

Léxico básico del español de Chile. Antecedentes, presentación y reflexiones

María Natalia Castillo Fadić¹

<https://doi.org/10.5565/rev/fraseolex.108>

Recibido: 17-7-2025 / Aceptado: 24-10-2025
Universitat Autònoma de Barcelona



Resumen

Este artículo presenta los antecedentes de la investigación que ha permitido determinar el núcleo estadístico del español de Chile y ha dado origen al diccionario *Léxico básico del español de Chile*, usando para ello herramientas de lingüística de corpus y estadística inferencial. Da cuenta de los aspectos centrales del estudio lexicoestadístico, sus límites, sus bases y las características principales del corpus estratificado de alrededor de 500.000 palabras sobre el que opera. Fundamenta la necesidad de un trabajo de esta naturaleza, especialmente desde una perspectiva sociolingüística y educativa. Sintetiza los resultados de la investigación y reflexiona sobre las múltiples aplicaciones del estudio en general y del diccionario estadístico obtenido como producto final en particular. Por último, plantea la necesidad de obtener los léxicos básicos de todos los lugares donde se habla español. Dada la envergadura del trabajo, el detalle de los métodos de análisis estadístico, de lingüística computacional y de corpus se recogen en Castillo Fadić 2020 y 2021b. Respecto de los resultados, se dividen en dos tipos: 1) los obtenidos directamente de la investigación, es decir, el léxico de mayor uso en el español de Chile, y 2) los ulteriores, centrados en la creación de un diccionario a partir de esos resultados (cf. Castillo Fadić, 2021a).

Palabras clave: lexicografía; corpus lingüístico; léxico básico; estadística léxica; sociolingüística.

Frequency Dictionary of Chilean Spanish. Background, Presentation and Reflections

Abstract

This article presents the background to the research that has allowed us to determine the statistical core of Chilean Spanish and that has led to the development of *Frequency Dictionary of the Chilean Spanish* (in Spanish, *Léxico Básico del Español de Chile*), using corpus linguistics and inferential statistics tools. The main aspects of the lexical-statistical study are described, along with its scope and limits, its foundations, and the main characteristics of the stratified corpus of approximately 500,000 words on which it is based.

¹ Pontificia Universidad Católica de Chile (Chile), mcastilf@uc.cl

The article justifies the need for research of this nature, particularly from sociolinguistic and educational perspectives. The main findings are summarized and the multiple applications of both the study as a whole and the resulting statistical dictionary are reflected upon. Finally, the importance of establishing core lexicons for all Spanish-speaking regions is argued. Given the magnitude of this research, detailed descriptions of the statistical, computational, and corpus linguistic methods employed can be found in Castillo Fadić, 2020 and 2021b. Regarding the results, these are divided into two categories: (1) those derived directly from the research, that is, the most frequently used lexicon of Chilean Spanish, and (2) subsequent results focused on the creation of a dictionary from these results (cf. Castillo Fadić, 2021a).

Keywords: lexicography; corpus linguistics; frequency dictionary; lexical statistics; sociolinguistics.

Sumario. 1. Introducción. 2. *Léxico básico del español de Chile*: antecedentes. 2.1. Estudios léxicos: una tarea de antigua data. 2.2. El caso de Chile. 2.2. Léxicos básicos. 3. *Léxico básico del español de Chile*: del estudio estadístico a la obra lexicográfica. 3.1. Lineamientos centrales. 3.2. Macroestudio interdisciplinario. 3.3. ¿Por qué crear un léxico básico? 3.4. Síntesis de los resultados. 4. Reflexiones finales. 5. Referencias bibliográficas.

1. Introducción

En este artículo, presentamos los antecedentes del *Léxico básico del español de Chile* (Castillo Fadić, 2021a), primer diccionario estadístico de su tipo en Chile y el Cono Sur, reseñado en Marín (2021) y en Gómez-Pablos (2022); para ello, pasamos revista a los principales estudios léxicos desde antes del surgimiento de la lingüística hasta nuestros días, con énfasis en los relativos al español de Chile. En este contexto, damos cuenta del tipo de investigación que implica la obtención de un léxico básico, de la importancia y proyecciones de llevar a cabo un estudio como este y de los objetivos específicos cumplidos de esta investigación en particular, para intentar motivar a los lingüistas de distintos lugares del mundo hispánico a generar investigaciones de este tipo en sus respectivas comunidades.

2. *Léxico básico del español de Chile: antecedentes*

2.1. Estudios léxicos: una tarea de antigua data

El léxico, justamente por ser la dimensión del lenguaje más cercana a la visión de mundo, ha interesado desde antiguo a diferentes estudiosos que, desde sus particulares perspectivas, orígenes y disciplinas, han visto en él un objeto de especial atractivo. En efecto, desde hace miles de años y en distintos continentes, el ser humano ha emprendido estudios léxicos, algunas veces con propósitos puros; otras con finalidades prácticas.

Ya Platón en “El Cratilo” aborda el tema léxico desde la filosofía, tocando la relación entre lo que hoy llamaríamos *significado*, *significante* y *referente*, cuando intenta distinguir entre la idea de un objeto y el objeto mismo. Aristóteles, en tanto, al proponer categorías gramaticales, opera de algún modo sobre el nivel léxico, sin llegar a centrarse específicamente en él. Posteriormente al 300 a. C., los estoicos intentan separar los estudios lingüísticos de la filosofía y, al concebir el lenguaje como un modo de entender la mente, desarrollan teorías semánticas. En paralelo a estos estudios teóricos, los gramáticos alejandrinos realizan otros empíricos, tendientes a expandir la lengua griega a los territorios que iban siendo anexados al Imperio por Alejandro Magno (cf. Robins, 1990).

La necesidad de comunicarse con hablantes de otras lenguas o de elaborar listados de productos para la venta o el trueque, en tanto, se manifiesta en atisbos precarios bilingües, trilingües o multilingües de lo que hoy llamamos *lexicografía*, surgidos en los pueblos sumerios y acadios al menos desde el año 2600 a. C. (Fernández-Sevilla, 1974, p. 13). El interés de los lingüistas por esta disciplina es relativamente reciente, pero esencial: la lexicografía, basada en el saber científico de la lingüística, se transforma en una ciencia aplicada del lenguaje, que procura generar diccionarios descriptivos y realistas, con definiciones sémicas; sus producciones lexicográficas se basan en la observación de la realidad, por lo que entregan juicios de existencia, no de valor, sobre los signos lingüísticos. La lexicografía, en este estadio, hace uso no solo de las herramientas de la lingüística contemporánea, sino que genera sus propias metodologías y teorías y recurre, además, a otras áreas del saber, especialmente a aquellas que favorecen el

