This is a repository copy of *Spanish-English Cognates and False Cognates in Academic Spoken Vocabulary: Lexical coverage, orthographic and phonological transparency, and frequency*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/227245/

Version: Accepted Version

Article:

Spanish-English Cognates and False Cognates in Academic Spoken Vocabulary: Lexical coverage, orthographic and phonological transparency, and frequency

Abstract

Despite the positive effect of cognate knowledge for L2 vocabulary acquisition and the importance of comprehending academic spoken English for academic success, little is known about the nature of cognates in academic spoken English. This study identified Spanish-English cognates among academic spoken English words, represented by Authors' (2017a) Academic Spoken Word List (ASWL). Results showed that half of these words were Spanish-English cognates, more than 49% were non-cognates, and less than 1% were false cognates. Cognates and false cognates covered 10.53% and 0.002% of the academic spoken corpus, respectively. The ASWL cognates were more transparent in spelling than pronunciation. There was a close relationship between orthographic and phonological transparency. Words that have high frequency in English but low frequency in Spanish, words that have high frequency in Spanish regardless of their frequency in English, and words that have low frequency in both languages accounted for 61.85%, 35.29%, and 2.85% of the ASWL cognates, respectively. The study offers further insights into the value of recognizing cognates and each kind of cognate for academic vocabulary acquisition and provides useful implications for vocabulary assessment and pedagogy.

Keywords: cognate; false cognate; academic speech; Spanish; Academic Spoken Word List; vocabulary

1. INTRODUCTION

Academic words (e.g., *evaluate*, *construct*) are lexical items that occur frequently in a wide range of academic texts (Coxhead 2020). Even though knowledge of academic words is crucial for students' success in English-medium university courses, the lack of knowledge of these words is frequently cited as one of the greatest challenges faced by second language learners (e.g., Evans and Morrison 2011). Fortunately, the similarities between their first language (L1) and second language (L2) can help learners from certain L1 backgrounds to overcome these hurdles to some extent. For example, Spanish and English are both part of the Indo-European language family. Research has found that Spanish-English cognates made up a considerable percentage of academic words (74.74%) (Lubliner and Hiebert 2011) as well as specialized words (57.59%-85.5%) (e.g., Reed et al. 2013) (see the Background section for

the definition of cognates adopted in the present study). Moreover, Vidal's (2003, 2011) experiments with Spanish learners in English for Specific Purposes (ESP) programs also showed that *predictability from word form and parts*, of which cognates were a sub-group, was the best predictor of vocabulary learning gains through listening to academic lectures. Given the percentage of Spanish-English cognates among academic words and their positive effect on vocabulary learning, it is expected that Spanish L1 students may have advantages over those from non-cognate L1 backgrounds when learning academic English words. It is important to note that previous studies only examined Spanish-English cognates in academic *written* English. As university students are also expected to comprehend academic spoken English (e.g., lectures and seminars) as part of their academic study, one question that arises is whether the findings of cognate research on academic written English also hold true for academic spoken English. Addressing this question is important for several reasons.

First, vocabulary in academic spoken English is distinctive from that in academic written English to a certain extent (e.g., Biber 2006; Simpson-Vlach and Ellis 2010; Authors 2020a). For example, Coxhead's (2000) Academic Word List, which represents academic written vocabulary, covers around 10% of the total number of running words in academic written English (see Coxhead 2016), but only 2.52%-4.6% in academic spoken English (Coxhead et al. 2017; Coxhead and Dang 2019; Dang and Webb 2014; Thompson 2006). Therefore, it is expected that the nature of cognates in academic spoken English would be different from that in academic written English.

Second, the differences in the way vocabulary is processed during listening or reading may lead to the difference in the difficulty of comprehending spoken and written discourse as well. Unlike written discourse, in spoken discourse, learners cannot refer to the orthographic overlap between English words and their Spanish cognates to support their comprehension. Moreover, they have to process the information spontaneously under real time circumstances (Biber 2006). Consequently, they may not have sufficient time to recognize cognates, especially those that have a greater overlap in orthographic forms than phonological forms. On the other hand, in spoken discourse, learners are exposed to the spoken forms of words whereas such information is not available to learners in written discourse. The opportunities to encounter the sound of cognates in spoken discourse may enable learners to better notice and recognize cognates that have a greater overlap in phonological forms than orthographic forms. Given the effects of modality on cognates, investigating the proportion of different kinds of cognates in academic spoken English would provide further insights into the relative

value of cognates for L2 learners, and this, in turn, may provide implications on how teachers can help these learners to improve their comprehension.

Another reason why it is worth investigating cognates in academic spoken English is related to the current view in defining academic vocabulary. Studies examining cognates in academic written English (e.g., Dalton 2008; Lubliner and Hiebert, 2011) followed Coxhead (2000) by defining academic words as those that are outside general high-frequency words. As a result, they only focused on academic words that are not general high-frequency words in English. However, recent studies have found that general high-frequency words accounted for a large percentage of words in academic spoken English (Authors 2020a), and a considerable number of L2 learners, including Spanish native speakers, have insufficient knowledge of these words (Olmos, 2009; Authors 2022d). To better understand the challenges of academic words for learners with different language proficiency levels, it is useful to examine cognate academic words regardless of whether they are general high-frequency words or not.

Taken together, it is important to determine the percentage of cognates in academic spoken English, but no studies have investigated this area. The present study aims to address this gap by examining the percentage of Spanish-English cognates among the most common words of academic spoken English and the nature of these cognates. Moreover, it further classified these cognates into different kinds in terms of orthographic and phonological transparencies and frequencies in both English and Spanish. Its findings would provide further insights into the value of being able to recognize cognates in general and each kind of cognates in particular for comprehension of academic spoken and written texts. Such information would allow teachers to be better aware of the challenges of learning academic words faced by students from English cognate and non-cognate L1 backgrounds, and thus to offer relevant support to each group of learners. The present study would also inform the design of vocabulary tests for students from English cognate and non-cognate L1 backgrounds. Although this study focused on Spanish-English cognates, to the best of our knowledge, it is the first to explore the nature of cognates in academic spoken English. Therefore, it would provide implications for further research examining cognates in other languages.

2. BACKGROUND

Cognates and false cognates

There are two approaches toward defining cognates. The narrow approach considers cognates as the words from two languages that share a common ancestor and have similar meanings and similar orthographic and/or phonological forms (Duñabeitia et al. 2010). This approach is useful for linguists who are interested in investigating the origins of words. However, it is not very relevant for research on second language acquisition (SLA). This approach does not acknowledge cognates resulting from borrowings nor does it reflect the fact that the similarities in forms and meanings of the words are more important for L2 learners than the word origins (Daulton 2008). Therefore, a common approach in SLA research is to consider cognates as the words with similar meanings and similar orthographic and/or phonological forms regardless of their ancestors (Rogers et al. 2015). In other words, loan words are also considered as cognates. This broad definition was adopted in the present study.

Cognates have a positive effect on vocabulary and language learning. English language learners from cognate backgrounds, including Spanish native speakers, tend to achieve much higher scores on vocabulary tests as well as tests on other language skills than those from non-cognate backgrounds (Laufer-Dvorkin 1991). For example, Vidal's (2011) study with Spanish ESP learners revealed that cognateness is one of the key factors contributing to vocabulary learning from listening to academic lectures and reading academic texts. Similarly, Urdaniz and Skoufaki (2022) examined the effects of cognateness, word frequency, and length on Spanish EFL learners' receptive knowledge of academic written words. They found that word frequency was the strongest predictor of item facility variance. Cognateness was the second strongest predictor. The frequency by cognateness interaction ranked third, indicating that word frequency was more predictive of item facility variance for cognates than non-cognates. Word length did not have a significant effect on item facility variance. Moreover, other studies have indicated that cognates are likely to be learned more easily and retained longer than non-cognates (e.g., Rogers et al. 2015). This is because the semantic and formal similarities between cognates in L1 and L2 may lighten the learning burden of cognates compared to non-cognates (Nation, 2022).

