

El Corpus l3xico del espa1ol de Colombia *CorlexCo* y su aplicaci3n a la investigaci3n sobre el espa1ol LE y L2

Gloria Viviana NIETO MARTÍN
Instituto Caro y Cuervo (Colombia)
viviana.nieto@caroycuervo.gov.co
<https://orcid.org/0000-0002-1560-1859>

Resumen: En este art3culo se presenta el Corpus l3xico del espa1ol de Colombia *CorlexCo* (<https://clicc.caroycuervo.gov.co/corpus/CorlexCo>) del Instituto Caro y Cuervo, herramienta que recopila combinaciones l3xicas con criterio integral, es decir, las propias del pa3s y tambi3n las compartidas con otros pa3ses hispanoamericanos y con el espa1ol general. No se limita a los colombianismos. Se trata de un macroproyecto iniciado en 2019 y desarrollado en fases anuales sucesivas. Por "combinaciones l3xicas" nos referimos a todo tipo de unidades pluriverbales (compuestos, colocaciones, construcciones con verbo soporte, locuciones, paremias y f3rmulas rutinarias). Los datos proceden de cinco fuentes de consulta: (1) diccionarios generales, escolares, dialectales, fraseol3gicos y combinatorios, (2) casi doscientos cincuenta repertorios l3xicos del espa1ol de Colombia, (3) m3s de ciento cincuenta investigaciones acad3micas sobre el l3xico de Colombia, (4) corpus lingüísticos y (5) el dominio <http://www.google.com.co>. Con respecto a la aplicaci3n del *CorlexCo*, por una parte, puede usarse en planeaci3n curricular, dise1o de materiales did3cticos y procesos de evaluaci3n de espa1ol como lengua extranjera y segunda lengua (para ind3genas y sordos); por otra, puede constituirse en obra de referencia de aprendientes, profesores, editores, traductores y, especialmente, de lingüistas que investiguen sobre el espa1ol LE y L2. En el apartado 1 se resume el estado actual de la lexicograf3a dialectal en Colombia. En el apartado 2 se describen las caracter3sticas principales del *CorlexCo*, la selecci3n del tipo de unidades l3xicas para cada fase anual, la selecci3n de las entradas y la anotaci3n, y se ilustra su uso con varios tipos de b3squeda. En el apartado 3 se muestra su aplicaci3n a la investigaci3n sobre espa1ol LE y L2. Este corpus es un recurso abierto que se actualiza peri3dicamente y que ofrece a la comunidad acad3mica una descripci3n ampliada del l3xico de Colombia.

Palabras clave: L2; lexicograf3a; combinatoria l3xica; espa1ol de Colombia

Català:

El Corpus Lèxic de l'espanyol de Colòmbia CorlexCo. Aplicacions a la recerca sobre l'espanyol com a LE i L2

Resum:

En aquest article es presenta el Corpus Lèxic del espa1ol de Colombia *CorlexCo* (<https://clicc.caroycuervo.gov.co/corpus/corlexco>) de l'Institut Caro y Cuervo, una eina que recopila combinacions l3xiques amb criteri integral, és a dir, les pr3pies del pa3s i tamb3 les compartides amb altres pa3sos hispanoamericans i amb l'espanyol general. No es limita als colombianismes. Es tracta d'un macroprojecte iniciat en 2019 i desenvolupat en fases anuals successives. Per "combinacions l3xiques" ens referim a tota mena d'unitats pluriverbals (compostos, col·locacions, construccions amb verb suport, locucions, parèmies i f3rmules rutinàries). Les dades procedeixen de cinc fonts de consulta: (1) diccionaris generals, escolars, dialectals, fraseol3gics i combinatoris, (2) gairebé dos-cents cinquanta repertoris l3xics de l'espanyol de Colòmbia, (3) m3s de cent cinquanta recerques acad3miques sobre el l3xic de Colòmbia, (4) corpus lingüístics i (5) el domini <http://www.google.com.co>. Respecte a l'aplicaci3 del *CorlexCo*, d'una banda, es pot utilitzar en la planificaci3 curricular, en el disseny de materials did3ctics i en els processos d'avaluaci3 d'espanyol com a llengua estrangera i segona llengua (per a ind3genes i sords); d'altra banda, pot constituir-se en obra de refer3ncia d'aprenents, professors, editors, traductors i, especialment, per a lingüistes

que investiguin sobre l'espanyol L1 i L2. En l'apartat 1 es resumeix l'estat actual de la Lexicografia dialectal a Colòmbia. En l'apartat 2 es descriuen les característiques principals del CorlexCo, la selecció del tipus d'unitats lèxiques per a cada fase anual, la selecció de les entrades i l'anotació, i s'il·lustra el seu ús amb diversos tipus de cerca. En l'apartat 3 es mostra la seva aplicació a la recerca sobre espanyol L1 i L2. Aquest corpus és un recurs obert que s'actualitza periòdicament i que ofereix a la comunitat acadèmica una descripció ampliada del lèxic de Colòmbia.

Paraules clau: L2; lexicografia; combinatòria lèxica; espanyol de Colòmbia

English:

Lexical corpus of Colombian Spanish CorlexCo and its application to research on TSFL and TSSL

Abstract: This article presents the Lexical Corpus of Colombian Spanish CorlexCo (<https://clicc.caroycuervo.gov.co/corpus/CorlexCo>) by Instituto Caro y Cuervo, a tool that compiles lexical combinations with integral criteria, that is, those only used in the country and also those shared with other Latin American countries and with general Spanish. It is not limited to Colombianisms. This is a macro project started in 2019 and developed in successive annual phases. By lexical combinations we refer to all types of pluriverbal units (compound words, collocations, support-verb constructions, idiomatic phrases, pemia and conversational routines). The data comes from five sources: (1) general, scholar, dialectal, phraseological and combinatorial dictionaries, (2) almost two hundred and fifty lexical repertoires of Colombian Spanish, (3) more than one hundred and fifty academic research on the lexicon used in Colombia, (4) linguistic corpus and (5) the domain <http://www.google.com.co>. With respect to the application of CorlexCo, on the one hand, it can be used in curricular planning, design of teaching materials and evaluation processes of Spanish as a foreign language and second language (for indigenous and deaf people) and, on the other hand, it can constitute a work reference for learners, teachers, editors, translators and linguists who research on TSFL and TSSL. Section 1 summarizes the current state of dialect lexicography in Colombia. Section 2 describes the main characteristics of CorlexCo, the selection of the type of lexical units for each annual phase, the selection of entries and annotation, and illustrates its use with various types of search. Section 3 shows its application to research on TSFL and TSSL. This corpus is an open resource that is periodically updated and that offers the academic community an expanded description of the Colombian lexicon, compared to dictionaries.

Keywords: 2L; lexicography; lexical combinations; Colombian Spanish

Introducción

Actualmente, después de México, Colombia es el segundo país con mayor número de hispanohablantes. Si bien contamos con amplios estudios descriptivos del español —muchos de ellos publicados por el Instituto Caro y Cuervo (ICC, <http://bibliotecadigital.caroycuervo.gov.co/>)—, aquellos dedicados al nivel léxico recogen principalmente usos regionales de palabras aisladas y compuestas, siguiendo el criterio diferencial con otras regiones del país y España. Debido a la ausencia de trabajos de recopilación y descripción lingüística de las unidades léxicas (univerbales y pluriverbales) del español integral de Colombia, los profesores de ELE/L2 no cuentan con obras de referencia que propongan la gradación del léxico por niveles de enseñanza.

Asimismo, como han señalado varios estudios durante las últimas décadas (Penadés Martínez, 1999; Olimpo de Oliveira, 2006; Gutiérrez Rubio, 2019-2020; Nieto Martín, 2022a, entre otros), existen muy pocos materiales didácticos para la enseñanza de ELE/L2 que incorporen unidades fraseológicas (UF) y muchos menos creados en Hispanoamérica. Por ello, en las publicaciones de la serie ELE/2 del ICC hemos procurado hacerlo, atendiendo a los principios básicos del enfoque léxico: *Hecho en Colombia: cultura colombiana para la clase de ELE* (Nieto Martín, 2017), *Viaje literario por Barranquilla. Lecturas guiadas para la clase de ELE* (Corredor, 2019), *Empelculados: cine en la clase de español como lengua extranjera y segunda lengua* (Ariza, Nieto y Cruz, 2020), *CántELE. Música colombiana para la clase de español como lengua extranjera. Volumen 1* (Hincapié, Nieto y León, 2021) y *Volumen 2* (Hincapié, Nieto, León y Ariza, 2022), y



Más fresco / a que una lechuga. Propuesta didáctica para la enseñanza de comparaciones estereotipadas en la clase de ELE/L2 (Nieto y Umaña, 2024).

