



Cross-Linguistic Overlap English-Spanish Vocabulary: A Data-Driven Approach to Lexical Similarity

Research Article

Correspondence:	Maria Isabel Maldonado Garcia < spanishprofessor1@gmail.com >	Incharge/Professor, Institute of Languages and Linguistics, University of the Punjab, Punjab, Pakistan
-----------------	---	--

Publication Details

Received: June 25, 2025**Accepted:** November 25, 2025**Published:** November 30, 2025

Abstract

This study presents a computational methodology for identifying shared vocabulary between English and Spanish to support second language acquisition. Similarity indexes (S.I.) were calculated to determine the percentage of orthographic and semantic overlap between the two languages. Both Spanish and English use the Roman script, which allowed for string-based comparison. English has received multiple loanwords from Latin, from which Spanish also derives. This study analyses the most frequent vocabulary in both languages to assess the similarity level and extract similar lexical items. The results show that a lexical similarity level of 53.13% corresponding to 1594 shared lexical terms was calculated using a lexico-statistical computational method to compare the 3000 highest frequency terms of the basic vocabulary of English and Spanish, identifying shared vocabulary as a pedagogical tool.

Keywords: Cross-language similarities, shared vocabulary, cognates, loanwords, second language acquisition

1. Introduction

The second language acquisition literature is extensive, especially in comparative studies of language pairs. Spanish and English, both members of the Indo-European language family, are considered language relatives. The lexical-orthographic comparison of the Spanish and English



languages is relevant since these languages have large numbers of speakers, who are not only native but also speak them as a second language. For example, in the USA, Spanish is the second most spoken language (Ethnologue, 2023). Due to the rise of English because of globalization, technology, and trade worldwide, many Spanish speakers need to teach it.

The Indo-European family is the most prominent family of languages by speaker numbers. According to Ethnologue, 439 languages are under its umbrella, and the speakers compose 46% of the world's population. They are located, in the majority, in Europe and South Asia (Ethnologue, 2023).

According to Instituto Cervantes (2022), the most spoken languages worldwide are Chinese, English, Spanish, Hindi-Urdu, and Arabic. Instituto Cervantes (2022) says about it:

"[...] the projections indicate that the number of potential users of Spanish will continue to increase in absolute terms until 2068, the year in which it will exceed 726 million people, with different degrees of language proficiency. [...] The driving force behind the growth of the Spanish-speaking community will be the increase in speakers with native proficiency in this language, which currently represents 6.3% of the world's population (496,573,842 people)."

The Spanish language belongs to the Italic branch of the Indo-European family. Forty-seven million three hundred seventy thousand five hundred forty-two speakers speak Spanish in Spain, and 559 million worldwide speakers in more than forty countries (Ethnologue, 2023).

In its classification, Ethnologue describes it as "Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, Ibero-Romance, West Iberian, Castilian." It describes it as "*SVO; prepositions; genitives, relatives after noun heads; articles, numerals before noun heads; adjectives before or after noun heads depending on whether it is evaluative or descriptive; question word initial; (C(C))V(C); nontonal. Silbo Gomero whistled a variety of Spanish used in the Canary Islands. Braille script. Latin script, primary usage.*"

On the other hand, in its classification, English is listed as having 582 million speakers in the United Kingdom, and 334,800,758 speakers in countries such as the USA, Australia, and Canada, where English is a native language (Lewis, 2015). According to Ethnologue (2019), it is the most spoken language in more than a hundred countries, spoken either as a native or SL. It is also the official language of many nations and has become an international language. Ethnologue describes it like this: "*Indo-European, Germanic, West, English,*" Its main features are: "*SVO, prepositions, genitives after noun heads, articles, adjectives, numerals before noun heads, question word initial, word order distinguishes subject, object, indirect objects, given and new information, topic and comment, active and passive, causative, comparative, consonant and vowel clusters; nontonal.*" (Ethnologue, 2019). English uses the Latin alphabet as a script. Ethnologue states: "*Braille script. Deseret Alphabet was developed in 1854 with limited usage until 1877. Latin script, primary usage. Shavian (Shaw) script, no longer in use*" (Ethnologue, 2019).

