Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras

e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**

UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Linguistic Legos: a Description of Spanish Syntagmatic Compounds in Three Mexican Corpora of Lexical Availability Task.

Legos lingüísticos: una descripción de los compuestos sintagmáticos españoles en tres cuerpos mexicanos de la tarea de disponibilidad léxica.

DOI: 10.32870/sincronia.axxvi.n82.35b22

Marco Antonio Pérez Durán
Facultad de Ciencias Sociales y Humanidades / Universidad Autónoma de San Luis Potosí (MÉXICO)
CE: marco.duran@uaslp.mx / **ID ORCID:** 0000-0003-0854-3109

Gabriela Silva Maceda
Facultad de Psicología / Universidad Autónoma de San Luis Potosí (MÉXICO)
CE: gabriela.silva@uaslp.mx / **ID ORCID:** 0000-0002-7313-4777

Recibido: 22/01/2022
Revisado: 08/03/2022
Aprobado: 13/06/2022

Abstract:

Syntagmatic compounds are a type of morphological compound that contributes to the increase of a learner's vocabulary. They are created when two or more lexemes join to form a new word. Considering they could have wide pedagogic implications due to the multiple combination possibilities from a single base, this analysis intends to describe the frequency and classification of syntagmatic compounds as well as their bases in three Mexican corpora of Lexical Availability Task of students from different times, regions and education levels. It also aims to describe the relationship between bases and lexical items to examine whether a few of the bases are needed for many of syntagmatic compounds found in Spanish. Results showed that syntagmatic compound are quite available (from 10% to 31% of total lexical items). Within the classification of these compounds, the most common were the prepositional ones. It was also found that only 50 bases are needed to form

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

more than a third of the compounds found in the lexical availability responses. A list of the top 20 bases is presented for its use among the teaching of Spanish as a second language. It is proposed that since studying a small quantity of words can enable learners to build a great number of compounds, they could be conceptualized as linguistic legos.

Keywords: Syntagmatic compound. Spanish. Lexical item. Base lexical item.

Resumen:

Los compuestos sintagmáticos son un tipo de compuesto morfológico que contribuye al aumento del vocabulario del alumno. Se crean cuando dos o más lexemas se unen para formar una nueva palabra. Considerando que podrían tener amplias implicaciones pedagógicas debido a las múltiples posibilidades de combinación desde una sola base, este análisis pretende describir la frecuencia y clasificación de compuestos sintagmáticos así como sus bases en tres corpus mexicanos de Tarea de Disponibilidad Léxica de estudiantes de diferentes épocas, regiones y niveles educativos. También tiene como objetivo describir la relación entre las bases y los elementos léxicos para examinar si algunas de las bases son necesarias para muchos de los compuestos sintagmáticos que se encuentran en español. Los resultados mostraron que los compuestos sintagmáticos están bastante disponibles (del 10% al 31% del total de elementos léxicos). Dentro de la clasificación de estos compuestos, los más comunes fueron los preposicionales. También se encontró que solo se necesitan 50 bases para formar más de un tercio de los compuestos encontrados en las respuestas de disponibilidad léxica. Se presenta una lista de las 20 bases principales para su uso entre la enseñanza del español como segunda lengua. Se propone que, dado que estudiar una pequeña cantidad de palabras puede permitir a los alumnos construir una gran cantidad de compuestos, podrían conceptualizarse como legos lingüísticos.

Palabras clave: Compuesto sintagmático. Español. Ítem léxico. Ítem léxico base.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Introduction

Within applied linguistics, a productive area of research has been the study of the available lexicon, mostly investigated through the lexical availability task. This specific methodology has been mostly associated to studies of teaching of a second language. Its inception is found in the French studies of the lexicon, as a way of overcoming the limitations of the frequency count. French researchers were most interested in the didactic implications of their methodology for they tried to identify a lexicon brought upon an immediate manner and leveled by difficulty to teach French as a Second Language. In 1953 Michéa was the first to distinguish between 'themed' and 'non-themed' (Michéa, 1953). At the same time, he defined the available lexicon as the set of lexical units that are available for a specific communicative event. These words belong to the participant and are only activated when required in a conversation. This lexicon is termed Lexical Availability (López, 2005).

Lexical availability is defined as the vocabulary likely to appear when the speaker is immersed in a specific communication situation; in other words, it is the vocabulary used to function when speaking about a specific topic (López, 2005). The *available lexicon* is the set of lexical units available for a definite communicative moment. To measure this available lexicon, the *lexical availability task* (Pérez, 2019), is used, which is applied in a specific time frame (for two or three minutes, depending on the methodology used) with a sample size contingent on the study´s purpose. Furthermore, this survey is comprised of usually 16 topics serving as a stimulus to elicit words related to these topics.

For the study of Spanish, Lopez Morales (1973) carried out an adaptation the French methodology to collect real-world community-specific lexicon in San Juan de Puerto Rico, in order to identify and study the students' lexicon and its influences of extralinguistic factors. The methodological framework developed by his work has been the basis for the lexicon research in dialectology, psycholinguistics, ethnolinguistics, sociolinguistics and the teaching of the first or mother language (L1) and second language (L2).

Another set of studies that highlighted the multiple uses of this methodology was the one carried out by Carcedo Gonzalez (1998). He used the lexical availability task to explore the available

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

lexicon in Spanish for 78 Finnish students at their last year of high school. This is the first time that lexical availability studies were used for Second Language Teaching. Carcedo Gonzalez's (1999) next set of studies covered other research lines within SLT, such as: 1) sociolinguistic variation by educational level, and its influence in the learning of Spanish; 2) the analysis of orthographic, phonic, morphologic and lexical errors; 3) contrast studies between different student populations learning Spanish or between different countries; and 4) stratified lexicon analysis from L1 to L2 for didactic purposes.

Regarding the learning of Spanish, most work is found in the studies of Spanish as a Second Language (SSL), while some work has been reported for Spanish as L1. Within the former and using sociolinguistic variables such as educational levels, there are published studies about the lexical availability of German (Hugo, 2003), Polish (Lopez Gonzalez, 2009), Slovac (Kranjc, 2009; Kalan, 2011), Portuguese (Fernández dos Santos, 2014) and Italian (Rubio, 2017) students, learning Spanish in their own countries. Another variant of SSL research study has been carried out with those students from other countries studying Spanish at a Spanish speaking-country, such as the ones by Samper Hernandez (2002), Perez Serrano (2009) and Lopez Casado (2013), commonly known as linguistic immersion contexts.

