

An Analysis of English-Spanish Cognates as a Source of General Academic Language

Shira Lubiner

California State University, East Bay

Elfrieda H. Hiebert

TextProject &
University of California, Santa Cruz

TextProject Article Series

July 2014

TextProject, Inc.

SANTA CRUZ, CALIFORNIA

[To appear in *Bilingual Research Journal*, 2011, 34(1), 76-93]

Abstract

Three analyses of Spanish-English cognates were conducted, with the purpose of identifying features that might facilitate or inhibit bilingual students' cognate recognition and cross-language transfer of vocabulary. Results revealed that both GSL and AWL corpora contain a substantial number of English-Spanish cognates, a high percentage of which can be categorized by one of 20 cognate patterns. Orthographic and phonological transparency was analyzed, suggesting that cognates are more transparent in terms of orthography than phonology. A frequency analysis indicated that most AWL cognates are more common in Spanish than in English. Results suggest that Spanish-speaking students may have a "cognate advantage" if they are taught to recognize cognates.

An Analysis of English-Spanish Cognates as a
Source of General Academic Language

Students who speak English as a second language face a daunting task on the road to English literacy. They must learn a vast number of English words in order to comprehend the texts they are required to read in school. Researchers estimate that English-speaking students learn approximately 3,000 words per year (Nagy, Herman, & Anderson, 1985) and know as many as 75,000 words by the end of high school (Snow & Kim, 2007). The vocabulary development of Spanish-speaking English learners lags behind that of native English speakers at every level, putting them at risk for academic under-achievement (August, Carlo, & Snow, 2005; Snow & Kim, 2007). According to the recent National Assessment of Educational Progress (NAEP), only 17% of Latino fourth grade children scored at the proficient or advanced level in reading, compared to 42% of White students (Education Trust, 2007). Although Latino achievement on NAEP has improved during the past decade, the achievement gap remains a major concern.

Despite the challenge that they face, Spanish-speaking students may have an advantage not available to all English learners. The Spanish and English languages share a common alphabet and 10,000- 15,000 cognates, words that are Latin-based, mean approximately the same thing, and share similar orthographic features (Nash, 1997). The influence of Latin on the two languages has provided people who speak English and Spanish with a common linguistic heritage - a potential "fund of knowledge" (Moll, Armanti, Neff, & Gonzalez, 1992) that bilingual students bring with them to American schools. (All references to bilingual students in this paper refer to those who speak Spanish as their first language, unless otherwise stated.)

English learners usually acquire words used for basic communication quickly; however academic vocabulary is often much more difficult to master (Cummins, 1994). Academic vocabulary is a term used to describe the vocabulary needed for academic discourse and comprehension of content area texts. It includes words that are used for general academic functions such as analyzing, interpreting, and evaluating information across disciplines—words such as *observe*, *conclude*, *system*, and *process*. Other forms of academic language consist of the technical, concept-laden words that are unique to each discipline and literary vocabulary (Hiebert & Lubliner, 2008). All three forms of academic language—general academic, content-specific, and literary—are part of a sophisticated linguistic register that is heavily Latin-based. In this study we focus on one of these vocabularies—general academic vocabulary. Unlike the content-specific vocabulary that is central to subject area instruction such as science and social studies (e.g., *terrarium*, *frigid zones*) or the literary vocabulary that is emphasized in reading/language arts programs (e.g., *tranquil*, *bayonet*), general academic vocabulary is not the focus of instruction in either subject areas or reading/language arts. Typically, knowledge of general academic vocabulary such as *form*, *model*, and *system* is assumed by authors of content-area texts, even though these words often change their meanings, parts of speech, and morphological forms in different subject areas. A high percentage of general academic vocabulary words are Latin-based, providing the possibility of a cognate advantage to Spanish-speaking students.

This potential cognate advantage has a historical explanation. Spanish descended directly from Vulgar Latin and Latin-based words are used for everyday communication purposes in Spanish. Corresponding Latin-based words in English are often more sophisticated than the more frequent German-origin vocabulary words. For example, *construct* and *construir* are cognates, descended from the same Latin word *construere*. However, *construir* is much more frequent than

construct and is used for everyday communication in Spanish. The asymmetrical relationship between academic vocabulary words in Spanish and English is due to the direct descent of Latin to Spanish (simple word to simple word) and the circuitous path that Latin words followed as they were incorporated into English (simple word to more complex word). Some Latin-based words entered English via French as a result of the French domination of England from 1066 through 1399. Other Latin words came directly into English during the Renaissance to meet demands for a sophisticated scientific and literary register that the English language lacked (Barber, 2000).

Despite the potential advantage that cognates offer, bilingual students often fail to notice cognate pairs even when they appear to be quite transparent (August, Carlo, & Snow, 2005, Feldman & Healy, 1998; Garcia, 1991; Nagy, 1995; Nagy, Garcia, Durgunoglu, & Hancin-Bhatt, 1993). Nagy et al. (1993) documented that fifth- and sixth-grade bilingual, biliterate Spanish-speaking students circled less than half of the known cognates that they encountered on a test of cognate identification. The reasons why students find cognate identification so difficult are not fully understood. It seems likely that cognate transparency is mediated by individual differences, exposure to cognate instruction, and by the complex array of semantic, orthographic, and phonological features that characterize particular sets of cognates (August et al., 2005). Cognate pairs rarely match in every way. The degree of orthographic, phonologic, and semantic overlap between cognates can be viewed as a set of inter-related continua, ranging in each dimension from identical to very different.

Semantic Factors

Semantic relatedness is the "gold standard" in terms of cognate status, determining whether orthographically similar words in Spanish and English can be used by bilinguals in

cross-linguistic transfer. However, semantic relatedness is not a simple construct. Spanish-English cognates share a common Latin root, but the languages have evolved over time and cognates do not always mean precisely the same thing in terms of contemporary usage. Trask (1996) identified seven categories that describe the relationship between cognates that differ semantically, four of which are relevant to this discussion: 1) generalization: the English word is more general than the Spanish word, 2) specialization: the English word is narrower and more specific than the Spanish word, 3) melioration: the English word has a more positive meaning than the Spanish word, 4) pejoration: the English word is more negative in meaning than the Spanish word. Table 1 contains examples of Spanish-English cognates that illustrate generalization, specialization, melioration, and pejoration.

The examples in Table 1 demonstrate the complexity of cognate relatedness. For example, the pair *molest/molestar* is an example of pejoration, a phenomenon that occurred over many centuries as Latin-based words entered English. The Spanish word *molestar* descended directly from the Latin word *molestare* (to bother or annoy) and retained the original meaning. The word, *molest* entered English via Old French around the twelfth century, gradually diverging from *molestare* and acquiring a deviant sexual connotation. Despite divergence in meaning, it is important to note that the semantic association between most of these cognate pairs is evident.