Unlike cognates, false cognates are words that have similar orthographic and/or phonological forms in the two languages but refer to different concepts (de Groot 2011). For example, *embarazada* (meaning 'pregnant' in Spanish) and *embarrassed* in English. Findings of

research on the effects of false cognates on learning are less conclusive. Several studies have found that learners produced more errors in L2-L1 translation tasks (e.g., Janke and Kolokante 2015), and exhibited lower accuracy and longer reaction time in L1-L2 translation tasks for false cognates than for non-cognates (Christoffels et al. 2015). These findings could be due to the interference between L1 and L2 meanings (Durlik et al. 2016). Conversely, other studies did not find that false cognates hindered vocabulary learning nor did they report evidence of the interference of L1 meaning on the acquisition of L2 cognates (Marecka et al. 2021). The mixed results indicate the complex nature of false cognates and the importance of further investigating their role in L2 learning.

Cognates in specialized vocabulary

Recognizing the importance of cognateness for learning and the significance of specialized vocabulary for academic study, several studies have examined the percentage of cognates in specialized vocabulary. Most of them have examined Spanish-English cognates among technical words in specialized texts: engineering magazines (Martínez, 1994), science texts (Bravo et al. 2007), and biology textbooks (Reed et al. 2013). In these studies, lists of technical words were created from these specialized texts. Then, items from these lists were classified into cognates, false cognates, and non-cognates. These studies found that cognates accounted for a considerable percentage of technical words: 62.26% (Martínez, 1994), 76% (Bravo et al. 2007), and 93% (Reed et al. 2013). In contrast, false cognates only made up a very small percentage: 2.33% (Martínez, 1994) and 1% (Reed et al. 2013). Among these studies, Martínez did not report the steps taken in the classification of cognates, false cognates, and non-cognates, but Bravo et al. and Reed et al. did. They both had items in the technical word lists translated into Spanish by native speakers of Spanish and then checked against a Spanish-English dictionary. Based on the translation, they then identified cognates and false cognates. Bravo et al. did not report whether they made these judgements in the classification of cognates, false cognates, and non-cognates themselves or involved other raters whereas Reed et al. had the classification task conducted by native speakers of Spanish.

While Martínez (1994), Bravo et al. (2007), and Reed et al. (2013) focused on technical vocabulary, Lubliner and Hiebert (2011) examined Spanish-English cognates in academic written English. They used Coxhead's (2000) Academic Word List (AWL) to represent academic words. Similar to previous studies, they asked a native speaker of Spanish and a proficient speaker of Spanish to translate the AWL headwords into Spanish and used this information to identify cognates. The results show that 74.74% of the AWL words were

Spanish-English cognates. While Lubliner and Hiebert provided useful information about cognates in academic English, they did not report the percentage of false cognates in the AWL. As previously mentioned, given the mixed findings related to L2 acquisition of false cognates, it would be useful to examine their percentage in academic vocabulary. Moreover, Lubliner and Hiebert did not report how they checked the frequency of the cognates in the two languages despite the fact that frequency in L1 and L2 plays a significant role in vocabulary learning (Peters 2020).

Compared to Spanish, studies examining cognates of English academic words in other languages are limited in number. For example, Daulton (2008) combined dictionary checking and the human rating to examine Japanese-English cognates in academic written English represented by Coxhead's (2000) AWL. He found that Japanese-English cognates accounted for about 27% of the AWL. This is lower than the percentage of Spanish-English cognates found by Lubliner and Hiebert (2011). This finding suggests that perhaps Spanish L1 learners may have advantage over Japanese L1 learners when learning the AWL and highlights the importance of examining the ratio of cognates in different L1s to anticipate the challenges faced by learners of English from different L1s when learning items from the same word lists.

Taken together, previous studies have indicated that a considerable number of specialized English words are cognates. However, they only examined the percentage of cognates in lists of specialized vocabulary. It would be useful to examine the lexical coverage of these cognates in corpora representing the target specialized texts as well. Such information would allow us to estimate the percentage of cognates among the total number of words in specialized texts, which then provides further information about the value of recognizing cognates for comprehending such discourse. Moreover, previous cognate studies have primarily focused on specialized vocabulary in written discourse. Investigating cognates in academic spoken English is also valuable because cognates in academic spoken English may differ from those in academic written English due to the linguistic variations between the two modalities.

The lack of cognate research on academic spoken discourse could be due to the central role of academic reading and writing in academic study and the lack of specialized spoken word lists. Another reason, as suggested by one anonymous reviewer, is the lack of available word lists for non-English languages. However, in recent years, the importance of engaging in academic spoken discourse (e.g., lectures, conference presentations) has been brought forward. Advances in technology have also made collecting academic spoken data and data

from languages other than English less challenging than it used to be. As a result, several word lists, such as the Academic Spoken Word List (Authors 2017a), have been created from academic spoken corpora to capture the core vocabulary in academic spoken English and several lists for non-English languages (e.g., Davies and Davies 2019) have been developed. The development of such lists makes it possible to examine cognates in academic spoken English. Yet no research has been conducted to fill this gap.

Effect of frequency and orthographic and phonological transparency on cognate learning

L2 vocabulary learning is affected by many factors, which are inter-related to each other (Peters 2020; Webb & Nation 2017). This section focuses on the effect of L1 and L2 frequency and orthographic and phonological transparency on the learning of cognates.

Frequency was chosen as the focus because it plays a significant role in the learning of L2 words in general and cognates in particular. Peters (2020) reviewed research on the effect of L1 and L2 frequency on L2 vocabulary learning and reported that L1 frequency had a small but significant impact on the learning of L2 words. The more frequently a word occurs in L1, the more likely L2 learners are familiar with the concept associated with this word. This familiarity may then facilitate these learners' acquisition of the L2 word related to this concept. Similarly, L2 frequency also positively contributes to the learning of L2 words in that the more learners encounter an L2 word, the more likely they will notice and learn it (e.g., Dang, Webb, and Coxhead 2022; Ellis 2002). Previous studies also provided evidence on the effects of word frequency on the learning of cognates. Stoeckel and Bennett (2013) tested the knowledge of general and academic words of Japanese and Korean EFL students and found that frequency of cognates in L1 predicted the direction of differential item functioning for cognates. Later, Urdaniz and Skoufaki (2022) tested Spanish EFL university students' knowledge of academic words and found that L2 frequency was the strongest predictor of item facility variance, followed by cognateness and then the interaction between L2 frequency and cognateness. Urdaniz and Skoufaki's finding suggests that while L2 frequency had a positive effect on L2 vocabulary learning, its effect may be weaker for cognates than non-cognates. The weaker effect of L2 frequency for cognates could be due the effect of L1 frequency. In fact, Bennett and Stoeckel (2014) analyzed Japanese EFL students' scores on a test of general and academic vocabulary and found that L1 frequency better predicted learners' vocabulary knowledge than L2 frequency. Similarly, Skjelde (2022) measured Norwegian EFL university students' knowledge of academic words which are also

cognates and found that nearly half of these target words were translated by all participants with a Graeco-Latin cognate, and that most of these words were also high-frequency academic words in the participants' L1. These findings suggested that L1 frequency tends to positively affect knowledge of cognates.

In addition to frequency, it is also important to examine the orthographic and phonological transparency of cognates. The similarities in forms and meanings of words in L1 and L2 may facilitate the learning of new words in L2. Previous studies (e.g., Otwinowska and Szewczyk 2019) found that the greater the degree of orthographic and phonological transparency, the easier it is for L2 learners to transfer word meaning across languages. In other words, when L2 users process cognates, the orthographic and phonological representations from both languages in their mental lexicon are activated in parallel, which may then activate their semantic representation and enable these users to recognise or recall the words faster (Dijkstra et al. 2018).