Within the studies of Spanish as L1, there are lexical availability studies of Spanish speakers in minority contexts or in Spanish-speaking countries. Lexical availability studies in the first have been reported for students in the USA (Verdeses, 2012; Moreno, 2012). Meanwhile, most of the reported research of Spanish as L1, has been carried out in Latin America.

In the literature, most of the lexical availability studies in Mexico is focused in the teaching of Spanish as L1, using different linguistic communities and diverse sociolinguistic variables for contrast analysis, such as the studies by Lopez Chavez (1994, 2004). However, it must be acknowledged that there is work reported about linguistic immersion of Spanish as L2 (Santos, 2015), where the L1 is an indigenous language, in this case, the Cora (naayeri) language.

The answers in the lexical availability task form the available lexicon, which can be nouns (the most frequent), verbal and syntagmatic compounds. Syntagmatic compounds are frequent in

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

lexical availability responses. They are a morphological compound that functions specifically to increase vocabulary, within the formation of words, and appears when two or more lexemes join to form a new word such as *abrelatas* [can opener], *sacacorchos* [corkscrew], *cubrecama* [duvet cover], *ecólogo* [ecologist], *hidrofobia* [hydrophobia], etc. These compounds allow the creation of new lexemes composed of the union or combination of two or more independent lexical units (words, roots, bases) into one (Casado, 2018: 51-59). Structurally, *compounds* are formed by a nucleus, which encompasses the meaning of the compound and a determinant that together generate a new word in the language. These new words can be of two types: a) lexical compounds, also called proper, univerbal or orthographic (i.e. *cuentagotas* [eyedropper]*, pelirrojo* [redhead]*, quitamanchas* [stain remover]); and b) syntagmatic compounds, which are also termed as improper or lexicalized syntagms (i.e. *lucha libre* [wrestling]*, orden del día* [agenda]). The latter are the focus of this paper for their possible practical implications in language teaching: multiple combinations could potentially be formed that could enhance a student's rate of learning.

Even when the published corpora does not list naturally occurring frequencies of each lexical entry, they do offer a window into the students lexical repertoire that could be analyzed for the presence of syntagmatic compounds.

Regardless of how frequent nouns and syntagmatic compound are commonly found in the answers, the analysis of specific grammatical categories has been focused elsewhere. In the literature, irrespective of whether the Spanish studies reported the lexicon in L1 or L2, the grammatical categories which have been analyzed are adjectives (Camarena Ortiz, 2010; Frey, 2007; Guerra, Paredes & Gomez, 2004; Lopez Chávez, 1994; Fuentes, 2014) and verbs (López, 1994). Therefore, a gap in the literature was identified for the organization, classification and analysis of syntagmatic compounds, which have remained, to the best of our knowledge, unexamined. Given the possibility of lexical bases functioning as sources for multiple syntagmatic compounds, the study of this grammatical category could have the most implications for the teaching of Spanish, particularly as L2.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

The present investigation aims to contribute to the field of cross-sectional studies across time in available lexicon in Spanish, alongside the work by Alba (2013) about the variation on Dominican available lexicon, or Hidalgo's (2017) about the variation on SSL students' available lexicon.

The purpose of this paper is to report the frequency and classification of syntagmatic compounds and their bases in students' lexical repertoires, through the lexical availability task appearing in three corpora from different time cross-sections designed to obtain lexical availability responses. This study also aims to describe the relationship between bases and lexical items to examine whether a few of the bases are needed to derive a significant number of syntagmatic compounds.

Before delving into the methods followed to examine syntagmatic compounds, an overview of this topic is presented next.

MERHODS

This section describes the procedure to select the three samples, their characteristics and the analysis plan.

Sample corpus selection procedure

To be able to select appropriate samples for comparison, eight criteria were established: 1) samples must be lexical availability studies in Spanish in Mexico, regardless of the time of data collection; 2) samples selected must have used the same topics for the task; 3) samples must follow the 3 minutes methodology for data collection; 4) studies must have applied a process of orthographic standarization; 5) studies must have applied the same mathematical formula for lexical availability by Lopez Chavez and Strassburger Frias (2000) for obtaining descriptive statistics; 6) studies must include full sample lists for analysis; 7) samples must include at least 100 participants and at most 300; 8) samples must be equally distributed among variables for sample segmentation, such as age, sex and educational level; 9) samples must come from a capital city.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Corpora characteristics

Three corpora were identified that followed the selection criteria from different age samples, cities and times: a primary students' sample from Mexico City in 1993 (López & Madrid Guillén, 1993); a secondary students' sample form Culiacan in the state of Sinaloa in 2014 (Velarde, 2014); and a preparatory students' sample from Zacatecas in 2010 (Perez Duran, n.d.).

Lope Blanch (cited in Butragueño, 2011) characterizes the sample from Mexico City as the "cultured norm" in the country as the reference, the sample from Culiacan belongs to the northwestern coastal zone, and the Zacatecas sample belongs to the Chihuahua variation in the transition zone.

Analysis plan

A mixed – qualitative and quantitative – analysis was carried out for the syntagmatic compounds of the three selected samples. These compounds were identified, classified, ordered and analyzed in order to describe: a) the lexical item frequency distribution by corpus; b) the lexical item frequency distribution by topic and corpus; c) the prevalence of syntagmatic compounds by corpus; d) the prevalence of syntagmatic compounds by topic in the pooled corpus; e) the classification of compounds by composition and by corpus; f) the distribution of compounds by number of words; and g) the description of the most frequent compound bases.

The basis for the syntagmatic compound categorization is Varela's (2009, 2018) classification into three groups: prepositional compounds, yuxtaposed compounds and noun+adjective compounds. Varela's full description is transcribed here in Table 1.