Luego de la colaboración del ICC y la editorial Difusión en la adaptación de la colección *Aula internacional Nueva edición* (Corpas, García y Garmendia, 2013) a América (*Aula América*, Ariza y otros: 2018) —proceso durante el que se comprobó que el nivel léxico requiere la mayor intervención (Ariza, Molina y Nieto, 2019)—, se hizo evidente la necesidad de contar con inventarios léxicos del español de Colombia gradados por niveles de enseñanza. Para atender esta necesidad, desde 2019, la línea de investigación de la Maestría en Enseñanza de Español como Lengua Extranjera y Segunda Lengua del ICC está construyendo el *CorlexCo* (<https://clicc.caroycuervo.gov.co/corpus/CorlexCo>), un macroproyecto que recopila y caracteriza las combinaciones léxicas (compuestos, colocaciones, construcciones con verbo soporte, locuciones, paremias y fórmulas rutinarias) más frecuentes del español usado en Colombia. Es decir, no nos limitamos al estudio de los colombianismos, sino que incluimos también las expresiones compartidas con el español de Hispanoamérica y el español general.

A la fecha, el *CorlexCo* cuenta con el registro de 20.534 combinaciones léxicas. Esta recopilación pretende convertirse en una obra de referencia para la selección objetiva del léxico, con aplicaciones en la planeación curricular, la creación de materiales didácticos y la evaluación en los procesos de enseñanza-aprendizaje de español como lengua extranjera y segunda lengua. Además, en fases posteriores, y tras el desarrollo del sistema gestor de diccionarios del ICC (LEXICC), tendrá la posibilidad de convertirse en un diccionario combinatorio del español de Colombia, con aplicaciones en español como lengua materna, segunda y extranjera.

1. La lexicografía del español de Colombia

La tradición lexicográfica de Colombia inicia con Rufino José Cuervo (1844–1911), cuyos dos primeros tomos del *Diccionario de construcción y régimen de la lengua castellana* fueron publicados por el ICC, en 1954. La segunda obra de referencia corresponde al *Atlas Lingüístico-Etnográfico de Colombia ALEC* (ICC, <https://alec.caroycuervo.gov.co/alec/>, 1982-1983) que recopila muestras de habla de cuatro de las seis regiones del país, en veintiséis de los treinta y dos actuales departamentos. Acerca del desarrollo de la disciplina lexicográfica en Colombia, destacamos aquí dos estudios recientes. Por una parte, Ciro (2022) presenta una cronología de sesenta y ocho repertorios lexicográficos publicados entre 1867 y 2020, de cuyo análisis resalta el interés por la documentación de los regionalismos por encima del léxico integral del país. Por otra parte, Ortiz Correa (2022) señala que la mayoría de los trabajos sobre fraseología se han enmarcado dentro de los estudios del folclor, esto es, no han sido elaborados por lingüistas, lo cual no deslegitima su valor en la documentación de los usos autóctonos de las hablas colombianas. Así, la relevancia de las publicaciones se halla en las compilaciones mismas, ya que no implementan criterios lexicográficos básicos, como la consistencia en la clasificación tipológica, en la lematización y en la definición de las UF.



En este sentido, en Colombia abundan los repertorios léxicos regionales; sin embargo, no existen aproximaciones sistemáticas ni exhaustivas a la combinatoria léxica ni a la fraseología de uso general en el país. Otra dificultad para avanzar en una obra lexicográfica de ámbito nacional radica en que los dos diccionarios de colombianismos disponibles (Haensch y Werner, 1993; ICC, 2018) —al igual que el *Diccionario de americanismos* (Asale, <https://www.asale.org/damer/>, 2010)—, tienen carácter diferencial, es decir, no registran los usos compartidos con España. Además, los diccionarios de colombianismos recopilan principalmente palabras simples y compuestas, sin prestar especial atención a las UF. Tal vez las únicas obras lexicográficas integrales de las que disponemos son las adaptaciones para Colombia de diccionarios escolares de editoriales españolas, aunque, por su propósito específico, tampoco registran UF de manera exhaustiva. Esta falta de atención a las UF en los diccionarios disponibles en el país puede explicar el interés por incluirlas en los repertorios léxicos regionales (a la fecha, el *CorlexCo* ha identificado y vaciado 242 fuentes).

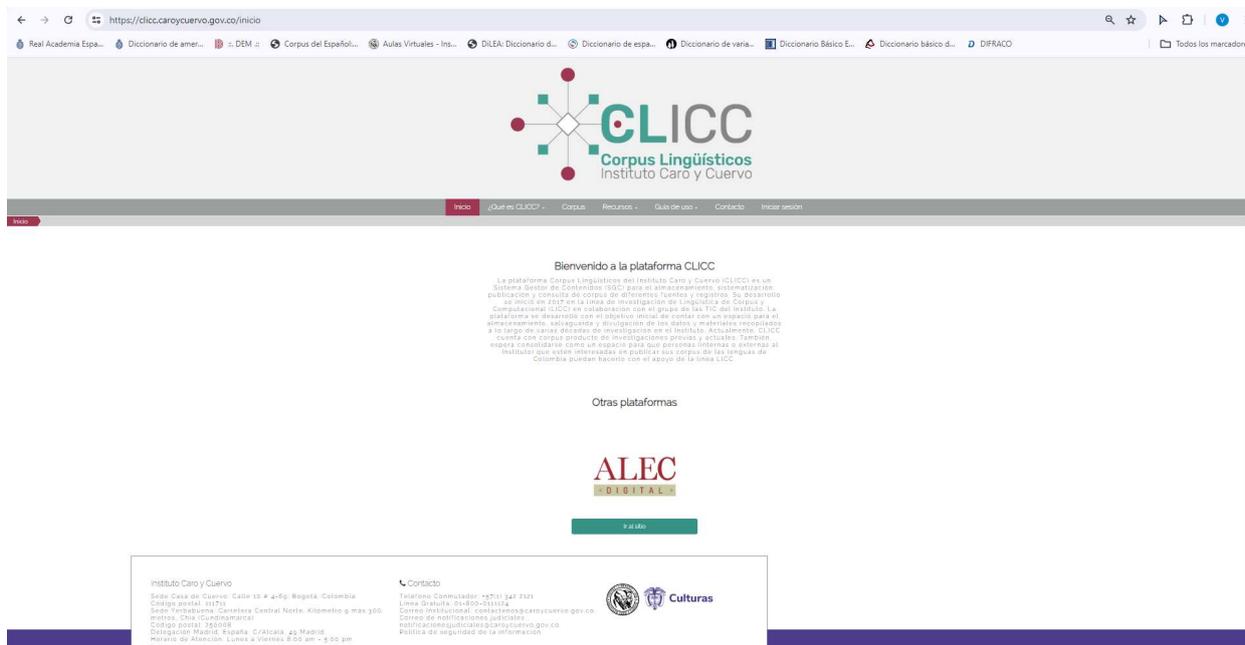
2. El corpus léxico del español de Colombia *CorlexCo*

2.1 Características

El *CorlexCo* forma parte de la plataforma Corpus Lingüísticos del Instituto Caro y Cuervo CLICC (<https://clicc.caroycuervo.gov.co/>), un sistema gestor de contenidos para el almacenamiento, la sistematización, la publicación y la consulta de corpus de diferentes fuentes y registros. Su desarrollo inició en 2017 en la línea de investigación de Lingüística de Corpus y Computacional (LICC), en colaboración con el grupo de las TIC del ICC. La plataforma se desarrolló con el objetivo inicial de contar con un espacio para el almacenamiento, la salvaguarda y la divulgación de datos y materiales recopilados a lo largo de varias décadas de investigación en el ICC. Actualmente, el CLICC cuenta con corpus producto de investigaciones tanto finalizadas como en curso. También espera consolidarse como un espacio para que personas (internas o externas al ICC) interesadas en publicar sus corpus de las lenguas de Colombia puedan hacerlo con el apoyo de la línea LICC. La Figura 1 muestra la página de inicio de CLICC.

Figura 1

Página de inicio de la plataforma CLICC del Instituto Caro y Cuervo.



Siguiendo la tipología descrita en Bernal e Hincapié (2018: 26-34), el *CorlexCo* es un corpus léxico (de combinaciones léxicas), escrito, monolingüe (en español), sincrónico (desde 1980), abierto (se actualiza periódicamente), anotado (con etiquetas lingüísticas) y no documentado (sin archivos relacionados). En la página de inicio del corpus (Figura 2) aparecen cuatro pestañas: (1) **Consultar el corpus**, a través de la cual se accede a los datos registrados; (2) **Metodología**, que ofrece una breve descripción metodológica y las referencias de las fuentes consultadas: diccionarios, repertorios léxicos del español de Colombia, blogs y grupos de Facebook, e investigaciones académicas sobre el español de Colombia; (3) **Equipo**, con los nombres de los participantes (investigadores, asistentes, pasantes, asesores) de cada fase y (4) **¿Cómo citar?**, con la referencia del corpus según las normas APA.