The term "false cognate" is often applied to any set of words that do not mean precisely the same thing in two languages, such as *molest/molestar* (Prado, 1996). However, this term should be reserved for words that are entirely unrelated such as rope (a braided cord) and ropa (clothes) or words that have diverged so greatly that *no* semantic overlap can be discerned, such as assist (help) and *asistir* (attend). Word pairs that are etymologically related, but share less than full meaning, can more accurately be labeled *partial-cognates*. The degree of semantic

overlap can be thought of as a continuum with “full” cognates which have identical meanings in the two languages (e.g., *art/arte*) at one end of the continuum and false cognates at the other end (e.g., *rope/ropa* (*rope/clothes*)). Between the two extremes are partial cognates with varying degrees of semantic overlap (e.g., *molest/molestar*).

An additional complication that must be considered in evaluating the semantic transparency of Spanish-English cognates is the polysemy of each word in its respective language. One set of meanings might share full cognate status, while alternate meanings of the same words may be completely unrelated. For example, the English word *letter* (letter of the alphabet) corresponds to the Spanish word *letra*, both of which descended from and retained the same meaning as the Latin word *littera*. But the alternate meaning of the English word *letter* (something that is written and mailed) is a non-cognate translation equivalent to the word *carta* in Spanish. The ubiquity of polysemous words in both languages means that many cognate pairs will have more than one place on the continuum of semantic relatedness. *Letter* and *letra* are simultaneously full cognates in the sense of the alphabet and false cognates in terms of something that is written and mailed.

The problem of polysemy in terms of cognate relatedness also extends to morphemes that have more than one possible match in the other language. For example, the morpheme *vac* in the English word *vacation* might be linked correctly to the Spanish word *vacación* or incorrectly to the Spanish word *vacante*. Although both words are originally descended from the common Latin root, *vacare*, meaning *empty*, the word *vacation* entered English via French and acquired the meaning *free from an activity*. The meanings of the words diverged so greatly, that the morphemes no longer carry related meaning, resulting in a false pairing of *vacation-vacante*.

Linguists call the problem of more than one possible cognate match onomasiological multiplication (Sales, 1998-1999).

Despite the problem posed by incomplete correspondence between cognates, studies have shown that more the 90% of Latin-based cognates (French-English and Spanish-English) are "true cognates", sharing substantial overlap in form and meaning (Granger, 1993; Moss, 1992).

Orthographic Factors

Cognates are not merely words that share meaning; they also share orthographic features that illustrate their common origin. Just as cognates vary along a continuum of semantic relatedness, they vary in orthographic overlap. The more similar the spelling of an English cognate is to its Spanish equivalent, the greater the degree of orthographic transparency of that particular cognate. Cristoffanini, Kirsner, & Milech (1986) categorized cognates into five groups according to their orthographic relationship. The categories include (1) orthographically identical cognates, (2) common stem cognates (cion-tion), (3) common stem cognates with regular suffix (dad-ty), (4) common stem cognates with irregular suffixes, and (5) morphologically unrelated translations.

Research has determined that orthographic transparency is a key factor in bilingual Spanish-speaking students' ability to benefit from cognates found in English texts (Nagy et al., 1993). Nagy et al. found that students were more successful in identifying cognates when words had clear orthographic overlap (e.g. *animal/animal*). They noted that even small spelling differences reduced students' ability to recognize English-Spanish cognate pairs. The importance of orthographic transparency in cognate recognition is underscored by cognate priming studies conducted in a variety of languages. This body of research measures students' response times to cognate and non-cognate pairs and has consistently documented faster responses to cognates. For

example, Dutch-English bilingual students in a variety of studies, recognized cognates more quickly on priming tasks, learned them more readily, and forgot them less frequently than non-cognate translation equivalents (de Groot & Keijzer, 2000; van Hell & de Groot, 1998).

Cristoffanini, Kirsner, and Milech (1986) found that subjects in their study responded to cognates at a similar rate of speed as inflections and derivations from the same language.

Bowers, Mimouni, and Arguin (2000) documented that French-English bilinguals responded more quickly to orthographically identical and highly similar cognates than to non-cognate translation equivalents. The researchers concluded that, "cognate relationships are explicitly coded within the orthographic system" (Bowers et al., 2000, p. 1292).

Phonological Factors

Phonological overlap also plays an important role in cognate identification and cross-linguistic transfer. In fact, some psycholinguists believe that cognate pairing is based almost entirely on phonological representations in memory (Carroll, 1992). According to Carroll, hearing a word in a second language automatically activates words in the first language that are acoustically similar. This phonological activation also applies to morphologically related words in the same language. Carroll explains that the degree to which semantic relatedness accompanies automatic phonological cognate pairing influences the amount of cross-linguistic transfer that occurs.

Pronunciation differences may obscure cognate relatedness, even when words appear very similar (Maldonado, 1997). Weak phonological correspondence between many cognate pairs complicates cognate recognition and makes it more difficult for bilingual students to transfer word meaning across languages. For example, Dressler (2000) examined fifth grade Latino students' cognate awareness and response to cognate strategy instruction and found that

the degree of phonological transparency was an important factor in bilingual students' ability to recognize cognates. August, Carlo, Dressler, and Snow (2005) suggested that phonological factors are particularly important in facilitating cross-language for bilingual students who are not literate in their native language and are unfamiliar with Spanish words in their written form.

Spanish and English share a large number of orthographically similar, etymologically related words; however the differing sound systems in the two languages can hinder cognate recognition if inappropriate phonological representations are automatically activated in response to print (Kroll & de Groot, 1997). Schwartz, Kroll, and Diaz (2007) noted that the efficiency of bilingual lexical processing results from a complex interplay of orthographic, phonological, and semantic mappings. When cognates do not match in each critical dimension, processing speed is slower and students' ability to utilize cognate information is reduced. The current investigation is designed to identify the factors that could facilitate or inhibit students' cognate recognition among general academic words.

Method

The investigation began with the identification of English-Spanish cognates in two corpora that are important to English learners: (a) one that consists of words based on high-frequency in written language overall—the General Service List (GSL) and (b) one that consists of words chosen for their appearance in numerous content areas—the Academic Word List (AWL).

The original GSL list (West, 1953) included 2,000 headwords that were identified as most useful to English learners because of their frequency and usefulness in written English. Baumann and Culligan (1995) updated the GSL including a total of 2,284 headwords ranked by

frequency, based on the Brown Corpus (Frances & Kucera, 1982). The current analysis used the updated GSL.