Table 1. Syntagmatic compound classification by Varela (2005, 2018)

Type of syntagmatic compound			Structure	Examples
1) Prepositional compound			N+Prep+N	Patas de gallo [crow's feet]
2) Yuxtaposed compounds	a) Nominals	1. Appositives (without dash)	N+N	Pájaro mosca [Bee hummingbird]
				Palabra clave [key word]

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

	2. Coordinating (with dash)	N+N N+A	Falda-pantalón [skirt-trousers] Salón-comedor [diningroom]
	b) Adjectives (In a coordinating mode)	A+A	Político-económico [Political-economic]
3) Noun+adjective compounds (Show the relationship Nucleus (N) modifier adjective (A) and the modifier's anterior or posterior position)		N+A	Llave inglesa [Stilson wrench] Buena fe [good faith] Hilo musical [musical thread]

For the description of the most frequent compound bases, the twenty most available compound-generating lexical item, or bases, were listed.

Syntagmatic compound and lexical availability

Answers to the lexical availability task are usually of three categories: nouns, verbs and syntagmatic compounds. When comparing different corpora, the position and the frequency of the lexical item are critical for analysis. Lopez Rivero (2008) makes a distinction between lexical item and word: *lexical item* and *word* are generally used as synonyms in everyday language; however, in statistical lexical studies, lexical item refers to each of the units in the lexicon or the different lexical entries in a frequency dictionary. In contrast, the word refers to each one of the basic primary units, perfectly distinguished by typography in an oral or written corpus. Lexical items, therefore, result from standardizing or lemmatizing participants' words in lexical availability studies.

Syntagmatic compounds are relatively common in available published studies. Sanchez-Saus Laserma (2011) states that, within the topic of lexical availability, syntagmatic compounds are syntagms with semantical and referential unity, which function as simple units and are very frequently used, such as *agujero negro* [black hole] and *aire acondicionado* [air conditioning].

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Syntagmatic compounds

Compounding is one of the most productive lexical procedures to form new words from the union of two or more independent lexical units into a single unit (Casado, 2018; Martinez, 2000; Varela, 2005, 2009, 2018, Val, 1999). Morphologically, compounds are classified into proper (or lexical, univerbal, orthographic) and improper (syntagmatic).

Syntagmatic compounds are syntagms whose parts are not orthographically joined but form an independent and cohesive semantic unit, with a new meaning. Val Alvaro (1999) and Garcia-Page (2011) state that syntagmatic compounds have one of the fuzziest boundaries between lexicon, morphology and syntax, since there is no clear frontier to distinguish them from similar structures, such as pseudocompounds, apposition nominal syntagms, locutions or phraseologisms. For the purpose of this study, a syntagmatic compound is a lexical item composed of a single lexical unit, even when its components are separate words, such as *fin de semana* [weekend]. Furthermore, they are syntagms that have acquired a new meaning which cannot be always deduced from the simple union of its components, such as *damas chinas* [Chinese checkers] (Luque, 2004).

Prepositional Compounds

This kind of compounds is represented by the structure N+ Prep+N, such as *puerta de acero* [steel gate], *caballo de mar* [seahorse], *máquina de escribir* [typewriter], *agua de Jamaica* [Jamaica drink], *garrafón de agua* [water container]. It has a single global meaning and a single referent, just as the orthographic compounds. Varela (2009), Pacanigni (2003) and Buenafuentes de la Mata (2007) state that constituents of the syntagmatic compound lack syntactic independence, that is, the syntactic criterion determines the lexical status of the syntagma, since it is impossible to substitute the nuclear constituent. Cruz Piñol (2015) explains this with the example *página de inicio (N+ prep +N)* [initial page] to explain its structure (nucleus *página*) and complement (de inicio). The nucleus cannot be substituted by *archivo* [file] or *plataforma* [platform] in **archivo de inicio* or **plataforma de inicio*. It is not possible either to switch any modifier or determinant between the bases, as in **página del inicio* nor **página del gran inicio.* Likewise, there is no flexibility for interchanging

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

positions between nucleus and subordinate, as in **inicio de página*, **página inicio*. Therefore, a syntagmatic compound acquires its own sentence syntactic function (subject, object, attribute, etc.).

For Valera (2009) this is the test to distinguish between compounds and syntagms, since the constituents of a syntagm still keep their syntactical independence, allowing to receive their own modifiers, as in *café con leche* as a syntagmatic compound, and opposed to *café con [leche fría]* as a syntagm. In addition, Casado Velarde (2018) states that, regardless of whether words maintain their own independence when forming the compound, its meaning is structured by semantic and syntactic cohesion. Cohesion refers to keeping one of the elements of the syntagm as a nucleus, which serves as the basis for the generation of new compounds. For example, within the topic *La escuela, muebles y útiles* [School, furniture and tools], a frequently available answer is *máquina de escribir* [typewriter] appearing in the primary school sample (1993), whose lexical item is *máquina de escribir* [computer]. When comparing lexical items in lexical availability studies, lexical variation is a critical resource in the diachronic development of language. Therefore, lexical variation consists of lexical differences between variants of the same language belonging to conceptual fields for which the speaker has diverse lexical alternatives designating a same extralinguistic reality, as in the case for *computadora* [computer].

Noun and Adjective Compound (N+A)

These compounds show a relationship between a qualifier and a nominal. Its structure is N+A or A+N, with the former being more frequent than the latter: *salon comedor* [dining room], *llave inglesa* [stilson wrench], *damas chinas* [Chinese checkers], etc. Varela (2009) describes this type of syntagmatic compound as very highly lexicalized, since its meaning cannot be deduced from its components.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Yuxtaposed Compound

This type of compound is the one formed by two nouns (N+N), where one functions as a nucleus and the other brings additional information. Varela (2009) classifies the yuxtaposed compound into nominal and adjectival.

Within the nominals, there are two subtypes of compounds: appositives and coordinating. In the appositive compounds, two nouns related by an appositive function, such as in *pez gato* [catfish], *pez lobo* [wolffish] and *pantalón campana* [taper trousers]. The noun in apposition denotes qualities of the noun functioning as nucleus. Meanwhile, for coordinating compounds, the coordinating nouns together signify a new meaning that is lost when the components are separated. Semantically they function as a coordination and the nouns are joined by a hyphen: *falda-pantalón* [*skirt-trousers], *vagón-comedor* [dining car], *poeta-escritor* [poet-writer], etc. In this same classification, there is the adjectival compound (A+A), which follows the same coordinating formula and combines semantically congruent adjectives, belonging to the same conceptual dominion and showing a strong degree of cohesion, reflected in the frequent use of the hyphen for the written form: *político-económico* [*political-economical], *léxico-semántico* [*lexical-semantic], *socioeconómico* [socioeconomic], etc.