In order to teach foreign languages swiftly, basic vocabularies were collected to aid students in learning languages. They include the most frequently used words in a particular language. Various basic vocabularies have been published. In the English language, Ogden and Richards' Basic English list was collected (1920-1968). For French, *Français Fondamental* was published in 1959.

For Spanish, *Vocabulario del Español Hablado* (1975) was published by Luis Marquez, although, previously, Víctor García Hoz wrote his *Vocabulario Usual, Vocabulario Común, Vocabulario Fundamental* in Madrid in 1953. (2004, pp. 55-56).

This study is conducted to identify the everyday lexical items in the high-frequency vocabularies of Spanish and English by comparing them to accelerate second language learning. Similarity indexes (S.I.) among languages are calculated to determine the percentage of similarities in the semantic structure of a language pair. Spanish and English use the Roman script, and English, a Germanic language, has received numerous loanwords from Latin, the language from which Spanish is derived. The present study focuses on the identification and quantification of shared vocabulary between English and Spanish to compare the 3000 most frequent words both languages to identify orthographic and semantic cognates. The resulting list is collected to support students of English and Spanish as a second language acquire core vocabulary efficiently.

1.2 Latin loanwords in English

It is widely known that English borrowed much of its vocabulary from Latin. According to Wollman (1993), Old English borrowed between 600 and 700 loanwords, 800 more were borrowed during various periods of the Brittonic languages (Cornish, Welsh, and Breton), and five hundred (early) Latin loanwords are commonly identified in West Germanic languages. Wollman (1993) considers the time of European Christianization, specifically the 6th century, a significant date for borrowing these loanwords.

Kavtaria (2011) states that the sixth century AD marked the peak of loanword borrowing due to Christianization. During this time, words like mass, monk, bishop, apostle, angel, altar, and other lexical items related to the workings of the church structure and services were borrowed. Kavtaria also mentions that other words entered the English language through French. However, loanwords from Latin were borrowed in the first century BC during the Roman and sub-Roman periods.

Another period of intense borrowing was after the Battle of Hastings (1066), which marked the Norman Conquest in the eleventh century, when Latin words entered the English language through French. Words like state, government, justice, crime, prison, parliament, council, power, court, library, lesson, autumn, dinner, uncle, table, plate, river, and others made their way into the language. England continued to borrow words during the Renaissance (16th and 17th centuries).

According to Kavtaria (2011), 65%-70% of English words are borrowed from various languages. In most cases, the reason for borrowing loanwords was the deficit theory, which means the absence of certain lexical items to denote specific realities. For example, potatoes and tomatoes were introduced to England by the Spanish, who had obtained them in the Americas. The English language adopted these words and adapted them from the Spanish (patata and tomate), which had borrowed them from Nahuatl (a language of South America) and other native-American languages.

1.3 Objectives of the Study

The study explores the orthographic overlap between two Indo-European languages: English and Spanish, to support second language acquisition. The methodology identifies shared high-

frequency terms from the 3,000 most common words in each language. For this reason, the objectives are:

1. To identify high-frequency shared vocabulary between Spanish and English.
2. To develop a shared vocabulary list based on orthographic and semantic similarity that supports SLA for English and Spanish language learners.

1.4 Research Questions

Accordingly, the research questions were formulated:

1. What is the percentage of similarity in English and Spanish?
2. Which high-frequency vocabulary is common in English and Spanish?

2. Literature Review

2.1 Lexical Similarity and Cognates

Cognate identification is essential in historical linguistics to establish genealogical relationships between languages (Campbell, 2004). Ethnologue defines lexical similarities: "The percentage of lexical similarity between two linguistic varieties is determined by comparing a set of standardized wordlists and counting those forms that show similarity in form and meaning. Percentages higher than 85% usually indicate a speech variant likely a dialect of the language with which it is being compared" (Lewis, 2013).

