



Comparative Analysis of Cognates in Two Datasets in Spanish Teaching in China

CHEN HAO  

Fudan University, China

YANG MING  

Shanghai International Studies University, China

Historia del artículo / Article Info

ARTÍCULO DE INVESTIGACIÓN

Recibido/Received

3 de junio de 2024

Aprobado/Accepted

20 de julio de 2024

Publicado/Published online

13 de agosto de 2024

<https://doi.org/10.19053/uptc.0121053X.n43.2024.17531>



Citación/Citation: Hao, C & Ming, Y. (2024). Comparative Analysis of Cognates in Two Datasets in Spanish Teaching in China. *Cuadernos de Lingüística Hispánica*, 43, 2024, e17531. <https://doi.org/10.19053/uptc.0121053X.n43.2024.17531>

✉ Correspondencia/Correspondence:

Chen Hao: Fudan University, China. h_chen@fudan.edu.cn

Yang Ming: Shanghai International Studies University, China. yangming@shisu.edu.cn

Abstract

This paper aims to conduct a comparative study of two glossary compilations between two programs of Spanish as a Foreign Language (SFL) teaching in China, through statistical analysis and a survey of familiarity with cognates among ELE students in China. Specifically, we investigate: (1) the lexical difference between two SFL teaching systems: one nationalized in China and the other at the international level (CEFR); (2) the awareness of bilingual cognates for Spanish learners in China. The results indicate that: (1) the vocabulary list nationalized in China has more high-frequency words than that of CEFR; (2) in L3 Spanish learners, the survey did not find an awareness of cognate, but rather an high-frequency effect in the lexicon. This study serves to emphasize the status of cognates for research and teaching in SFL, as well as to provide a reference for future lexical studies of SFL both in China and internationally.

Keywords: cognate awareness, lexical dataset, word frequency, lexical coverage, cognate coverage.

Análisis comparativo de cognados en dos bases de datos en la enseñanza de español en China

Resumen

El objetivo de este artículo es realizar un estudio comparativo de dos compilaciones de glosarios entre dos programas de enseñanza de Español como Lengua Extranjera (ELE) en China, a través de un análisis estadístico y una encuesta sobre la familiaridad con cognados para estudiantes de ELE en China. En concreto, se investiga: (1) la diferencia léxica entre dos sistemas de enseñanza de ELE: uno nacionalizado en China y el otro a nivel internacional (MCER); (2) la conciencia de los cognados bilingües para los estudiantes de español en China. Los resultados indican que: (1) la lista de vocabulario nacionalizada en China tiene más palabras de alta frecuencia que la del MCER; (2) en estudiantes de español como L3 no se descubrió ninguna conciencia de cognados en la encuesta, sino más bien un efecto de alta frecuencia en el léxico. Este estudio sirve para enfatizar el estado de los cognados para la investigación y la enseñanza en ELE, así como para proporcionar una referencia para futuros estudios léxicos de ELE tanto en China como a nivel internacional.

Palabras clave: conciencia de cognados, base de datos léxica, frecuencia de palabras, cobertura léxica, cobertura de cognados.

Analyse comparative des apparentés dans deux bases de données de l'enseignement de l'espagnol en Chine

Résumé

L'objectif de cet article est de réaliser une étude comparative de deux compilations de glossaires entre deux programmes d'enseignement de l'Espagnol comme Langue Étrangère (ELE) en Chine, à travers une analyse statistique et une enquête sur la familiarité avec les apparentés chez des étudiants ELE en Chine. Bref, nous étudions : (1) la différence lexicale entre deux systèmes d'enseignement ELE : l'un nationalisé en Chine et l'autre au niveau international (CECR) ; (2) la conscience aux apparentés bilingues chez les apprenants de l'espagnol en Chine. Les résultats indiquent que : (1) la liste de vocabulaire nationalisée en Chine contient davantage de mots à haute fréquence que le CECR ; (2) chez les apprenants d'espagnol comme L3, aucune conscience d'apparentés n'a été découverte dans l'enquête, mais plutôt un effet à haute fréquence dans le lexique. Cette étude sert à souligner le statut des apparentés à la recherche et à l'enseignement de l'ELE, ainsi qu'à fournir une référence pour les futures études lexicales de l'ELE en Chine et à l'échelle internationale.

Mots clés: conscience apparentée, base de données lexicale, fréquence des mots, couverture lexicale, couverture apparentée

Análise comparativa de cognatos em duas bases de dados no ensino de espanhol na China

Resumo

O objetivo deste artigo é realizar um estudo comparativo de duas compilações de glossários entre dois programas de ensino de Espanhol como Língua Estrangeira (ELE) na China, através de uma análise estatística e de uma pesquisa sobre familiaridade com cognatos para alunos ELE na China. Especificamente, investigamos: (1) a diferença lexical entre dois sistemas de ensino ELE: um nacionalizado na China e outro em nível internacional (QECR); (2) conscientização de cognatos bilíngues para alunos de espanhol na China. Os resultados indicam que: (1) a lista de vocabulário nacionalizado na China tem mais palavras de alta frequência do que a do QECR; (2) nos alunos de espanhol L3 não foi descoberta nenhuma consciência de cognatos na pesquisa, mas sim um efeito de alta frequência no léxico. Este estudo serve para enfatizar o status dos cognatos para a pesquisa e o ensino do ELE, bem como para fornecer uma referência para futuros estudos lexicais do ELE, tanto na China como internacionalmente.