In sum, syntagmatic compounds are a type of responses within the Lexical Availability Task that are yet to be examined. Considering they could have wide pedagogic implications due to the multiple combination possibilities from a single base, the aim of this analysis is to report the frequency and classification of syntagmatic compounds as well as their bases and describe the relationship between bases and lexical items to examine whether a few of the bases are needed for a large number of syntagmatic compounds.

Results

It must be noted that the results presented here originate from the published corpora listing unique lexical items (compounds) with their corresponding lexical availability index. In this sense, each

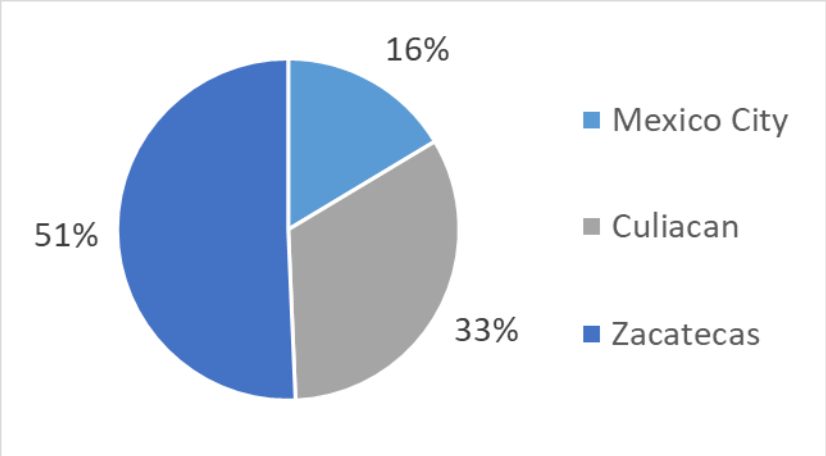
Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

lexical item might be underrepresented from its frequency in the original sample, but these data are not published.

Lexical item frequency distribution by corpus

First, the totality of lexical items by corpus is described. The three corpora contained 16, 479 lexical items, with the smallest, the Mexico City primary corpus, having 2,700; the Culiacan middle-school corpus had 5,428; and the Zacatecas high-school corpus had 8,351 lexical items. The relative frequencies are illustrated in Figure 1.

Figure 1. Lexical items distribution by corpus

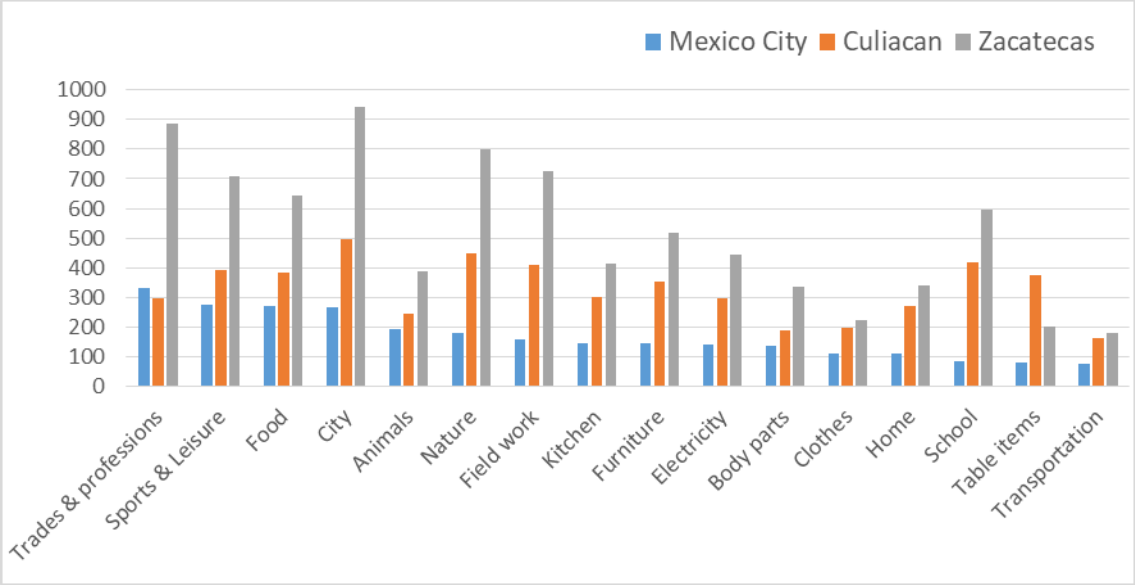


Lexical item frequency distribution by topic and corpus

Next, the total lexical items by corpus and topic are shown in Figure 2, ordered from most productive to least productive in the reference corpus, that is, the Mexico City corpus. Since the Zacatecas corpus is the largest in lexical items, their frequencies are higher in almost all topics.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

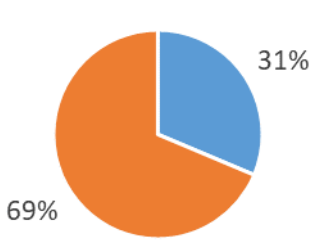
Figure 2. Lexical items by topic and by corpus



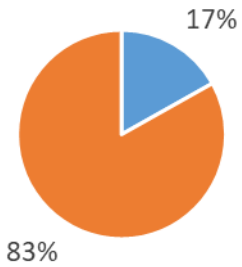
Prevalence of syntagmatic compounds by corpus

To address the question of how many syntagmatic compounds are present in each corpus, Figure 3 shows the percentage of syntagmatic compounds as proportion of lexical items by corpus. There is diversity among the corpora, ranging from 10% in the Zacatecas corpus to 31% in the Mexico City corpus.