The AWL was developed by Coxhead (2000) as a means of providing university students, who were learning English as a second language, with words that were critical in reading academic texts in a variety of disciplines. The process began with the development of the Academic Corpus, a body of over 3,513,330 running words found in 414 academic texts in different subject areas. After developing the Academic corpus, Coxhead identified 570 headwords representing 3110 words not included in the GSL and likely to be found in academic texts. The criteria that Coxhead used for inclusion in the AWL were (a) specialized occurrence: the word does not appear on the GSL word list, (b) range: a member of the word family occurs at least 10 times in each of the four main sections of the Academic Corpus and in 15 of 28 subject areas, and (c) frequency: word family members must occur 100 or more times in the Academic Corpus. According to Coxhead, the GSL corpus accounts for 90% of the words found in fiction and 75% of the words found in nonfiction texts. According to Coxhead, the combination of the GSL and AWL corpora covers approximately 86% of the words found in the Academic Corpus (2000).

The first author (a proficient but not native speaker of Spanish) translated the GSL and AWL headwords into Spanish and identified cognates in each corpus. The cognate lists were then compared to those of Rubén Morán Molina, an educator at the International Bénédict Schools of Languages Entrerios in Guayaquil – Ecuador (available at <http://www.cognates.org>). The first author and Moran agreed on 91% of the GSL cognates and 85% of the AWL cognates. A native Spanish-speaking professor who was born in Mexico evaluated the list of discrepant words, determining which should be characterized as cognates. The cognate identification

process resulted in a final cognate corpus consisting of 426 AWL cognates accounting for 75% of the AWL total corpus and 772 GSL cognates, accounting for 34% of the GSL corpus. The consolidated cognate corpus, the basis for the analyses in this study, had a total of 1198 cognates.

Three analyses were conducted on the cognate corpus: (a) the pattern analysis was developed to classify cognates based on high-frequency orthographic shifts, (b) the transparency analysis examined the orthographic and phonological transparency of selected cognates from the GSL and AWL cognate corpora, (c) the frequency analysis examined the relative frequency of cognates in Spanish and English.

Pattern Analysis

A cognate scheme was developed to classify cognates according to orthographic patterns. The first author began by examining the cognate corpus. Predictable orthographic shifts between Spanish and English word pairs were identified and the cognate corpus was sorted by pattern. Three native Spanish-speaking teachers reviewed the list generated by the first author and suggested additional patterns. A revised list, including the patterns suggested by bilingual teachers was developed and a classification protocol was developed to facilitate the sorting of cognates into pattern groups. When cognates could be classified in more than one way, the most specific pattern possible was selected. For example, the cognate pair *natural/natural* was categorized as *pattern 2 (al/il)*, based on the specific *al* ending, rather than the more general "*same*" category. The classification protocol also limited the number of letter shift in patterns. For example, cognate pairs sorted into the *add/change* category could have no more than two letter shifts. (e.g. *group/grupo* has two shifts from English to Spanish - the deletion of the first *o* and the addition of the final *o*.) Cognate pairs with more than two letter shifts were classified as

"*other*", a general category designated for cognate pairs that did not fit into any of the specific patterns.

The first author (non-native Spanish-speaker) and a Mexican-American bilingual teacher independently sorted the cognates by pattern, using the classification protocol. The percentage of agreement between the two raters was 91.14% for the total cognate corpus. Ratings completed by a third rater (a Puerto Rican-American Bilingual teacher) were used to classify the cognate pairs when the first two raters disagreed. The orthographic shifts described in Table 2 are based on English words, because this investigation examines bilingual students' ability to recognize cognates found in English language texts.

Transparency Analysis

The transparency analysis examined the degree of orthographic and phonological transparency exhibited by cognates belonging to different patterns. Orthographic transparency was evaluated by calculating the *Longest Common Subsequence Ratio* (LCSR). This statistical method entails dividing the longest sequence of letters shared by two words by the total number of letters of the longer word (Kondrak, 2001). The resulting cognate coefficients are then compared to determine the relative transparency of cognate pairs. For example, the longest sequence of letters in the cognate pair *problem/problema* is p-r-o-b-l-e-m (7 letters), was divided by 8 (the number of letters in *problema*, the longer word), resulting in a coefficient of .88. The cognate pair *chemical/química* is much less orthographically transparent. The two words have a common four-letter sequence m-i-c-a, divided by 8 letters in the longer word produced a LCSR coefficient of .5.

Phonological transparency was determined by calculating the *Common Phoneme Ratio* (CPR). This method, developed by the first author, entails dividing the number of common

phonemes in the cognate pair by the number of phonemes in the longer word. For example, the words *problem* [p-r-ah-b-l-eh-m] and *problema* [p-r-oh-b-l-ay-m-ah] share five phonemes representing the sounds /p/, /r/, /b/, /l/, /m/ in the words. When the common phonemes (5) are divided the total phonemes in the longer word (8), the resulting coefficient (.63) provides an estimate of phonological transparency. It is important to note that, unlike LCSR, CPR is subjective and ratings are influenced by local and regional dialects in both languages.

Two sets of 42 cognate pairs, including two or more representatives of each cognate pattern (except 17 and 18), were selected from the cognate corpus. The first author (non-native Spanish-speaker) and two native Spanish-speaking teachers (one Mexican American and one Puerto Rican American) who have Reading Specialist certificates, independently calculated the number of phonemes in the English words, the number of phonemes in the Spanish words, and the number of common phonemes. Inter-rater reliability (Cronbach's alpha) for the CPR analysis ranged from .85 to .91 on the sets of words.

Frequency Analysis

The final analysis was a comparison of English and Spanish word frequency in terms of cognate pairs. The cognate corpora were divided into four word frequency zones in each language. The first Word Zone (A) includes the first 1000 words; Word Zone B represents words ranked 1001-3000; Word Zone C includes words ranked 30001-5000; and Word Zone D includes words with ranking of 5000 or higher. Word frequency zones were listed on a matrix (Table 4), allowing for a comparison of relative frequency between Spanish and English cognates. The analysis of relative word frequency was limited to words that appear uniquely on the AWL list, since words on the GSL are, by definition, highly frequent in English.

Results

Pattern Analysis

The pattern analysis revealed that the GSL and AWL lists vary in terms of the numbers of total cognates and distribution among the cognate patterns. Cognates comprise 34% of the words in the GSL and are distributed across 18 of the 20 cognate patterns. (No GSL cognates were identified as belonging to the *ing* or *ed* patterns.) The *add/change* pattern is the most frequent, comprising 25% of the cognates found in the GSL cognate corpus. The next most frequent categories include infinitives (19%) and *ion* (5%). Approximately 84% of the GSL cognates can be categorized into one of the patterns listed on the chart (all patterns except *other*).