Figura 2

Página de inicio del CorlexCo del Instituto Caro y Cuervo.



Se planea habilitar tres tipos de búsqueda: por palabra, por metadatos y por combinaciones. Por el momento, solo está disponible la búsqueda por palabra. Las Figuras 3.1 y 3.2 muestran un ejemplo de búsqueda de combinaciones léxicas con la palabra ‘pelo’, que arroja veinte (20) resultados.

Figura 3.1

Resultados 1-10 de búsqueda de combinaciones léxicas con la palabra ‘pelo’ en el CorlexCo.

No	Archivo		Resultados	Opciones
1	CLEC.CL_NV_0183	et	PELO se rompe	
2	CLEC.CL_NV_0182	et	PELO se revienta	
3	CLEC.CL_NV_0181	et	PELO se reseca	
4	CLEC.CL_NV_0180	et	PELO se quiebra	
5	CLEC.CL_NV_0159	et	PELO se parte	
6	CLEC.CL_NV_0158	et	PELO se ondula	
7	CLEC.CL_NV_0157	et	PELO se horquilla	
8	CLEC.CL_NV_0156	et	PELO se friza	
9	CLEC.CL_NV_0155	et	PELO se enreda	
10	CLEC.CL_NV_0154	et	PELO se engrasa	

Figura 3.2

Resultados 11-20 de búsqueda de combinaciones léxicas con la palabra 'pelo' en el CorlexCo.



The screenshot shows the CLICC (Corpus Lingüístico Instituto Caro y Cuervo) web interface. At the top, there is a navigation menu with links for 'Inicio', '¿Qué es CLICC?', 'Corpus', 'Recursos', 'Guía de uso', 'Contacto', and 'Iniciar sesión'. Below the menu, there is a search bar and a table of results. The table has columns for 'No.', 'Archivo', 'Resultado', and 'Opciones'. The results show combinations like 'se encrespa', 'se enchurca', 'se encana', 'se daña', 'se cae', 'se achila', 'sale', 'nace', 'crece', and 'brilla'.

No.	Archivo	Resultado	Opciones
11	CLEC_CL_NV_0153	PELO se encrespa	[Opciones]
12	CLEC_CL_NV_0152	PELO se enchurca	[Opciones]
13	CLEC_CL_NV_0151	PELO se encana	[Opciones]
14	CLEC_CL_NV_0150	PELO se daña	[Opciones]
15	CLEC_CL_NV_0149	PELO se cae	[Opciones]
16	CLEC_CL_NV_0148	PELO se achila	[Opciones]
17	CLEC_CL_NV_0147	PELO sale	[Opciones]
18	CLEC_CL_NV_0146	PELO nace	[Opciones]
19	CLEC_CL_NV_0145	PELO crece	[Opciones]
20	CLEC_CL_NV_0144	PELO brilla	[Opciones]

Adicionalmente, haciendo clic en el ícono de ojo de la columna 'opciones', la plataforma permite descargar la ficha de metadatos de cada combinación léxica en formato PDF, como lo muestra la Figura 4.

Figura 4

Descarga de ficha de metadatos de la combinación léxica 'el pelo se encrespa' en el CorlexCo.



Ficha de metadatos

Tipo de combinación léxica: colocación N+V
 Combinación léxica: el pelo se encrespa
 Significado: N/A
 Grado de opacidad: bajo
 Unidad léxica plena: N/A
 Unidad léxica 1: pelo
 Unidad léxica 2: encresparse
 Unidad léxica 3: N/A
 Unidad léxica 4: N/A
 Unidad léxica 5: N/A
 Unidad léxica 6: N/A
 Unidad léxica 7: N/A
 Unidad léxica 8: N/A
 Campo temático 1: N/A
 Campo temático 2: N/A
 Extensión de uso: español general
 Registro de uso: no marcado
 Estructura sintagmática 1: Dsg+Nsuj+Vpr
 Estructura sintagmática 2: Dsg+Nsuj+Vpr+Ci (el pelo se [le] encrespa [a alguien])
 Otras combinaciones 1: V+NCI
 Otras combinaciones 2: N/A
 Formas nominales 1: el encrespamiento del pelo
 Formas nominales 2: el pelo encrespado, el pelo crespo
 Campo léxico 1: Individuo: dimensión física
 Campo léxico 2: Vello corporal
 Campo léxico 3: Pelo
 Campo léxico 4:
 Variantes 1: el pelo se enchurca
 Variantes 2: el cabello se encrespa
 Sinónimos: N/A
 Antónimos 1: N/A
 Antónimos 2: N/A
 Combinaciones afines: N/A

En la Tabla 1 se consolidan las fases del CorlexCo desarrolladas entre 2019 y 2024. A la fecha, contamos con el registro de 20534 combinaciones léxicas. Durante las tres primeras fases, nos centramos en la recopilación de

construcciones verbales, que corresponden a 14488 combinaciones léxicas, esto es, el 70,7 % de todo el corpus. En 2019, iniciamos la recopilación de locuciones verbales (3626; 17,7 %); en 2020, vimos la necesidad de diferenciarlas de las colocaciones verbonominales (4405; 21,5 %) y, en 2021, de las construcciones con verbo soporte (6460; 31,5 %). En 2022, atendiendo a la relación sintáctica entre verbo y adverbio, pasamos a la recolección de construcciones adverbiales (3899; 18,9 %), dentro de las que predominan las locuciones (3294; 16 %), seguidas en menor medida por las colocaciones (605; 2,9 %). En 2023, recopilamos dos estructuras distintas: locuciones adjetivas (1712; 8,3 %) y compuestos sintácticos con gentilicios (435; 2,1 %). En 2024, estamos realizando un análisis contrastivo de las locuciones adverbiales y las locuciones adjetivas recopiladas en las dos fases anteriores, debido a que muchas coinciden en su forma.

Tabla 1

Fases del CorlexCo desarrolladas de 2019 a 2024. Elaboración propia.

Fase	Tipo de unidades léxicas	No. de unidades léxicas	%	No. de unidades léxicas totales	%
2019	Locuciones verbales	3626	17,7	3623	17,7
2020	Colocaciones Nsubj.+V	1432	7	4405	21,5
	Colocaciones V+Ncd	2973	14,5		
2021	Construcciones con verbo soporte	6460	31,5	6460	31,5
2022	Locuciones adverbiales	3294	16	3899	18,9
	Colocaciones V+loc. adv.	432	2		
	Colocaciones V+adv.	121	0,6		
	Colocaciones adv.+adj./part.	52	0,3		
2023	Locuciones adjetivas	1712	8,3	2147	10,4
	Compuestos sintácticos con gentilicios	435	2,1		
2024	Locuciones adverbiales y locuciones adjetivas	-	-	-	-
		Total	100	20534	100

2.2 Procedencia de los datos

Considerando que, como hemos mencionado, Colombia aún no cuenta con un diccionario nacional o integral, los datos del *CorlexCo* proceden de cinco fuentes de consulta: (1) diccionarios generales, escolares, dialectales, fraseológicos y combinatorios, (2) casi doscientos cincuenta repertorios léxicos del español de Colombia, (3) más de ciento cincuenta investigaciones académicas sobre el léxico de Colombia, (4) corpus lingüísticos y (5) el dominio <http://www.google.com.co>.

(1) Diccionarios consultados como fuente del *CorlexCo*

La Tabla 2 presenta la lista de dieciocho diccionarios consultados para la construcción de la herramienta: generales (3), escolares (2), dialectales (6), fraseológicos (5) y combinatorios (2). Los tres diccionarios generales (DLE

[<https://www.rae.es/>], DUE y Clave) fueron seleccionados por constituirse en obras de referencia en todo el mundo hispánico. Los dos diccionarios escolares (DIDEC Primaria y Secundaria) corresponden a adaptaciones para Colombia de diccionarios españoles de Ediciones SM. En cuanto a los diccionarios dialectales, uno comprende todo el ámbito hispánico (DVE), uno Hispanoamérica (DA), uno México (DEM) y tres Colombia (NDCol, BDC y Dicol). Tanto los diccionarios fraseológicos (DFDEA y DiLEA) como los combinatorios (Redes y Práctico) abarcan unidades léxicas del español de Iberoamérica.

Tabla 2

Diccionarios consultados como fuente del CorlexCo. Elaboración propia.