Figure 3. Percentage of syntagmatic compounds as proportion of lexical items by corpus

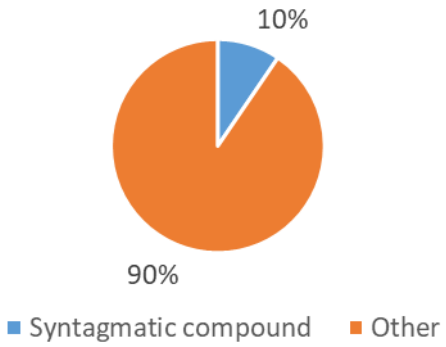


a. Mexico City corpus



b. Culiacan corpus

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES



c. Zacatecas corpus

Individually, the Mexico City generated the most syntagmatic compounds in relative terms, 31% (but 846 syntagmatic compounds in absolute terms); then the Culiacan corpus generated a middle relative frequency of 17% (but the most syntagmatic compounds in absolute terms 916); finally, the Zacatecas corpus contained the least syntagmatic compounds in relative terms, only 10% (and also the least in absolute terms: 795 syntagmatic compounds). Even when the corpora are very different in size, the absolute numbers of syntagmatic compounds are similar.

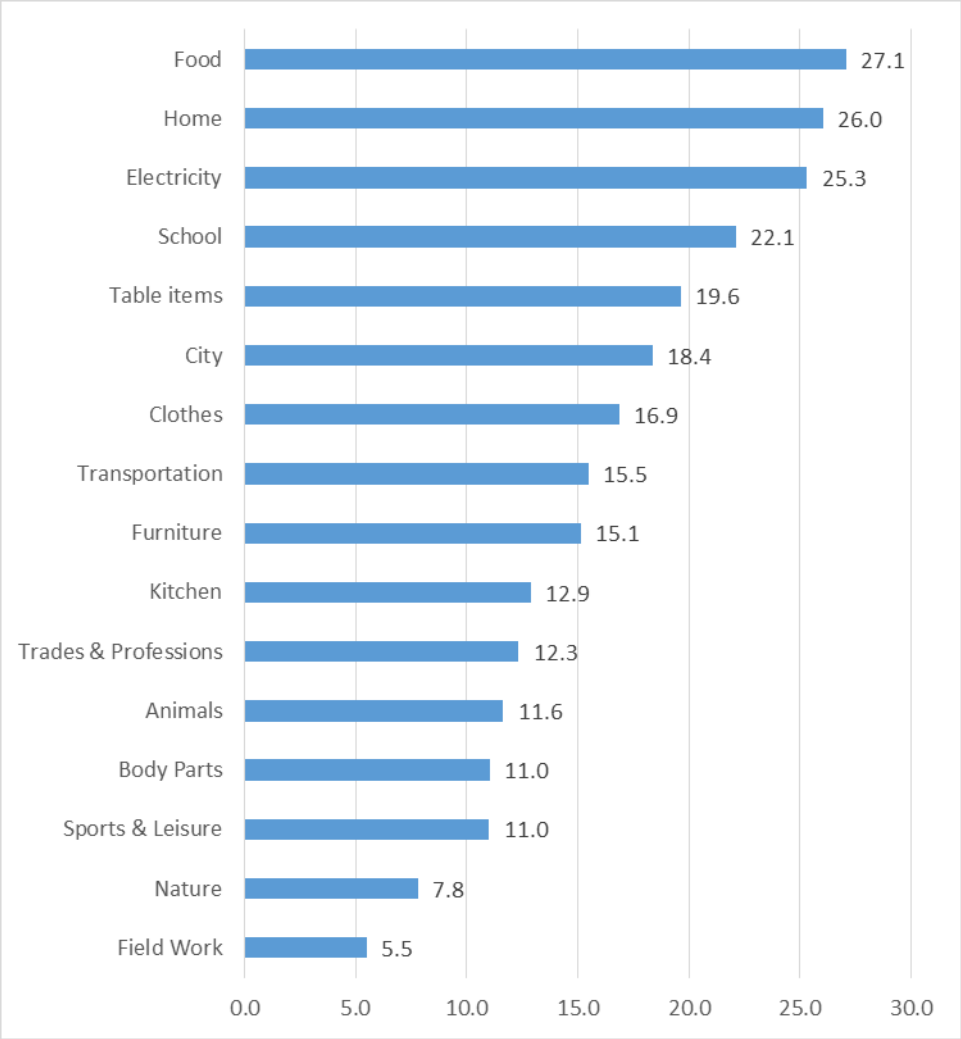
Since the three corpora were different in time, location and education level, contrast analyses were not carried out and lexical items were all pooled together. When forming a single new corpus, syntagmatic compounds account for 15.5% of total lexical items, reflecting the weight of the largest Zacatecas corpus.

Prevalence of syntagmatic compounds by topic in the pooled corpus

Next, the relative frequency of syntagmatic compounds as a proportion of the lexical items within each topic is shown in Figure 4 where, within the topics of Food, Home and Electricity at least a quarter of lexical items are syntagmatic compounds. In contrast, the topics of Nature and Field Word show scarcity in this kind of compounds.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Figure 4. Percentage of Syntagmatic Compounds as Proportion of Lexical Items within Topic



Classification of compounds by composition and by corpus

Next the frequency of syntagmatic compounds by grammatical category, according to Varela's classification (2005, 2018), is presented in Table 2.

Table 2. Classification of syntagmatic compounds by corpus.

	Prepositional compound	Yuxtaposed compounds		Noun+Adjective compounds	Total Syntagmatic
		Nominals	Adjectives		