The preliminary studies on S.I. calculation were conducted by Swadesh in 1948 using a lexicostatistical method in various studies (Swadesh, 1948, 1950, 1952, 1955). Those studies were conducted to find family relations between languages. However, in this paper, the similarity analysis was conducted by recognizing cross-language cognates to identify lexical items that are common or similar in both languages. The varieties used for the comparison were the Mexican variety of Spanish and the American English, which was utilized since they have the most significant number of speakers—similarity levels. Between English and Spanish, the answer has yet to be revealed. Nevertheless, other figures are already extracted; for example, the S.I. between Portuguese and Spanish is 89%, while between English and Portuguese, it is 20.4% (Maldonado-Garcia & Borges, 2014). These indices were calculated by comparing word lists.

2.2 Methods for Measuring Lexical Similarity

To determine similarity levels between terms, we look at their level of overlap (Mackay & Kondrak, 2005, pp. 40-47). Brew, C., and McKelvie, D. (1996) had already utilized a corpus of word pairs for extraction for lexicography. Costa et al. (2005) and Sherkina-Lieber (2004:108-121 & 2008: 192-200) also researched the facilitating effects of cognate words in bilinguals later.

Allen D., & Conklin, K. (2013), who studied phonetic and semantic overlap (synonymy) between English and Japanese, Djistra et al. (2010) who investigated bilingual word recognition, and

language similarity between Dutch and English and Costa et al. who did not study lexical similarity, but the resistance in cognates between Spanish and English and the facilitating effects of cognate in bilingual speech production. The method utilized in this research was set by Blair (1990) and Levenshtein (1965, pp. 707-710). Furthermore, other manners of cognate identification were employed by Holmes and Ramos (1993), List (2012), and Rama, Kolachina and Kolachina (2013), and other scholars. Word similarity calculation methods are usually based on orthography (Wagner & Fischer, 1974), which is possible only when a language pair uses a particular script, or phonetic (Hall & Dawling, 1980), usually based on the IPA when languages use different scripts. These lexical similarities are mainly calculated based on genetic relations (Maldonado Garcia, 2013a, 2013b). However, identification can also occur through strings and synonymy similarities (Maldonado-Garcia & Yapici, 2014). String analysis of the alignment of the word symbols can be used to calculate the similarity. When orthography is utilized, the symbols' basic order and alignment are identical in the compared pair of languages.

In this manner, word similarity is compared by substituting, inserting, and deleting symbols. For our analysis, these symbols are orthographic characters. The figure shows some Indo-European languages and their similarity indexes.

Figure 1: Some Indo-European Languages and their Similarity Indexes

	Russian S.I.	French S.I.	Portuguese S.I.	German S.I.	Spanish S.I.
English	24%	27%	20.4%	60%	-

Source: Lewis (2013), English-Portuguese calculated by Maldonado Garcia & Borges (2014)

2.3 Latin Loanwords and their role in English

The similarity index between English and German is higher than that of other languages because English and German belong to the same family of languages, the Germanic family. However, Portuguese and French (languages that derive from Latin) have a lower S.I. The reason is that they do not belong to the Germanic family but to the Romance family. In this regard, the S.I. between them is comparatively lower, even with the numerous loanwords from Latin that English has received. The Russian language belongs to the East Slavic family, so the S.I. is also lower. Higher S.I.s indicate that the languages are genetic relatives. This study aims to determine S.I., or the similarity percentage between English and Spanish.

2.4 Existing gaps and study motivation

The motivation behind this methodology is the scarcity of studies investigating similarities in terms of percentages (mainly conducted by Ethnologue). Another is that basic vocabulary lists are an effective aid for second language learning that teachers can use in their language classes, as shown in Ramirez, Chen and Pasquarella (2013), Hancin-Bhatt and Nagy (2004), and Dressler et al. (2011).

The Oxford Dictionary lists cognate as having the exact linguistic derivation as another' (and 'formally related; connected: related to or descended from a common ancestor'), (Oxford Dictionary, 2016). For the List, cognate identification is calculated through the edit distance based on similarity. This edit distance is calculated through the alignment matches or mismatches (List, 2012, p. 125).

This study responds to the above-mentioned gap by applying computational similarity measures to achieve the identification of shared vocabulary of both languages to support teachers and learners by highlighting useful, high-utility terminology that is easily recognizable.