procesamiento y análisis de corpus: la informática y la estadística (cf., por ejemplo, Lavid, 2005, pp. 135-142). Los grandes avances y el cúmulo de herramientas disponibles en la actualidad han ido erradicando paulatinamente del ámbito lexicográfico la intuición como medio de selección y procesamiento de unidades léxicas, criterio tan poco científico como habitual en los primeros estudios de la disciplina, sustituyéndola por el empleo de modernas metodologías lingüísticas y lexicográficas y por herramientas computacionales de análisis léxico-estadístico. Y pese a que los lexicógrafos con formación lingüística saben de sobra que en la elaboración de diccionarios “no basta con la intuición del lexicógrafo” (Lara, 1990, p. 25), aún muchos de “los diccionarios tradicionales no están suficientemente formalizados ni estructurados para que la información que poseen pueda ser tratada computacionalmente” (Lavid, 2005, p. 135), lo que los deja un paso atrás de las nuevas tendencias: esta realidad se sostiene en gran medida en las percepciones de algunos usuarios que se resisten a las obras descriptivas, por la asentada creencia de que los diccionarios tienen no solo la misión, sino también la potestad de “autorizar” usos lingüísticos, lo que los lleva muchas veces a valorar aquellas obras que reproducen los prejuicios lingüísticos más arraigados en la sociedad.

Pero no todo es lexicografía en el ámbito de los estudios léxicos de las últimas décadas. Desde mediados del siglo XX se han configurado equipos multidisciplinarios que han posibilitado el nacimiento de la léxico-estadística, disciplina ligada a la lingüística matemática y a la lingüística computacional y sustentada en los aportes de la lingüística de corpus. Esto se ha traducido en un fuerte impulso a los estudios del léxico atemático de mayor uso en las distintas comunidades, llamado *léxico básico* (cf., por ejemplo, Morales, 1986) y del léxico temático de mayor disponibilidad, llamado *léxico disponible* (cf., por ejemplo, López Morales, 1999); juntos, el léxico básico y el disponible permiten acceder al *léxico fundamental* de una comunidad. Estos estudios, que no encuentran su finalidad principal ni terminal en la elaboración de diccionarios definitorios, sino que, de hecho, se traducen en diccionarios estadísticos, no definitorios, son altamente relevantes desde una perspectiva sociolingüística, psicolingüística y pedagógica.

2.2. El caso de Chile

En Chile, como en el resto del mundo, los primeros trabajos centrados en el léxico tienen un afán práctico y un alcance lexicográfico; son bilingües y se relacionan directamente con la evangelización de los pueblos amerindios: sacerdotes y frailes ven la necesidad de traducir la Palabra de Dios del español a las lenguas indígenas, para lo que los misioneros procuran aprender los idiomas de los nativos o, al menos, conocer algunas correspondencias, organizadas en pequeños listados de palabras (cf., por ejemplo, Febrés, 1765); todo esto, en paralelo a las políticas lingüísticas que apuntaban a expandir la lengua española por sobre las hablas vernáculas (cf. Solano, 1991). Más adelante, esta necesidad se plasma en una obra lexicográfica más acabada, como es el *Diccionario araucano-español y español-araucano* del misionero capuchino Fray Félix José de Augusta (1916).

La lexicografía hispánica monolingüe, en tanto, como ha sido habitual en la tradición lexicográfica hispanoamericana, ha tomado como modelo y punto de

referencia al diccionario mayor de la Real Academia Española y posteriormente de la Asociación de Academias de la Lengua Española (desde el *DRAE* que sucedió al *Diccionario de Autoridades* hasta el actual *DLE*)², y ha originado diccionarios semasiológicos diferenciales o contrastivos, llamados comúnmente de “chilenismos”. Siempre en esta línea, la lexicografía de autor del país ha avanzado considerablemente desde el prescriptivismo extendido en el siglo XIX y parte importante del siglo XX, representado por lexicógrafos como Rodríguez (1875), Ortúzar (1893) o Román (1901-1918), ha tendido a reflejar cada vez con mayor fidelidad la realidad de la lengua y ha alcanzado su punto más alto en Morales Pettorino, Quiroz Mejías, & Peña Álvarez (1984-1987) y Morales Pettorino (1998 y 2006), con su recurso a bases de datos de mayor consistencia y amplitud, posibilitado por los avances informáticos; los progresos han sido evidentes en lo relativo a lematización, distinción de acepciones, definición y empleo de marcas sistémicas, de variación, pragmáticas y otras.

Los primeros diccionarios chilenos, en efecto, fueron elaborados por hombres cultos preocupados por el cambio lingüístico, ya que lo percibían como un riesgoso deterioro del idioma español, que podía conducir a la fragmentación de la lengua, como había sucedido en el pasado con el latín: el diccionario era para ellos un medio para contribuir al perfeccionamiento de la lengua. Estos autores trabajaban preferentemente en solitario; tomaban como modelo la España castellana – representada, por lo general, aunque no exclusivamente, por la Real Academia Española (RAE)– y tendían a desestimar las creaciones locales, pese a que, en esta etapa, la subjetividad imperante no admitía mayores sistematizaciones.

Así, el abogado y parlamentario Zorobabel Rodríguez (1875, p. 7), autor del primer diccionario de chilenismos, considera que “La incorrección con que en Chile se habla i escribe la lengua española es un mal tan jeneralmente reconocido como justamente deplorado”. Y luego, s.v. *abutagarse*, *abutagado*, señala que “No son chilenismos, sino muestras palpables del descuido con que miramos cuanto atañe a la recta pronunciación de las palabras. Claro se está que lo que manda el Diccionario es *abotagarse*, *abotagado*, *a*”.

En la misma línea, el sacerdote salesiano Camilo Ortúzar (1893, p. 5) afirma que

Se nos tilda a los hispano-americanos de hablar cierta jerigonza y de ser como contrabandistas del idioma español: tantas son las locuciones viciosas que tienden entre nosotros a convertirlo en un revuelto fárrago, ya que no en miserables dialectos.

Y luego, añade, bajo el título “Advertencias”, que

Van escritas con letra gruesa mayúscula las expresiones incorrectas, y con versalilla las palabras ó frases castizas. Las abreviaturas de que nos servimos son las mismas del Diccionario de la Real Academia.

En el también sacerdote Manuel Antonio Román (1901-1918), se aprecia asimismo un explícito afán moralizador: la pureza de los hablantes no debe verse mancillada por el empleo de términos impropios, ya sea por su carácter malsonante, ya sea por tratarse de voces locales no “reconocidas” por la RAE. Por ello, no incluye “las voces ni las acepciones que expresan un objeto ó idea obscena” (p. 7),

² Los diccionarios de la Real Academia Española de los siglos XVIII al XX están disponibles para consulta en el Nuevo Tesoro Lexicográfico de la Lengua Española: <https://apps2.rae.es/ntle/SrvltGUISalirNtle>

arguyendo que no pertenecen al lenguaje general, y precisando que quedan “fuera de lo que, por otra parte, prescriben las leyes de la moral cristiana, del decoro y de la buena educación” (p. 7). Por lo mismo, y al igual que Ortúzar, de quien se declara admirador, Román (1901-1918, p. 1) distingue los vocablos estigmatizados de los aceptables: “Las palabras escritas con letra VERSALITA y con cursiva, son correctas; las escritas con **negrita** son viciosas”. Este carácter prescriptivo se aprecia, en mayor o menor medida, en la mayor parte de los diccionarios chilenos hasta apenas las últimas décadas del siglo XX. La novedad de este autor es que identifica la limpieza de la lengua con la del alma, en una obra repleta de asociaciones religiosas, que considera que “el lenguaje está destinado por su naturaleza a expresar no solo los sentimientos del alma, sino también todas las ideas que guardan relación con el mundo” (Román, 1901-1918: p. 7), con una tesis que parece basada en observaciones bíblicas (cf., por ejemplo, Lc 6, 45; Mt 12, 34-37; o Santiago 3, 1-12), fiel testimonio de que las obras lexicográficas suelen ser reflejo de la cosmovisión de sus autores.