Tipo de diccionario	Diccionarios consultados
Generales	Diccionario de la lengua española (RAE-Asale, 23ª edición, 2020) DLE Diccionario de uso del español (Gredos, 2016) DUE Diccionario Clave. Diccionario de uso del español actual (Ediciones SM, 2010) Clave
Escolares	Diccionario didáctico escolar DIDEC. Primaria (Ediciones SM, 2013) Diccionario didáctico escolar DIDEC. Secundaria (Ediciones SM, 2013)
Dialectales	Nuevo diccionario de colombianismos (ICC, 1993) NDCol Diccionario de americanismos (Asale, 2010) DA Breve diccionario de colombianismos (ACL, 2012) BDC Diccionario de variantes del español (Capdevila, 2015) DVE Diccionario de colombianismos (ICC, 2018) Dicol Diccionario del español de México (Colmex, 2019) DEM
Fraseológicos	Diccionario de locuciones verbales para la enseñanza del español (Arco/Libros, 2002) Diccionario de locuciones adverbiales para la enseñanza del español (Arco/Libros, 2005) Diccionario de locuciones nominales, adjetivas y pronominales para la enseñanza del español (Arco/Libros, 2008) Diccionario fraseológico documentado del español actual (JdeJ Editores, 2018) DFDEA Diccionario de locuciones idiomáticas del español actual DiLEA (UAH y UCA, 2019)
Combinatorios	Diccionario combinatorio del español contemporáneo. Redes (Ediciones SM, 2005) Diccionario combinatorio práctico del español contemporáneo (Ediciones SM, 2011)

(2) Repertorios léxicos del español de Colombia

La Tabla 3 muestra la distribución geográfica de los 242 repertorios léxicos consultados como fuente del *CorlexCo* y corrobora el hecho de que Colombia es un país de regiones consolidadas por sus hablas. Las referencias de las publicaciones pueden consultarse en el botón *Metodología* del sitio web del corpus. Aunque algunos de los repertorios llevan la denominación ‘diccionario’, ‘lexicón’ o ‘glosario’, en realidad, integran principios lexicográficos rudimentarios.

Tabla 3

Distribución geográfica de los repertorios léxicos del español de Colombia consultados como fuente del CorlexCo. Elaboración propia.

Regiones	Departamentos	No. de fuentes	% de fuentes
Andina	Antioquia (51), Cundinamarca (15), Norte de Santander (9), Caldas (7), Nariño (7), Quindío (6), Santander (6), Boyacá (4), Huila (4), Tolima (4), Cauca (2), Risaralda (1)	116	48 %
Todas	-	68	28 %
Caribe	Todos (12), Atlántico (6), Córdoba (6), Guajira (5), Bolívar (2), Cesar (2), Sucre (1), Magdalena (-)	34	14 %
Pacífico	Valle del Cauca (12), Chocó (4), todos (3)	19	7,9 %
Orinoquia	Meta (1), Casanare (1), todos (1), Arauca (-), Vichada (-)	3	1,2 %
Amazonia	Caquetá (1), Vaupés (1), Amazonas (-), Guainía (-), Guaviare (-), Putumayo (-)	2	0,9 %
Insular	-	-	-
		242	100 %

Como se evidencia, el número de repertorios de la región Andina (116; 48 %) es muy superior al de aquellos que intentan abarcar el territorio nacional (68; 28 %); y el de Antioquia (51; 21 %) es mucho mayor que el de los demás departamentos. Vale la pena aclarar que diecisiete (17) de estas publicaciones llevan en el título la denominación ‘paisa’, que incluye a personas originarias de Antioquia, Caldas, Quindío, Risaralda, norte del Tolima, y norte y occidente del Valle. Destacamos que las identidades culturales del Caribe y el Pacífico se reflejan en la existencia de repertorios léxicos regionales (“todos”). Infortunadamente, continúan existiendo muy pocas fuentes sobre las hablas de la Orinoquia y el Amazonas, como lo había advertido hace cinco décadas el *Atlas Lingüístico-Etnográfico de Colombia ALEC* (ICC, 1982-1983). No hemos hallado ninguna fuente sobre el léxico de la región insular (archipiélago de San Andrés y Providencia en el mar Caribe, e islas Malpelo y Gorgona en el océano Pacífico).

Dentro de los repertorios léxicos, hemos incluido fuentes digitales, como blogs y grupos de Facebook, pues consideramos que la aparición de las combinaciones léxicas en memes, ilustraciones e interacciones en grupos públicos se constituye en indicio de su frecuencia de uso y actualidad. La Figura 5 muestra ilustraciones de comparaciones estereotipadas con alimentos usadas en Colombia, creadas por el australiano Eddie White, compartidas en sus cuentas de Facebook e Instagram y recopiladas en su libro *Caldo de ojo* (2023). Interpretamos su interés por ilustrar estas comparaciones como una necesidad de comprender el contenido lingüístico e idiosincrásico que aportan al aprendizaje del español y a la asimilación de la cultura colombiana por parte de los extranjeros.

Figura 5

Ilustraciones de comparaciones estereotipadas con alimentos. Eddie White: <https://www.facebook.com/eddiewhiteart/photos> by



La Figura 6 muestra una interacción del grupo de Facebook *CULTURA COSTEÑA: palabras-dichos-costumb[r]es y creencias* (Martínez Montenegro, 2018-), en la que el administrador propone la pauta sintáctica “más duro que...”, para que los seguidores la completen y comenten. Esta fuente nos ha resultado especialmente útil en la recopilación de variantes léxicas y regionales de las combinaciones léxicas.

Figura 6

Interacción sobre comparaciones estereotipadas con la pauta sintáctica “más duro que...”.
CULTURA COSTEÑA: palabras-dichos-costumbres y creencias: <https://www.facebook.com/laculturadelacosta>



(3) Investigaciones académicas sobre el español de Colombia

Más de ciento cincuenta investigaciones académicas sobre el español de Colombia se constituyen en otra fuente de consulta relevante del *CorlexCo*. Dentro de estas se destacan artículos publicados en revistas académicas, capítulos de

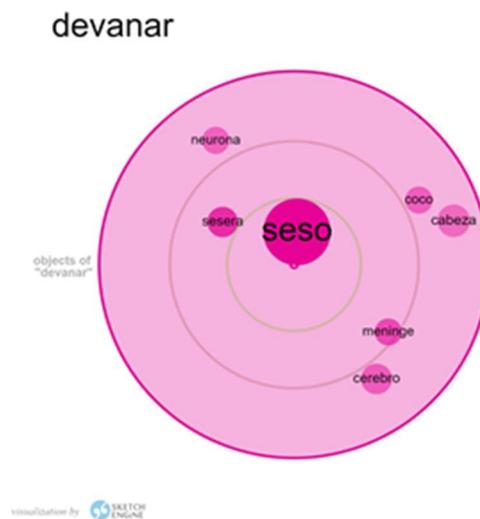
libro y libros resultado de investigación. Además, disponemos, en menor medida, de artículos en otras publicaciones seriadas (boletines, anales); de trabajos de grado de pregrado, maestría y doctorado; y de memorias de eventos académicos. Estas referencias también pueden consultarse en el botón *Metodología* del sitio web del corpus.

(4) Corpus lingüísticos

Con el fin de establecer la frecuencia de uso y la extensión de uso de las combinaciones léxicas registradas en el *CorlexCo*, y de hallar otras combinaciones, consultamos dos corpus lingüísticos: *CORPES XXI* de la RAE (<https://apps2.rae.es/corpes/>) y *Spanish Web 2018 (esTenTen18)* de *Sketch Engine* (<https://www.sketchengine.eu/>). Teniendo en cuenta la naturaleza combinatoria del *CorlexCo*, *Sketch Engine* nos ha resultado especialmente útil por sus herramientas de coocurrencias léxicas (concordancias y colocaciones). La Figura 7 muestra la frecuencia y la tipicidad de la combinación léxica ‘devanar + lexemas somáticos’ tomada de *Sketch Engine*. Solamente la locución ‘devanarse los sesos’ aparece registrada en diccionarios (DLE, DUE, DFDEA, DiLEA, DEM), mientras que en Colombia usamos también ‘devanarse el cerebro’, ‘devanarse el coco’, ‘devanarse la cabeza’ y ‘devanarse las neuronas’, con lo cual constatamos el escaso registro de las variantes léxicas en todo tipo de obras lexicográficas (generales, dialectales y fraseológicas). Como explicamos en Nieto Martín (2024a), ‘sesera’ y ‘meninge’ no son componentes de las locuciones verbales somáticas usadas en Colombia.

Figura 7

Tipicidad y frecuencia de ‘devanar’ + lexemas somáticos. Sketch Engine.



2.3 Anotación

Cada fase anual del corpus se desarrolla en cuatro etapas: 1) recopilación de las combinaciones léxicas, mediante la revisión de las fuentes de consulta; 2) lematización de las combinaciones léxicas; 3) descripción lingüística de las

combinaciones léxicas, según los metadatos establecidos; y 4) carga de las combinaciones léxicas y sus metadatos en la plataforma del *CorlexCo* en CLICC. Las tres primeras etapas se registran de manera manual en un archivo de Excel, como se muestra en la Figura 8.