In comparison to the GSL list, the AWL includes a higher percentage of cognates overall (nearly 75% of the AWL headwords are cognates), more of which can be categorized by pattern (93%). The largest number of AWL cognates can be categorized as infinitives (41%) followed by the *add/change* pattern (20%). Table 3 includes the percentages of cognates in the GSL and AWL corpora that can be categorized according to each cognate pattern.

The combined cognate corpus includes 1198 cognates, 87% of which can be categorized by a specific pattern. Four patterns (*ous*, *ly*, *ing*, *ed*) include less than 1% of the cognate corpus, suggesting that they might be dropped from future analyses. The remaining 16 patterns are grouped into four clusters: *same*, *add-change*, *verbs*, *es*. All of the cognate patterns, with the exception of *es*, entail consistent orthographic shifts in word endings.

Same Cluster. The four *same* cluster patterns (*same-misc.*, *same-al/il*, *same-ar/or*, *same-able/ible*) represent a large number of cognates in both corpora. The GSL cognate corpus includes 73 cognates (9%) that can be categorized according to one of the *same* patterns. Forty-one AWL cognates, representing 10% of the corpus, are orthographically same-cluster cognates.

Add-Change Cluster. The *add-change* cluster includes a wide range of patterns, the most frequent of which is the *add-change* pattern. This pattern includes a large number of words from both lists (25% of the GSL and 20% of AWL). Orthographic shifts in words in this pattern are simple, usually entailing the presence of an additional letter (*a*, *e*, *o*) at the end of the Spanish word (*art/arte*). In some cases the silent *e* in the English word is replaced by a voiced vowel in Spanish (*motive/motivo*). Other orthographic differences in this category include vowel diagraphs such as *ou* in English that are not present in Spanish (*group/grupo*) and letter shifts such as the presence of double consonants in English, but not Spanish (*effect/efecto*). Other add-change patterns such as *ory/ary*, *ty*, *ic/ical*, *ant/ent*, *ance/ence*, *ure*, *ous*, *ive*, and *ly* are relatively infrequent, appearing less in less than 5% of the cognates in either corpus.

Verb Cluster. The verb cluster consists primarily of *infinitives*, the highest frequency pattern on the AWL list (41%) and the second highest frequency pattern on the GSL (19%). The fact that the AWL analysis was limited to headwords, a large percentage of which are infinitives, may have inflated the percentage on this list. The *infinitive* pattern is quite complex because

Spanish infinitives can be constructed with *ar*, *er*, or *ir* ending. Within word letter shifts are common as the following examples demonstrate: *to charge/cargar*, *to include/incluir*, *to mark/marcar*, *to establish/establecer*. These orthographic differences substantially reduce the transparency of infinitive cognate pairs.

Es Cluster/Pattern. The *es* pattern had to be categorized as a distinct cluster because it is the only set of words that are characterized by a first letter shift. Words that begin with *sc*, *sp*, or *st* in English contain consonant blends that are difficult for Spanish-speakers to pronounce. Consequently, these words are spelled with an *e* before the *s* in Spanish (*student/estudiante*). *Es* pattern words are low frequency in the cognate corpus, comprising approximately 2% of GSL and AWL words.

Other Cluster/Pattern. All of the cognate pairs that do not fit one of the 20 patterns described above, are categorized as other. Most of these words are orthographically opaque as the following examples demonstrate: *paragraph/párrafo*, *technique/técnica*, *cell/célula*. Sixteen percent of the GSL and 7% of the AWL cognates are categorized as *other*.

Transparency Analysis

The second analysis examined the degree of orthographic and phonological transparency exhibited by cognates belonging to different patterns. Table 4 includes a representative sample of the cognates and their LCSR and CPR coefficients. Several patterns can be observed in the data in Table 4. The most obvious point is that cognates differ a great deal, both in terms of comparison to other cognates and in terms of the orthographic and phonological relatedness of one cognate to its pair. The correlation between LCSR and CPR coefficients is .22 (not significant), suggesting little relationship between orthographic and phonological transparency. The LCSR coefficients (mean .73) are generally much larger than the CPR coefficients (mean

.49) demonstrating that cognates are more substantially more transparent in terms of orthography than phonology. Four cognate pairs are spelled identically and an additional five sets had LCSR coefficients above .80, however none of the cognate pairs have a CPR - phonemic correspondence greater than .71.

Frequency Analysis

The final analysis, a comparison of word frequency between cognate pairs, was limited to words that appear uniquely on the AWL list, since words on the GSL are, by definition, highly frequent in English. Sixty-six cognate pairs could not be evaluated because the Spanish word ranking was unavailable, leaving 360 AWL cognate pairs out of a total of 426 to be evaluated in terms of relative frequency. The analysis (Table 5) revealed that 277 AWL cognate pairs (77%) were more frequent in Spanish than English; 66 cognate pairs (18%) were of equal frequency in the two languages, and 17 cognate pairs (.05%) were more frequent in English than Spanish. One hundred and thirty-seven cognates (38%) were substantially more common in Spanish than English, varying by two or three frequency zones. This category of cognates includes words such as *acquire*, *demonstrate*, *interpret*, and *motive* that are part of the academic register in English, while the corresponding cognates (*adquirir*, *demostrar*, *interpretar*, *motivo*) are everyday words in Spanish. The results of the frequency analysis demonstrate that a large percentage of AWL cognates are everyday words in Spanish, suggesting that Spanish-speaking students could be taught to use their Spanish word knowledge to comprehend academic texts in English.

Discussion

As students get older, their academic texts include an increasing number of conceptually complex words, a corpus of general academic and content vocabulary words that are essential to comprehension. Fortunately, a substantial number of these words are English-Spanish cognates.

Bravo et al. (2007) found that 76% of the words identified for instruction in the fourth-grade science units they reviewed were English-Spanish cognates. Carlo, August, McLaughlin, Snow, Dressler, Lippman, et al. (2004) concluded that 68% of the words judged to be difficult in middle-grade texts were cognates. The percentage of cognates in adult texts appears to mirror that found in texts designed for children. Martinez (1994) examined 257 sub-technical vocabulary words found in adult texts and found that two thirds of the words were cognates.