Palavras-chave: consciência cognata, base de dados lexical, frequência de palavras, cobertura lexical, cobertura cognata.

Introduction

Cognates can be defined as two words that come from the same etymological origin or are derived from the same word (Aske, 2015). In the multilingual realm, there are bilingual and trilingual cognates. It is worth mentioning that cognates can be classified into different categories based on different criteria. Taking Spanish-English bilingual cognates as an example, there may be homographic cognates, quasi-homographic cognates and cognates with minimal or no homographs based on morphological classification (Ortega, 2007; Chen, 2023). It is also allowed to categorize identical cognates, cognates with one letter difference, cognates with more than one letter difference, cognates with different affixes (Nash, 1997; Rodríguez, 2009). In accordance with the semantic classification, cognates can be categorized into cognates of total semantic similarity, partial similarity and false cognates (Ortega, 2007); based on grammatical function criteria, they can be classified as monofunctional, bifunctional, and multifunctional cognates (Ortega, 2007). Furthermore, within the field of different lexical sources, Spanish-English cognates can be divided into 13 main categories (Aske, 2015).

Additionally, based on semantic equivalence, cognates can be categorized as synonymous cognates (true friends) and false friends (Izquierdo Gil, 2003), with the former ones further be classified into interlinguistic cognates by para-synthesis, prefixation, and suffixation, which generates two derivations *per se*: heterogeneous and homogeneous (Gonzalo Pérez, 2015). False friends or *faux amis* are sometimes referred to as interlinguistic homographs (Dijkstra et al., 1999; Schröter & Schroeder, 2016). Terminologically, it is important to note that the concept of “false cognate” is different from that of “false friend” (Moss, 1992). The latter was proposed by Keossler and Derocquigny (1928). In terms of lexical differences, the main distinction between “false cognate” and “false friend” lies in the fact that the former are not cognates, lacking a common etymology, but sharing similar word forms (e.g., “mucho” in Spanish and “much” in English). In the process of generating false friends, their meanings have undergone partial or complete changes, to an extent that they can be classified as false friends and semi-false friends, also known as partially false friends (Aske, 2015). In addition to the aforementioned classification of cognates, loanwords, true cognate families, toponymy, anthroponymy and abbreviations all contain cognates (Ortega, 2007; Aske, 2015; Gonzalo Pérez, 2016; Chen, 2023). Furthermore, in early research on bilingual acquisition and translation, cognates were generally considered to be translation equivalents in bilingualism that are orthographically and/or phonologically similar, whilst non-cognate words were defined as translation equivalents in bilingualism that are not similar in form (Lotto & de Groot, 1998; Costa et al., 2000; Comesaña et al., 2012).

Different approaches in cognate research have led to the divergence of concepts of cognates. In general, the definition of cognates is twofold to serve both qualitative and quantitative research methods (Chen, 2023). Qualitative research on cognates has underpinned the ontological basis of cognate, aligned to etymology, lexicography and historical linguistics (García, 2013; Aske, 2015), whilst the quantitative research focuses

on the applicative and experimental perspectives of cognate use, the extent to which cognates are applied, on a regular basis, as objects or stimuli of research material in empirical research of applied linguistics, psycholinguistics, cognitive translation and interpretation studies, with a definition widely approved by respective researchers (Dijkstra et al., 1999; Dijkstra & Van Heuven, 2002; Comesaña et al., 2010, 2012; Poort & Rodd, 2017, 2019). The ontological research of cognates is primarily based on the phonological, morphological and semantic elements of the vocabulary in monolingual or multilingual contexts, focusing mainly on the origin, differentiation, and evolution of cognates; the empirical research, in the meantime, begins with the functional and practical aspects of cognates, examining their similarity and equivalence in language acquisition and translation. In specific research domains, applied linguistics is mainly concerned with the use of cognates in lexical acquisition and teaching, while psycholinguistics focuses on the influence and status of cognates in lexical processing, wherein the investigations revolve around lexical input, memory, storage and production in the brain through the accessibility of cognates and their sublexical features in the mental lexicon.