En general, esté o no presente la palabra “chilenismo” en los títulos de los diccionarios chilenos, el criterio de selección de unidades léxicas es diferencial y contrastivo respecto del DRAE; el uso, en tanto, suele determinarse sobre la base del conocimiento de mundo de los lexicógrafos, algunos de los cuales procuran ratificar este conocimiento previo mediante el hallazgo a posteriori de alguna documentación de la lexía, que servirá también, si el diccionario lo admite, como ejemplo real de uso. Sobra precisar que lo que en estos diccionarios se llama *uso* no debe confundirse con el *uso estadístico* al que nos referiremos más adelante.

Algunos diccionarios, lejos de promover la norma peninsular, buscan retratar con afán patriótico lo que consideran propio de la nación o de una localidad a través de su léxico. Un buen ejemplo a este respecto es Lagos Altamirano (2005), obra de carácter enciclopédico que contiene lexías simples y complejas, pasando por artículos de interés turístico, como aquellos en los que se enumeran vastamente las plazas locales, con sus respectivas historias y ubicaciones.

Un término medio es el de diccionarios que emplean criterios mixtos, oscilando entre lo lingüísticamente diferencial y lo que, sin ser diferencial, da cuenta de referentes extralingüísticos que los autores estiman propiamente chilenos: un ejemplo es la consideración en diccionarios diferenciales de voces como *cóndor*; en este caso, aunque la unidad léxica pertenece al español general, el hecho de que el ave figure en el escudo nacional le confiere al animal un valor de símbolo patrio, aun cuando su hábitat se extienda más allá de las fronteras del país.

A partir de la década de 1970 y, con mayor fuerza desde fines de la década de 1990, a los diccionarios de autor se han sumado valiosos diccionarios corporativos elaborados por la Academia Chilena de la Lengua. Actualmente, esta institución lleva a cabo en paralelo diversos proyectos lexicográficos.

En las últimas décadas, el interés por los estudios léxicos ha llevado a diversos lingüistas a emprender análisis metalexicográficos de obras chilenas, lo que sin duda constituye un gran aporte; en esta línea, se encuentran los valiosos trabajos de Soledad Chávez (2022a y 2022b) y Darío Rojas (cf., por ejemplo, Rojas 2020), entre otros. Para una visión panorámica de los trabajos sobre léxico, lexicografía y metalexicografía desarrollados en Chile, puede revisarse Rabanales (2004-2005), sin duda la primera lectura indispensable para quien desee iniciarse en el estudio de esta área en nuestro país. Más de veinte años antes, Rabanales & Contreras (1987) habían registrado el léxico de hablantes cultos de Santiago de

Chile, ordenándolo por áreas temáticas y utilizando cuestionarios onomasiológicos, en el marco del *Proyecto de estudio coordinado de la norma lingüística culta del español hablado en las principales ciudades de Iberoamérica y de la península Ibérica*. Este trabajo, enmarcado en la lingüística de campo y la lingüística de la variación, ofrece resultados y métodos que aún hoy resultan de interés y que pueden además servir de base para múltiples estudios contrastivos de enfoque diacrónico.

Adentrándose de lleno en la lexicología y la sociolingüística, en Chile se han desarrollado trabajos señeros en el campo de la estadística léxica y de la lingüística computacional aplicada a la lexicoestadística, entre cuyos precursores podemos destacar a Valencia (1994a, 1994b y 1997), Valencia *et al.* (1991), Echeverría (1990 y 1991), Echeverría, Herrera, Moreno, & Pradenas (1987) y Echeverría, Vargas, Urzúa & Ferreira (2008), que se centran en el léxico disponible de los escolares, con una finalidad eminentemente aplicada a la planificación lingüística. Gracias a estos estudios, podemos conocer los vocablos de mayor disponibilidad entre los estudiantes y apreciar sus deficiencias en términos de competencia léxica, las que dan cuenta del abismo que existe en la actualidad entre el léxico que conocen los alumnos y el que realmente necesitan para desenvolverse adecuadamente en sus interacciones comunicativas orales y escritas.

El propio Estado ha tomado conciencia de esta realidad y ha buscado caminos para uniformar criterios en torno al léxico que debe enseñarse en las escuelas. Así, el Ministerio de Educación Pública promulga y publica a principios de 1996 (24 de enero y 03 de febrero respectivamente) la Norma DTO-40, modificada por última vez el 20 de julio de 2009, según consta en el DTO-280. Allí establece “objetivos fundamentales y contenidos mínimos obligatorios para la Educación Básica y Media y fija normas generales para su aplicación”. A esta norma se adjunta, a modo de anexo, un “Vocabulario de uso frecuente en el español de Chile que debería ser comprendido al fin de NB2”, vale decir, al terminar el primer ciclo básico (1°, 2°, 3° y 4° años de primaria). Este

[...] vocabulario comprende aproximadamente 1.500 palabras de uso frecuente en el español de Chile y un conjunto de palabras especializadas de los distintos subsectores curriculares, que deberían ser comprendidas (vocabulario pasivo) por los egresados de NB2.

Respecto de las bases léxico-estadísticas del listado, el documento explicita que

Las palabras de uso frecuente se han definido sobre la base de un estudio realizado por la OEA en Puerto Rico en 1952, que fue actualizado y adaptado para Chile por Felipe Allende en 1994, a partir del análisis de 200 textos utilizados en la enseñanza escolar de Chile.

Y aunque el Ministerio señala que “Los estudios de vocabulario requieren de bases mucho más amplias para tener un valor científico, por esto la presente lista tiene un valor solo referencial y pragmático” (Ministerio de Educación Pública, 1996), estimamos que la elaboración de este listado ha sido un útil primer paso en el reconocimiento de la necesidad de uniformar criterios en torno a la planificación lingüística. En este sentido, los múltiples estudios de disponibilidad léxica que se han desarrollado en el país a partir de los trabajos señeros de Max Echeverría, Alba Valencia y otros, varios de ellos insertos en el marco del Proyecto Panhispánico de

Léxico Disponible, pueden ser un referente para la planificación del léxico temático tanto en adquisición de la lengua materna como en la enseñanza-aprendizaje de segundas lenguas. Para la planificación del léxico atemático, en tanto, el de mayor uso en el español de Chile con independencia del tema de que se trate o de la situación, que conforma el núcleo estadístico de la lengua, se requiere conocer el léxico básico.