The high percentage of cognates in academic texts suggests that cognates might provide a powerful tool for bilingual students; however the advantage cognates might confer has yet to be documented in research. Two questions appear to be salient in terms of bilingual students' ability to identify and transfer cognate information from language to language: 1) Does the student *know* the meaning of the Spanish word that corresponds to the English word? 2) Can the student *access* the Spanish word meaning based on the English orthographic and phonological features? In response to the first question, bilingual students' semantic word knowledge in Spanish and English does not overlap nearly as much as we might expect. Young bilingual children appear to learn many words uniquely in Spanish or English, rather than learning words for the same concept in both languages (Oller, Pearson, & Cobo-Lewis, 2007; Umbel, Pearson, Fernandez, & Oller, 1992). According to Oller et al. (2007) the uneven distribution of bilingual vocabulary knowledge is related to the locus of language acquisition – whether words are learned at home or at school. Bilingual children are more likely to know words related to household activities such as sewing or cooking uniquely in Spanish, while classroom-related words such as *blackboard* are likely to be known exclusively in English. Vocabulary assessments given to bilingual students support the belief that bilingual vocabulary is distributed across languages. Bilingual Latino children typically score substantially lower than monolingual children when their vocabulary is

tested in either Spanish or English. When vocabulary test scores in both languages are compiled, however, the gap between bilingual and monolingual children narrows (Pearson, Fernández, & Oller, 1993; Umbel et al., 1992).

Despite the incomplete overlap of Spanish-English word knowledge, the frequency analysis suggests that there is a large body of everyday Spanish words that corresponds to a corpus of general academic vocabulary in English. The frequency analysis revealed that seventy-five percent of the AWL headwords are cognates, most of which are more common in Spanish than in English. For example, the AWL word, *terminate* is very rare in English, with a ranking of 16697. However, the cognate *terminar* is extremely common in Spanish, with a ranking of 219. Bilingual students are likely to know the meaning of common Spanish words such as *terminar*, providing them with the means to comprehend many academic English words. This simplifies the instructional task substantially. Rather than trying to teach a large corpus of completely unknown general academic vocabulary words, teachers can focus on the development of bilingual students' strategic skills and morphological and metalinguistic awareness needed to recognize and make use of cognates (Berninger & Nagy, 2008).

An answer to the second question, "can the student *access* the Spanish word meaning based on the English orthographic and phonological features," involves several considerations. The first issue that must be considered is whether bilingual students store vocabulary words in one lexicon or two. According to Smith (1997), bilinguals' languages are represented in separate lexicons. Studies demonstrating slower reading times for mixed language passages are cited as evidence of separate lexical access (Macnamara & Kushnir, 1971; Obler & Albert, 1978). However, research has also provided evidence for a common semantic representation across languages (Dufour & Kroll, 1995). In studies conducted by Caramazza and Brones (1980) and

Potter, Von Eckardt & Feldman (1984) fluent bilinguals categorized words in two languages into semantic categories with equal speed and accuracy. These studies provide evidence that bilingual students may have equally rapid access to semantic representations of words they know in both languages and that cognates are particularly likely to be stored in a common lexicon (Cunningham & Graham, 2000). The question of a common or separate lexicon is salient to this discussion in that it may explain the degree to which bilingual students are able to access cognate information efficiently when reading in a second language.

The differing orthographies of Spanish and English may also be a factor in bilingual students' ability to access cognates. According to the orthographic depth hypothesis (Smith, 1997), the process of reading in languages with shallow orthographies such as Spanish is distinctly different from reading in languages with deep orthographies such as English. When reading in languages with shallow orthographies, readers can rely on highly consistent spelling and may be able to access word meaning directly from the phonological coding (Frost, Katz, & Bentin, 1987). While reading in languages with deep, opaque orthographies such as English, readers must rely on lexical and semantic factors to pronounce words correctly and access meaning. Because shallow orthographies allow the reader to bypass the internal lexicon, researchers suggest that factors such as word frequency and semantic relatedness are less salient when reading these languages (Katz & Feldman, 1983; Frost et al., 1987). The disparate processes involved in accessing words in Spanish (a shallow orthography) and English (a deep orthography) might help to explain students' inconsistent recognition of Spanish-English cognates.

Numerous studies have demonstrated that bilingual students are more likely to recognize cognates that are orthographically similar (Caramazza & Brones, 1979; Crisoffanini et al., 1986).

But research has not determined whether bilingual students notice regular cognate patterns such as words ending in *ent/ente* and *ence/encia* in English and Spanish or recognize cognates belonging to these patterns more readily. Cognitive psychologists suggest that pattern recognition is key factor in reasoning and memory (Rips, 1994) and heightened ability to recognize patterns differentiates expert performance from that of novices (Bereiter & Scardamalia, 1993). Bilingual students' well-documented inconsistency in recognizing cognates that they encounter in texts may reflect lack of proficiency in detecting patterns. Helping them become familiar with high frequency cognate patterns and gain expertise in classifying cognates based on these patterns may make cognates easier to recognize and remember.

Pattern instruction based on AWL headwords may provide an effective vehicle for accelerating bilingual students' vocabulary growth, as each AWL headword represents more than five morphologically related words in English (Coxhead, 2000). Morphological awareness, the ability to notice that words are comprised of meaningful parts, may be particularly important to bilingual students because it facilitates cognate recognition and contributes to reading comprehension achievement, independent of vocabulary knowledge (Hancin-Bhatt & Nagy, 1994; Nagy et al., 2006). The pattern analysis conducted in this study suggests that cognates with similar orthographic features can be grouped for instruction. Systematically teaching students to recognize the orthographic shifts that characterize these patterns may help them develop the ability to identify cognates in texts. Cognates with lesser degrees of overlap, such as those belonging to the *other* pattern, may require more instruction. Cross-language transfer is not automatic for many bilingual students, emphasizing the need for increasingly explicit pattern instruction, in relationship to cognate opaqueness.

A limitation of the pattern analysis was the use of AWL headwords, rather than the complete AWL corpus. As a result, cognate patterns consisting of words with the inflected endings *ed* and *ing* were under-represented (only one cognate in the corpus followed either of these patterns). However, a preliminary review of the extended AWL word family list confirms the inclusion of large numbers of inflected words that end in *ed* and *ing*. English speaking children usually master words with inflected endings before they enter school and acquire words with derivation endings at a later point (Anglin, 1993; Carlisle & Fleming, 2003; Tyler & Nagy, 1989). While it is not clear which words are mastered first by bilingual students, a reasonable assumption is that bilingual students follow a similar trajectory in learning words with inflected and derivational endings as their English-speaking peers. Teachers of bilingual students may want to emphasize cognate patterns that include words with inflected endings first, before moving on to patterns that include more complex derivational endings such as *ous* (pattern 13) or *ive* (pattern 14).