Figura 8

Etapas de recopilación y descripción lingüística de las combinaciones léxicas del *CorlexCo*.

En este sentido, en la Tabla 4 presentamos el inventario de metadatos usados para la anotación de las combinaciones léxicas recopiladas en el *CorlexCo*. Hay dos metadatos que no proporcionamos (ejemplos con estructura sintagmática 1 y 2), debido a que, en fases futuras, CLICC los generará automáticamente a partir de los datos de todos los corpus disponibles.

Tabla 4

Inventario de metadatos usados para la anotación de las combinaciones léxicas en el *CorlexCo*.

identificador: código que identifica la combinación léxica en el corpus
tipo de combinación léxica: locución verbal / colocación Nsubj.+V / colocación V+Ncd / construcción con verbo soporte / locución adverbial / colocación V+loc. adv. / colocación V+adv. / colocación adv.+adj./part. / locución adjetiva / compuesto sintáctico
combinación léxica: lematizada en su forma básica
unidad léxica plena: solo aplica para construcciones con verbo soporte (ej., <i>informar</i> < <i>dar una información</i>)
significado: solo aplica para combinaciones léxicas con grado de opacidad medio y alto
grado de opacidad: bajo (transparente) / medio (más o menos transparente) / alto (opaco)
nivel de español: básico bajo / básico alto / intermedio bajo / intermedio alto / avanzado bajo / avanzado alto / superior
lengua de origen: español / extranjerismo / indigenismo / latinismo
extensión de uso: español de Colombia / español de América / español general
registro de uso: informal / no marcado / formal
restricción de uso: ninguna / vulgar / despectivo / eufemístico



frecuencia de uso: muy baja / baja / media / alta / muy alta

rasgos de formación léxica: diminutivo / aumentativo / compuesto / parasíntesis (ej., *como una sedita, al carisellazo*)

rasgos sintácticos: coordinación / disyunción / pronombre enclítico / construcción comparativa

unidad léxica 1-10: registra hasta diez componentes en el orden de aparición en la combinación léxica. No se incluyen artículos, preposiciones ni conjunciones.

estructura sintagmática 1: categorías gramaticales y argumentos sintácticos básicos

ejemplo con estructura sintagmática 1: ejemplo en contexto (En fases futuras, CLICC generará automáticamente este dato)

estructura sintagmática 2: categorías gramaticales y argumentos sintácticos extendidos

ejemplo con estructura sintagmática 2: ejemplo en contexto (En fases futuras, CLICC generará automáticamente este dato)

formas nominales 1: solo aplica para las construcciones verbales (ej., *la transmisión de una enfermedad* < *transmitir una enfermedad*)

formas nominales 2: solo aplica para las construcciones verbales (ej., *la transmisión de una enfermedad* < *una enfermedad transmitida*)

otras combinaciones 1: solo aplica para combinaciones léxicas que pueden pertenecer a distintos tipos: colocación verbal/locución verbal, locución adverbial/locución adjetiva, etc. (ej., *meter un gol*: marcarlo [colocación V+N] / *meterle un gol a alguien*: obtener un triunfo sobre él, a veces con engaño [locución adverbial])

otras combinaciones 2: Idem

clase semántica 1: afirmación / cantidad / duda / frecuencia / lugar / modo / negación / orden / tiempo

clase semántica 2: afirmación, cantidad (exceso/abundancia/escasez, mayoreo/menudeo), duda (probabilidad/improbabilidad), frecuencia (nulidad/ocasionalidad/regularidad), lugar (cercanía/lejanía, orientación, trayectoria/permanencia, ubicación), modo (actualización/desactualización, ánimo/desánimo, ascenso/descenso, bicolor/multicolor, bien/regular/mal, bienestar/malestar, causalidad, certeza, confianza/desconfianza, concesión, constancia/inconstancia, continuidad/discontinuidad, cuidado/descuido, cumplimiento/incumplimiento, dedicación, desnudez/abrigo, equilibrio/desequilibrio, exactitud/inexactitud, exceso/abundancia/escasez, facilidad/dificultad, forma, fortuna/infortunio, hacia la luz/contra la luz, honor/deshonor, individual/grupal, inhibición/desinhibición, inmediatez/mediatez, inocencia/malicia, instrumento, intensidad, justificación/sin justificación, lealtad/deslealtad, libertad/restricción, lógica/absurdidad, medio, movilidad/inmovilidad, obediencia/desobediencia, osadía/temor, orden/desorden, posición precaución/sin precaución, previsión/imprevisión, privacidad/publicidad, profundidad/superficialidad, puntualidad/impuntualidad, rapidez/lentitud, relevancia/irrelevancia, satisfacción/insatisfacción, sinceridad/engaño, sorpresa, terminación/no terminación, totalidad/parcialidad/progresión, tranquilidad/intranquilidad, triunfo/fracaso, unicidad/tipicidad, vanidad/humildad, violencia, voluntad/involuntariedad/obligación, voracidad/inapetencia), negación (rotundez), orden (prioridad/no prioridad/alternancia), tiempo (actualidad, cercanía/lejanía, permanencia, determinación/indeterminación, inicio/finalización, inmediatez/mediatez, prontitud/tardanza, puntualidad/impuntualidad, simultaneidad/sucesión)

noción 1: individuo: dimensión afectiva / individuo: dimensión cognitiva / individuo: dimensión comunicativa / individuo: dimensión física / individuo: dimensión social / naturaleza / sociedad / universo

noción 2: actividad mental / alimentación / ciclo de vida / ciencia y tecnología / comunicación / creencias y valores / cuerpo / cultura / economía / elemento de la naturaleza / embarazo / emoción / fauna / fenómeno astronómico / fenómeno atmosférico / fenómeno biológico / fenómeno físico / fenómeno geológico / fenómeno hidrológico / fenómeno meteorológico / fenómeno químico / flora / gobierno / gobierno y política / identidad colectiva / interacción social / lengua / ley y justicia / naturaleza / ocio / percepción / prácticas sociales / salud / servicios / sistema excretor / suerte / transporte / vello corporal / vestuario

noción 3: acidez / alimentos / armamento / armonía / bebidas / brillo / calzado / ciclo de vida / circunstancia desfavorable / circunstancia desfavorable / clima / combustión / comportamiento animal / comunicación / conflicto / conflicto social / conocimiento / construcción / creatividad / desastre natural / difracción / dureza / electricidad / emoción / estación del año / estado de la materia / estética / estrella / explosivos / fenómeno astrológico / fenómeno astronómico / fenómeno físico / fenómeno hidrológico / finanzas / fuego / función vital / gestión del conocimiento / iluminación / industria y comercio / informática / infraestructura / intensidad / interacción social / investigación / juegos / juegos de azar / legislación / lengua / luminosidad / maquinaria /

mar / materiales tecnológicos / medicamentos / memoria / moneda / movilidad / oxidación / pelo / petróleo / piel / producto medioambiental / productos, bienes y servicios / propagación / relaciones personales / relaciones públicas / resonancia / reverberación / río / ropa / salud / seguridad / servicios / servicios domiciliarios / sistema circulatorio / sistema digestivo / sistema fonador / sistema muscular / sistema óseo / sistema respiratorio / sensorial urinario / sistemas de medición / solidificación / suerte / tecnología / telecomunicaciones / tiempo / trabajo / tradición / transporte acuático / transporte aéreo / transporte terrestre / vello facial / vía láctea / volcán

noción 4: accidente / arma de fuego / ave / bienes / cantidad / cicatrización / comunicación / conflicto / defensa / emoción / erupción cutánea / funcionamiento / iluminación / incidente / infección / infraestructura / insecto / inversión / lesión cutánea / lluvia / locomoción / moneda / movilidad / oído / olfato / pez / picazón / pigmentación / presión / productos, bienes y servicios / reptil / resonancia / roca / supuración / tierra

noción 5: anfibio / ave / cerámica / cicatrización / clima / combustión / dinero / emoción / erupción cutánea / fibra textil / infección / insecto / inversión / lesión cutánea / lluvia / madera / mamífero / ortografía / pigmentación / reptil / supuración / tierra / viento

noción 6: puede corresponder a alguna de las incluidas en las anteriores nociones

variantes: otras combinaciones léxicas registradas en el corpus con el mismo significado, con modificaciones fonéticas (ej. *por si acaso* > *por si acas*), morfológicas (*hacer un favor* > *hacer un fa*), léxicas (por motivos diafásicos, ej., *como una patada en el culo/rabo/trasero*), en la preposición (ej., *a/en blanco y negro*), en la conjunción (ej., *a cara y sello/a cara o sello*) o por adición/eliminación de componentes (ej., *a la de Dios/a la de Dios y a la Virgen*)

parasinónimos: otras combinaciones léxicas registradas en el corpus con el mismo significado, en las que varían uno o varios componentes por otro(s) perteneciente(s) al mismo campo léxico (ej., *echar/botar/tirar la casa por la ventana*)

geosinónimos: otras combinaciones léxicas registradas en el corpus con el mismo significado, con variación geográfica (ej., *voltear la arepa* < *voltear la tortilla*)

sinónimos: otras combinaciones léxicas registradas en el corpus con el mismo significado, sin ser variantes, parasinónimas ni geosinónimas (ej. *a la bulla de los cocos/a la bulla de los tarros, tomar el pelo/mamar gallo*)

antónimos: otras combinaciones léxicas registradas en el corpus con el significado contrario (ej., *a gran escala/a pequeña escala*)

fuentes de consulta: (ej. DLE, DUE, Clave, DFDEA, DiLEA, Práctico, Redes, DA, DEM, NDCol, DiCol, Google.com.co, etc.)