2.3. Léxicos básicos

A este respecto, importa precisar que los conceptos sociolingüísticos y lexicoestadísticos de *uso* y de *léxico básico* difieren de los que se emplean en lengua común: de allí que no todos los registros léxicos que incluyen el calificativo “básico” sean léxicos básicos en el sentido explicado más arriba, y que tampoco lo que suele entenderse como léxico de alto uso lo sea realmente desde una perspectiva científica.

En efecto, muchas veces con motivación pedagógica, distintos autores del mundo hispánico han buscado acceder a lo que llaman *léxicos* o *vocabularios básicos*. Es el caso de Blásquez Fabián *et al.* (1986), que incluye en su listado mayoritariamente vocablos que “figuran en el **Diccionario de la Lengua Española** (del año 1.984) de la Real Academia Española”, obtenidos en su mayor parte de encuestas realizadas a estudiantes y, en menor medida, de la revisión de libros de autores contemporáneos; los razonamientos estadísticos no son compartidos con el lector “por razones de espacio” (Blásquez Fabián *et al.*, 1986, p. 21), de modo que no podemos evaluarlos. Butiñá Jiménez (1985, p. 3), por su parte, presenta un “vocabulario básico” de 786 vocablos en español, ordenados alfabéticamente a partir de “la confrontación de los vocabularios básicos castellano, catalán y gallego. Es decir, de los conceptos de mayor frecuencia de uso que coinciden en las tres lenguas”, con sus equivalentes en catalán y gallego; manifiesta, como resulta evidente, una confusión entre vocablos y conceptos. Neggers (1980), por su parte, entrega en sucesivas ediciones de motivación pedagógica “una especie de antología de palabras” extraídas de fuentes secundarias y obtenidas “Utilizando ciertos criterios lexicográficos” no especificados; como es lógico, las cuatro mil palabras aquí registradas nada tienen que ver con un léxico básico en el sentido sociolingüístico o lexicoestadístico.

Para estas disciplinas, los *léxicos básicos* corresponden a los aproximadamente cinco mil vocablos más usados en una comunidad y se obtienen por medio de la aplicación de fórmulas de *frecuencia*, *dispersión* y *uso*. En términos simples, la frecuencia da cuenta del número de veces que aparece la unidad léxica en un corpus; la frecuencia será relativa si se observa en relación con la muestra, o absoluta si no considera parámetros de referencia. La dispersión, según Ávila (1999, p. 78), “debe observar cómo se distribuyen las apariciones de un mismo elemento en distintos ámbitos o tipos y si se diferencia en mucho o en poco el número de apariciones para cada elemento”, de modo que la dispersión de un término aumenta cuando su distribución es equilibrada y disminuye cuando una unidad léxica es exclusiva de un ámbito. El uso, en tanto, es el producto de la frecuencia por la dispersión. De allí se desprende que unidades léxicas de alta frecuencia en contextos restringidos (dispersión mínima) tengan bajo uso. Baste considerar que si

la dispersión mínima es = 0, cualquier índice de frecuencia, por elevado que sea, al multiplicarse por 0 tendrá como producto un uso = 0.

El antecedente directo de los léxicos básicos se encuentra en los *diccionarios estadísticos*, también llamados *diccionarios de frecuencia*, debido a que dan cuenta específicamente del índice de frecuencia con que cada unidad léxica es empleada en una lengua determinada. En español, Rodríguez Bou (1952a y 1952b) y Juillard & Chang-Rodríguez (1964) son los precursores. El primero –en cuya obra se basa Felipe Allende para su propuesta al MINEDUC–, realiza su estudio preferentemente en torno al español de Puerto Rico, mientras que los segundos se concentran en el español peninsular contenido en textos escritos en el periodo comprendido entre las dos guerras mundiales (p. 14). Más tarde, Rodríguez Bou (1967) presenta además un inventario del léxico de los preescolares.

En Hispanoamérica, el trabajo que ha marcado pauta, pese a no ser el primero de su tipo, es el realizado en Puerto Rico por Morales (1986), quien, metodológicamente, optó por trabajar sólo con fuentes escritas. Aunque el léxico básico elaborado en República Dominicana (González, Cabanes, & García, 1982) es anterior, sus deficiencias en la clasificación de obras por mundo, en la segmentación de unidades y en el procesamiento estadístico, así como sus contradicciones, lamentablemente impiden, y en ello concordamos con Morales (1986, p. 36), “todo intento de labor comparativa”. A modo de ejemplo, aunque los autores definen el uso como “la proporción que guardan entre sí la frecuencia y la dispersión” y entregan una fórmula para calcularlo (González, Cabanes, & García, 1982, p. 33), señalan en otro apartado que

En total, el LEXICO BASICO [sic] lo componen 8244 palabras que poseen una frecuencia mayor de cuatro, o bien una dispersión no inferior a 38.80. Para pertenecer, pues, al LEXICO BASICO [sic] es suficiente que una palabra cumpla con cualquiera de estos dos requisitos; y no es necesario que posea los dos a la vez (González, Cabanes, & García, 1982, p. 7).

Se ha desarrollado también una investigación sobre léxico básico en Venezuela, dirigida por María Eugenia Martínez, cuyos resultados no han sido publicados y que esperamos que pronto pueda ver la luz. En Costa Rica, en tanto, existe un léxico básico restringido a los niños pre-escolares (cf. Sánchez & Murillo, 1997 y 2002). En Cataluña se elaboró también un listado de palabras (Serra *et al.*, 2000, pp. 533-550) ordenadas por frecuencia según las edades de los informantes (entre 12 y 47 meses); la limitación del corpus está dada porque se trata solo de diez informantes y estos no conforman un grupo homogéneo: uno es monolingüe español; cuatro son monolingües catalanes; cinco son bilingües catalán-español; esto impide basar en este estudio otros posteriores realizados solo en relación con el español. En Andalucía, Ávila (1999) procesa un corpus de 500.000 palabras obtenido de fuentes orales y, mediante la aplicación de fórmulas estadísticas, determina los 5228 vocablos de mayor uso en el español hablado de la ciudad de Málaga (cf. también Ávila, 1998), usando para ella una fórmula de dispersión distinta de la empleada en investigaciones previas. Una obra más reciente es la de López Chávez (2013), que da cuenta del léxico básico del español de México.

En Chile, ya bien entrado el siglo XXI, no existía una investigación centrada en determinar el léxico básico del español de Chile desde una perspectiva lexicostadística y sociolingüística. La primera investigación al respecto ha sido

desarrollada por Castillo Fadić gracias al patrocinio de Humberto López Morales. A partir de los resultados de esta investigación, pionera en el país se ha elaborado un diccionario homónimo (Castillo Fadić, 2021a), que corresponde también al primer diccionario de autor del español de Chile elaborado íntegramente por una mujer.