Cognate transparency is quite complex and understanding factors that help or inhibit cognate recognition may be important in helping bilingual students access cognates. The transparency analysis demonstrates that a majority of English-Spanish cognates are more similar in terms of orthography than phonology. Although orthographic differences reduce cognate transparency, they are often predictable and can be identified and taught to bilingual students: (1) Consonant doubling: consonants are more likely to be doubled in English than in Spanish (*accept/aceptar, intelligence/inteligencia*), (2) Vowel complexity: vowels are much more complex in English than Spanish and often include digraphs, several of which may be used to spell the same sound. For example all of the following digraphs are used to spell *long u* in English: *oo, ou, ui, ew*. Spanish vowel sounds, on the other hand, are more likely to be spelled

with a single vowel. For example, the word *fruit* in English includes a digraph (*ui*), while the corresponding Spanish word *fruta* has a single intermediary vowel (*u*). The word *group* in English (*ou* digraph) corresponds to the Spanish word *grupo* (same intermediary vowel - *u*). (3)

Letter shifts: a letter or digraph in one language corresponds to a different letter or digraph in the other language. For example, the consonant digraph *ch* in English shifts to *qu* in Spanish, as can be observed in the following pairs (*machinery/maquinaria*, *chemical/química*). The letter *y*, when used as a vowel in English, shifts to *i* in Spanish (*style/estilo*, *cyle/ciclo*). Another common shift is *ph/f* as in *phase/fase*, *elephant/elefante*. A group of words derived from Greek via Latin exhibit often exhibit consonant shifts (*ph-f*) and/or vowel shift (*y-i*) in English and Spanish cognates. Examples include *physical/físico*, *philosophy/filosofía*, *phenomenon/fenómeno*.

Teaching these regular letter shifts between cognates may reduce confusion and help bilingual students recognize cognates more efficiently.

The transparency analysis revealed relatively low levels of phonological transparency in the cognate corpus. This finding can be explained by two factors: vowel pronunciation and syllable stress. Spanish vowels are highly regular and rarely correspond to their English equivalents in terms of pronunciation. The cognates *decide/decidir* illustrate how differing vowel sounds reduce the phonemic correspondence between cognate pairs. *Decide* is pronounced [dee-**sahyd**] and *decidir* is pronounced [day-see-**dir**]. Note that there is *no* correspondence between vowels in this set of words. The final syllable *dir* is stressed in the Spanish word, further accentuating pronunciation differences. Even orthographically identical cognates may sound very different in the two languages. For example, the word *animal* is spelled the same in both languages, but the English word is pronounced [**an**-uh-muhl] while the Spanish word is pronounced [ah-nee-**mal**]. Another example is the large group of cognates that end in /tion/ in

English and /ción/ or /sión/ in Spanish. These words are orthographically similar, but the final syllable is pronounced [shuhn] in English and [see-**ohn**] in Spanish (e.g. *nation* and *nación* are pronounced [**nay**-shun] and [nah-see-**ohn**]). The silent /h/ in Spanish is another example of the divergence between orthographic and phonologic transparency. Cognate pairs such as *human* [**hyoo**-muhn] and *humano* [oo-**ma**-noh]; *hero* [**hear**-oh] and *héroe* [**áy**-roh-ay] look very similar in Spanish and English but are pronounced very differently.

Nagy, Beringer, and Abbott (2006) point out that phonological complexity makes it more difficult for students to detect morphological relationships between words. The many phonological differences between English-Spanish cognates revealed by the transparency analysis may help to explain the weak cognate identification skills that that Nagy et al. documented in their studies (Garcia & Nagy, 1993; Nagy, Garcia, Durgunoglu, & Hancin-Bhatt, 1992; Nagy et al., 1993).

The transparency analysis also revealed a small, insignificant correlation (.22) between the orthographic and phonological coefficients, suggesting a lack of symmetry in terms of cognate overlap. Inconsistent mappings of sound and spelling across languages may confuse students and inhibit their ability to recognize cognates (Schwartz, Kroll, & Diaz, 2007). This issue may be addressed by teaching bilingual students to recognize phonological shifts between cognate pairs, particularly if they are not literate in Spanish and lack familiarity with Spanish orthography. Prompting students to evaluate whether an English word "looks or sounds" like a word they know in Spanish is an important facet of cognate instruction (Author A & others, 2008).

Cognates differ in multiple dimensions and may be more or less related in terms of orthography, phonology, and semantics. The incomplete semantic correspondence of many

cognates is of particular concern to educators and underscores the importance of strategic processes in cross-linguistic transfer. Careful instruction is needed to help bilingual students evaluate cognates in terms of the context in which they appear. The degree to which context supports comprehension and the student's skill at inferring word meaning from context are important factors in comprehension. Polysemous words are particularly challenging for bilingual students (August, Carlo, & Snow, 2005). Several studies have demonstrated that bilingual students' word knowledge was limited to only one meaning of polysemous words (August, Carlo, Lively, McLaughlin, & Snow, 2006; August, Carlo, Lively, Dressler, & Snow, 2005). Nagy, McClure and Mir (1995) noted that inferring word meaning from context was difficult for the bilingual middle school students they studied, due in part to the large volume of unknown words in texts. Word difficulty was related to conceptual difficulty, word length, morphological complexity, concreteness or abstractness, richness of context, and word frequency (Nagy, et al., 1995). Key factors that facilitate inferring word meaning from context include linguistic knowledge, world knowledge, and strategic knowledge. When linguistic knowledge is limited, heightened world knowledge and strategic knowledge may compensate, helping English learners acquire new English vocabulary from context (Nagy et al., 1995).

Learning to infer word meaning entails a complex interplay of cognate information in both languages and English textual clues. When cognates are closely related in each dimension – orthography, phonology, and semantics - the task of cross-linguistic transfer is facilitated. The greater the differences between cognates, the more challenging the task of inferring English word meaning. It is important that bilingual students acquire tools to infer the meaning of English cognates of varying levels of orthographic and phonological transparency and semantic relatedness. Rather than dismissing words as "false cognates" when they differ in contemporary

meaning, students can be challenged to figure out how partial cognate information can be used to construct meaning of a text. The processes that bilingual students use in identifying cognates and inferring word meaning from partial cognates builds cognitive flexibility, a key competency in skilled reading (Berninger & Nagy, 2008). Research has demonstrated that students with weak vocabulary development score significantly higher on reading comprehension tests when they have high levels of cognitive flexibility (Cartwright, Hodgkiss, & Isaac, 2008). Teaching students flexible cognate use entails breaking down the process of cognate identification, crosschecking context, and determining whether the meaning makes sense. The cognate strategy is similar to other cognitive strategies used to enhance comprehension. Students are likely to benefit from scaffolds such as cue cards, modeling, coaching, and gradual release of responsibility (Author A & Others, 2008; Rosenshine, 1997).

