismos usuarios de la zona dialectal. Por otro lado, el predominio de locuciones verbales es evidente en los resultados, esto supone que el uso de frases y expresiones coloquiales está constituido por acciones y la realización de actividades que sustituyen verbos estándares o utilizados en un discurso de registro medio-alto, lo que a su vez aporta un acervo léxico nuevo a la construcción del imaginario social juvenil en B.C.

Gracias a este trabajo es posible remarcar que las palabras y expresiones sirven como una construcción representativa del mundo y la percepción que el usuario tiene de este, por lo que el uso de metáforas y metonimias en el discurso cotidiano es un fenómeno con cierta innovación léxica arraigada a una brecha generacional y que permea en la sociedad joven, es decir, dichas innovaciones son transmitidas a las generaciones más jóvenes consolidando así un estatus de permanencia en el léxico disponible de los hablantes en los registros menos elevados del habla y la estigmatización de su uso en el registro culto. Estas realidades se encuentran plasmadas en la propuesta lexicográfica de la investigación. La utilización de herramientas de la lingüística de corpus hace posible la identificación de palabras desconocidas o poco utilizadas por generaciones de mayor edad o pertenecientes a una clase social alta y con un registro de habla elevado.

Como se aprecia al inicio, este trabajo es la primera parte de un proyecto mayor, pero ha funcionado para vislumbrar otros enfoques de análisis. La continuación del estudio del HC en la zona debe ser analizada desde la óptica de la semántica y sus postulaciones, como la teoría de prototipos que ha auxiliado la creación y formulación de acepciones en las entradas léxicas de algunos diccionarios de renombre. La aplicación de esta teoría en cada entrada podría determinar nuevas

acepciones en las entradas léxicas existentes, las cuales enriquecerían la propuesta lexicográfica del presente trabajo.

Referencias

- Andrade, J. y Cortez, J. (2017). “Nuevas culturas, bilingüismo y glocalidad en la construcción de la jerga urbana cachanilla”. *Congreso Internacional de Idiomas 2017*. UABC: Facultad de Idiomas, Tijuana B.C.
- Anglada, E. (2005). *Lexicografía española*. Barcelona: Publicacions i Edicions, Universitat de Barcelona.
- Anthony, L. (2017). Laurence Anthony’s AntConc. Recuperado de <https://www.laurenceanthony.net/software/antconc/>
- Atkins, S. y Rundell, M. (2008). *The oxford guide to practical lexicography*. New York: Oxford University Press.
- Beinhauer, W. (1996). *El Español Coloquial* (3ra. ed.). España: Editorial Gredos.
- Berrissoul, A. (2012). *La elaboración de diccionarios*. Madrid: Universidad de Madrid.
- Blake, B. (2010). *Secret Language: Codes, Tricks, Spies, Thieves, and Symbols*. Oxford: University Press.
- Bolinger, D. (2008). “Defining the indefinable”. En T. Fontenelle (Ed.), *Practical lexicography, a reader* (pp. 193-196). Gran Bretaña: Oxford University Press.
- Briz, A. (2019). *El español coloquial: situación y uso* (8va. ed.). Madrid: Arcos Libros.
- Durkin, P. (2015). *The Oxford Handbook of Lexicography*. Atlantic City: Oxford University Press.
- Faber, P. y Mairal, R. (1999). *Constructing Lexicon of English Verbs*. Berlín: Mouton.
- Fitch, R. (2006). *Jergas de habla hispana*. Madrid: BookSurge.
- Fitch, R. (2011). *Diccionario de Coloquialismos y Términos Dialectales del Español*. Ciudad de México: Arcos Libros.
- Grzega, J. y Schöner, M. (2007). “English and General Historical Lexicology; Materials for Onomasiology Seminars”. *Onomasiology Online Monographs* (Vol. 1). Alemania: Katholische Universität Eichstätt-Ingolstadt.
- Hernández, R., Fernández, C., y Baptista, P. (2014). *Metodología de la Investigación* (6ta. ed.). Ciudad de México: Mc Graw Hill Education.
- Hualde, J., Olarrea, A., Escobar, A., & Travis, C. (2010). *Introducción a la lingüística hispánica*. Cambridge: Cambridge University Press.
- Islas-Carmona, O. (2015), *Cifras sobre jóvenes y redes*, Entretextos, año 7 n. 19, pp. 01-16.

- Lara, L. (2006). *Curso de Lexicología*. Ciudad de México: El Colegio de México: Centro de Estudios Literarios y Lingüísticos.
- Lara, L. (2010). *Diccionario del Español de México*. Ciudad de México: El Colegio de México: Centro de Estudios Literarios y Lingüísticos.
- Lara, L. (2016). *Teoría semántica y método lexicográfico*. Ciudad de México: El Colegio de México: Centro de Estudios Literarios y Lingüísticos.
- Martínez, J., Muñoz, J., Mora, S. y Ángel, M. (2011). *Clases de Palabra (III)*. El verbo. Lengua Castellana y Literatura. Madrid: Akal Sociedad Anónima.
- Martínez, M. (2003). *Definiciones del concepto de campo en semántica: antes y después de la lexemática de E. Coseriu*. Almería: Universidad de Almería.
- Patridge, E. (2015). *Slang: To-Day and Yesterday*. New York City: Routledge Revivals.
- Rábago, A y Saldívar, R. (2008). “Los medios impresos en la transformación del lenguaje: anglicismos en diarios bajacalifornianos”. *Memorias del IV Foro Nacional de Estudios en Lenguas (FONAEEL 2008)*. Universidad de Quintana Roo.
- Rico, A. (2014). *De vulgaridades, insultos, malsonancias. El diccionario del subestándar mexicano*. Mexicali, B.C.: Universidad Autónoma de Baja California.
- Ruano, F. (2011). *La Lexicografía: el arte de hacer diccionarios*. Recuperado de <https://ruanofaxas.com/article/la-lexicografia-el-arte-de-hacer-19j6x763f3uf8-176>.
- Saldívar, R. y Rodríguez, I. (2018). “Análisis del léxico en diferentes registros textuales en la construcción del imaginario social del narcotráfico en México”. *Literatura y Lingüística*, (37), pp. 381-400.
- Saldívar, R. y Rodríguez I. (2015). “El narcolenguaje en el habla actual de Baja California, México”. *Dialectología: revista electrónica*, 14, p. 97-114. Recuperado de <https://www.raco.cat/index.php/Dialectologia/article/view/293932>
- Saldívar, R. (2014). *Análisis lexicológico del narcolenguaje en Baja California*. Mexicali, Baja California: Universidad Autónoma de Baja California.
- Sebastián-Gallés, N. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Universitat de Barcelona.
- Terrádez, M. (2001). *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo*. Valencia: Universitat de València.
- Zanettin, F. (2013). “Corpus Methods for Descriptive Translation Studies”. *Procedia - Social and Behavioral Sciences*, 95, pp. 20-32.