3. Léxico básico del español de Chile: del estudio estadístico a la obra lexicográfica

3.1. Lineamientos centrales

Nuestra investigación se ha centrado en la determinación del léxico básico del español de Chile, conformado por los alrededor de cinco mil vocablos de mayor uso. Para ello, opera sobre el *Corpus Básico del Español de Chile* © (Castillo Fadić, 2012), corpus de referencia que considera solo fuentes escritas, por sus ventajas metodológicas sobre las orales en un estudio lexicoestadístico de esta naturaleza:

[...] la lengua escrita es la que ofrece más posibilidad de tratamiento adecuado y riguroso, puesto que ésta es la única que ofrece, a priori, unas categorías generales (drama, ensayo, narrativa, etc.), fáciles de identificar y con suficiente autonomía entre ellas [...]. (Morales, 1986, p. 12).

Sin estas categorías, es imposible realizar un análisis estadístico como este, de modo que el requisito de *clasificación* es fundamental (cf. Morales, 1986, p. 14 y Julland & Chang-Rodríguez, 1964, p. 15).

La investigación es *sincrónica* y *exhaustiva*; nuestras fuentes son libros y diarios publicados entre 1981 y 2006 por autores chilenos, obtenidos mediante un muestreo proporcional aleatorio estratificado sin reemplazo a partir del total de obras consignadas en la Biblioteca Nacional de Chile, lo que quiebra la tradicional tendencia de los estudios léxicos –y especialmente lexicográficos– del mundo hispánico, de elegir sus fuentes considerando parámetros estéticos, ideológicos, o comerciales. Concordamos con Ávila (1998, p. 246) en que esta decisión aumenta la representatividad de la muestra, por cuanto el lenguaje de las obras canónicas suele presentar grandes diferencias cualitativas y cuantitativas con el uso lingüístico real.

Para que el material recolectado fuese *homogéneo* y, por ende, procesable estadísticamente, nuestras fuentes representan el español general, lo que deja fuera publicaciones regionales, diccionarios, catálogos y obras que contienen lengua estereotipada o artificial, lo que incluye desde libros de poesía o de oraciones, hasta textos escolares y literatura infantil y juvenil, donde adultos emplean el lenguaje que estiman será cercano a su público objetivo, sin pertenecer ellos mismos a ese público; en esto, coincidimos con Morales (1986, p. 13). Traducciones y adaptaciones de libros o de noticias procedentes de agencias internacionales son también excluidas, en lo que concordamos además con Julland & Chang-Rodríguez (1964, p. 14):

To insure stylistic as well as socio-cultural homogeneity, only samples of running text were admitted, to the exclusion of glossaries or list of isolated words [...] ; only prose to the exclusion of verse; only texts reflecting mature language, to the exclusion of

infant language; and only text originally written in Spanish to the exclusion of translations.

Por el requisito de *clasificación*, las fuentes se dividieron en cinco mundos, homólogos a los de Juillard & Chang-Rodríguez (1964) y a los de Morales (1986). Estos son:

- Drama: obras dramáticas.
- Narrativa: textos predominantemente narrativos, ficción y no ficción.
- Ensayo: ensayos, textos de opinión, preferentemente argumentativos.
- Técnico-científico: textos de especialidad, predominantemente descriptivos, expositivos y explicativos.
- Prensa: diarios de circulación nacional.

En el caso de la prensa, además, las secciones se clasificaron en *Opinión, Nacional, Deportes* y *Espectáculos*, lo que supuso un muestreo proporcional aleatorio estratificado sin reemplazo para determinar la sección que se consideraría de cada diario sorteado inicialmente y evitar desbalances.

Para un adecuado funcionamiento de las fórmulas estadísticas, buscamos contar con el mayor número posible de obras por mundo (al menos un 15 % de cada universo relativo). Cuidamos también que el número de palabras por mundo diera validez estadística a la muestra: así, cada uno incluye algo más de 100 mil palabras, lo que arroja un corpus total de al menos 500 mil palabras, en concordancia con la tradición establecida por Juillard & Chang-Rodríguez (1964)³ y seguida por Morales (1986) y Ávila (1999)⁴ y que consideramos consistente para obtener resultados estadísticamente válidos, más aún cuando nuestro corpus cumple con los requisitos de *sincronía, exhaustividad, homogeneidad* y *clasificación*. Concordamos por tanto con Morales (1986, pp. 14-15) en que

el grado de rigor metodológico en la selección del texto y la homogeneidad del mismo, sumado a la clasificación posterior que el texto presente, condicionan su volumen. En esto se basan lingüistas como Juillard cuando favorecen la reducción del tamaño de la muestra en oposición a algunos recuentos anteriores que manejaban millones de palabras.

Cabe precisar que Juillard & Chang-Rodríguez (1964, p. 22) estiman que el tamaño de la muestra no debe considerar únicamente aspectos lingüísticos y matemáticos, sino también el propósito de la investigación, su tipo y el nivel en el

³ Juillard & Chang-Rodríguez (1964, p. 25) estiman que “a universe of 500,000 exhaustively counted words was considered amply sufficient”.

⁴ Ávila (1998, p. 122) afirma que “no son suficientes unos cuantos cientos de palabras para obtener una distribución estable, pero a la vez está demostrado, y de hecho así lo corroboramos nosotros, que a medida que avanzamos en la realización de un recuento, el número de tipos léxicos o lemas nuevos es progresivamente menor. Es decir, cada vez necesitamos considerar un mayor número de palabras para disponer de un nuevo tipo léxico, de manera que si el léxico es finito, llegaría un momento en que el número de palabras nuevas que añadiésemos al recuento no aportaría nuevos tipos léxicos. Parece por tanto que establecer índices de frecuencia léxica sobre corpus de varios millones de palabras no tiene un especial significado”. Esto concuerda con Lara (2006, p. 156), quien, al comparar la suficiencia de su *Corpus del Español Mexicano*, de menos de dos millones de ocurrencias, con el del *Trésor de la langue française*, de 70 millones, concluye que su corpus alcanza la suficiencia con menos datos, lo que lo hace “tan representativo como el francés, sólo que menos costoso; es decir, resultó más eficaz”.

que se centra. Así, en un estudio del nivel fonológico, con bajo número de invariantes, es posible operar sobre una muestra pequeña; en un estudio del nivel léxico, por el contrario, el elevado número de invariantes exige una muestra de mayor amplitud; ahora bien, ante distintos estudios centrados en un mismo nivel, los descriptivos necesitan muestras menores que los explicativos o comparativos. En términos de Julland & Chang-Rodríguez (1964, p. 23),

In taking a universe of 500,000 words, we chose a size suited to the need of the highest levels of structuration and to the more complex comparative, typologic, and diachronic objectives, aiming to satisfy the statistically more demanding among our objectives.