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

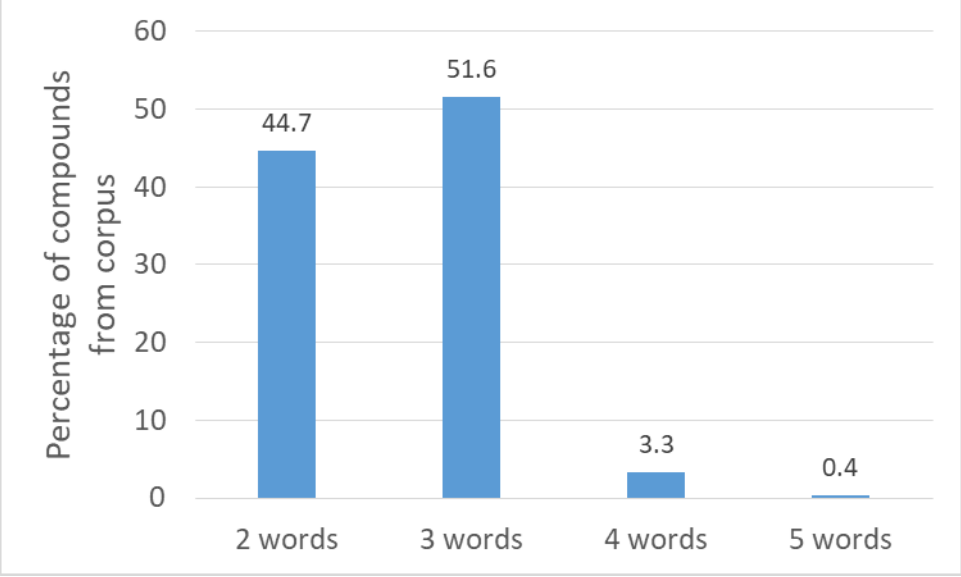
		Appos.	Coord.			compound
Mexico City	461	21	4	2	284	846
Culiacan	450	39	8	6	380	916
Zacatecas	492	67	11	5	327	795
Total	1,403	127	23	13	991	2,557

These data show that, in all corpora, the prepositional compound is the most common, followed by the Noun+Adjective, and the least frequent are the yuxtaposed ones.

Distribution of compounds by number of words

The compounds can also be classified by the number of words. Although it could be expected that two-word compounds to the most frequent, in the corpora from the lexical availability task it is the three-word compound the one that is the most common. This distribution is also shown in Figure 5, which also illustrates that 4- and 5-word compounds are relatively rare.

Figure 5. Percentage of compounds by number of words.



Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

In addition, the most common second word in all compounds, irrespective of the amount of words was the preposition *de* [of], which is present in 41.5% of pooled syntagmatic compounds.

Description of the most frequent compound bases and their relationship to syntagmatic compounds

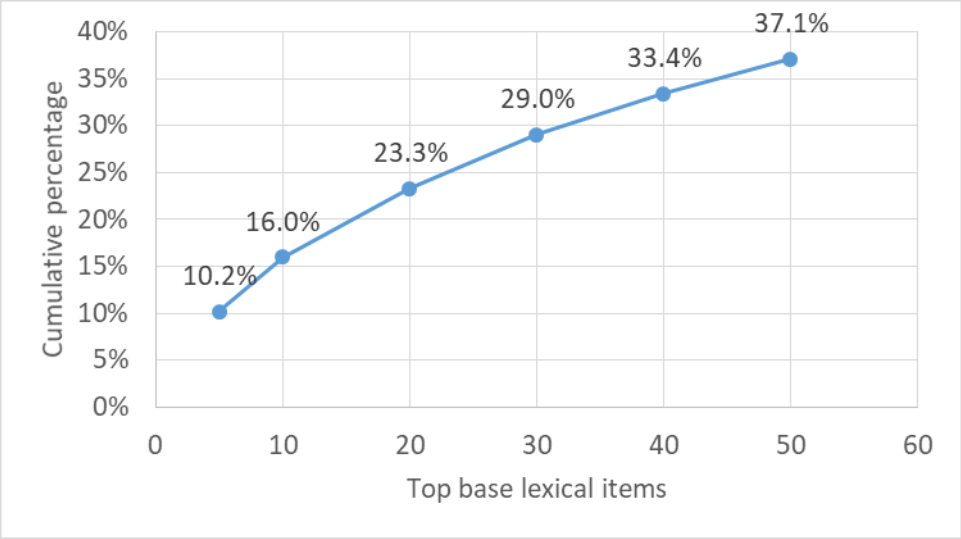
When forming a syntagmatic compound, there is a group of lexical items termed as lexical base, whose main function is to serve as conceptual nucleus for the syntagmatic compound, generating a specific meaning for it. For example, the lexical item *agua* [water] tends to specify its meaning when an adjective is added, such as in *clara* [clear] in *agua clara* [clear water] or when a prepositional syntagm is added such as in *de Jamaica* [made of hibiscus] in *agua de Jamaica* [hibiscus water]. Lopez Rivero (2008) proposes that when a speaker produces a list of lexical items it does not necessarily means that he knows how to apply (or structure) all these words in conversational language. The purpose of obtaining lexical items lists is to observe how the participant creates other representations of the same object from the repertoire they already have.

The most frequent compound-generating lexical items or *base items* found in the three corpora are reported next.

Very few base items account for a significant number of syntagmatic compounds. For example, the top 5 bases account for 10.2% of all syntagmatic compounds, while the top 50 bases account for 37.5% of all the same compounds. The relationship between the bases and their cumulative proportion of all compounds is illustrated in Figure 6, showing a trend of diminishing returns for additional bases.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Figure 6. Relationship between top base lexical items and cumulative percentage as a proportion of all syntagmatic compounds



Finally, to illustrate this phenomenon, and in order to provide a list that could be used for learners of Spanish as a Second Language, Table 3 presents the most available bases found in the three corpora.

Rank	Base lexical item	Mean LAI*	Five most common examples
1	Cuarto [room]	0.031	Cuarto de lavado Cuarto de televisión Cuarto de servicio Cuarto de estudio Cuarto de baño
2	Agua [water]	0.212	Agua de horchata Agua de jamaica Agua de piña

Rank	Base lexical item	Mean LAI*	Five most common examples
11	Tienda [store]	0.177	Tienda de autoservicio Tienda de ropa Tienda de abarrotes Tienda de muebles Tienda de zapatos
12	Cancha [field/ court]	0.032	Cancha de futbol Cancha de basquetbol

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: 1562-384X
Año XXVI, Número 82 julio-diciembre 2022
DOI: 10.32870/sincronia.axxvi.n82
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