With regard to its particular lexical status, the synergy of cognate related research in multiple fields and concepts has always been a hot topic. In the field of language teaching, many studies have emphasized the importance of cognate awareness and its status (Chen, 2021, 2023; Cenoz et al., 2022). The present study focuses on the connection between cognates and multilingual teaching as well as lexical content in multilingual teaching materials. We conduct a comparative analysis of the cognate vocabulary in the structured vocabulary lists for intermediate-level Spanish learners in China and the CEFR B2 vocabulary list. The study aims to clarify attributes such as proportion and frequency of cognates in both lists, as well as the similarities and differences in students' lexical input and potential lexical availability in classroom through a lexical familiarity survey. This reflection further highlights the differences between local and international teaching materials in the development of vocabulary lists. By identifying the importance of cognates in vocabulary teaching, the study also explores vocabulary issues in Spanish teaching materials localized in China. The specific research questions addressed in this paper are as follows:

RQ1: What is the basic lexical property and cognate status of the vocabulary list in the localized Chinese lexical list of the Spanish (Intermediate) learning program compared to CEFR A1 to B2 vocabulary?

RQ2: How familiar are students with cognates and non-cognate vocabulary in their vocabulary repertoire?

Literature Review

Cognate Effect and Lexical Competence

The facilitation or inhibition effects of cognates reflect cross-linguistic influence within second or multiple languages. Cognate effects lead to a greater advantage for cognate

acquisition in second or multilingual vocabularies. Through a cognate recognition study of 74 English-Spanish bilinguals, Nagy et al. (1993) found that students who identified more cognates achieved better results in English. In a comparison of cognates (true/false friends) and non-cognates in second language translation research, Otwinowska and Szweczyk (2017) pointed out that both true and false cognates have stronger learnability compared to non-cognates, which is related to their frequency and orthographic similarity. In one of trilingual acquisition studies, Chen and Song (2024) conducted a longitudinal experiment on pedagogical intervention with cognate vocabulary and found evidence of cross-lexical transfer from the second to the third language, demonstrating the facilitating role of cognates in trilingual vocabulary learning, as observed in a similar bilingual pedagogical intervention study by Dressler et al. (2011). Therefore, we believe that cognates can enhance the lexical competence of students.

Cognates in Pedagogical Translanguaging

As highlighted by Cenoz and Gorter (2020), training in cognate recognition is a new trend in the development of multilingual skills. Through pedagogical intervention research on cognates in Basque, Spanish and English, Cenoz et al. (2022) found that teaching interventions on cognates help students enhance their metalinguistic awareness. They also emphasized that through cognate recognition exercises, students improve their cognitive understanding of vocabulary orthographic and semantic transparency. Thus, fostering cognate awareness, or the development of metalinguistic awareness, contributes to an upcoming research approach under the umbrella term of pedagogical translanguaging. However, current research in this area focuses primarily on cognate recognition, with limited analysis of cognates in relation to vocabulary lists in teaching materials, while teachers' beliefs indicate that cognates do not require as much attention in lexical teaching as non-cognates, since the former are more learnable (Sánchez-Gutierrez, 2022).

Lexical Availability

The concept of lexical availability first appeared in the research of French scholars (López, 1995; Izquierdo Gil, 2003; Song, 2009; Hidalgo Gallardo, 2017). They developed vocabulary lists based on the selection of vocabulary used by native speakers in communication, measured by concrete semantic load and notional character, aiming to improve native language vocabulary teaching. However, this concept has gradually been applied to second language vocabulary teaching (Carcedo, 2000; Izquierdo Gil, 2003) and even multilingual vocabulary teaching (Song, 2009). Morales (1984, p. 62) defines lexical availability as "the lexical stock that is activated and can be used in a specific communicative situation", meaning the units that come to mind first when presented with a thematic stimulus or category.

The research paradigm commonly used to assess lexical availability involves asking participants to write words related to a specific domain within a given time frame

(with a specified or unspecified number of words), followed by further analysis of the written words by researchers. Furthermore, the traditional definition of lexical availability has been expanded from a cognitive perspective (Hidalgo Gallardo, 2019). Therefore, lexical availability may encompass factors such as familiarity, typicality, imaginability, age of acquisition, cognitivity, frequency, and length (Hampton & Gardiner, 1983; Hernández-Muñoz et al. 2006), with varying effects depending on whether it is an L1 or an L2, that influence the structure of the category itself and the lexical choices made in relation to it. Hence, it is related to cross-linguistic lexical units: bilingual or trilingual cognates.

Martínez-Adrián and Gallardo-del-Puerto (2017) conducted an analysis of lexical availability among Spanish and Chinese learners of English as a Foreign Language (EFL). Their findings revealed variations among the learner groups, with Spanish learners exhibiting greater levels of lexical availability and spelling accuracy compared to Chinese learners. They attribute these differences to typological proximity, which indicates that in second language acquisition, lexical availability is related to the lexical association between the two languages, and this cross-lexical influence primarily manifests as cognateness effects (Otwinowska, 2016). Since cognates exist not only in second language acquisition but also in multilingual contexts, further research is needed to explore the similarities and differences between multilingual and second language acquisition contexts.