3. Methodology

3.1 Research Design

The approach employed in the study is a quantitative, computational, corpus-based approach intended to identify shared high-frequency vocabulary between English and Spanish. Various studies by Maldonado Garcia and Borges (2014), Maldonado Garcia and Gavrishyk (2016), Chohan and Maldonado Garcia (2019), and Chohan and Maldonado Garcia (2022) use a string similarity algorithm as used in this study.

The fundamental objective is to calculate lexical similarity utilizing the edit distance and then categorize the word pairs according to their level of orthographic similarity. The analysis produced a pedagogical list of useful common lexical items to support the teaching and learning of English and Spanish.

3.2 Data Collection

In this case, the corpora utilized have two sets of 3000 high-frequency words, one for each language. The data collection was facilitated through an internet search. The search yielded a variety of results. *The Oxford 3000 American English (most frequent words)* list, an existing corpus, was selected for the study. It constituted Corpus A. The words in this corpus were later correlated with the equivalent terms in Spanish obtained from Corpus B, for which another existing corpus was selected. This is *A Frequency Dictionary of Spanish: Core Vocabulary for Learners* (Davies, 2006).

3.3 Steps of the Methodology

3.3.1 Algorithm Selection

After the word list compilation, a high-frequency words corpus of 3000 words in English its equivalent list in Spanish were compared using the Levenshtein algorithm, used by Kessler (1995) and Rama, Kolachina, and Kolachina (2013) as well as other researchers for the computational dialect comparison. The algorithm calculates the edit distance, which refers to the minimum number of processes necessary to transform one string into another simply through a character's substitution, insertion, or deletion. The edit distance between the symbols in a string a and b is given by $lev\ a,b (|a|, |b|)$:

Figure 2: The Algorithm (Levenshtein)

$$\text{lev}(a, b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \text{lev}(\text{tail}(a), \text{tail}(b)) & \text{if } \text{head}(a) = \text{head}(b), \\ 1 + \min \begin{cases} \text{lev}(\text{tail}(a), b) \\ \text{lev}(a, \text{tail}(b)) \\ \text{lev}(\text{tail}(a), \text{tail}(b)) \end{cases} & \text{otherwise} \end{cases}$$

Source: (Levenshtein 1965)

With the above-mentioned formula, the following computed example was drawn:

Figure 3: Computed illustration of Levenshtein distance between the sequences ‘gota’ & ‘goto’

		G	O	T	A
	0	1	2	3	4
G	1	1	2	3	4
O	2	2	1	2	3
T	3	3	2	1	2
O	4	4	3	2	1

Cognates can now be identified through orthographic and string alignment.

Categories were set according to the level of edit distance. This is significant because the strings' length may differ and present the same edit distance. This would manifest in different levels of similarity even when the distance is the same. The higher the similarity, the lower the edit distance. An edit distance of zero means that the pairs are exactly the same. In this regard, the lower the edit distance and the longer the string, the higher the similarity percentage.

3.3.2 Distance Computation

Initially, a 3000-word high-frequency word list of English was obtained, followed by the Spanish equivalents. Then, the English language high-frequency words were placed in column 1 on an Excel spreadsheet for symbol comparison with Spanish in column 2. Finally, the edit distance calculation was placed in column 3.

3.3.3 Analytical Model and Category Assignment

At this point, several categories for assessing similarity were mapped with an adapted method that combines Blair (1990) and a model mentioned in Mackay and Kondrak (2005), from which we drew inspiration for the alignment of the vocabularies and the setting of categories. The following categories were implemented.

C.1.: High similarity

- a. Equal terms. Those strings present a distance of zero mean a 100% similarity index.
- b. Terms that present one difference, and the remaining characters are positioned in the exact location in the strings.
- c. Terms with distance two because they differ on two characters.