			Agua de sabor Agua potable				Cancha de tenis Cancha deportiva Cancha de volibol
3	Plato [plate]	0.272	Plato hondo Plato chico Plato extendido Plato para postre Plato plano	13	Máquina [machine]	0.044	Máquina de coser Máquina de escribir Máquina de soldar Máquina eléctrica Máquina de baile
4	Sala [room]	0.208	Sala de juegos Sala de tv Sala de descanso Sala de tele Sala de televisión	14	Salón [room]	0.195	Salón de juegos Salón de belleza Salón de tv Salón de visitas Salón de danza
5	Centro [center]	0.246	Centro comercial Centro nocturno Centro de cómputo Centro de entretenimiento Centro de copiado	15	Chile [pepper]	0.096	Chile relleno Chile morrón Chile ancho Chile en nogada Chile poblano
6	Mesa [table]	0.149	Mesa de centro Mesa de trabajo Mesa de noche Mesa comedor Mesa de juego	16	Licenciado [person who holds a Bachelor degree]	0.204	Licenciado en derecho Licenciado en economía Licenciado en matemáticas Licenciado en comercio Licenciado en mercadotecnia
7	Ingenie	0.292	Ingeniero civil	17	Oso	0.282	Oso hormiguero

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

	ro [engineer]		Ingeniero electricista Ingeniero en computación Ingeniero agrónomo Ingeniero industrial		[bear]		Oso panda Oso polar Oso pardo Oso blando
8	Carne [meat]	0.376	Carne asada Carne de cerdo Carne de puerco Carne de res Carne molida	18	Mueble [furniture]	0.020	Mueble de sala Mueble de comedor Mueble de televisión Mueble para aparato eléctrico Mueble para libros
9	Cuchara [spoon]	0.294	Cuchara sopera Cuchara grande Cuchara chica Cuchara de madera Cuchara de postre	19	Bote	0.006	Bote de basura Bote de remos Bote de basura inorgánica Bote de basura orgánica Bote de lámina
10	Juego [game]	0.035	Juego de mesa Juego de video Juego de azar Juego geométrico Juego al azar	20	Sopa	0.269	Sopa de pasta Sopa de letras Sopa de espagueti Sopa de haba Sopa de verduras

Table 3. Twenty most frequent base lexical items, their mean Lexical Availability Index and five of the most frequent examples. Only bases are translated

N.B.: Mean LAI=Mean Lexical Availability Index

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Table 3 shows the mean Lexical Availability Index (LAI) computed from the three corpora. While the rank was derived from the processed corpora, that is, from unique or different entries in the three corpora, it should be emphasized that in each original (unpublished) corpus, these base lexical entries have larger frequencies or prevalence, which can only be globally inferred from the Mean Lexical Availability Index. In other words, among these 20 bases, the base *carne* has a high Lexical Availability Index of 0.376, making it highly available in the natural, unprocessed corpora. Likewise, in second and third places by availability, *cuchara* and *Ingeniero* have LAIs of .294 and .292, respectively.

Discussion

This study analyzed three different corpora from students who are Native speakers of Spanish in three different times, regions and educational levels, obtained through the Lexical Availability Task. The aim was to report the frequency and classification of syntagmatic compounds and their bases in students' lexical availability responses. This study also aimed to describe the relationship between bases and lexical items to examine whether a few of the bases are needed for large number of syntagmatic compounds.

The analyses showed that syntagmatic compounds are relatively common in lexical availability responses in Native speakers of Spanish of three regions, times and education levels, ranging from 10% in the Zacatecas (high school) corpus to 31% in the Mexico City (primary) corpus. However, the absolute numbers of syntagmatic compounds were similar across corpora, so the smaller proportion of syntagmatic compounds in that corpus seems to be more a function of the largest size of the Zacatecas corpus in terms of lexical items.

The analyses also showed that food, home and electricity were the topics with a larger share of syntagmatic compounds. This is interesting in the sense that even when lexical availability responses might be biased as they are not words within conversation or discourse, food and home

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

are still topics that may be important in language learning. That electricity is so productive for compounds might be an artifact since many objects are described by attaching the word *eléctrico* [electrical] to several lexical bases.

Regarding the compounds classification using Varela's system of categorization, the most common types of syntagmatic compounds were the prepositional ones, a fact reinforced by the finding that *de* [of] is present in 41.5% of all syntagmatic compounds. Also interesting is the finding that more than half of syntagmatic compounds are comprised of three (51.6%) words, followed by those comprised of only two (44.7%).

When analyzing the relationship between the top bases and the cumulative percentage they accounted for, it was possible to glimpse at the learning potential of syntagmatic compounds since learning only 50 bases are needed to form more than a third of all these types of compounds.

Finally, for the use of teachers of Spanish as a Second Language, the 20 most common bases are listed. It is worth noting that, within the top 20 bases, there are *cuarto*, *sala* and *salón*, all different variations of "room". Also interesting is the fact that 5 of these bases are related to eating and food: *agua, plato, carne, cuchara*, and *chile*.

It should be emphasized that all analyses come from lexical entries in the processed corpora, that is, the list of unique entries with their corresponding Lexical Availability Index, which accounts for their frequencies. Therefore, it is possibly safe to assume that many of the syntagmatic compounds might even be underrepresented since those lexical entries with higher indices are supposed to be more frequent in the lexical availability responses than their occurrence in the published corpora with unique entries.

Altogether, these results suggest practical implications for both Native speakers of Spanish and learners of Spanish as a Second Language, which was the area that the Lexical Availabity Task was originally designed for in France to teach immigrant children. These findings can enrich the teaching of Spanish as easy and effective means of expanding the learners' vocabulary. Research into lexical availability in Spanish-speaking populations has not been harnessed yet in educational settings in a substantial and effective way, despite the vast body of work in contrast and dialectal

e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

variation studies in both L1 and L2, whose original aim was pedagogic. In that sense, a new line of research is needed that integrates Lexical Availability results with well desgined language interventions. For example, Rodriguez Muñoz and Muñoz Hernandez (2008), suggest developing this line of research while considering the students' social characteristics such as age, other language knowledge, learning context, education level and others. Only when the connection between the potential of the lexical availability studies and their vocabulary learning outcomes is fully evaluated can the promise of these studies be fulfilled. In this way the teaching of language in general and, specifically of vocabulary, could be centered around the student and contribute to their communication skills in socially and culturally diverse contexts.

Limitations

Even when there is representation of diverse students from different times, regions and educational levels, the study reported here has several limitations. First, all samples belong the Mexican variant of Spanish, and other variants could produce different findings. For example, it is quite possible that the base *chile* might be idiosyncratic to Mexican Spanish and conversely, it could be less available in other Spanish variants. Second, even with the familiarity of topics, the Lexical Availability Task's scope is limited to lists of lexical entries that could potentially differ if the syntagmatic compounds were extracted from conversation, or oral or even written discourse.