Lexical Selection and Word Frequency

In consideration of vocabulary coverage and word frequency, lexical selection for vocabulary learning and teaching seems to favor the prioritized words over others (e.g. low-frequency words over high-frequency words), because high-frequency words are less cognitively demanding and easier to learn. Lexical processing studies have demonstrated that cognates are mainly high-frequency words, henceforth the cognate effect and word frequency effect are practically identical (Strijkers et al., 2010). In the meantime, lexical frequency has been one of the most used in the L2 acquisition literature (Sánchez-Gutiérrez et al., 2019). Therefore, the present study aims to compare the cognate status between the intermediate level of vocabulary list of the Chinese national Spanish teaching syllabus and the DELE B2 vocabulary list. We seek to infer the differences in vocabulary learning between Chinese-Spanish students and the DELE requirements in terms of word frequency, concreteness, word length, and a particular impact on cognate status.

Methodology

Lexical Coverage and Cognate Status Analyses

In the study of lexical selection, we adopted the analytical approach proposed by Sánchez-Gutiérrez et al. (2022). Rather than conducting a study on textbooks, our analysis consists of the lexical properties in two datasets based on glossaries, mainly the proportion of vocabulary required to be mastered among 5,000 high-frequency words and the cognate status. The first one of the datasets is the National Spanish Major Level Four Vocabulary Glossary (hereafter EEE4G), which has a semi-official status. By “semi-

official”, we mean that although this vocabulary list is not officially published by the National Spanish Language Teaching and Research Group, it is compiled based on the official syllabus of the Spanish Major Level Four Examination (intermediate level) and is widely used by Chinese Spanish language teachers, as stated in the introduction of the manual (Li & Cheng, 2006). The second lexical dataset was constructed by incorporating several glossaries from vocabulary textbooks used for DELE preparation. We collected all the alphabetic glossaries of six vocabulary books from levels A1 level to B2 available on the current market to achieve one dataset (hereafter B2G). All materials are indicated in Table 1.

Table 1. Two datasets of the present study

	Vocabulary books in question
EEE4G	<i>Spanish Major Level Four Vocabulary</i> . (2006), Beijing: People’s Education Press
B2G	<i>Vocabulario ELE (A1-B1)</i> . (2016), Shanghai Foreign Language Education Press <i>Vocabulario. Nivel avanzado B2: Vocabulario</i> . (2023), Anaya ELE <i>Vocabulario. Nivel avanzado A1-A2: Vocabulario</i> . (2008), Anaya ELE <i>Viva el Vocabulario - intermedio (B1-B2)</i> . (2010), enClave-ELE <i>Vocabulario ELE B1: Léxico fundamental de español de los niveles A1 a B1</i> . (2014), SGEL <i>Vocabulario ELE B2</i> . (2018), SGEL

Familiarity of Spanish Cognates

From all the Spanish cognates aforementioned in EEE4G dataset, we randomly selected 30 cognates and 30 non-cognates to ensure lexical status as followed: word length ranging from 6 to 13 with an average word length of 8.07; syllable number ranging from 2 to 5, with an average of 3.28. All the word properties of these 60 words are shown in Table 2. The criteria of the cognate selection based on 3 steps: 1. Using the English-Spanish cognate dictionary (Nash, 1997) to identify 2070 Spanish cognates; 2. Using three websites to determine the cognate status for the rest of the dataset, consulting the etymological resource of each word. The three websites are RAE (<https://www.rae.es/>) for Spanish words, OED (<https://www.etymonline.com/>) for English words and SpanishCognates.org (<https://spanishcognates.org/>) for reference. Finally, we obtained a total of 2634 cognates (see Table 2). The Levenshtein Distance (LD) was calculated for English-Spanish cognates and for non-cognates and their translation equivalents: LD can be understood as the minimal number of insertions, deletions or substitutions required to change one word into the other (Zhu & Mok, 2018): the LD between *product* and *producto* is one, as *product* can be derived from *producto* by deleting the letter *o*; and LD between *versión* and *version* is also one, for the accent mark is considered as well.

Based on this, 41 Spanish major students (Chinese L3 Spanish learners) with an average age of 19.47 (Freshmen: 17; Sophomores: 18; Juniors: 3; Seniors: 3) were recruited to assess the familiarity of cognates and noncognates on a scale of 1-5 via an anonymous questionnaire filled out through scanning QR codes. Participants also rated their multilingual proficiency of Listening, Speak, Comprehension and Writing (on a scale of 1-10) in Chinese (L1), English (L2) and Spanish (L3): the majority (39:2) considered that their English proficiency is higher than their Spanish proficiency in the four terms mentioned above: L1 (Chinese): 8.28; L2 (English): 6.69; L3 (Spanish): 3.93. Preliminary masked priming experiment of lexical decision task (LDT) on Chinese L3 learners of Spanish with the 60 words revealed cognate effects (Chen & Liang, 2023), however, the study remains several questions to be further discussed. One of them indicates that many participants have shown low accuracy rates in the LDT. The evaluation of the familiarity of these 30 cognates and 30 non-cognates by first-year (beginner Spanish learners) and second-year (intermediate Spanish learners) students serves two purposes: firstly, to understand if there are differences in the familiarity of cognates and non-cognates among students at different learning stages, with L2 proficiency much higher than L3, and secondly, to provide inference for subsequent lexical processing research on L3 Spanish learners. Due to the scarce number of third and fourth-year students, and their generally high familiarity with the vocabulary (mostly scoring above 4), their ratings were discarded from the present analysis.