De acuerdo con la naturaleza de los metadatos, podemos agruparlos en cuatro categorías:

- 1) *identificación de la combinación léxica:* identificador, tipo de combinación léxica, combinación léxica.
- 2) *descripción lingüística de la combinación léxica:* semántica (significado, grado de opacidad, clases semánticas 1 y 2, nociones 1-6, parasinónimos, sinónimos, antónimos), etimológica (lengua de origen), morfológica (rasgos de formación léxica), sintáctica (rasgos sintácticos, estructuras sintagmáticas 1 y 2, ejemplos con estructura sintagmática 1 y 2), léxica (unidad léxica plena, unidades léxicas 1-10, formas nominales 1 y 2, otras combinaciones 1 y 2), dialectal (extensión de uso, geosinónimos), diafásica (registro de uso, restricción de uso).
- 3) *gradación para la enseñanza de la combinación léxica:* nivel de español.
- 4) *referencias de la combinación léxica:* fuentes de consulta.

La Tabla 5 muestra la anotación de la colocación verbonominal ‘el tráfico circula’ registrada en la ficha de metadatos descargable.

Tabla 5

Ejemplo de anotación de la combinación léxica 'el tráfico circula' en el CorlexCo.

identificador: CLEC_CL_NV_0208
tipo de combinación léxica: colocación Nsubj.+V
combinación léxica: el tráfico circula
unidad léxica plena: N/A
significado: N/A
grado de opacidad: bajo
nivel de español: intermedio alto
lengua de origen: español
extensión de uso: español general
registro de uso: no marcado
restricción de uso: N/A
frecuencia de uso: media
rasgos de formación léxica: N/A
rasgos sintácticos: N/A
unidad léxica 1: tráfico
unidad léxica 2: circular
unidad léxica 3-10: N/A
estructura sintagmática 1: Dsg + Nsubj. + V
ejemplo con estructura sintagmática 1: N/A
estructura sintagmática 2: N/A
ejemplo con estructura sintagmática 2: N/A
formas nominales 1: la circulación del tráfico
formas nominales 2: N/A
otras combinaciones 1: N/A
otras combinaciones 2: N/A
clase semántica 1: N/A
clase semántica 2: N/A
noción 1: Sociedad
noción 2: Transporte
noción 3: Movilidad
noción 4: N/A
noción 5: N/A
noción 6: N/A
variantes: el tránsito circula, el tránsito corre, el tránsito fluye, el tránsito se mueve
parasinónimos: el tráfico corre, el tráfico fluye, el tráfico se mueve
geosinónimos: N/A
sinónimos: N/A
antónimos: el tráfico se congestiona, el tráfico se estanca, el tráfico se paraliza, el tráfico se tranca, el tránsito se congestiona, el tránsito se estanca, el tránsito se paraliza, el tránsito se tranca
fuentes de consulta: Redes, Google.com.co

Para asignar el nivel de enseñanza a una unidad léxica, en Nieto Martín (2022a), proponemos tomar como criterios iniciales sus características lingüísticas: frecuencia de uso, grado de opacidad, estructura sintagmática y número de valencias (sujeto y complementos), la extensión de uso, el registro de uso y las restricciones de uso. De esta manera, deberíamos priorizar la enseñanza de combinaciones léxicas con frecuencias de uso alta y media, con mayor transparencia, con estructuras sintácticas simples y menor número de valencias, de uso en el español general, con registro de lengua no marcado y menores restricciones de uso.

3. Aplicación del *CorlexCo* a la investigación sobre español LE y L2

Los primeros beneficiarios del *CorlexCo* han sido los estudiantes de la Maestría en Enseñanza de Español como Lengua Extranjera y Segunda Lengua del ICC que han desarrollado sus trabajos de grado sobre enseñanza del léxico, tomando como base los datos recopilados en el corpus y aportando nuevos datos. Asimismo, la participación de los maestrantes, como auxiliares de investigación del proyecto, promueve su formación investigativa en lexicología y fraseodidáctica. A continuación, presentamos un resumen de los seis trabajos de grado derivados, hasta la fecha, del *CorlexCo*.

- *Inventario léxico para la enseñanza de español de negocios en América. Niveles A1, A2 y B1* (Urrego Jiménez, 2020). Propone un repertorio de 3412 unidades léxicas (univerbales y pluriverbales). La selección objetiva del léxico se logra a partir de la revisión de estudios previos y la búsqueda en corpus lingüísticos; el contraste con manuales de enseñanza; la consideración de la sinonimia, la antonimia y la polisemia, y la especificación de la transparencia u opacidad.
- *Locuciones verbales con léxico de comida más frecuentes en el español de Colombia. Propuesta de gradación para su enseñanza en el aula de ELE* (Umaña Buitrago, 2021). Propone la gradación de 96 locuciones, para niveles A2-C2, basándose en tres criterios: la frecuencia de uso, la extensión de uso y la motivación semántica.
- *Propuesta didáctica para el desarrollo de la comprensión e interpretación de las expresiones humorísticas usadas en Bogotá en aprendientes de ELE* (Hernández Quezada, 2022). Presenta el proceso de diseño y pilotaje (con aprendientes extranjeros) de una guía sobre comparaciones estereotipadas con núcleos verbal y adjetivo, y propone recomendaciones para su enseñanza en la clase de ELE.
- *Análisis del tratamiento didáctico de unidades fraseológicas somáticas en cuatro manuales de ELE de nivel intermedio B1* (Castro Santa, 2022). Examina los tipos de presentación (explícita o implícita) y las actividades de práctica de las UF en los cuatro manuales de ELE más usados para la enseñanza de ELE en Colombia.
- *Propuesta de adaptación del manual «Lola y Leo 2» al contexto hispanoamericano* (Montoya Vargas, 2024). Presenta las modificaciones lingüísticas y socioculturales que deberían incorporarse en el manual para su uso en la enseñanza de español en México y Colombia.
- *Las comparaciones estereotipadas con animales y alimentos en el español de Colombia y el chino mandarín: recomendaciones para su enseñanza a sinohablantes* (Zhang, 2024). A partir de la experiencia de pilotaje del material *Más fresco / a que una lechuga. Propuesta didáctica para la enseñanza de comparaciones estereotipadas en la clase de ELE/L2* (Nieto y Umaña, 2024) con aprendientes sinohablantes, contrasta las comparaciones usadas en los dos países y proporciona recomendaciones para facilitar su comprensión lingüística y sociocultural.

En el siguiente apartado presentamos un caso concreto de aplicación los datos registrados en el *CorlexCo* en la investigación del español como lengua extranjera. Se trata, en particular, de investigar sobre de la enseñanza de las comparaciones estereotipadas con núcleo adjetivo usadas en Colombia.

3.1 Aplicación del *CorlexCo* en la investigación sobre la enseñanza de comparaciones estereotipadas

Durante la quinta fase de desarrollo del *CorlexCo* (2023), recopilamos 1727 locuciones adjetivas, de las cuales 545 corresponden a comparaciones estereotipadas. La Tabla 6 muestra los campos léxicos presentes en las comparaciones, cuya frecuencia difiere del número total de estas, porque algunas contienen dos (ej., *más perdido que Adán* [persona] *el Día de la Madre* [evento]) y hasta tres referentes (ej., *más encartado/a que gallina* [fauna] *criando* [acción] *patos* [fauna]).

Tabla 6

Campos léxicos de las comparaciones estereotipadas recopiladas en el CorlexCo. Elaboración propia.