Still, even when considering those limitations, the findings presented here allow for a conceptualization of syntagmatic compounds bases, at least in the teaching of Spanish, as linguistic legos, since they could, at least theoretically, significantly expand a learners' vocabulary. Future work testing a compound bases vocabulary didactic intervention could examine this assumption.

References

Alba, O., (2013). Variación diacrónica del léxico disponible dominicano, en *Lingüística española actual*, 35.1: 149-180.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Buenafuentes, C., (2007). *Procesos de gramaticalización y lexicalización en la formación de compuestos en español*. PhD thesis. Universidad Autónoma de Barcelona (http://www.tdx.cesca.es/TESIS_UAB/AVAILABLE/TDX- 0321107-172834//cbm1de1.pdf)

Butragueño, P. M., (2011). Dialectología de nuevos mundos: una lectura variacioncita del Atlas Lingüístico de México. In *13th Hispanic Linguistics Symposium*, Cascadilla Proceedings Project, pp.17-32.

Camarena, E. D. (2010). La interlengua en el léxico disponible de un grupo alumnos de portugués en México. In I. Moskowich-Spiegel Fandiño, B. Crespo García, I. Lareo Martín y P. Lojo Sandino (eds.), *Language Windowing through Corpora. Visualización del lenguaje a través de corpus*, vol. I. La Coruña: Universidad de la Coruña.

Carcedo, A., (1998), "Sobre las pruebas de disponibilidad léica para estudiantes de español como lengua extranjera", *RILCE* 14/2, (número monográfico "Español como lengua extranjera: investigación y docencia"): 205-224.

Carcedo, A., (1999), "Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad", *Pragmalingüística* 5-6: 75-94.

Casado, M. (2018). *La innovación léxica en el español actual*, Madrid: Gredos.

Cruz, M., (2015). Base de datos para conocer (y dar a conocer) la investigación en ELE, *Doblele,* (1), 77-100.

Hidalgo, M. (2017). Evolución diacrónica del léxico disponible de estudiantes sinohablantes de ELE, en *Revista de enseñanza del Español a Sinohablantes*,

Frey, M. L. H. (2007). Disponibilidad léxica y escritura del español como lengua extranjera: propuesta de comparación de dos corpus, *Interlingüística, 17*, 366-373.

Fuentes, C. (2014). Identificando y trabajando con el léxico disponible: experiencias iniciales, ponencia pronunciada en el *V Simposio para el mejoramiento de la enseñanza y el aprendizaje de las lenguas, una necesidad formativa urgente de atender para el logro de la educación omnilateral en el siglo XXI*, México: UNAM.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Guerra, L., F. Paredes y M. E. Gómez Sánchez (2004). Disponibilidad léxica en la Comunidad de Madrid, oral paper presented at the XXXIV Simposio de la Sociedad Española de Lingüística, Madrid, 13- 16 de diciembre de 2004.

Hugo, N. M. (2003). *Disponibilidad léxica en español como lengua extranjera*, Trabajo de investigación inédito, Saarbrücken: Universität des Saarlandes.

Kalan, M. Š. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera, *Studia Romanica Posnaniensia*, *41*(1), 63-85.

Kranjc, M. (2009). *Disponibilidad léxica en alumnos eslovenos en español como lengua extranjera*, Memoria de licenciatura, *Ljubljana: Universidad de Ljubljana*.

López, J. (1994). Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica, en *Revista de Estudios de Adquisición de la Lengua española*, 1, pp. 67–84.

López, J. y M. T. Madrid Guillén (1993). *Léxicos disponibles de escolares mexicanos,* Alhambra Mexicana, UNAM.

López, H. (2005). Un nuevo corpus para el estudio del español: la disponibilidad léxica, en *Oralia: Análisis del discurso oral*, (8), pp. 141-160.

Luque, L. (2004). La formación de compuestos en español: el caso de los sustantivos, *Verba hispánica: anuario del Departamento de la Lengua y Literatura Españolas de la Facultad de Filosofía y Letras de la Universidad de Liubliana,* N.º. 12, Facultad de Filosofía y Letras de la Universidad de Ljubljana, pp. 143-154.

Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage, *Les langues modernes*, 47, 338-344

Moreno, F. (2012). Disponibilidad léxica: cuestiones metodológicas: A propósito de disponibilidad léxica de los estudiantes hispanos de Redwood City, CA, *Revista Nebrija de Lingüística Aplicada*, 11.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 82 julio-diciembre 2022
DOI: **10.32870/sincronia.axxvi.n82**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Pérez, M., (2019), "Un estudio sobre los macrocentros de interés (McI). Una aportación teórica para los estudios actuales de disponibilidad léxica", *Cuadernos de Lingüística de El Colegio de México*, 6(1). 1-42.

Rubio, R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de español como lengua extranjera, *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía. Venecia: Edizioni Ca'Foscari*, 143-164.

Sánchez-Saus, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*, Tesis doctoral, Cádiz.

Santos, S., (2015). *Disponibilidad léxica en náayeri (cora) y español de estudiantes bilingües de primaria: un estudio contrastivo*. *CONACYT*.

Samper, M., (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*, Málaga, ASELE.

Varela, S. (2018). *Morfología léxica: la formación de palabras*, Madrid: Gredos.

Varela, S. (2009). *Morfología léxica: la formación de palabras*, Madrid: Gredos.

Varela, S. (2005). Lexicalización y cambio categorial, En *Palabras, norma, discurso: en memoria de Fernando Lázaro Carreter*, Ediciones Universidad de Salamanca, pp. 1133-1146.

Val, J. F., (1999). La composición. In *Gramática descriptiva de la lengua española*, I. Bosque and V. Demonte, (Eds), (Tome III). Madrid: Espasa, 4757-4842.

Velarde, J. J., (2014). *Análisis de léxico disponible en hablantes escolares de Culiacán Sinaloa*, Masters Thesis, México: Universidad Autónoma de Sinaloa.

Verdeses, R., (2012). Disponibilidad léxica de los estudiantes hispanos de Redwood City, California, *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, (11), 1.