Table 2. Descriptive statistics of 60 words

	Total	Cognates	Noncognates
Log_frq (EsPal)*	M=1.28, SD=0.54	M=1.43, SD=0.40	M=1.14, SD=0.62
Concreteness	M=4.77, SD=0.93	M=4.71, SD=0.94	M=3.23, SD=0.56
Length	M=8.01, SD=1.00	M=8.10, SD=0.91	M=8.03, SD=1.08
Syllabus	M=3.28, SD=0.55	M=3.33, SD=0.54	M=3.23, SD=0.56
LD	-	M=2.37, SD=1.08	M=6.87, SD=1.38

*Log_frq in EsPal is the log10 (frequency+1) of Spanish words in EsPal database and the latter is a database of Spanish word properties based on written and subtitle corpora: <https://www.bcbl.eu/databases/espal/wordidx.php>.

Results

Lexical Frequency Comparison

The comparison of lexical selection between two lexical datasets is reflected through word frequency. We compare EEE4G and B2G based on the 5000 high-frequency Spanish words band of the frequency dictionary (Davies & Davies, 2017) to examine both lexical coverage and cognate coverage. In coordination with the 5000 high-frequency Spanish

words, the percentage of words contained in the EEE4G and B2G within the 5000-word list reflects the lexical coverage of high-frequency words and the proportion of words from each list within the 5000-word list reflects the distribution of high-frequency words. Using the same approach, we can also obtain cognate coverage and cognate distribution in high-frequency word band. The data from both vocabulary lists are presented in Table 3: there are 2587 words that completely overlap between the two vocabulary lists, accounting for 59.79% of EEE4G and 51.81% of B2G. Among these completely overlapping words, there are 1584 cognates, accounting for 36.61% of EEE4G and 31.72% of B2G. Additionally, there are 1003 non-cognate words overlapped completely, accounting for 23.18% of EEE4G and 20.09% of B2G. The B2G list contains 2406 exclusive words, among which there are 1286 cognates and 1120 non-cognates.

Table 3. Frequency data of the vocabulary datasets

	EEE4G			B2G		
	Total	Cognate	noncognate	Total	Cognate	noncognate
N	4327	2634	1693	4993	2870	2123
% in dataset	100	60.87	39.13	100	57.48	42.5
% within 5000	81.35	84.24	76.85	55.64	48.05	75.51
(N)	(3520)	(2219)	(1301)	(2778)	(1379)	(1603)
% of 5000	70.4	44.38	26.02	55.56	27.58	32.06

Familiarity of Cognates and Non-cognates

We have conducted lexical analyses and comparisons focusing on word properties and familiarity by using R software. The results of the analysis of cognates and non-cognates are as follows: a t-test on the LD revealed a significant difference between the two groups ($t = -2.0649$, $df = 50.48$, $p = 0.04409$). In the meantime, there was no significant difference found between cognates and non-cognates in terms of concreteness ($t = 0.88647$, $df = 41.854$, $p = 0.3804$) and word length ($t = 0.25444$, $df = 54.386$, $p = 0.8001$). Additionally, there was no significant difference observed between the selection of words in Spanish and English in terms of Zipf values ($t = -0.33791$, $df = 105.26$, $p = 0.7361$).

To our surprise, the comparison of familiarity between initial and intermediate level of L3 learners for the 60 words in this study did not reveal higher familiarity for cognates than for non-cognates. Instead, we found that participants were more familiar with high-frequency words (≥ 18 per million) than low-frequency words (< 18 per million) ($t = 2.6572$, $df = 57.323$, $p = 0.01019$), which manifest a frequency effect. Moreover, as the frequency cutoff criterion decreased, a frequency effect was also observed for initial level when the frequency was ≥ 11 per million ($t = 2.3951$, $df = 41.371$, $p =$

0.02122). In terms of different academic years, sophomores were more familiar with all words compared to freshmen as expected ($t = 5.3734$, $df = 113.48$, $p = 4.171e-07$).