Recognizing the important role of L1 and L2 frequency and orthographic and phonological transparency in cognate learning, several studies have analysed specialized vocabulary from these perspectives. All of them examined Spanish-English cognates. Three studies examined L1 and L2 frequency of cognates in specialised languages (Bravo et al. 2007; Lubliner and Hiebert 2011; Reed et al. 2013). They all used corpora for information about the frequency of cognates. Lubliner and Hiebert (2011) did not provide detailed information about the corpora from which the frequency information was derived, whereas Bravo et al. and Reed et al. did. Bravo et al. reported using Zeno et al.'s (1995) frequency database and the *Corpus del Español* (Davies 2001) to check the frequency of the cognates in English and Spanish, respectively. Reed et al. checked the English frequency of cognates in a more current corpus of English (Corpus of Contemporary American English, COCA; Davies 2008). They also used Davies's *Corpus del Español*. While Bravo et al. used the whole corpus of texts from 1970s to 1990s (100 million words) for the analysis, Reed et al. only used texts from the 1990s (20 million words) to make the Spanish corpus comparable to COCA. Together, the findings of Bravo et al. (2007), Lubliner and Hiebert (2011) and Reed et al. (2013) indicate that the frequency patterns of cognates may vary depending on the kind of vocabulary. In the case of technical vocabulary, both Bravo et al. and Reed et al. found that cognates having low frequency in both languages accounted for the largest percentage of cognates: 46.97% (Bravo et al. 2007) and 88.36% (Reed et al. 2013). Next came those having high frequency in both languages (13.64% in Bravo et al. and 33.36% in Reed et al.) and those having higher

frequency in Spanish but low frequency in English (36.36% in Bravo et al. and 3.66% in Reed et al.). Words with high frequency in English but low frequency in Spanish accounted for the smallest percentage: 3.03% (Bravo et al. 2007) and 1.77% (Reed et al. 2013). The fact that cognates having low frequency in both languages accounted for the largest percentage of cognates in technical vocabulary indicated that explicit teaching of these cognates may be useful because learners may not be able to refer to their prior vocabulary knowledge in L1 or L2 to infer the meaning of these words.

Different results were found in the case of academic vocabulary, Lubliner and Hiebert's (2011) found that of the 360 AWL cognates pairs whose Spanish frequency information was available, 49.72% were high-frequency words in Spanish but had low frequency in English, 26.40% were high-frequency words in both English and Spanish, 21.66% were low frequency words in both English and Spanish, and 2.22% were high-frequency words in English but low frequency in Spanish. These findings indicate that explicit teaching of cognates which have high frequency in Spanish is probably not worth the limited classroom time because learners may be able to refer to their L1 knowledge to infer the meaning of these words. One possible reason for the discrepancy between Lubliner and Hiebert's and the other two studies is that Lubliner and Hiebert examined academic words whereas Bravo et al. and Reed et al. investigated technical words.

Compared to studies on L1 and L2 frequency, few studies have investigated the transparency of cognates in specialized vocabulary. Martínez (1994) classified cognates into actual cognates and potential cognates. Actual cognates were those having 'very close similarity between the Spanish and English forms' (e.g., *total/total, additional/adicional*), including borrowing from other languages (e.g., *software, rate*) (Martínez 1994: 87). On the other hand, potential cognates had a lower degree of similarities between the Spanish and English forms (e.g., *offer/ofrecer*). This approach has a high level of subjectivity because it is unclear what was considered as lower and higher degrees of similarities. Therefore, Lubliner and Hiebert used the longest common subsequence ratio (LCSR) method and the common phoneme ratio (CPR) to represent orthographic transparency and phonological transparency, respectively. As Lubliner and Hiebert's method based on the ratio of shared letters (in the case of LCSR) and shared phonemes (in the case of CPR) of the words in two languages, it is more transparent than Martínez's (1994). Lubliner and Hiebert (2011) found that the mean of LCSR (0.73) was higher than that of CPR (0.49), suggesting that the AWL cognates were more transparent in terms of spelling than pronunciation. However, no significant correlation

was found between the LCSR and CPR ratios, which indicates no significant association between orthographic and phonological transparency.

Taken together, focusing on Spanish-English cognates, previous studies provide useful insights into the nature of cognates in specialized vocabulary. It was found that the frequency patterns of cognates may vary depending on the kind of vocabulary. In the case of academic vocabulary, words having high frequency in Spanish but low frequency in English accounted for the largest percentage of cognates. They were followed by those having high frequency in both languages and then those having low frequency in both languages. Word having high frequency in English but low frequency in Spanish accounted for the smallest percentage of cognates. Previous research also found that academic words are more transparent in terms of spelling than pronunciation and orthographic transparency and phonological transparency were not likely to associate with each other. Despite their insights, previous studies focused on academic *written* discourse. It is unclear whether their findings hold true for vocabulary in academic *spoken* discourse. Given the different nature of academic spoken and written discourses (Biber 2006), the findings may be different for academic spoken discourse. First, general high-frequency words make up a larger proportion of academic spoken vocabulary than academic written vocabulary (Dang 2018). It is possible that cognates with high frequency in L2 may account for a larger percentage of cognates in academic spoken vocabulary than in academic written vocabulary. Second, given its spoken nature, academic spoken words may be more transparent in terms of pronunciation than spelling. However, no previous studies have tested these hypotheses. Such studies are important because their findings can provide further insights into the similarities and differences between cognates in spoken and written English. This may enable teachers to estimate the learning burden that each kind of cognates may pose to L2 learners and plan their support accordingly.

3. RESEARCH QUESTIONS

To address the gaps identified in the literature review, this study aims to identify Spanish-English cognates among academic spoken English words represented by Authors' (2017a) Academic Spoken Word List and determine the transparency and frequency patterns of these cognates. It will address four research questions:

1. How many lexical headwords in the Academic Spoken Word List (ASWL) are cognates and false cognates with Spanish?

2. What is the coverage of these cognates and false cognates in academic spoken English?

3. How can the identified cognates be classified in terms of orthographic and phonological transparencies?

4. How can the identified cognates be classified in terms of frequency in English and Spanish?

4. METHODOLOGY

4.1. Word list

The list of ASWL lexical flemmas (Authors 2017a) was used for cognate analysis. It consists of items that occur frequently in an English academic spoken corpus of 13-million words. This corpus represents the language in academic lectures, seminars, lab sessions, and tutorials from 28 different subject areas. The ASWL provides coverage of about 90% of words in different kinds of academic spoken English (Authors, 2019; 2022a, 2022b), which indicates that it represents academic spoken words in English to a large extent.

One innovative feature of the ASWL compared to the AWL is that it does not assume that learners already have knowledge of high-frequency words. Therefore, general high-frequency words are still included in the ASWL if they meet the selection criteria. In particular, the ASWL words are classified into four levels based on their frequency in general English. Items at Levels 1-3 of the ASWL are also general high-frequency words (1st, 2nd, and 3rd 1,000 most frequency general words, respectively). Items at Level 4 are words at lower frequency levels (outside the most frequent 3,000 words of general English). Although the ASWL was originally a word family list, only ASWL flemma headwords[1] (i.e., not inflected forms) were used in the analysis of this study. We focused on flemma headwords rather than word family headwords because not all members of the ASWL word families are cognates. Moreover, only ASWL lexical flemmas were analyzed because preliminary analyses indicated that none of the ASWL function words were cognates. In brief, the list of 4,761 ASWL lexical flemma headwords was used for the analysis of Spanish-English cognates in the present study.