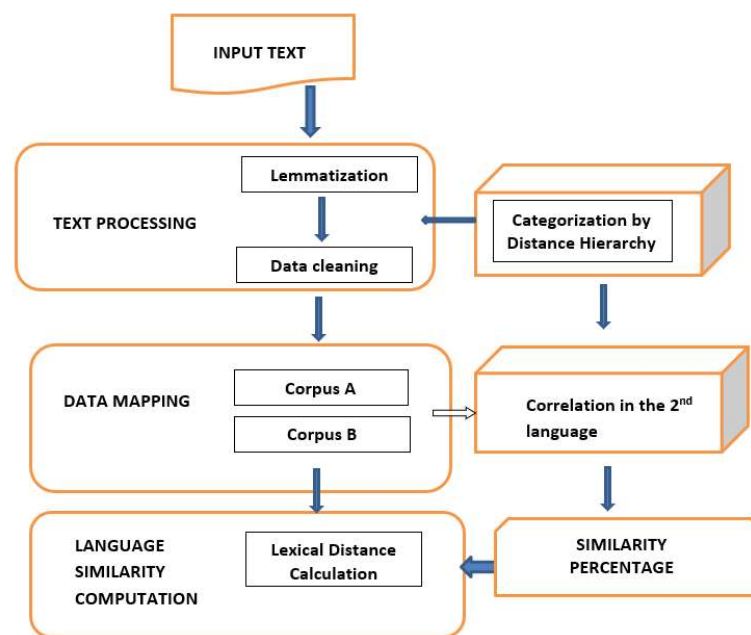
C. 2: Moderate similarity

- d. Terms that present three or more characters that are different.

C. 3: No similarity

- e. Terms not similar orthographically.

Figure 4: Process for determining the similarity level



3.3.4 Ethical Considerations

As this study involved computational analysis of public language corpora, human participants were not involved. Therefore, an ethical review committee approval was not required. This study was conducted using publicly available linguistic data. No personal, or proprietary data were accessed or used at any stage. The process adhered to international ethical standards for data usage, ensuring that all materials were freely accessible, non-identifiable and intended for linguistic and educational research purposes. The corpora used followed the principles of size, representativeness and balance across semantic domains and representativeness related to vocabulary usage in everyday language.

This approach strengthens the fairness, reliability and pedagogical relevance of the outcomes obtained.

3.3.5 Limitations of the Study

This study presents various key limitations:

1. Focus on orthographic similarity: As both, English and Spanish share the same (Roman) script, some similar words that differ in spelling may not have been identified as cognates.
2. High frequency vocabulary only: While this selection aids in pedagogical relevance, it excludes semantically essential vocabulary.
3. No contextual or syntactic analysis: The study does not account for part-of-speech differences or grammatical roles which may affect the usage of certain words.
4. No machine or AI data was used.

4. Analysis of the Data

4.1 Corpora Comparison

In Maldonado Garcia and Borges (2014), the S.I. calculation was performed by collecting 500 high-frequency words in English. In this case, the corpora utilized have two sets of 3000 high-frequency words, one for each language, and an internet search facilitated data collection. The search yielded a variety of results. *The Oxford 3000 American English (most frequent words)* list was selected for the study. This corpus is a high-frequency list that constituted corpus A. The words in this corpus were later correlated with the equivalent terms in Spanish obtained from Corpus B (*A Frequency Dictionary of Spanish. Core Vocabulary for Learners*, Davies, 2006).

The list contains the highest frequency 3000 English words. Their equivalents in Spanish are included in the first 5000 words. Corpus A was set in the first column, while Corpus B was in the second column. For example, the closest equivalent of the word *century* is *centuria*, but this word was not present in Corpus B as it is not a high-frequency word; rather, the word *siglo*, with the same meaning, was, and so this term was placed in the second column even though *centuria* is more similar to *century* than *siglo* is. In the third column, the edit distance was calculated and documented like this:

Table 2: Calculation of the edit distance between English and Spanish

English	Spanish	Edit Distance
a	un	2
abandon	abandonar	2
ability	capacidad	7
able	capaz	4
about	sobre	5
above	arriba	5
abroad	en el extranjero	15
absolute	absoluto	1
absolutely	absolutamente	5
academic	académico	1
accept	aceptar	3
acceptable	aceptable	1
access	acceso	1
accident	accidente	1
accompany	acompañar	4
according	de acuerdo a	8
account	cuenta	5
accurate	preciso	8
accuse	acusar	3
achieve	lograr	7
achievement	logro	11
acknowledge	reconocer	8
acquire	adquirir	3

Following the steps, Corpus A (*The Oxford 3000 American English*) was mapped on an Excel sheet, and the data cleaning was carried out. In this regard, all P.O.S. classifications (like nouns, verbs, adverbs, and others) mentioned in the file were removed, and only the lexical items remained. At the end of the data cleaning of corpus A, 3000 words of the English language, American variety, remained in the Excel spreadsheet.