El hecho de trabajar con similar número de palabras por mundo –alrededor de 100 mil, como en Morales (1986) y Julland & Chang-Rodríguez,(1964)–, supone valores estadísticamente similares en los errores muestrales de cada submuestra, lo que facilita la realización de comparaciones entre mundos, hace más sencillo el cálculo de la dispersión y permite manejar frecuencias absolutas (Julland & Chang-Rodríguez, 1964, p. 26), más amigables que las frecuencias relativas para los usuarios del diccionario que desde el comienzo de la investigación proyectamos crear.

After experimenting with a number of weighted words and after testing the results, dangers and disadvantages appeared so considerable that the chose to operate with equal, unweighted samples, thus dividing the 500,000 words of our universe into five worlds of 100,000 words each.

(Julland & Chang-Rodríguez, 1964, p. 26)

3.2. Macroestudio interdisciplinario

Se trata de un macroestudio de réplica (cf. Morales, 1986) que se centra en el nivel léxico de la lengua y tiene un primer enfoque descriptivo y cuantitativo, en tanto procura establecer cuáles son los aproximadamente cinco mil vocablos más usados en la nación. Para ello, analiza un corpus de más de quinientas mil palabras, extraído de fuentes primarias escritas de autores chilenos, en un marco natural, y publicadas en el período comprendido entre enero de 1981 y diciembre de 2006 (cf. Castillo Fadić, 2021b). El estudio, sincrónico, sintópico, de naturaleza empírica, finalidad pura e implicancias aplicadas, trabaja con distintos géneros o tipos textuales –narrativa, ensayo, drama, técnico-científico y prensa– y establece comparaciones entre el léxico de cada uno de ellos, usando herramientas de estadística inferencial. De aquí se desprende una segunda mirada explicativa y cualitativa, de alcance sociolingüístico. Siendo un estudio de réplica respecto de otros elaborados en otros lugares del mundo hispánico, tiene proyecciones dialectológicas. El procesamiento del corpus usa herramientas de lingüística de corpus, lingüística computacional y léxico-estadística y considera aspectos lexicológicos, semánticos, morfológicos y sintácticos, principalmente en la lematización de las unidades léxicas y en la distinción de homónimos (Castillo Fadić, 2020). La posterior creación de un diccionario estadístico (Castillo Fadić, 2021a) implica además el empleo de teorías y metodologías lexicográficas.

3.3. ¿Por qué crear un léxico básico?

Los léxicos básicos permiten establecer políticas lingüísticas basadas en la realidad de la lengua. Al determinar científicamente cuáles son los vocablos de mayor uso, podemos precisar también qué palabras requiere adquirir o aprender un individuo para desenvolverse de manera competente en la comunidad, tanto para comprender como para producir textos. El *Léxico básico del español de Chile* aspira, en este sentido, a contribuir a la enseñanza-aprendizaje del español como primera o segunda lengua, facilitando el acceso de las personas a los bienes culturales. Conocer el léxico básico y enseñarlo en las escuelas, permitiría disminuir el abismo entre el vocabulario que emplean los estudiantes y el que efectivamente necesitan para decodificar un texto; esto posibilitaría, a su vez, diagnosticar con mayor precisión a los estudiantes con dificultades de comprensión lectora, deslindándolos de aquellos que no entienden lo que leen porque está escrito con palabras que les son ajena (Castillo Inzulza, 2007, p. 38).

En otro ámbito, nuestro trabajo podría ser de utilidad para la lingüística computacional, al presentar las unidades no solo más frecuentes sino también más estables del español de Chile, necesarias para alimentar los diccionarios internos de múltiples aplicaciones informáticas, muchas de ellas de uso cotidiano. Pensamos, por ejemplo, en los predictores de texto de teléfonos móviles y ordenadores, en los programas para mejorar el acceso de personas con discapacidad auditiva o visual a textos orales y escritos, en la comunicación de seres humanos con máquinas mediante indicaciones orales (Siri, Waze y otros), etc. En esto concordamos con Ávila (1998, p. 8), quien considera que

Las tecnologías del habla, [...], interesadas en la conversión de textos escritos en su equivalente oral, en la transformación del habla en texto escrito, en la traducción automática de conversaciones, la identificación o verificación de hablantes en servicios telefónicos o en el desarrollo de sistemas que permitan la comunicación oral hombre-máquina, demandan a la lingüística actual cada vez con más fuerza la realización de índices de frecuencias de lengua. Estos han de ser válidos como sustento de las bases de datos que deben incluirse en las compilaciones informáticas y, asimismo, deben establecer unos límites, fundamentados en la estadística, al grueso del corpus con el que se ha de operar.

3.4. Síntesis de los resultados

A la fecha, podemos afirmar que todos los objetivos de esta investigación se han cumplido. Nuestro objetivo general ha sido elaborar el *Léxico básico del español de Chile* (Castillo Fadić, 2021), obra ya concluida que consta de 4831 entradas.

Nuestros objetivos específicos han sido:

a) Precisar la frecuencia, dispersión y uso de cada una de las unidades léxicas contenidas en nuestro corpus de español de Chile, lo que se ha hecho no solo para todos los vocablos del corpus, sino también para todos los tipos, lo que nos permite disponer de un corpus estratificado, etiquetado y con datos estadísticos (Castillo Fadić, 2012).

b) Determinar cuáles son los vocablos de mayor frecuencia, dispersión y uso en el español de Chile. De esto damos cuenta en Castillo Fadić (2021a).

c) Analizar los resultados obtenidos por medio del contraste entre los índices de frecuencia léxica de cada mundo y los índices de dispersión y uso.

d) Contribuir al estudio del español de Chile. Junto con la creación de un corpus de referencia estratificado (Castillo Fadić, 2012) y un diccionario estadístico (Castillo Fadić, 2021a), los resultados obtenidos nos han permitido generar nuevos estudios sobre el español de Chile, centrados en la combinatoria de verbos de alto uso (Castillo Fadić, 2015b), en la distribución y en las funciones de unidades léxicas específicas (Castillo Fadić y Sologuren Insua, 2017), en la variación morfológica de verbos en tiempos y modos verbales particulares de acuerdo con diferentes variables (Castillo Fadić y Sologuren Insua, 2018), en las predicciones de vocablos relevantes desde una perspectiva de género (Castillo Fadić, 2019), etc.

e) Sentar las bases para seleccionar las unidades léxicas pertinentes para elaborar distintos tipos de diccionarios del español de Chile. Por lo pronto, esperamos que Castillo Fadić (2021a) a sea la base del primer diccionario básico del español de Chile.