4.2. Identifying cognates, non-cognates, and false cognates

The identification of cognates included three main steps. The first step involved identifying the Spanish translations of the ASWL headwords. To begin with, the second author, who is

an advanced speaker of Spanish, and the third author, who is a native speaker of Spanish, translated all the English headwords into Spanish. Both had a background in applied linguistics and were also advanced speakers of English. Various online English-Spanish dictionaries were consulted (Linguee, Cambridge Dictionary, Oxford, Diccionario de la Lengua Española – Real Academia Española). Multiple steps were taken in the translation stage to ensure the reliability of the translation. First, maximum effort was made to avoid any regional specific vocabulary. Instead, the focus was on the Spanish equivalents that were widely used across the different Spanish speaking countries. This criterion ensured that the results of the present study were relevant to a large number of L2 learners. Second, if an English flemma functioned both as a noun and verb, or a verb and adjective, both forms were translated into Spanish. This criterion helped us to identify all possible cognates of the ASWL headwords. Finally, when all the headwords were translated, a professional translator of Spanish and English performed a final check and made a few changes if he came up with a more appropriate term.

The second step involved checking the pronunciation of the ASWL English headwords and their Spanish translations. The first author, who is a proficient speaker of English, checked the English pronunciation (i.e., phonetic transcription) of the ASWL headwords in the *Oxford Learner's Dictionary of English*. This dictionary is popular among L2 learners and is informed by updated evidence, including corpus-driven evidence, from one of the largest language research centers in the world. In the meantime, the second and third authors transcribed all the Spanish equivalents. The Spanish phonetic transcription was conducted using the website EasyPronunciation.com[2], more specifically the option 'Convert text to phonetic transcription'.

The third step included identifying cognates. The second and third authors independently classified the ASWL headwords into three groups based on their spelling, pronunciation, and meaning. The first group was cognates (words that have similar meaning and are similar in either spelling or pronunciation, or both). The second group was false cognates (words which have different meanings but have similar spelling or pronunciation or both). The last group was non-cognates (words which are different in spelling, pronunciation, and meanings). For example, the English-Spanish pair *able/capaz* was coded as 0, whereas the pair *campus/campus* was coded as 6. The rules of cognate analysis were initially based on Lubliner and Hiebert (2011) and then adapted and simplified based on the results of an

analysis of a small sample of data (see Appendix 1 for the detailed coding protocol and Appendices 2 and 3 for the rules of the cognate analysis).

During the classification of cognates, non-cognates and false cognates, we found that while the majority of English words had only one translation that was used in a wide range of Spanish-speaking regions, a number of them had more than one Spanish translation (i.e., they were polysemous). To deal with these polysemous items, we examined all of these translations. If a word had one or more of these translations that were cognates, it would be considered cognate no matter how frequent the translations were. For example, *arm* has several translations that have been used frequently in a wide range of Spanish speaking regions: *brazo, arma*, and *armar*. As *arma* and *armar* were rated as cognates, *arm* was rated as cognate[3]. This approach was employed to make our method more straightforward and allow future studies to replicate our study. One limitation of this approach is that it may underestimate the learning burden of polysemous words whose non-cognate and/or false cognate sense (translation) is more frequent in academic spoken English than their cognate sense (translation). However, our further analysis indicated that this is unlikely a major issue in our study. Of the 2,417 ASWL cognates identified in the present study, 115 were polysemous (having more than one translation) (4.76%). The majority of these polysemous items (57.39%; 66 items) had translations as cognates only. Of the remaining items, in addition to the cognate translation, 30.43% (35 items) also had another translation as non-cognates, 10.43% (12 items) had another translation as false cognates, and 1.74% (2 items) had one translation as cognate and another as a false cognate.

When the rating of the two raters was compared, interrater reliability using Cohen's kappa showed that there was near perfect agreement across all four ASWL levels (Level 1: $\kappa = .95$, $p < .000$; Level 2: $\kappa = .930$, $p < .000$; Level 3: $\kappa = .960$, $p < .000$; Level 4: $\kappa = .936$, $p < .001$). The two raters discussed the areas of disagreement and finalized their rating. The list of cognates was then used for the transparency analysis and the frequency analysis. Information from these analyses is important because words that are less transparent and less frequent in the two languages are likely to have greater learning burdens than the others.

4.3. Lexical coverage analysis

Once the cognates and false cognates had been identified, we then used RANGE (Heatley et al. 2002) to analyze the lexical coverage of these items in academic spoken English. The list of identified cognates and false cognates were set as the baseword lists in RANGE. Authors'

(2017a) Academic Spoken Corpus was used to represent academic spoken English. To the best of our knowledge, this is the largest academic spoken corpus of English that has been created.

4.4. Transparency analysis

Although there are different approaches to measure transparency (e.g., Levenshtein distance), the transparency analysis of the cognates in the present study followed Lubliner and Hiebert's (2011) approach because it has been used to analyze the transparency of cognates in academic written vocabulary. Following the same approach would allow us to compare our findings with Lubliner and Hiebert's (2011) to provide insights into the differences and similarities between the transparency of cognates in spoken and written discourses. The longest common subsequence ratio (LCSR) and the common phoneme ratio (CPR) were used to determine the orthographic and phonological transparency of each cognate, respectively. The LCSR was calculated by dividing the number of shared letters between the two words by the number of letters of the longer words. For example, *aspect* (6 letters) and *aspecto* (7 letters) had 6 letters in common and the longer word (*aspecto*) had 7 letters. Therefore, the LCSR ratio was 0.86 (= 6 ÷7). Meanwhile, the CPR was calculated by dividing the shared phonemes of the two words by the number of phonemes of the longer word. For example, *aspect* /ˈæspekt/ had six phonemes and *aspecto/* as.ˈpek.to/ had seven phonemes. As a result, the CPR was 0.86 (= 6 ÷7). To conduct the analysis, for each cognate, the first author compared the spelling of the English word and its Spanish translation to identify the number of shared letters between the two words and the number of letters of the longer words. At the same time, the second and third authors compared the pronunciation of the English word and its Spanish translation to determine the number of shared phonemes between the two words and the number of phonemes of the longer words. Once the initial analysis has been completed, the three authors checked the whole data set for consistency. After that, the first author calculated the LCSR and the CPR, and conducted Pearson correlation analysis with SPSS 27.0 to determine the relationships between these ratios.

4.5. Frequency analysis

The frequency of the ASWL words in general English was checked against Nation's (2012) BNC/COCA 25,000-word lists. These lists were developed from two large corpora of English, the British National Corpus (BNC) and Corpus of Contemporary American English (COCA), which represent various kinds of spoken and written discourse in English. Following Schmitt and Schmitt (2014), items from the 1st to 3rd 1000 BNC/COCA word

family lists are considered as general high-frequency English words. Those from the 4th 1,000 onwards are considered as general low-frequency English words. The BNC/COCA lists were chosen because they are the most recent and largest frequency lists of general English. Moreover, previous studies (Authors 2017b; Authors 2022c) suggested that the high-frequency words in the BNC/COCA lists better represent high-frequency words in English than other high-frequency word lists.

The frequency of the Spanish translations of the ASWL cognates was checked against the Spanish frequency word lists (Davies and Davies 2019) which are available via Finlayson et al.'s (2023) MultilingProfiler at https://www.multilingprofiler.net/. These frequency lists are the most recent and comprehensive frequency lists of Spanish. They were developed from a large corpus of Spanish that represents texts from different kinds of spoken and written discourse. Moreover, similar to the BNC/COCA list, Davies and Davies's (2019) Spanish frequency lists, which are available in the MultilingProfiler, are classified into 1,000-word frequency levels. Items from the 1st to 3rd 1,000-word lists are considered as general high-frequency Spanish words while those outside these lists are general low-frequency Spanish words. Davies and Davies's lists were purely based on frequency while frequency and other criteria were used to select items for Nation's BNC/COCA lists. Therefore, these lists are not perfectly comparable. Despite this limitation, to the best of our knowledge, Davies and Davies's (2019) lists are the only available Spanish frequency lists that have similar components as the BNC/COCA lists. Therefore, using MultilingProfiler for Spanish frequency checking against Davies and Davies' lists is an innovation of our study. It ensured that the classification of items in the Spanish frequency lists would be consistent with that in the English frequency lists (BNC/COCA lists), which then enabled us to make a transparent comparison of the frequency of these cognates in the two languages.