Campos léxicos	Frecuencia de aparición	% de aparición
Naturaleza (fauna: 124, otros: 55)	179	23,8
Alimentos	152	20
Objetos	128	17
Personas	102	13,5
Cuerpo humano y animal	36	4,7
Acciones	35	4,6
Eventos	35	4,6
Lugares	26	3,4
Edificaciones	15	2
Unidades de medida	12	1,6
Otros	7	0,9
Medios de transporte	6	0,8
Partes del día	5	0,7
Música	4	0,5
Medicamentos	3	0,4
Colores	2	0,3
Comodines	2	0,3
Divinidades	2	0,3
Números	2	0,3
Sensaciones	2	0,3
Total	755	100 %

Como vemos, los campos léxicos con mayor frecuencia de aparición son los alimentos (152; 20 %), la fauna (124; 16,4%), los objetos (128; 17 %) y las personas (102, 13,5 %). A partir de estos resultados, elegimos las categorías y las comparaciones más frecuentes para la elaboración del material *Más fresco/a que una lechuga. Propuesta didáctica para la enseñanza de comparaciones estereotipadas en la clase de ELE/L2* (Nieto y Umaña, 2024), como se muestra en la Tabla 7. Si bien las comparaciones pueden adoptar tres estructuras sintácticas: *como X* (ej., *como un tomate*), *(tan) ADJ como X* (ej., *[tan] rojo como un tomate*) y *más ADJ que X* (ej., *más rojo que un tomate*), la tercera es la más usada.

Tabla 7

Comparaciones estereotipadas incluidas en «Más fresco/a que una lechuga».

Propuesta didáctica para la enseñanza de comparaciones estereotipadas en la clase de ELE/L2». Elaboración propia.

Comparaciones con alimentos de consumo general

como un tomate, (tan) rojo/a como un tomate, más rojo/a que un tomate

como el agua, (tan) claro/a como el agua, más claro/a que el agua

como una lechuga, (tan) fresco/a como una lechuga, más fresco/a que una lechuga

como el pan, (tan) bueno/a como el pan, más bueno/a que el pan

Comparaciones con alimentos consumidos en Colombia

como un tamal, (tan) amarrado/a como un tamal, más amarrado/a que un tamal

como un kumis, (tan) preparado/a como un kumis, más preparado/a que un kumis

como la panela, (tan) viejo/a como la panela, más viejo/a que la panela

Comparaciones con animales

como una cabra, (tan) loco/a como una cabra, más loco/a que una cabra

como una mula, (tan) terco/a como una mula, más terco/a que una mula

como una tortuga, (tan) lento/a como una tortuga, más lento/a que una tortuga

como un camarón, (tan) rojo/a como un camarón, más rojo/a que un camarón

como una lombriz, (tan) feliz como una lombriz, más feliz que una lombriz

Comparaciones de situaciones con animales

como un mico en un bonsái, (tan) aburrido/a como un mico en un bonsái, más aburrido/a que un mico en un bonsái
como un caballo en un balcón, (tan) aburrido/a como un caballo en un balcón, más aburrido/a que un caballo en un balcón

como marrano estrenando lazo, (tan) contento/a como marrano estrenando lazo, más contento/a que marrano estrenando lazo

como la piraña mueca, (tan) de malas como la piraña mueca, más de malas que la piraña mueca

como gallina criando patos, (tan) enredado/a como gallina criando patos, más enredado/a que gallina criando patos

como un perro a cuadros, (tan) raro/a como un perro a cuadros, más raro/a que un perro a cuadros

Comparaciones con objetos

como un carro por debajo, (tan) feo/a como un carro por debajo, más feo/a que un carro por debajo

como arbolito de Navidad, (tan) prendido/a como arbolito de Navidad, más prendido/a que arbolito de Navidad

como una tortuga, (tan) lento/a como una tortuga, más lento/a que una tortuga

como un hongo, (tan) solo/a como un hongo, más solo/a que un hongo

como un Alka-Seltzer, (tan) saludable como un Alka-Seltzer, más saludable que un Alka-Seltzer

como un brasier, (tan) mentiroso/a como un brasier, más mentiroso/a que un brasier

como una moneda de cuero, (tan) falso/a como una moneda de cuero, más falso/a que una moneda de cuero

Comparaciones con personajes bíblicos

como Adán y Eva el Día de la Madre, (tan) perdido/a como Adán y Eva el Día de la Madre, más perdido/a que Adán y Eva el Día de la Madre

como Caín, (tan) malo/a como Caín, más malo/a que Caín

como Matusalén, (tan) (viejo/a) como Matusalén, más viejo/a que Matusalén

como Judas, (tan) falso/a como Judas, más falso/a que Judas

Adicionalmente, incluimos la categoría *Comparaciones con referentes colombianos*, cuyos adjetivos son gentilicios nacionales (ver Tabla 8). Hemos hallado la mayoría de estas locuciones en el dominio <http://www.google.com.co> y no en los repertorios léxicos.

Tabla 8

Comparaciones estereotipadas «más ADJ que X» con referentes colombianos. CorlexCo.

Comparación estereotipada	Fuente de consulta
Lugares más samario que El Morro más cartagenero que el cerro de La Popa más bogotano que el cerro de Monserrate más pastuso que el Galeras	Red Red Red Red
Platos típicos más tolimense que el tamal más cachaco/a que la changua más cachaco/a que el ajiaco santafereño más barranquillero/a que el arroz de lisa más pastuso/a que el cuy con papas	Red Red Red Caribeñol, Dichos costeños... Red
Arepas más paisa que la arepa más costeño/a que la arepa de huevo más santandereano/a que la arepa amarilla	Eddie White Cultura Costeña Red
Amasijos más caleño/a que el pandebono más opita que las achiras más manizaleño/a que los dedos [de queso] de la 23 más costeño/a que el bollo de yuca	Red Red Red Cultura Costeña
Otros más boyacense que la papa más pereirano/a que las papitas de la Lucerna más santandereano/a que la hormiga culona más costeño/a que el suero más cartagenero/a que la Kola Román más chochoano/a que el Platino	Red Red Red Cultura Costeña Red Red
Otricos más caleño/a que la salsa más costeño/a que el vallenato más llanero/a que el joropo más boyacense que la ruana	Red Red Red Red

Conclusiones

El Corpus léxico del español de Colombia *CorlexCo* recopila, con criterio integral, combinaciones léxicas usadas en Colombia, y se encuentra a disposición de la comunidad académica a través de la plataforma CLICC del Instituto Caro y Cuervo. Proporciona información lingüística (ortográfica, fonética, morfológica, sintáctica, semántica, pragmática) y sociolingüística de las combinaciones léxicas y la fraseología del español de Colombia; y se constituye en una herramienta que respalda los procesos investigativos de enseñanza-aprendizaje de español como lengua extranjera y segunda lengua (para aprendientes indígenas y sordos) en el país, al tener aplicaciones en la selección léxica objetiva para la planeación curricular, el diseño de materiales didácticos y los procesos de evaluación de español.

De esta manera, procuramos que el *CorlexCo* se convierta en una herramienta de referencia para orientar la selección léxica basada en las necesidades de enseñanza-aprendizaje del español en/de Colombia –siguiendo principios didácticos, fraseológicos y lexicográficos–, según diversos criterios, como el nivel de lengua (básico, intermedio, avanzado, superior), el tipo de combinación léxica (compuestos, colocaciones, construcciones con verbo soporte, locuciones, paremias, fórmulas rutinarias), la extensión de uso (español general, español de América, español de Colombia), la frecuencia de uso (alta, media, baja), el fin específico (negocios, jurídico, médico, turismo, etc.), entre otros.

Como acciones futuras de mejora, además de la alimentación periódica del corpus, se prevé habilitar la búsqueda por metadatos y por combinaciones, ya que actualmente solo está disponible por palabra. A largo plazo, el *CorlexCo* se propone convertirse en un diccionario combinatorio, tras el desarrollo del sistema gestor de diccionarios del ICC (LEXICC).