Corpus A was then matched with Corpus B. Corpus B consisted of the equivalents in Spanish (from *A Frequency Dictionary of Spanish. Core Vocabulary for Learners* (Davies, 2006)) of the lexical items of Corpus A, which were then mapped on the spreadsheet. The corpus for this study was then compiled with 6000 words.

4.2 Edit Distance Calculation

The distance between Corpus A and Corpus B was calculated at this point. The edit distance was calculated using the computational application Planetcalc Math Levenshtein Distance Calculator (2015), and the results were then mapped in column three of the spreadsheet, as in Table 2.

The corpora comparison yielded the following similarity results:

Category 1:

A) Terms with distance zero (S.I. 100%): actor-actor, actual-actual, ah-ah, album-álbum, alcohol, alcohol, animal-animal, area-área, artificial-artificial, bacteria-bacteria, balance-balance, banana-banana, bar-bar base-base, blog-blog, cable-cable, camping-cámping, cancer-cáncer, capital-capital, CD-CD, central-central, chip-chip, chocolate-chocolate, civil-civil, club-club, color-color, conclusion-conclusión, control-control, criminal-criminal, crisis-crisis, crucial-crucial, cruel-cruel, cultural-cultural, debate-debate, decision-decision, detective-detective, digital-digital, director-director, division-división, doctor-doctor, drama-drama, DVD-DVD, editor-editor, error-error, euro-euro, explosion-explósión, extra-extra, factor-factor, familiar-familiar, favor-favor, federal-federal, festival-festival, final-final, flexible-flexible, formal-formal, fundamental-fundamental, gas-gas,

general-general, global-global, golf-golf, hockey-hockey, honor-honor, horrible-horrible, horror-horror, hospital-hospital, hotel-hotel, humor-humor, idea-idea, ideal-ideal, individual-individual, industrial-industrial, instructor-instructor, internet-internet, jazz-jazz, junior-júnior, legal-legal, local-local, marketing-marketing, material-material, mental-mental, menu-menú, metal-metal, mineral-mineral, model-modelo, monitor-monitor, moral-moral, motor-motor, multiple-multiple, musical-musical, natural-natural, no-no, normal-normal, nuclear-nuclear, oh-oh, opinion-opinión, original-original, panel-panel, particular-particular, personal-personal, piano-piano, plan-plan, popular-popular, pose-pose, poster-póster, principal-principal, radio-radio, real-real, región-región, regional-regional, regular-regular, religión-religión, robot-robot, rock-rock, role-role, rural-rural, sandwich-sándwich, sector-sector, senior-sénior, series-series, sexual-sexual, similar-similar, simple-simple, social-social, solar-solar, taxi-taxi, televisión-televisión, terrible-terrible, total-total, tropical-tropical, unión-unión, versión-versión, via-vía, video-video, virtual-virtual, virus-virus, visión-visión, visual-visual, vital-vital, web-web. (144 pairs)