Bilingual students need to acquire a vast array of words, more quickly than other students, if they are to catch up to their monolingual peers (Ordonez, McLaughlin, & Snow, 2002). Cognates, particularly those that are related to general academic words in English, provide a potentially rich source of vocabulary growth for Spanish-English bilingual students, a population whose under-achievement is of serious concern to educators and policy-makers (Proctor, Carlo, August, & Snow, 2005; Snow & Kim, 2007). When bilingual students learn to infer the meaning of the 426 AWL headword cognates described in this study, they gain access to thousands of general academic words likely to be found in texts and used in academic discourse in a variety of content areas. The analyses included in this paper were designed to help educators understand the nature of English-Spanish cognates so that they can provide a more nuanced approach to cognate instruction.

References

- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58 (10), Serial #238.
- Author A & Others (2008)
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice* 20(1), 50-57.
- Berninger, V. & Nagy, W. (2008). Flexibility in word reading: Multiple levels of representations, complex mappings, partial similarities, and cross-modality connections. In K. Cartwright (Ed.). *Literacy processes: Cognitive flexibility in learning and teaching* (pp. 114-139). New York, NY: Guilford Press.
- Bowers, J., Mimouni, Z., & Arguin, M. (2000). Orthography plays a critical role in cognate priming: Evidence from French/English and Arabic/French cognates. *Memory & Cognition* 28(8), 1289-1286.
- Bravo, M, Author B, & Pearson, P. D. (2007). Tapping the linguistic resources of Spanish-English bilinguals: The role of cognates in science. In R.K. Wagner, A. Muse, & K. Tannenbaum (Eds.) *Vocabulary development and its implications for reading comprehension* (pp. 140-156). New York City, NY: Guilford.
- Caramazza, A. & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology*, 34, 77-81.
- Carlisle, J.F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading*, 7, 239-253.
- Carroll, S. (1992). On cognates. *Second Language Research* 8 (2) 93-119.

- Cartwright, K., Hodgkiss, M., & Isaac, M. (2008). Graphophonological-semantic flexibility: Contributions to skills reading across the lifespan. In K. Cartwright (Ed.) *Literary processes: Cognitive flexibility in learning and teaching*. New York, NY: Guilford Press
- Cristoffanini, P., Kirsner, K., & Milech, D. (1986). Bilingual lexical representation: The status of Spanish-English cognates. *The Quarterly Journal of Experimental Psychology* 38A, 367-393.
- Cummins, J. (1994). In C. Leyba (Ed.) *School and language minority students: A theoretical framework*. Sacramento, CA: California State Department of Education..
- Cunningham, T. & Graham, C. (2000). Increasing native English vocabulary recognition through Spanish immersion: Cognate transfer from foreign to first language. *Journal of Educational Psychology* 92, 37-49.
- de Groot, A. & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign language vocabulary learning and forgetting. *Language Learning* 50(1), 1-56,
- Frances, W., & Kucera, H. (1982). *Frequency Analysis of English Usage*, Houghton Mifflin, Boston
- Frost, R. Katz, L. Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 104-115.
- Garcia, G., & Nagy, W. (1993). Latino students' concept of cognates. In D. Leu & C.K. Kinzer (Eds.), *Examining central issues in literacy research, theory, and practice. Forty-Second Yearbook of the National Reading Conference*.

- Hancin-Bhatt, B. & Nagy, W. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics* 15, 289-310.
- Hiebert, E.H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In S.J. Samuels & A. Farstrup (Eds.), *What research has to say about vocabulary* (pp. 106-129). Newark, DE: International Reading Association.
- Hopstock, P.J., & Stephenson, T. G. (2003). *Descriptive study of services to LEP students and LEP students with disabilities. Special Topic Report #1: Native Languages of LEP Students*. Washington, DC: U.S. Department of Education OELA. Retrieved from http://www.ncela.gwu.edu/resabout/research/descriptivestudyfiles/voll_research_fulltxt.pdf
- Jiménez, R., García, G., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly* 31(1), 90-112.
- Katz, I. & Feldman, L. b. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9, 157-166.
- Kroll, J. & de Groot, A. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. de Groot & J. Kroll (Eds.) *Tutorials in bilingualism*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kuo, L, & Anderson, R. (2006). Morphological awareness and Learning to read: A cross-language perspective. *Educational Psychologist*, 41(3), p. 161-180.

- Maldonado, C. (1997). Lexical processing in uneven bilinguals: An exploration of English-Spanish activation and meaning. *Edinburgh Working Papers in Applied Linguistics*, 8. p. 76-97.
- Macnamara, J. & Kushnir, S. (1971). The bilingual's linguistic performance: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 10, 480-487.
- Nagy, W., Garcia, G., Durgunoglu, A., Hancin-Bhatt, B. (1992). *Cross-language transfer of lexical knowledge: Bilingual students use of cognates*. Champagne, IL: Technical Report 558.
- Nagy, W., Garcia, G., Durgunoglu, A., Hancin-Bhatt, B. (1993). English-Spanish bilingual students' use of cognates in English reading. *Journal of Reading Behavior*, 25(3), 241-259.
- Nagy, W., Berninger, V.W., & Abbott, R.B. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology*, 98(1), 134-147.
- Nagy, W. McClure, & Mir, M. (1995). *Linguistic transfer and the use of context by Spanish-English bilinguals*. Champagne, IL.: Technical Report 616.
- Obler, L. & Albert, M. (1978). A monitor system for bilingual language processing. In M. Paradis (Ed.), *Aspects of bilingualism* (pp. 105-113). Columbia, SC: Hombeam Press.
- Oller, K., Pearson, B., Cobo-Lewis, A. (2005). Profile effect in early bilingual language and literacy. *Applied Psycholinguistics* 28, 191-230.
- Potter, M., So, K., Von Eckardt, B., & Feldman, L. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.

- Prado, M. (1996). NTC's Dictionary of False Cognates. NY: McGraw-Hill/Contemporary.
- Rips, L. (1994). Deduction and its cognitive basis. In R. Sternberg (Ed.) *Thinking and problem solving*. San Diego, CA: Academic Press.
- Rosenshine, B. (1997). *The case for explicit, teacher-led cognitive strategy instruction*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Sales, S. (1998-1999). False friends in English for Spanish-speaking student of English: Morphology, syntax and lexis as sources of false friendship. *Jornades de Forment de la Investigació*, 4, 1-6.
- Schwartz, A., Kroll, J., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processing*, 22(1), 106-129.
- Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28, 649-667.
- van Hell, J. & de Groot, A. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition* (1), 3, 193-211.
- West, M. (1953). *A general service list of English words*. London: Longman.