After the frequency of the cognates in English and Spanish had been determined, these cognates were categorized into four groups as:

(a) Having low frequency in both English and Spanish
(b) Having high frequency in English but low frequency in Spanish
(c) Having high frequency in both English and Spanish
(d) Having low frequency in English and high frequency in Spanish

5. RESULTS

5.1. Cognates, false cognates and non-cognates among ASWL lexical words

Table 1 presents the percentage of cognates, non-cognates, and false cognates in the ASWL. Cognates accounted for 50.77% of the total number of ASWL lexical words while non-cognates made up 49.17%. Only 0.06% of the ASWL lexical words were false cognates. However, the proportion of cognates, non-cognates, and false cognates slightly varied across the ASWL levels. The percentage of cognates significantly increased from Level 1 to Level 4 while the percentage of non-cognates significantly decreased across these levels. Nearly 73% of the ASWL level 1 words were non-cognates, leaving more than 27% of the ASWL words at this level being cognates. In contrast, from Level 2 to Level 4, cognates always accounted for a larger percentage than non-cognates, and the gap in percentage became larger from Levels 2 to 4. At Level 2, the percentage of cognates (56.42%) was only slightly higher than that of non-cognates (43.42%). However, the percentage of cognates was nearly four times as large as the percentage of non-cognates at Level 3 and six times larger than the percentage of non-cognates at Level 4. In fact, cognates accounted for nearly 80% and more than 85% of the ASWL lexical words at Levels 3 and 4, respectively. Regardless of the ASWL level, false cognates accounted for the smallest number of ASWL lexical words, and all of them were at Levels 1 and 2 of the ASWL.

<INSERT TABLE 1 ABOUT HERE>

TABLE 1.

Percentage of Different ASWL Lexical Words in Terms of Cognateness (%)

Group	Level 1	Level 2	Level 3	Level 4	All levels
Number of lexical flemmas	*1,957*	*1,480*	*1,166*	*158*	*4,761*
Cognates	27.08%	56.42%	78.64%	85.44%	50.77 %
Non-cognates	72.87 %	43.45%	21.36%	14.56%	49.17%
False cognates	0.05%	0.14 %	0.00%	0.00%	0.06%

Table 2 presents the percentage of three kinds of cognates in the ASWL: those having similar spelling and similar pronunciation (e.g., *campus* in English and *campus* in Spanish), those having similar spelling but different pronunciation (e.g., *micro* /maɪkrəʊ/ in English and *micro* /ˈmi.kɾo/ in Spanish), and those having different spelling but similar pronunciation

(e.g., *physics* /ˈfɪzɪks/ in English and *física* / ˈfi.si.ka / in Spanish). The last column of this table shows that cognates with similar spelling and pronunciation accounted for the largest percentage of cognates (nearly 50%), followed by cognates with similar spelling but different pronunciation (more than 30%), and then cognates with different spelling but similar pronunciation (less than 20%). The second to fifth columns show that regardless of the ASWL levels, cognates with similar spelling and pronunciation always accounted for the largest percentage (46.42%-57.78%). Moreover, the percentage of these words among the cognates increased from Level 1 to Level 4. Cognates with similar spellings but different pronunciation always accounted for a larger percentage of cognates at Level 1 to Level 3 than those with different spelling but similar pronunciation. However, at Level 4, the percentage of these words is roughly the same (around 20%).

<INSERT TABLE 2 ABOUT HERE>

TABLE 2.

Percentage of Different Kinds of Cognates in the ASWL Lexical Words (%)

Sub-group	Level 1	Level 2	Level 3	Level 4	Total
Similar spelling, similar pronunciation	46.42	46.71	51.04	57.78	48.90
Similar spelling, different pronunciation	34.15	34.49	34.02	20.74	33.47
Different spelling, similar pronunciation	19.43	18.80	14.94	21.48	17.63
Total	100	100	100	100	100

5.2. Coverage of cognates and false cognates in the academic spoken corpus

Our analysis also showed that overall, the ASWL cognates covered 10.53% of the total number of words in Authors (2017a) academic spoken corpus while false cognates covered only 0.002%. Of the three kinds of cognates, those with similar spelling and similar pronunciation accounted for the highest coverage (5.47%). Next came cognates with similar spelling and different pronunciation (2.57%), and then cognates with different spelling and similar pronunciation (2.49%).

5.3. Transparency analysis

Table 3 presents the results of the analysis of spelling and pronunciation transparency of the ASWL cognates. Regardless of the levels, these cognates were largely transparent in terms of spelling (LCSR ratio of around 0.7-0.8) and relatively transparent in terms of pronunciation (CPR ratio of around 0.5). The higher transparency in terms of spelling may be because

cognates with similar spelling but different pronunciation accounted for a larger percentage of ASWL cognates than those with different spelling but similar pronunciation.

<INSERT TABLE 3 ABOUT HERE>

TABLE 3.

Mean and Standard Deviation of the Longest Common Subsequence Ratio (LCSR) Ratio and the Common Phoneme Ratio (CPR)

ASWL level	Level 1	Level 2	Level 3	Level 4	All
LCSR	0.73 (0.17)	0.75 (0.16)	0.77 (0.15)	0.76 (0.16)	0.76 (0.16)
CPR	0.49 (0.15)	0.52 (0.15)	0.53 (0.14)	0.52 (0.15)	0.52 (0.15)

Note: SD values are in brackets

Pearson correlation analysis revealed a medium[4] and positive correlation between the LCSR and CPR ratios of the ASWL cognates for the whole list ($r = 0.51$; $p < 0.001$). Significant correlations were also found at each ASWL level. The correlations were large at Level 1 ($r = 0.56$; $p < .001$) and medium at Level 2 ($r = 0.51$; $p < .001$), Level 3 ($r = 0.48$; $p < .001$) and Level 4 ($r = 0.37$; $p < .001$).

5.4. Frequency analysis

Table 4 presents the results of the frequency analysis. Cognates that have high frequency in English but low frequency in Spanish accounted for the largest percentage of ASWL cognates (nearly 62%). The next group were cognates that have high frequency in both English and Spanish (32.56%). As for cognates that have low frequency in both English and Spanish and cognates that have low frequency in English but high frequency in Spanish, each group made up less than 3% of the ASWL cognates. It should be noted that among the 135 ASWL cognates that are low frequency words in English (Level 4 ASWL words), 48.89% had high frequency in Spanish and 51.11% had low frequency in Spanish.

<INSERT TABLE 4 ABOUT HERE>

TABLE 4.

Frequency of the ASWL Cognates in English and Spanish

Group	%
High frequency in English but low frequency in Spanish	61.85
High frequency in both English & Spanish	32.56
Low frequency in both English and Spanish	2.85
Low frequency in English but high frequency in Spanish	2.73
Total	100

6. DISCUSSION

This study is the first to explore cognates in academic spoken English. Based on the most recent lists of academic spoken English words (ASWL), general English words (BNC/COCA lists), and general Spanish words (MultilingProfiler lists) and a large and representative academic spoken corpus of English, the study identified Spanish-English cognates among the core academic spoken words of English and analyzed their transparency and frequency. Therefore, it provides new and deep insights into the nature of cognates and the value of recognizing cognates and their sub-kinds for L2 learners.