B) Terms with distance 1: *absolute-absoluto, academic-académico, acceptable-acceptable, access-acceso, accident-accidente, act-acto, action-acción, active-activo, apt-pto, administration-administración, adult-adulto, advance-avance, agent-agente, air-aire, alarm-alarma, alcoholic-alcohólico, alternative-alternativa, ambition-ambición, análisis-análisis, anual-anual, apart-aparte, apartment-apartamento, April-abril, argument-argumento, arms-arms, art-arte, artista-artista, artistic-artístico, aspecto-aspecto, autor-autor, automatic-automático, band-banda, basic-básico, billion-billón, bomb-bomba, border-borde, calm-calma, camera-cámara, candidate-candidato, capture-captura, case-caso, cast-casta, cause-causa, celebration-celebración, carácter-carácter, class-clase, classic-clásico, click-clic, client-cliente, columna-columna, combination-combinación, comercial-comercial, comfort-confort, competitive-competitivo, competitor-competidor, complete-completo, component-componente, concentration-concentración, concept-concepto, condition-condición, conduct-conducta, conflicto-conflicto, congress-congreso, consideration-consideración, constant-constante, construction-construcción, contact-contacto, context-contexto, continente-continente, contrast-contraste, contribution-contribución, convenient-conveniente, conversation-conversación, correct-correcto, cost-costo, creation-creación, creative-creativo, credit-crédito, crime-crimen, criterion-criterio, critic-crítico, culture-cultura, curve-curva, decade-decada, decent-decente, decoration-decoración, defense-defensa, definition-definición, demand-demanda, dentist-dentista, department-departamento, description-descripción, diagram-diagrama, diet-dieta, direct-directo, direction-dirección, discipline-disciplina, discusión-discusión, distribution-distribución, document-documento, dollar-dólar, domestic-doméstico, double-doble, dramatic-dramático, economic-económico, edition-edición, education-educación, elect-electo, election-elección, electric-eléctrico, electronic-electrónico, element-elemento, emotion-emoción, emotional-emocional, event-evento, exact-exacto, except-excepto, exhibit-exhibir, exhibition-exhibición, experiment-experimento, expert-experto, exploration-exploración, expresión-expresión, extreme-extremo, false-falso, fantastic-fantástico, favorite-favorito, figure-figura, firm-firme, form-forma, fortune-fortuna, function-función, future-futuro, garage-garaje, generation-generación, habit-hábito, he-él, helicopter-helicóptero, hero-héroe, hey-ey, historic-histórico, honest-honesto, human-humano, illegal-ilegal, image-imagen, imagination-imaginación, impact-impacto, important-importante, imposible-imposible, impresión-impresión, in-en, incident-incidente, incredible-increíble, indirect-indirecto, infection-infección, informal-informal, information-información, ingredient-ingrediente, initial-inicial, initiative-iniciativa, insect-insecto, Institute-instituto, institution-institución, instruction-instrucción, instrument-instrumento, intelligent-inteligente, intense-intenso, intention-intención, interest-interés, international-internacional, introduction-introducción, invention-inención, investigation-investigación, invitation-invitación, just-justo, lemon-limón, limit-límite, line-línea, lion-león, liquid-líquido, list-lista, literatura-literatura, lot-lote, magic-magia, map-mapa, mass-masa, may-mayo, military-militar, million-millón, mine-mina, minor-menor, minute-minuto, misión-misión, modern-moderno, moment-momento, move-*

mover, much-mucho, music-música, my-mi, narrative-narrativo, nation-nación, national-nacional, native-nativo, negative-negativo, net-neto, north-norte, not-no, note-nota, notion-noción, obligation-obligación, observation-observación, occasion-ocasión, ocean-océano, oficial-oficial, on-en, operation-operación, option-opción, or-o, order-orden, ordinary-común, organ-órgano, organization-organización, origin-origen, pair-par, paper-papel, part-parte, participant-participante, passion-pasión, per-por, perfect-perfecto, period-período, permanente-permanente, person-persona, perspective-perspectiva, pilot-piloto, planet-planeta, plant-planta, plastic-plástico, plate-plato, poem-poema, poet-poeta, politics-política, position-posición, positive-positivo, possible-posible, potential-potencial, practice-práctica, prediction-predicción, preparation-preparación, present-presente, presentation-presentación, president-presidente, princess-princesa, prison-prisión, problema-problema, process-proceso, producto-producto, production-producción, profesión-profesión, profesional-profesional, professor-profesor, program-programa, progress-progreso, protection-protección, public-público, publication-publicación, pure-puro, range-rango, rapid-rápido, rare-raro, reaction-reacción, reception-recepción, recommendation-recomendación, reduction-reducción, refrigerator-refrigerador, regulation-regulación, relation-relación, relative-relativo, remote-remoto, reporter-reportero, reputation-reputación, resident-residente, resolve-resolver, responsible-responsable, rest-resto, restaurant-restaurante, revolution-revolución, rich-rico, rock-roca, roll-rollo, romantic-romántico, salt-sal, satellite-satélite, secret-secreto, section-sección, selection-selección, senator-senador, session-sesión, severe-severo, sex-sexo, sincere-sincero, situation-situación, ski-esquí, solid-sólido, solution-solución, species-especies, spiritual-espiritual, stable-estable, staff-personal, studio-estudio, sum-suma, tablet-tableta, talent-talento, tea-té, temperature-temperatura, tennis-tenis, text-texto, tomato-tomate, tone-tono, tradition-tradición, traditional-tradicional, transition-transición, transport-transporte, tube-tubo, tunnel-túnel, turn-turno, ugly-feo, uniform-uniforme, universe-universo, unusual-inusual, urban-urbano, valley-valle, various-varios, vast-vasto, victim-victima, violent-violento, visit-visita, vitamin-vitamina, volumen-volumen, zero-cero, zone-zona. (347pairs).