Levenshtein Distance of Cognates in Two Datasets

According to the differences in Levenshtein distance of 0, 1, and 2, this study further categorized the English-Spanish cognates from the two datasets into three groups, as shown in the table 4 below:

Table 4. Levenshtein distance for cognates of two datasets

EEE4G (cognates = 2634)			B2G (cognates = 2870)		
LD = 0	LD = 1	LD > 1	LD = 0	LD = 1	LD > 1
227	502	1905	335	511	2024
8.6%	19.1%	72.3%	11.7%	17.8%	70.5%

Discussion

In the present study, we conducted a comparative analysis of intermediate level of Spanish vocabulary glossaries in the teaching of Spanish as a foreign language in China, based on concepts or criteria such as word frequency, lexical coverage, cognate status, and so on. Additionally, we further classified cognates morphologically based on differences in LD, as there are differences in the bilingual lexical processing of identical and non-identical cognates (Comesaña et al., 2015). We observed that approximately 10% of cognates in both datasets are identical, and almost 1/3 of them have a LD less than or equal to 1, which can facilitate the learning of bilingual and multilingual vocabulary from beginner to intermediate levels. However, the study does not analyze the semantic classification of cognates. Therefore, within the context of vocabulary instruction, multilingual teachers should address the semantic similarities of cognates between English and Spanish. This entails providing guidance on the specific usage of words and fostering an understanding of concepts such as false friends.

From the perspective of lexical competence and the cognate effect, the proportion of cognates in the two datasets, which accounts for more than half of the total (60.87% and 57.48%, respectively), can provide Chinese multilingual students with a lexical advantage during the learning process. Multilingual teachers need to leverage the high proportion of cognates to effectively employ corresponding strategies under the pedagogical translanguaging framework to enhance students' metalinguistic awareness. Cognate status requires greater attention in teaching activities or interventions.

The lexical frequency analysis of the two datasets revealed that the EEE4G dataset has a higher proportion of high-frequency words compared to the B2G dataset, reaching 81.35%. This suggests that the vocabulary required for the Chinese Spanish teaching syllabus is more aligned with the pragmatic issues at the universal level. We believe that the

emphasis on fewer high-frequency words in the CEFR-based B2G vocabulary highlights the uniqueness of Spanish vocabulary. For instance, it includes common low-frequency words such as “abrelatas” (can opener), “verdulero” (greengrocer), and “paragüero” (umbrella stand), which are not covered in either the EEE4G or the 5000 high-frequency words. Perhaps for native speakers, these words are more familiar and easier to master. However, from the perspective of vocabulary frequency, the DELE vocabulary curriculum may need to consider the mastery of more practical high-frequency words. In fact, frequently used words can facilitate learning through lexical association to the extent to which the three previously mentioned words can be easily learned through association with common Spanish words like “abrir” (to open), “lata” (can), “verdura” (vegetable), and “paraguas” (umbrella), which is particularly beneficial for foreign students studying in Spain via everyday experience. From the perspective of the proportion of cognates, EEE4G contains a higher proportion of high-frequency cognates, which aligns with the viewpoints mentioned earlier in present study regarding how cognates facilitate vocabulary acquisition and how to effectively employ the advantages of cognates in vocabulary teaching. Moreover, an increased prevalence of cognates is highly beneficial for multilingual learners in their vocabulary acquisition from beginner to intermediate levels. This is because the majority of cognates are high-frequency words (Strijkers et al., 2010). In the EsPal (database of Spanish word properties based on written and subtitle corpora, cf. Duchon et al., 2013), there are 1863 and 1490 high-frequency cognates with frequencies exceeding 11 per million, respectively, constituting 43.06% and 29.84% of the two lexical datasets.

In the research conducted by Song (2009), a comparative analysis of lexical availability among Chinese, English, Italian, Japanese and other languages learners of Spanish revealed that Chinese students manifested weaker lexical competence in certain semantic domains with lower acquisition difficulty, such as “clothing”, “furniture”, and “kitchen utensils”. Moreover, they exhibited slightly inferior ability for lexical diversity compared to students from the other countries. She attributed the differences to teaching materials. The results of the present study partially support this viewpoint, the measurement of lexical familiarity uncovered a frequency effect. This suggests that the inclusion of high-frequency words in the EEE4G of Chinese national Spanish syllabus may be more practical than that of the SFL syllabus of Spain. Furthermore, Martínez-Adrián and Gallardo-del-Puerto (2017) conducted a comparative study on the lexical availability between Spanish and Chinese learners with English as their L2. Their research suggests that language proximity (primarily manifested in cognates at the lexical level) plays a significant role in lexical competence, particularly in lexical availability and spelling accuracy. Therefore, aligning with their findings, the present study argue that high-frequency cognates can effectively enhance the proficiency of learners whose L2 is English and L3 is Spanish in the aforementioned two facets of lexical competence.

The cognate familiarity study did not reveal any significant statistical differences, which may be due to the small sample size and the limited number of words in the questionnaire. It could also be related to the lack of prior training in cognate recognition

among the participants, as suggested by Cenoz and Gorter (2020). Further research is needed to understand the cognate awareness in multilingual individuals.

Conclusions

As a special category of cross-linguistic lexis, cognates often serve as research materials in multilingual acquisition and lexical processing. The present study has collected two datasets: the intermediate level of Spanish glossaries for Chinese learners and foreign learners of SFL in Spain. Furthermore, it compares two datasets in terms of word frequency, lexical coverage, and cognate coverage, revealing distinct emphases in each. Irrespective of the dataset, a higher proportion of cognates in a lexical dataset can provide multilingual learners with the cognate advantages while building their mental lexicon. This research will advance the investigation into cognate status within the vocabulary size of multilingual individuals. Specifically, considering the intermediate level Spanish vocabulary size for Chinese learners, the research of English-Spanish cognates among Chinese students will yield more reliable findings.