6.1. Nature of cognates in academic spoken English

The first and second research questions enquire about (a) the percentage of Spanish-English cognates and false cognates among the ASWL lexical words and (b) the lexical coverage of each group of words in academic spoken English. Regarding cognates, Spanish-English cognates accounted for more than half of the ASWL lexical words. Moreover, the percentage of cognates increased according to the ASWL level, from more than 27% in Level 1 to more than 56% in Level 2, nearly 80% in Level 3, and more than 85% in Level 4. The percentage of cognates among the ASWL Level 3 and Level 4 words found in this study is higher than the percentage of AWL cognates in academic written English reported by Lubliner and Hiebert (2011) (nearly 75%)[5]. Both Levels 3 and 4 of the ASWL and the AWL represent academic words that are outside the most frequent 2,000 English words. Therefore, the findings of this study indicate that cognates are likely to make up a larger percentage of academic spoken words than academic written words. Moreover, this study found that the ASWL cognates accounted for more than 10% of the total number of words in the academic spoken corpus of English. This is another new finding of the present study. No previous

cognate studies have examined the percentage of cognates in corpora representing the target specialized discourse despite the fact that such information can provide further insights into the value of cognates for comprehension of that particular discourse.

The large percentage of cognates among academic spoken words and their reasonable coverage in academic spoken English suggests that Spanish-English students have a greater advantage when learning academic spoken English words than learners from non-cognate L1 backgrounds. This finding supports previous studies (Martínez 1994; Bravo et al. 2007; Lubliner and Hiebert 2011; Reed et al. 2013). The present study extends these earlier studies by providing further insights into the importance of cognates in academic spoken discourse. In particular, the findings shed light on the learning burden of different groups of cognates. Nearly half of the Spanish-English ASWL cognates have similar spelling and pronunciation in the two languages. Spanish EAP learners may be likely to guess the meaning of these words when listening to academic spoken English. Therefore, the learning burden of these words may be lower than other kinds of cognates. In addition, more than 30% of the ASWL cognates have similar spelling but different pronunciation in the two languages. This finding suggests that learners may have more difficulty recognizing these words in oral communication than in written communication. In contrast, less than 20% of the ASWL cognates have similar pronunciation but different spelling in the two languages. Learners may find it more difficult to recognize these cognates in written communication than in oral communication (see further below for pedagogical implications).

Unlike cognates, false cognates only accounted for 0.06% of the ASWL lexical words and covered only 0.002% of the academic spoken corpus. This figure is consistent with the previous studies that showed that false cognates only made up 2.33% of technical words in engineering magazines (Martínez 1994) and 1% in biology texts (Reed et al. 2013). While previous research examined cognates of technical words and focused on written discourse, the present study advances this research agenda by investigating academic words and spoken discourse. It was also found that all of these false cognates are high-frequency words in English (at Levels 1 and 2 of the ASWL). The modest percentage of false cognates and the fact that they are all high-frequency words in English indicate that this group of words would unlikely cause problems for Spanish EAP learners.

The third research question concerns the transparency of the Spanish-English ASWL cognates. The ASWL cognates are largely transparent in terms of spelling (LCSR ratio of

0.76 for all levels) and relatively transparent in terms of pronunciation (CPR ratio of 0.52 for all levels). These figures are slightly higher than those found by Lubliner and Hiebert (2011) with the AWL (LCSR ratio of 0.73 and CPR ratio of 0.49)[6]. Moreover, while Lubliner and Hiebert did not find a significant relationship between the spelling and pronunciation of the AWL cognates, this study found medium and positive correlations between the LCSR and CPR coefficients of the ASWL words, which indicates a close relationship between orthographic and phonological transparency. The contrasting findings of the two studies provide further evidence on the distinctive difference between vocabulary in spoken and written English and highlight the importance of researching vocabulary in academic spoken discourse. Despite these differences, the present study found that the ASWL cognates are more transparent in terms of spelling than pronunciation, which is consistent with Lubliner and Hiebert's analysis of the AWL cognates. The greater similarities of cognates of academic words in terms of spelling when compared to pronunciation are probably due to the fact that 'Spanish vowels are highly regular and rarely correspond to their English equivalents in terms of pronunciation' (Lubliner and Hiebert 2011: 89).

The last research question is related to the frequency in English and Spanish of the Spanish-English ASWL cognates. The present study found that most of the ASWL cognates (nearly 62%) are high-frequency words in English but have low-frequency in Spanish. Their low frequency in Spanish means that Spanish-English learners are unlikely to know these words in their L1. However, as these words are general high-frequency words in English, if learners already know these words in English, they may not have much difficulty learning them in academic English. A considerable percentage of ASWL cognates are either high-frequency words in both English and Spanish (32.56%), or low-frequency words in English but are high-frequency words in Spanish (2.73%). This finding is consistent with Lubliner and Hiebert's (2011) finding that a large number of cognates in academic written English are everyday vocabulary in Spanish[7]. Given their high frequency in Spanish, these words are likely to be known by Spanish EAP learners and may not cause much learning burden for them. Less than 3% of the ASWL cognates have low frequency in both English and Spanish. This figure was much smaller than the percentage of the AWL cognates that have low frequency in both English and Spanish found by Lubliner and Hiebert (2011) (21.66%)[8]. The different findings may be because academic spoken English does not consist of as many general low-frequency words as academic written English (Authors 2020a). ASWL cognates that have low frequency in both English and Spanish are unlikely to be known by Spanish

native speakers. Therefore, teachers should provide learners with explicit instruction on these words.

6.2. Challenging the conventional definition of academic vocabulary

This study challenges a common assumption related to academic vocabulary. It is often assumed that EAP learners may not know academic words that are not general high-frequency words in English due to their low frequency, and thus need more support with these words than with general high-frequency words. This study shows that this is unlikely the case for Spanish EAP learners. More than 85% of the Level 4 ASWL words, which represent academic words that are not general high-frequency words, are Spanish-English cognates. This percentage is much larger than the percentage of cognates in Levels 1-3 words, which contain academic words that are also general high-frequency words in English. Moreover, nearly 50% of the Level 4 cognates have high frequency in Spanish. Level 4 also has the largest percentage of cognates with similar spelling and similar pronunciation (nearly 58%). In addition, none of the Level 4 ASWL words are false cognates. Together these findings suggest that the learning burden of academic words that are not general high-frequency words in English for Spanish EAP learners is probably not as high as expected.

This study also calls for more attention to academic words that are also general high-frequency words. The frequency analysis shows that most of the ASWL cognates are high-frequency words in English but have low frequency in Spanish. The low frequency of these cognates in Spanish means that Spanish EAP learners may not be able to rely on their L1 knowledge to facilitate the acquisition of these words. However, as these cognates are high-frequency words in English, if learners already have a solid knowledge of general high-frequency English words, such knowledge would help them to learn these cognates. Yet research has showed that a number of EAP learners, including Spanish L1 learners (Oslo 2009), have insufficient knowledge of high-frequency English words. Therefore, teachers should not assume that learners already know these words. Instead, they should test learners' knowledge of high-frequency English words to estimate their vocabulary knowledge.