Table 1. Categories of Semantic Changes in Spanish-English Partial Cognates

	Semantic Change	Spanish Word	Spanish Meaning	English Word	Latin Word	Derivation
Generalization	English meaning is more general	crimen miserable	crime of murder poor	crime miserable	criminis miserabilis	O. French O. French
Specialization	English meaning is more specific	campo parientes educación	field, country relatives upbringing	camp parents education	campus parens educatus	Latin O. French Latin
Melioration	English meaning is more positive	fracaso suceso	disaster outcome	fracas success	fragere/ quassare successus	French Latin
Pejoration	English meaning is more negative	desgracia molestar disgusto	mistake bother displeasure	disgrace molest disgust	dis+gratia molestare dis+gustare	M. French O. French M. French

Table 2 Cognate Clusters and Patterns

Cluster	Pattern	Differences Permitted	Examples
I SAME	(1) same - misc.	<u>no</u> differences (except accent)	area/área
	(2) al, il	<u>one</u> letter may be different	animal/animal
	(3) ar, or	<u>one</u> letter may be different	popular/popular, color/color
II Add- Change	(4) able, ible	<u>one</u> letter may be different	visible/visible
	(5) ion	up to <u>two</u> letters may be different plus ending & accent	nation/nación
	6 add-change	up to <u>two</u> letters may be different	fruit/fruta, group/grupo, art/arte
	7 ary, ery, ory	up to <u>two</u> letters may be different plus ending	necessary/necesario
	8 ty	up to <u>two</u> letters may be different plus ending	activity/actividad
	9 ic, ice, ical	up to <u>two</u> letters may be different plus ending & accent	intrinsic/intrínseco medical/médico
	10 ant, ent	up to <u>two</u> letters may be different plus ending	experiment/experimento instant/instante
	11 ance, ence	up to <u>two</u> letters may be different plus ending	influence/influencia, importance/importancia
	12 ure	up to <u>two</u> letters may be different plus ending	adventure/aventura
	13 ous	up to <u>two</u> letters may be different plus ending	famous/famoso
	14 ive	up to <u>two</u> letters may be different plus ending	active/activo
	15 y	up to <u>two</u> letters may be different plus ending	dictionary/diccionario
	16 ly	up to <u>two</u> letters may be different plus ending	finally/finalmente
III VERBS	17 ing	up to <u>two</u> letters may be different plus ending	pasando
	18 ed	up to <u>two</u> letters may be different plus ending	accepted/aceptado decided/decidido
	19 Infinitives	up to <u>two</u> letters may be different plus ending	to cost/costar to move/mover to decide/decidir
IV ES	20 Es (beginning)	letters may be different plus beginning es	student/estudiante
V OTHER	21 Other	any word that doesn't fit the other patterns or has too many differences	coffe/café

Table 3. Cognates in Clusters and Patterns

GSL						AWL						TOTAL					
Pattern	#Cogs	%	Cluster	#Cogs	%	Pattern	#Cogs	%	Cluster	#Cogs	%	Pattern	#Cogs	%	Cluster	#Cogs	%
1	16	2	I	73	9	1	5	1	I	41	10	1	21	2	I	114	10
2	31	4	II	412	53	2	21	5	II	171	40	2	52	4	II	583	49
3	19	2	III	144	19	3	10	2	III	174	41	3	29	2	III	318	27
4	7	1	IV	18	2	4	5	1	IV	10	2	4	12	1	IV	28	2
5	75	10	V	125	16	5	16	4	V	30	7	5	91	8	V	155	13
6	193	25				6	85	20				6	278	23			
7	13	2				7	7	2				7	20	2			
8	13	2				8	6	2				8	19	2			
9	19	2				9	13	3				9	32	3			
10	30	4				10	17	4				10	47	4			
11	27	3				11	5	1				11	32	3			
12	8	1				12	2	0				12	10	1			
13	5	1				13	0	2				13	5	0			
14	12	2				14	8	2				14	20	2			
15	16	2				15	12	3				15	28	2			
16	1	0				16	0	0				16	1	0			
17	0	0				17	0	0				17	0	0			
18	0	0				18	1	0				18	1	0			
19	144	19				19	173	41				19	317	26			
20	18	2				20	10	2				20	28	2			
21	125	16				21	30	7				21	155	13			
Total	772					Total	426					Total	1198				

Table 4. Analysis of Cognate Transparency

English	Spanish	Corpus Pattern	LCSR	CPR
idea	idea	GSL 1 (same)	1.00	0.50
civil	civil	AWL 2 (al/il)	1.00	0.53
nuclear	nuclear	AWL 3 (ar/or)	1.00	0.71
visible	visible	AWL 4 (able/ible)	1.00	0.48
nation	nación	GSL 5 (ion)	0.50	0.34
problem	problema	GSL 6 (add/change)	0.88	0.63
machinery	maquinaria	GSL 7 (ary/ery)	0.20	0.48
difficulty	dificultad	GSL 8 (ty)	0.60	0.50
music	música	GSL 9 (ic/ical)	0.83	0.50
patient	paciente	GSL 10 (ant/ent)	0.50	0.42
science	ciencia	GSL 11 (ance/ence)	0.71	0.52
culture	cultura	AWL 12 (ure)	0.86	0.38
precious	precioso	GSL 13 (ous)	0.75	0.25
active	activo	GSL 14 (ive)	0.83	0.50
economy	economía	AWL 15 (y)	0.75	0.50
founded	fundado	AWL 18 (ed)	0.43	0.43
evaluate	evaluar	AWL 19 (infinitive)	0.75	0.38
specific	específico	AWL 20 (es)	0.80	0.60
cycle	ciclo	AWL 21 (other)	0.40	0.65
MEAN			0.73	0.49

Table 5

Analysis of Cognate Frequency in English and Spanish

AWL	High Frequency	Moderate Frequency		Rare
	SPANISH Word Zones 0-2 First 1000	SPANISH Word Zone 3 1001-3000	SPANISH Word Zone 4 3001-5000	SPANISH Word Zones 5-6 5001+
ENGLISH Word Zones 0-2 First 1000	6	9	2	0
ENGLISH Word Zone 3 1001-3000	*	40	6	0
ENGLISH Word Zone 4 3001-5000	**	*	19	0
ENGLISH Word Zones 5-6 5001+	***	**	*	0

* More frequent in Spanish (one zone)

** More frequent in Spanish (two zones)

*** More frequent in Spanish (three zones)