Como vemos, cada uno de los objetivos específicos se ha cumplido. Aunque, por la extensión de los resultados no podemos dar cuenta detallada de todos ellos aquí, por lo pronto, presentamos los 50 vocablos de mayor uso en el español de Chile, con una representatividad acumulada superior al 50 %:

Rango	Lema	Categoría
1	el	ART DEF
2	de	PREP
3	y	CONJ COORD
4	a	PREP
5	en	PREP
6	un	ART DEF
7	que	PRON REL
8	ser	V SEMI AUX
9	que	CONJ SUB
10	se	PRON
11	su	ADJ POS
12	por	PREP
13	con	PREP
14	no	ADV
15	para	PREP
16	como	CONJ SUB
17	este	ADJ DEM
18	más	ADV
19	lo	PRON PERS
20	estar	V AUX
21	tener	V
22	se	PRON PERS
23	poder	V
24	hacer	V

25	o	CONJ COORD
26	todo	ADJ IND
27	le	PRON PERS
28	pero	CONJ COORD
29	ese	ADJ DEM
30	dicir	V
31	si	CONJ SUB
32	me	PRON PERS
33	él	PRON PERS
34	sin	PREP
35	otro	ADJ IND
36	entre	PREP
37	ir	V
38	dar	V
39	haber	V
40	mismo	ADJ CAL
41	año	SUST
42	deber	V
43	sobre	PREP
44	nos	PRON PERS
45	ver	V
46	cuando	CONJ SUB
47	también	ADV
48	este	PD
49	ya	ADV
50	hasta	PREP

Tabla 1. Primeros 50 vocablos de mayor uso en el español de Chile.
Fuente: Elaboración propia.

4. Reflexiones finales

Hemos presentado los principales antecedentes del macroproyecto de investigación que ha dado origen al diccionario estadístico titulado *Léxico básico del español de Chile* (Castillo Fadić, 2021). Hemos procurado presentar los aspectos fundamentales de la investigación, incluyendo lo relativo a la conformación del *Corpus Básico del Español de Chile* © (Castillo Fadić, 2012), creado expresamente para este estudio. Hemos hecho hincapié en el carácter ampliamente interdisciplinario de la tarea emprendida y también en la posibilidad de que nuestros resultados contribuyan al establecimiento de políticas lingüísticas basadas en el uso real del léxico del español de Chile. Manifestamos, también, nuestro deseo de que el diccionario estadístico obtenido como producto final de la investigación sea una herramienta eficaz para educadores y estudiantes.

Aunque los resultados de esta investigación podrían ser aplicados directamente en el ámbito académico y pedagógico y en el establecimiento de nuevas políticas lingüísticas, especialmente por medio de nuestro diccionario estadístico (Castillo Fadić, 2021), nos alienta la posibilidad de crear además, *a posteriori*, a partir de nuestro léxico básico y de los resultados obtenidos del análisis estadístico del *Corpus Básico del Español de Chile* ©, un diccionario fundamental exemplificado del español de Chile, semasiológico e integral, monolingüe para quienes aprenden español como primera lengua y bilingüe para quienes lo aprenden como lengua extranjera. Pensando de antemano en este proyecto, cada una de las citas que constituyen nuestro corpus fue subclasificada por materia y cuenta con referencias completas, para funcionar eventualmente como ejemplo real de uso al interior de un artículo lexicográfico. Como Lara (1990, p. 234), consideramos que el trabajo de los lexicógrafos hispanoamericanos no puede seguir siendo periférico o complementario y que resulta de vital importancia crear diccionarios de lengua orientados a las necesidades específicas de nuestras comunidades.

Esperamos que a este estudio que ha buscado replicar trabajos señoseros como el de Morales (1986), lo sigan muchos otros, para que todo el mundo hispánico pueda contar con descripciones realistas de la lengua, de modo de favorecer el desarrollo de políticas lingüísticas locales y panhispánicas que faciliten la adquisición-aprendizaje del español como lengua materna y como lengua extranjera en todos los países donde se habla español.

5. Referencias bibliográficas

- Augusta, J. d. (1991, 1^a ed. 1916). *Diccionario araucano*. 2^a ed. Íntegra. Editorial Kushe.
- Castillo Fadić, M. N. (2012). *Corpus Básico del Español de Chile* ©.
- Castillo Fadić, M. N. (2015). “El verbo <hacer> en el español de Chile: tipos y combinaciones frecuentes en el género ensayo”, *E-Aesla*, 1, 1-9. <https://cvc.cervantes.es/lengua/eaesla/pdf/01/50.pdf>
- Castillo Fadić, M. N. & Sologuren Insúa, E. (2017). El reformulador <es decir> en el español de Chile: una propuesta de clasificación funcional. *Revista Lenguas Modernas*, 49, 77-92.
- Castillo Fadić, M. N. & Sologuren Insúa, E. (2018). Pretérito imperfecto de subjuntivo en el español de Chile: ¿existe alternancia libre entre las desinencias -ra y -se? *Onomázein*, 42, 153-171.
- Castillo Fadić, M. N. 2019. ¿Qué se dice de la mujer y el hombre en el español de Chile?: estudio exploratorio de las combinaciones frecuentes de los vocablos mujer y hombre en un corpus de referencia estratificado. *Boletín de Filología*, 54(1).
- Castillo Fadić, M. N. (2020). Corpus Básico del Español de Chile ©: metodología de procesamiento y análisis. *Lexis*, 44(2), 483-523.
- Castillo Fadić, M. N. (2021a). *Léxico básico del español de Chile*. Liberalia Ediciones.
- Castillo Fadić, M. N. (2021b). Corpus Básico del Español de Chile ©: metodología de obtención, revisión y constitución definitiva. En Rojas Abelardo San Martín, D. y Chávez, S. (eds.), Anejo N°3 del *Boletín de Filología* (Estudios en homenaje a Alfredo Matus Olivier) (pp. 219-252).
- Ávila Muñoz, A. M. (2006). *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Universidad de Málaga.
- Ávila Muñoz, A. M. (1999). *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Universidad de Málaga.