This study has several limitations that could be addressed by future studies. First, it looked at the benefit of cognates from the perspective of comprehension (receptive knowledge). Research examining this issue from the perspective of production (productive knowledge) would be useful. Second, this study focused on general academic words. A growing number of specialized spoken word lists have been created such as the Hard Science Spoken Word

List (Authors, 2018a), the Soft Science Spoken Word List (Authors 2018b), the Medical Spoken Word List (Authors 2020b), and the list of spoken technical legal vocabulary (Bancroft-Billings 2020). Research examining cognates in these discipline-specific lists would be useful. Third, this study examined only Spanish-English cognates in academic spoken English. Research on cognates in other languages would be useful. Fourth, this study determined the learning burden of cognates based on analysis of transparency and frequency. Experimental studies with learners would provide further insights into the actual burden of different kinds of ASWL cognates to learners. Fifth, the cognate analysis involved cross-checking from various sources to identify the Spanish translations of the ASWL words, their spelling, pronunciations, and frequencies in both languages. Given this laborious work, we only analyzed flemma headwords because it resulted in a more manageable number of items for the analysis (4,761 items) than examining all flemma headwords and their members (13,448 items). However, one limitation of this approach is that analyzing only headwords misses to reflect the fact that headwords of some flemmas may be cognates but not their members, and vice versa. To give readers a better idea of the scale of this problem, we randomly sampled 10% of the lexical flemmas from each ASWL level; that is, a total of 476 flemmas (196 from Level 1, 148 from Level 2, 117 from Level 3, and 16 from Level 4). Table 5 shows that regardless of the ASWL levels, there is not much difference in the percentage of cognates among these flemmas when flemma headwords and flemma headwords plus members were examined. However, to provide a fuller picture of cognates in the ASWL, future studies could analyze both headwords and their members. Sixth, although we have tried our best to maximize the reliability of the pronunciation checking, there may be subjectivity involved in our rating.

TABLE 5.

Percentage of cognates among flemma headwords and flemma headword plus members in the 10% sample (%)

ASWL level	Flemma headwords	Flemma headwords plus members
Level 1	27.08	26.37
Level 2	56.42	52.82
Level 3	78.64	78.38
Level 4	85.44	89.19
All levels	50.77	47.88

Lastly, we did not use parallel corpora because some of the dictionaries that we used (e.g., Cambridge Dictionary) were derived from corpus analyses. However, future studies could consider using parallel corpora in their analyses.

7. IMPLICATIONS

Corpus-based specialized word lists such as the ASWL are useful resources for students in EAP and English-medium university programs. However, most of them are only available as lists of word forms. To maximize the usefulness of these lists, supplementary information about items in these lists is needed (Skoufaki and Petrić 2021). By identifying cognates in the ASWL and their patterns, this study effectively responds to this need and provides useful implications. As for assessment, given the positive effect of cognate knowledge on L2 vocabulary acquisition, cognates should be considered when designing vocabulary tests (Nation and Webb 2011; Read 2023). However, due to the lack of lists of academic spoken vocabulary and spoken corpora, most vocabulary tests tend to rely on lists of academic written words and the analysis of written texts. Relying solely on written texts to develop tests of spoken vocabulary may be misleading. As evidenced in this study, the nature of cognates in academic spoken discourse is different from that in academic written discourse. Therefore, the features of cognates in academic spoken English found in this study and the list of ASWL cognates (Appendices 4-7) are valuable to researchers who endeavor to design tests of academic spoken vocabulary (e.g., selecting target items, determining the sample ratio of cognates versus non-cognates in these tests). As for pedagogy, the considerable percentage of Spanish-English cognates in academic spoken English, especially those with the same spelling and pronunciation in two languages, suggests that if Spanish EAP learners are aware of the similarities between their L1 and English, they can make good use of them to expand their vocabulary knowledge. This would then support their comprehension of academic spoken English and enhance their confidence when participating in academic spoken interactions. However, it does not mean that teachers should take cognates for granted. The variation in transparency and frequency in the two languages means that students may not always be able to recognize them intuitively. Therefore, teachers should spend time helping learners to develop strategies and morphological and metalinguistic awareness to make good use of these cognates (Lubliner and Hiebert 2011; Rodriguez 2010). The list of cognates with information about the transparency and frequency developed in this study could inform teachers in setting the learning goals, selecting which cognates should be

the focus of explicit instruction and organizing learning activities for Spanish EAP learners effectively learn each kind of cognates. Unlike cognates, false cognates only account for 0.06% of the ASWL items and cover only 0.002% of the total number of words in the ASWL corpus. Given this tiny proportion, it may not be necessary for teachers to spend a significant amount of time teaching false cognates. However, it does not mean that false cognates should be ignored. Teachers could briefly mention the false cognates to raise learners' awareness of these items.

8. CONCLUSIONS

This study is the first to explore cognates in academic spoken English. Taking Spanish-English cognates as a case, the study shows that more than half of the academic spoken lexical words are Spanish-English cognates, less than 50% are non-cognates, and only less than 1% are false cognates. Cognates also made up a reasonable percentage of words in academic spoken English (more than 10%) while false cognates accounted for a very small percentage (0.002%). ASWL cognates are more transparent in terms of spelling than pronunciation, but there is a close relationship between orthographic and phonological transparencies. In terms of frequency, words that have high frequency in English but low frequency in Spanish accounted for the largest percentage of ASWL cognates, followed by words that have high frequency in Spanish regardless of their frequency in English. ASWL cognates that have low frequency in both English and Spanish made up the smallest percentage. The transparency and frequency analyses of these cognates indicate that learning the ASWL words may not be very challenging for Spanish EAP learners if they already have a solid knowledge of general high-frequency words in both Spanish and English. Overall, the study offers further insights into the value of cognates for academic vocabulary acquisition and provides useful implications for L2 vocabulary assessment and pedagogy. We hope that this study will open the door for more research on cognates in other L1 backgrounds.

NOTES

[1] The word family is made up of a base form, its inflected forms, and closely related derived forms whereas the flemma only includes the base form and its inflected forms. For instance, the flemma *view* includes *view, views, viewing,* and *viewed.* Meanwhile, in addition to these members, the word family *view* also consists of *viewable, viewer, and viewers.*

[2] To address the concern about the accuracy of the output from EasyPronunciation, we sampled 20% of the words from each ASWL level and asked a native speaker in Spanish, who had a PhD in Phonetics, to check the pronunciation of the sampled words. The accuracy rate was high across the four levels: 91.50% accuracy at Level 1, 90.50% of accuracy at Level 2, 93.50% of accuracy at level 3, and 89% of accuracy at Level 4. Most problems encountered with the sampled words were related to wrongly placed stress marks and incorrect syllable division. These issues would not have affected the accuracy of our analysis (when we counted phonemes).

[3] One concern with this approach is that the translations (senses) rated as cognates (e.g., *arma, armar*) may not occur in academic spoken English. To address this concern, we have conducted further analysis by checking concordance lines of all polysemous ASWL words in the academic spoken corpus. We found that all translations (senses) that were cognates occurred in the corpus. For instance, *arm* with the sense related to weapon (*arma*, *armar*) is considered as cognate and this sense indeed occurred in academic spoken English:

> *.. what if one state begins to **arm** itself more than it has done previously… (Political Sciences)*

[4] The interpretation of correlation size followed Plonsky and Oswald's (2014) guideline that *r* values "close to .25 be considered small, .40 medium, and .60 large" (p. 889).

[5] Lubliner and Hiebert adopted the word family headwords as the lexical unit while the present study employed flemma headwords. To directly compare our findings with Lubliner and Hiebert's, we converted the flemma headwords of the ASWL cognates into word family headwords and found similar results. 49.44% of the lexical ASWL word family headwords are cognates. There was an increase in the percentage of cognates from Level 1 (26.27%) to Level 2 (58.67%), Level 3 (72.37%), and Level 4 (90.67%).

[6] Results of the word family headword analysis are consistent with those of the flemma headword analysis. The ASWL cognates were largely transparent in spelling (LCSR ratio of 0.77 for all levels, 0.76 for Level 1, 0.76 for Level 2, 0.79 for Level 3, and 0.80 for Level 4). They were relatively transparent in pronunciation (CPR ratio of 0.53 for all levels, 0.51 for Level 1, 0.52 for Level 2, 0.55 for Level 3, and 0.55 for Level 4).