Authors' contributions

Hao Chen: conceptualization, research, methodology, data curation, software, visualization, resources, writing (original draft), writing (review of the draft and revision/correction); Ming Yang: conceptualization, formal analysis, resources, supervision, writing (original draft), writing (review of the draft and revision/correction), administration of the project.

Funding: No fundings

Conflicts of interest: The authors declare no conflict of interest that could affect the impartiality, objectivity or information presented in this research article.

References

- Aske, J. (2015). *Spanish-English Cognates, an Introduction to Spanish Linguistics*. World Languages and Cultures. SalemState University.
- Carcedo, A. (2000). *Disponibilidad léxica en español como lengua extranjera: el caso finlandés* (estudio del nivel preuniversitario y cotejo con tres fases de adquisición). Turun Yliopisto.
- Cenoz, J., & Gorter, D. (2020). Teaching English through Pedagogical Translanguaging. *World Englishes*, 39(2), 300-311.
- Cenoz, J., Leonet, O., & Gorter, D. (2022). Developing Cognate Awareness through Pedagogical Translanguaging. *International Journal of Bilingual Education and Bilingualism*, 25(8), 2759-2773.

- Chen, H. 陈豪. (2021). 西班牙语词汇研究综述和展望——兼议多语习得中的同源词研究 Overview and Prospect of Spanish Lexical Researches: A discussion of cognate words research in multilingual language acquisition. 外语教育研究 *Foreign Language Education & Research*, (4), 44-50. doi:10.16739/j.cnki.cn21-9203/g4.2021.04.008.
- Chen, H. 陈豪. (2023). 跨语言影响视域下英西同源词习得研究 *A Study on the Acquisition of English-Spanish Cognates from the Perspective of Cross-linguistic Influence*. 浙江大学出版社 Zhejiang University Press.
- Chen, H. 陈豪, & Liang, Q. 梁倩. (2023). 西班牙语三语习得者的语素意识研究 A Study on Morphological Awareness of Spanish Trilingual Learners. 语言学 Research *Linguistic Research*, 33, 155-167.
- Chen, H. 陈豪, & Song, C. Q. 宋楚翘. (Manuscript submitted for publication). Learning Spanish (L3) and Forgetting English (L2)? Intervention of Cognates Teaching in Lexical Production in China: A Longitudinal Comparative Study.
- Comesaña, M., Ferré, P., Romero, J., et al. (2015). Facilitative Effect of Cognate Words Vanishes when Reducing the Orthographic Overlap: The Role of Stimuli List Composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 614–635. <https://doi.org/10.1037/xlm0000065>
- Comesaña, M., Soares, A. P., & Lima, C. (2010). Semantic Representations of New Cognate vs. Noncognate Words: Evidence from Two Second Language Learning Methods. *Procedia-Social and Behavioral Sciences*, 5, 199-203.
- Comesaña, M., Soares, A. P., & Sánchez-Casas, R. & Lima, C. (2012). Lexical and Semantic Representations in the Acquisition of L2 Cognate and Noncognate Words: Evidence from Two Learning Methods in Children. *British Journal of Psychology*, 103(3), 378-392.
- Costa, A., Caramazza, A., & Sebastián-Galles, N. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283.
- Davies, M., & Davies, K. H. (2017). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Routledge.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, 41(4), 496-518.

- Dijkstra, T., & Van Heuven, W. J. (2002). The Architecture of the Bilingual Word Recognition System: From Identification to Decision. *Bilingualism: Language and Cognition*, 5(3), 175-197.
- Dressler, C., Carlo, M. S., Snow, C. E. et al. (2011). Spanish-Speaking Students' Use of Cognate Knowledge to Infer the Meaning of English Words. *Bilingualism: Language and Cognition*, 14(2), 243-255.
- Duchon, A., Perea, M., Sebastián-Galles, N. et al. (2013). Espal: One-stop Shopping for Spanish Word Properties. *Behavior Research Methods*, 45(4), 1246-1258.
- García, M. I. M. (2013). Estudio etimológico de cuatro pares de cognados en español y urdu. *Revista Iberoamericana de Lingüística (RIL)*, 8: 61-74.
- Gonzalo Pérez, A. (2015). *Los cognados sinonímicos como facilitadores de la adquisición y el aprendizaje del léxico español por alumnos angloparlantes*. [Tesis doctoral], Universidad Complutense de Madrid.
- Hampton, J., & Gardiner, M. (1983). Measures of Internal Category Structure: A Correlational Analysis of Normative Data. *British Journal of Psychology*, 74, 491-516.
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive Aspects of Lexical Availability. *European Journal of Cognitive Psychology*, 18(5), 730-755.
- Hidalgo Gallardo, M. (2017). La disponibilidad léxica como herramienta de evaluación interventiva del vocabulario en ELE: aplicación en un contexto universitario sinohablante. *SinoELE*, 16, 43-63.
- Hidalgo Gallardo, M. (2019). El perfil léxico del manual de ELE Español Moderno 1 desde la óptica de la disponibilidad léxica: posibles mejoras para la selección del vocabulario. *Ogigia. Revista Electrónica de Estudios Hispánicos*, 25: 233-257. [7https://doi.org/10.24197/ogigia.25.2019.233-257](https://doi.org/10.24197/ogigia.25.2019.233-257)
- Izquierdo Gil, M. C. (2003). *La selección del léxico en la enseñanza del español como lengua extranjera*. [Tesis doctoral] Universidad de Valencia.
- Koessler, M., & Derocquigny, J. (1928). *Les faux amis: ou, Les trahisons du vocabulaire anglais (conseils aux traducteurs)*. Vuibert.
- Li, X. K. 李晓科, & Cheng, L. L. (2006). 西班牙语专业四级词汇 *Spanish Major Level Four Vocabulary*. People's Education Press.
- López, H. (1995). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología*, 35(1), 245-259.