- Blásquez Fabián, M. et al. (1986). Experiencias sobre la enseñanza del vocabulario. Vocabulario básico para preescolar y E.G.B. Universidad de Oviedo-Instituto de Ciencias de la Educación.
- Butiá Jiménez, J. (1985). *Vocabulario básico del español*. Gaycesa.
- Castillo Inzulza, G., (2007). *La escuela de anticipación: una escuela básica universalista, inclusiva*. CPEIP - MINEDUC.
- Chávez Fajardo, S. (2022a). *Diccionarios del fin del mundo*. Fondo de Cultura Económica.
- Chávez Fajardo, S. (2022b). *Elementos de lexicografía hispanoamericana fundacional. Acerca del Diccionario de chilenismos y de otras voces y locuciones viciosas de Manuel Antonio Román (1901-1918)*. Doble Pleca.
- Echeverría, M. (1990). Crecimiento del vocabulario pasivo en comunidades escolares. En *Actas del VIII Seminario de Investigación y Enseñanza de la Lingüística* (pp. 272-282), Central de Publicaciones de la Universidad de Chile.
- Echeverría, M. (1991). Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. En *Actas del II Seminario Internacional sobre aportes de la lingüística a la enseñanza del español como lengua materna* (pp. 61-78), Editorial de la Universidad de Puerto Rico.
- Echeverría, M., Herrera, M. O., Moreno, P., & Pradenas, F. (1987). Disponibilidad léxica en Educación Media. *RLA*, 25, 55-115.
- Echeverría, M., Vargas, R., Urzúa, P., & Ferreira, R. (2008). Dispografo: una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 81-91. <https://dx.doi.org/10.4067/S0718-48832008000100005>
- Frébrés, Andrés. (1765). *Arte de la lengua general del Reyno de Chile, con un diálogo chileno-hispano muy curioso: a que se añade la doctrina christiana, esto es, rezo, catecismo, coplas, confesionario, y pláticas, lo más en lengua chilena y castellana: y por fin un vocabulario hispano-chileno, y un calepino chileno-hispano mas copioso*. Calle de la Encarnación-Lima.
- Fernández-Sevilla, J. (1974). *Problemas de lexicografía actual*. Caro y Cuervo.
- Gómez-Pablos, Beatriz. (2022). Castillo Fadić, María Natalia. Léxico básico del español de Chile. *Études romanes de Brno*, 43(2), 330-331.
- González, A., Cabanes, S., & García, F. (1982). *Léxico básico de la lengua escrita en la República Dominicana. Informe de la investigación del léxico básico de la lengua escrita en República Dominicana y Diccionario de uso, frecuencia y dispersión*. Universidad Nacional Pedro Henríquez Ureña.
- Juillard, A., & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words* (The Romance Languages and their Structures, First Series SI). Mouton & Co.
- Lagos Altamirano, D. (Dir.). (2005). *Diccionario lingüístico del habla de Valparaíso*. Universidad de Playa Ancha de Ciencias de la Educación.
- Lara, L. F. (1990). *Dimensiones de la lexicografía. A propósito del Diccionario del español de México*. El Colegio de México.
- Lavid, J. (2005). *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Cátedra.
- López Chávez, J. (2013). *Léxico básico del español de México*. Instituto de Educación y Formación Atenea SC-Edere, SA de CV.
- López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Arco Libros.
- Marín, Pablo. (2021, 12 de junio). Palabras que nunca faltan. En *La Tercera*. <https://www.latercera.com/la-tercera-domingo/noticia/palabras-que-nunca-faltan/Y4ZY7OHSNVA2PIZWYKAKODCPLQ/>
- Ministerio De Educación Pública. (1996, 3 de febrero). *Norma: DTO-40. Ley Chile-Biblioteca del Congreso Nacional*: <http://www.leychile.cl/Navegar?idNorma=8043&idVersion=2012-01-28>

- Morales Pettorino, F. (1998). *Diccionario ejemplificado de chilenismos y de otros usos diferenciales del español de Chile* (Vol. V). Universidad de Playa Ancha de Ciencias de la Educación.
- Morales Pettorino, F. (2006). *Nuevo diccionario ejemplificado de chilenismos y de otros usos diferenciales del español de Chile*, Vols. I, II y III. Universidad de Playa Ancha-Editorial Puntágeles-Publicaciones Bicentenario.
- Morales Pettorino, F., Quiroz Mejías, Ó., & Peña Alvarez, J. (1984-1987). *Diccionario ejemplificado de chilenismos y otros usos diferenciales del español de Chile* (Vols. I, II, III y IV). Ed. Universitaria.
- Morales, A. (1986). *Léxico básico del Español de Puerto Rico*. Academia Puertorriqueña de la Lengua-Editorial La Muralla, S. A.
- Neggers, G. (1980). *Vocabulario culto*, 2^a ed. Playor.
- Ortúzar, C. (1893). *Diccionario manual de locuciones viciosas y de correcciones de lenguaje con indicación del valor de algunas palabras y ciertas nociones gramaticales*. Impr. Salesiana-S. Benigno Canavese.
- Rabanales, A. (2004-2005). Temática de las obras lexicográficas chilenas y estudios afines. Una visión panorámica. *BFUCH*, 40, 137-166.
- Rabanales, A., & Contreras, L. (1987). *Léxico del habla culta de Santiago de Chile*. Universidad Nacional Autónoma de México.
- Robins, R. H. (1990). Historia de la lingüística. En Newmeyer, F. J. (comp.), *Panorama de la lingüística moderna de la Universidad de Cambridge* (Vol. I. Teoría Lingüística: Fundamentos) (pp. 529-551), Visor Distribuciones S. A.
- Rodríguez Bou, I. (1952a). *Recuento de vocabulario español*, Vol. I. Serie II, Núm. XII. Consejo Superior de Enseñanza-Universidad de Puerto Rico-Publicaciones Pedagógicas.
- Rodríguez Bou, I. (1952b). *Recuento de vocabulario español*, Vol. II, Partes I y II. Serie II, Núm. XII. Consejo Superior de Enseñanza-Universidad de Puerto Rico-Publicaciones Pedagógicas.
- Rodríguez Bou, I. (dir.). (1967). *Recuento de vocabulario de preescolares*. Consejo Superior de Enseñanza-Universidad de Puerto Rico-Publicaciones Pedagógicas.
- Rodríguez, Z. (1875). *Diccionario de chilenismos*. Imprenta "El Independiente".
- Rojas, D. (2020). Lexicografía académica, hispanismo y dictadura: el *Diccionario del habla chilena* (1978) de la Academia Chilena de la Lengua. *Études romanes de Brno*, 41(2), 215-234. DOI: 10.5817/ERB2020-2-13
- Román, M. A. (1901-1918). *Diccionario de chilenismos y de otras voces y locuciones viciosas* (Vols. I, II, III, IV y V). Impr. "La Revista Católica".
- Sánchez, V., & Murillo, M. (1997). Léxico básico de los niños preescolares de la sub-región educativa. *Káñina, Rev. Artes y Letras*, 22(2), 77-124.
- Sánchez, V., & Murillo, M. (2002). *Léxico básico de los niños preescolares costarricenses*. Editorial Universidad de Costa Rica.
- Serra, M. et al. (2000). *La adquisición del lenguaje*. Ariel.
- De Solano, F. (1991). *Documentos sobre política lingüística en Hispanoamérica (1492-1800)*. Consejo Superior de Investigaciones Científicas
- Valencia, A. (1994a). *El léxico de los estudiantes de 4º año de Educación Media*. Santiago de Chile: Centro de Interés "Procesos mentales". Serie *Documentos de Estudio*, 26. C.P.E.I.P.
- Valencia, A. (1994b). Disponibilidad léxica en Educación Media: III y IV. En *X Encuentro en Torno a la Admisión Universitaria*. Talleres Gráficos DIFEST.
- Valencia, A. (1997). Disponibilidad léxica. Muestreo y estadísticos. *Onomázein*, 2, 197-226.
- Valencia, A. et al. (1991). Evaluación de la riqueza léxica en estudiantes chilenos de cuarto medio. *Estudios Filológicos*, 26, 71-81.