[7] Similar results were found in the word family headword analysis. The majority of word family headwords of the ASWL cognates are either high-frequency words in both English

and Spanish (64.34%), or low-frequency words in English but are high-frequency words in Spanish (4.82%).

[8] Further analyses using the word family headwords had similar results. A total of 3.81% of the word family headwords of the ASWL cognates have low frequency in both languages, which was much lower than the figure reported by Lubliner and Hiebert (2011).

REFERENCES

Bancroft-Billings, S. 2020. 'Identifying spoken technical legal vocabulary in a law school classroom'. *English for Specific Purposes*, *60*: 9–25.

Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Bennett, P., & Stoeckel, T. (2014). Word frequency and frequency of loanwords as predictors of word difficulty. *Verb*, *3*(2), 4–5.

Bravo, M., E., Hiebert and P. Pearson. 2007. 'Tapping the linguistic resources of Spanish/English bilinguals: The role of cognates in science' in Wagner,R.K, A. Muse, & K. Tannenbaum (ed.), *Vocabulary development and its implications for reading comprehension* (pp. 140–156). Guilford.

Christoffels, I., K. Timmer, L. Ganushchak and W. La Heij 2015. 'On the production of interlingual homophones: Delayed naming and increased N400'. *Language, Cognition and Neuroscience*, *31*/5: 628–638.

Coxhead, A. 2000. 'A new academic word list'. *TESOL Quarterly*, *34/*2: 213–238.

Coxhead, A. (2016). Reflecting on Coxhead (2000), 'a New Academic Word List'. *TESOL Quarterly*, *50*(1), 181–185.

Coxhead, A. 2020. 'Academic vocabulary' in S. Webb (ed.), *The Routledge handbook of vocabulary studies* (pp. 97–110). Routledge.

Coxhead, A., & Dang, T. N. Y. (2019). Vocabulary in university tutorials and laboratories: Corpora and word lists. In K. Hyland & L. L. C. Wong (Eds.), *Specialised English: New Directions in ESP and EAP Research and Practice* (pp. 120–134). Routledge.

Coxhead, A., Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, *30*, 66–78.

Dang, T. N. Y. (2018). The nature of vocabulary in academic speech of hard and soft sciences. *English for Specific Purposes*, *51*, 69–83.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76.

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). The relationship between lexical coverage, learner knowledge, and teacher perceptions of the usefulness of high-frequency words. *Foreign Language Annals*, *55*(4), 1212–1230. https://doi.org/10.1111/flan.12663

Daulton, F. E. 2008. *Japan's built-in lexicon of English-based loanwords*. Multilingual Matters.

Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Davies, M. and K. Davies. 2019. *A frequency dictionary of Spanish: Core vocabulary for learners* (2nd ed.). Routledge.

Dijkstra, T., Wahl, A., Buytenhuijs, F., van Halem, N., Al-jibouri, Z., de Korte, M., and Rekke, S. 2018. Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*: 657–679

de Groot, A. M. B. 2011. *Language and cognition in bilinguals and multilinguals: An introduction*. Routledge.

Duñabeitia, J., M. Perea, and M. Carreiras. 2010. 'Masked translation priming effects with highly proficient simultaneous bilinguals'. *Experimental Psychology*, *57*: 98–107.

Durlik, J., J. Szewczyk, M. Muszyński, and Z. Wodniecka. 2016. 'Interference and inhibition in bilingual language comprehension: Evidence from Polish-English interlingual homographs'. *PLoS One*, *11/*3.

Evans, S., and B, Morrison. 2011. 'Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong'. *English for Specific Purposes*, *30*: 198–208.

Finlayson, N., E., Marsden and L. Anthony (2023). 'Introducing MultilingProfiler: An adaptable tool for analysing the vocabulary in French, German, and Spanish texts'. *System.118*: 103122. https://doi.org/10.1016/j.system.2023.103122

Heatley, A., I. S. P., Nation. and A. Coxhead. (2002). *Range: A program for the analysis of vocabulary in texts* [Computer software]. http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx

Janke, V., and M. Kolokante 2015. 'False cognates: The effect of mismatch in morphological complexity on a backward lexical translation task'. *Second Language Research*, *31/*2: 137–156.

Jarvis, S. 2009. Lexical transfer. In Pavlenko, A. (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches*. Multilingual Matters

Laufer-Dvorkin, B. 1991. *Similar lexical forms in interlanguage*. Gunter Narr Verlag.

Lubliner, S. and E. H. Hiebert. 2011. 'An analysis of English–Spanish cognates as a source of general academic language'. *Bilingual Research Journal, 34/*1: 76–93.

Martínez, M. E. 1994. 'Spanish–English cognates in the subtechnical vocabulary found in Engineering magazine texts. *English for Specific Purposes*, *13/*1: 81–91.

Marecka, M., J. Szewczyk, A. Otwinowska, J. Durlik, M. Foryś-Nogala, K. Kutyłowska, and Z. Wodniecka. 2021. 'False friends or real friends? False cognates show advantage in word form learning'. *Cognition*, *206*: 104477.

Nation, I. S. P. 2007. 'The four strands'. *Innovation in Language Learning and Teaching*, *1/*1: 1–12.

Nation, I. S. P. 2012. *The BNC/COCA word family lists* [Computer software]. http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P. 2022. *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nation, I. S. P. and S. Webb. 2011. *Researching and analyzing vocabulary*. Heinle, Cengage Learning.

Olmos, C. 2009. 'An assessment of the vocabulary knowledge of students in the final year of secondary education. Is their vocabulary extensive enough?' *Journal of English Studies*, *Special issue*, 3–90

Peters, E. 2020. 'Factors affecting the learning of single-word items' in *Routledge handbook of vocabulary studie*s (pp. 125–142). Routledge.

Plonsky, L. and F. L. Oswald. 2014. 'How big is "big"? Interpreting effect sizes in L2 research' *Language Learning*, *64/*4: 878–912.

Read, J. 2023. 'Towards a new sophistication in vocabulary assessment', *Language Testing*, 40/1: 40-46.

Reed, D. K., L. A. Medina, N.A. Martínez. and L. G. Veleta. 2013. 'The accessibility of academic vocabulary to Spanish speaking high school biology students', *The High School Journal*, *97*/2: 80–91.

Rodriguez, T. A. 2010. 'Teaching cognates in isolation and context: a practical note on Montelongo et al. (2009)'. *Psychological Reports,*107/2: 564–566.

Rogers, J., S. Webb, and T. Nakata. 2015. 'Do the cognacy characteristics of loanwords make them more easily learned than noncognates?' *Language Teaching Research*, *19/*1: 9–27.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503.

Simpson-Vlach, R. and R. Ellis. 2010. 'An Academic Formulas List: New methods in phraseology research'. *Applied Linguistics*, *31/*4: 487–512.

Skjelde, K. M. (2022). Exploring L2 English students' knowledge and conceptions of academic vocabulary. *16*, *3*(4), 1–33. https://doi.org/10.5617/adno.9238

Skoufaki, S. and B. Petric. 2021. 'Exploring polysemy in the Academic Vocabulary List: A lexicographic approach' *Journal of English for Academic Purposes*, *54*: 101038.

Stoeckel, T., & Bennett, P. (2013). Sources of differential item functioning between Korean and Japanese examinees on a second-language vocabulary test. *Vocabulary Learning and Instruction*, *2*(1), 47–54.

Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253–270). Peter Lang.

Urdaniz, R. P., & Skoufaki, S. (2022). Spanish L1 EFL learners' recognition knowledge of English academic vocabulary: The role of cognateness, word frequency and length. *Applied Linguistics Review*, *13*(4), 661–703.

Vidal, K. 2003. 'Academic listening: A source of vocabulary acquisition?' *Applied Linguistics*, *24/*1: 56–89.

Vidal, K. 2011. A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61/*1: 219–258.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.