- Lotto, L., & De Groot, A. M. (1998). Effects of Learning Method and Word Type on Acquiring Vocabulary in an Unfamiliar Language. *Language Learning*, 48(1), 31-69.
- Martínez-Adrián, M., & Gallardo-del-Puerto, F. (2017). The Effects of Language Typology on L2 Lexical Availability and Spelling Accuracy. *International Journal of English Studies*, 17(2), 63-79.
- Morales, H. L. (1984). *La enseñanza de la lengua materna: lingüística para maestros de español*. Editorial Playor.
- Moss, G. (1992). Cognate Recognition: Its Importance in the Teaching of ESP Reading Courses to Spanish Speakers. *English for Specific Purposes*, 11(2), 141-158.
- Nagy, W. E., García, G. E., Durgunoğlu, A. Y. et al. (1993). Spanish-English Bilingual Students' Use of Cognates in English Reading. *Journal of Reading Behavior*, 25(3), 241-259.
- Nash, R. (1997). *NTC's Dictionary of Spanish Cognates: Thematically Organized*. National Textbook Company.
- Ortega, V. M. (2007). *La cognación entre el inglés y el castellano*. Universidad Nacional Experimental Politécnica.
- Otwinowska, A. (2016). *Cognate Vocabulary in Language Acquisition and Use: Attitudes, Awareness, Activation* (Vol. 93). Multilingual Matters.
- Otwinowska, A., & J. M. Szweczyk. (2017). The More Similar the Better? Factors in Learning Cognates, False Cognates and Non-cognate Words. *International Journal of Bilingual Education and Bilingualism*, 22(8), 974-991. <https://doi.org/10.1080/13670050.2017.1325834>
- Poort, E. D., & Rodd, J. M. (2017). The Cognate Facilitation Effect in Bilingual Lexical Decision is influenced by stimulus list composition. *Acta psychologica*, 180, 52-63.
- Poort, E. D., & Rodd, J. M. (2019). Towards A Distributed Connectionist Account of Cognates and Interlingual Homographs: Evidence from Semantic Relatedness Tasks. *PeerJ*, 7, e6725.
- Rodríguez, S. M. (2009). *Cognados y falsos cognados: su uso en la enseñanza del inglés*. Universidad de Guadalajara.
- Sánchez-Gutiérrez, C., Marcos Miguel, N., & Olsen, M. (2019). Words and Textbooks: An Analysis of Vocabulary Coverage and Lexical Characteristics in L2 Spanish Textbooks. In P. Ecke & S. Rott (org.), *Understanding Vocabulary Learning and Teaching: Implications for Language Program Development* (pp. 78-98). Cengage.

- Sánchez-Gutierrez, C.H., Robles-García, P., & Pérez Serrano, M. (2022). L2 Spanish Vocabulary Teaching in US Universities: Instructors' Beliefs and Reported Practices. *Language Teaching Research*, 1–23. <https://doi.org/10.1177/13621688221074443>
- Schröter, P., & Schroeder, S. (2016). Orthographic Processing in Balanced Bilingual Children: Cross-language Evidence from Cognates and False Friends. *Journal of Experimental Child Psychology*, 141, 239-246.
- Song, J. D. 宋尽冬. (2009). 中国西班牙语学习者可用词汇研究初探 A Tentative Study of Chinese Spanish Learners's Available Vocabulary. *外语与外语教学 Foreign Languages and Their Teaching*, (6), 7-11.
- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex*, 20(4), 912-928.
- Zhu, Y., & Mok, P.P.K. (2018). Visual Recognition of Cognates and Interlingual Homographs in Two Non-native Languages: Evidence from Asian Adult Trilinguals. *Linguistic Approaches to Bilingualism*, 10(4), 441-470.