Referencias

- Ariza Herrera, Emma; Corpas, Jaime; García, Eva; Garmendia, Agustín; Nieto Martín, Gloria Viviana; Molina, Guillermo; Soriano, Carmen (2018). *Aula América*. Difusión.
- Ariza Herrera, Emma; Molina Morales, Guillermo; Nieto Martín, Gloria Viviana (2019). Adaptación de materiales para la enseñanza de español como lengua extranjera en Latinoamérica: el caso de Aula América 1. *Forma y Función*, 1 (32), 101-123. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/77418>
- Ariza Herrera, Emma; Nieto Martín, Gloria Viviana; Cruz Arcila, Ferney (2020). *Empelculados: cine en el aula de español para extranjeros y sordos*. Instituto Caro y Cuervo. <https://spanishincolombia.caroycuervo.gov.co/materiales-didacticos-ele/empelculados-cine-en-la-clase-de-espanol-como-lengua-extranjera-y-segunda-lengua-2020/>
- Asociación de Academias de la Lengua Española (2010). *Diccionario de americanismos*. Santillana. <https://www.asale.org/damer/>
- Bernal Chávez, Julio Alexander; Hincapié Moreno, Diana Alejandra (2018). *Lingüística de corpus*. Instituto Caro y Cuervo. <http://bibliotecadigital.caroycuervo.gov.co/1703/1/Linguistica-de-corpus-2018.pdf>
- Capdevila, Lauro (2015). *Diccionario de variantes del español*. <https://xn--diccionariovariantespaol-4rc.org/>
- Castro Santa, José Félix (2022). *Análisis del tratamiento didáctico de unidades fraseológicas somáticas en cuatro manuales de ELE de nivel intermedio B1* [Trabajo de grado de maestría]. Instituto Caro y Cuervo. <http://bibliotecadigital.caroycuervo.gov.co/id/eprint/1821>
- Ciro, Lirian Astrid (2022). Repertorios lexicográficos en Colombia de los siglos XIX al XXI. En Adriana María Ortiz Correa y Sergio Alonso Lopera Medina (Eds.), *Aportes teóricos e investigativos en lingüística: una mirada desde el contexto colombiano* (pp. 103-119). Peterlang.
- Corpas, Jaime; García, Eva; Garmendia, Agustín (2013). *Aula internacional Nueva edición*. Difusión.
- Corredor Romero, Liliana (2019). *Viaje literario por Barranquilla. Lecturas guiadas para la clase de ELE*. Instituto Caro y Cuervo. <https://spanishincolombia.caroycuervo.gov.co/materiales-didacticos-ele/viaje-literario-por-barranquilla-lecturas-guiadas-para-la-clase-de-ele-2019/>



- Cuervo, Rufino José (1954). *Diccionario de construcción y régimen de la lengua castellana. Tomos I y II*. Instituto Caro y Cuervo.
- Gutiérrez Rubio, Enrique (2019-2020). Materiales didácticos y fraseología en el discurso oral coloquial. *Linred: Lingüística en la Red*, 17, 1-18. https://ebuah.uah.es/dspace/bitstream/handle/10017/48818/materiales_gutierrez_LINRED_2019_17.pdf?sequence=1&isAllowed=y
- Haensch, Günther y Werner, Reinhold (Dir.) (1993). *Nuevo diccionario de colombianismos*. Instituto Caro y Cuervo.
- Hernández Quezada, Mónica Paola (2022). *Propuesta didáctica para el desarrollo de la comprensión e interpretación de las expresiones humorísticas usadas en Bogotá en aprendientes de ELE* [Trabajo de grado de maestría]. Instituto Caro y Cuervo. <http://bibliotecadigital.caroycuervo.gov.co/id/eprint/1822>
- Hincapié Moreno, Diana Alejandra; Nieto Martín, Gloria Viviana; León Wintaco, Luz Dary (2021). *CántELE. Música colombiana para la clase de español como lengua extranjera Volumen 1*. Instituto Caro y Cuervo. <https://spanishincolombia.caroycuervo.gov.co/materiales-didacticos-ele/cantele-musica-colombiana-para-la-clase-de-ele-vol-1-2021/>
- Hincapié Moreno, Diana Alejandra; Nieto Martín, Gloria Viviana; León Wintaco, Luz Dary; Ariza Herrera, Emma (2022). *CántELE. Música colombiana para la clase de español como lengua extranjera Volumen 2*. Instituto Caro y Cuervo. <https://spanishincolombia.caroycuervo.gov.co/materiales-didacticos-ele/cantele-musica-colombiana-para-la-clase-de-ele-vol-2-2022>
- Instituto Caro y Cuervo (1982-1983). *Atlas Lingüístico-Etnográfico de Colombia ALEC*. Editorial Arco.
- Instituto Caro y Cuervo (2018). *Diccionario de colombianismos*. Legis SA.
- Martínez Montenegro, Tomás (2009-). *Cultura Costeña: palabras-dichos-costumbres y creencias*. <https://www.facebook.com/laculturadelacosta>
- Montoya Vargas, Mayra Alejandra (2024). *Propuesta de adaptación del manual «Lola y Leo 2» al contexto colombiano* [Trabajo de grado de maestría]. Instituto Caro y Cuervo. En prensa.
- Nieto Martín, Gloria Viviana (Coord.) (2017). *Hecho en Colombia: cultura colombiana para la clase de ELE*. Instituto Caro y Cuervo. <https://spanishincolombia.caroycuervo.gov.co/materiales-didacticos-ele/hecho-en-colombia-2017/>
- Nieto Martín, Gloria Viviana (2020). Variación léxica de las locuciones verbales de Colombia y España. *Revista Estudios de Lingüística Universidad de Alicante ELUA*, anexo 7, 201-215. <https://rua.ua.es/dspace/handle/10045/109692>
- Nieto Martín, Gloria Viviana (2022a). La motivación semántica como criterio de gradación de las locuciones verbales para su enseñanza en español LE/L2. En Yolanda Pasuy y María Clemencia Silva (Eds.), *Enseñanza, lengua y cultura en ELE* (pp. 69-93). Editorial Universidad de Caldas y Universidad EAN.
- Nieto Martín, Gloria Viviana (2022b). ‘Abrir la boca’ y ‘abrir el pico’: locuciones verbales somáticas usadas en Colombia y España. En Miguel Ibáñez Rodríguez; Carmen Cuéllar Lázaro (Eds.), *De la hipótesis a la tesis: traductología y lingüística aplicada* (pp. 187-208). Editorial Comares. <https://www.comares.com/media/comares/files/toc-139100.pdf>
- Nieto Martín, Gloria Viviana (2024a). Especialización léxica diatópica de las locuciones verbales somáticas de Colombia y España: ¿una cuestión lingüística o idiosincrásica. En Lucía Navarro Brotons; Adelina Gómez González-Jover; Eva Navarro Martínez (Eds.), *De la hipótesis a la tesis: la dimensión cultural de la traducción*. Editorial Comares. <https://www.comares.com/media/comares/files/toc-139100.pdf>



- Nieto Martín, Gloria Viviana (2024b). Variación léxica de las locuciones verbales somáticas del español de Colombia. Diatopía, diafasia y polisemia. *Islas*. En prensa.
- Nieto Martín, Gloria Viviana (2024c). Comparaciones estereotipadas con alimentos en el español de Colombia. Relevancia en la enseñanza de ELE/L2. *LETRAS*. En prensa.
- Nieto Martín, Gloria Viviana; Buitrago Umaña, Diana Rocío (2024). *Más fresco/a que una lechuga. Propuesta didáctica para la enseñanza de comparaciones estereotipadas en la clase de ELE/L2*. Instituto Caro y Cuervo. En prensa.
- Nieto Martín, Gloria Viviana; Labrada Hernández, Geisy (2022). Variación diatópica de las unidades fraseológicas zoomorfas en el español de Cuba y Colombia. *Chakiñan, Revista de Ciencias Sociales y Humanidades*, 18, 38-49. <https://doi.org/10.37135/chk.002.18.02>
- Olimpo de Oliveira, María Eugenia (2006). *Fraseología y enseñanza de español como lengua extranjera*. Universidad de Alcalá. <https://www.educacionfpydeportes.gob.es/mc/redele/biblioteca-virtual/numerosanteriores/2006/memoriamaester/1-semester/olimpio-d.html>
- Ortiz Correa, Adriana María (2022). Fraseografía en Colombia: 67 años entre el *folclor* y la lingüística. Una aproximación al estado de la cuestión. En Adriana María Ortiz Correa y Sergio Alonso Lopera Medina (Eds.), *Aportes teóricos e investigativos en lingüística: una mirada desde el contexto colombiano* (pp. 121-144). Peterlang.
- Penadés Martínez, Inmaculada (1999). *La enseñanza de las unidades fraseológicas*. Arco/Libros.
- Real Academia Española y Asociación de Academias de la Lengua Española (2009-2011). *Nueva gramática de la lengua española*. Espasa. <https://www.rae.es/gram%C3%A1tica/>
- Umaña Buitrago, Diana Rocío (2021). *Locuciones verbales con léxico de comida más frecuentes en el español de Colombia. Propuesta de gradación para su enseñanza en el aula de ELE*. [Trabajo de grado de maestría inédito]. Instituto Caro y Cuervo.
- Urrego Jiménez, Yenny Esmeralda (2020). *Inventario léxico para la enseñanza de español de negocios en América. Niveles A1, A2 y B1* [Trabajo de grado de maestría]. Instituto Caro y Cuervo. <http://bibliotecadigital.caroycuervo.gov.co/id/eprint/1722>
- White, Eddie (2018-). *Eddie White*. <https://www.facebook.com/eddiwhiteart>
- White, Eddie (2023). *Caldo de ojo. 100 dichos y palabras colombianas*. Quillango Editores.
- Zhang, Dandan (2024). *Las comparaciones estereotipadas con animales y alimentos en el español de Colombia y el chino mandarín: recomendaciones para su enseñanza a sinohablantes* [Trabajo de grado de maestría]. Instituto Caro y Cuervo. En prensa.