C) Terms with distance 2: *abandon-abandonar, addition-adición, additional-adicional, admire-admirar, admit-admitir, adopt-adoptar, adventure-aventura, agency-agencia, agenda-agenda, alter-alterar, angle-ángulo, application-aplicación, apparent-aparente, arrest-arrestar, article-artículo, association-asociación, athlete-atleta, attention-atención, attitude-actitud, attraction-atracción, attractive-atractivo, audience-audiencia, baby-bebé, bank-banco, bank-banco, barrier-barrera, based-basado, basis-base, bath-baño, be-ser, blank-blanco, boat-bote, boot-bota, brilliant-brillante, button-botón, buy-comprar, calculate-calcular, cancel-cancelar, Captain-capitán, career-carrera, cat-gato, category-categoría, celebrate-celebrar, cell-celula, center-centro, ceremony-ceremonia, champion-campeón, characteristic-característica, circle-círculo, cite-citar, clear-claro, climate-clima, clue-clave, coast-costa, colored-coloreado, combine-combinar, comedy-comedia, commission-comisión, common-común, communication-comunicación, company-compañía, compare-comparar, compete-competir, complex-complejo, concéntrate-concentrar, concert-concierto, conclude-concluir, conference-conferencia, confirm-confirmar, consider-considerar, consist-consistir, construct-construir, consume-consumir, continue-continuar, continuous-continuo, contract-contrato, contribute-contribuir, convert-convertir, convince-convencer, courage-coraje, course-curso, court-corte, cream-crema, create-crear, creature-criatura, critical-crítico, cure-curar, curved-curvo, cycle-ciclo, data-datos, day día, decide-decidir, declare-declarar, decorate-decorar, defend-defender, define-definir, definite-definido, delicious-delicioso, depend-depender, describe-describir, desert-desierto, detail-detalle, detect-detectar, determine-determinar, determined-determino, dialogue-diálogo, diary-diario, different-diferente, dishonest-deshonesto, disk-disco, distance-distancia, distribute-distribuir, district-distrito, divide-dividir, divorced-divorcio, documentary-documental, dominate-dominar, donate-donar, dozen-docena, drug-droga, east-este, economy-economía, edit-editar, educate-educar, effect-efecto, efficient-eficiente, electrical-eléctrico, emergency-emergencia, energy-energía, enter-entrar, enthusiasm-entusiasmo, entire-entero, episode-episodio, equal-igual, escape-escapar, essay-ensayo, essential-*

esencial, estimate-estimar, evaluate-evaluar, evidence-evidencia, exam-examen, examine-examinar, excellent-excelente, executive-ejecutivo, exist-existir, existence-existencia, expand-expandir, expectation-expectativa, experience-experiencia, explore-explorar, export-exportar, expose-exponer, express-expresar, extend-extender, external-externo, extraordinary-extraordinario, family-familia, famous-famoso, fault-falla, fiction-ficción, float-flotar, flower-flor, frequency-frecuencia, fresh-fresco, front-frente, fruit-fruta, fund-fondo, garden-jardín, generate-generar, generous- generoso, genre-género, giant-gigante, go-ir, graduate-graduado, grain-grano, gray-gris, Group-grupo, guard-guardia, guitar-guitarra, hall-sala, hand-mano, historical-histórico, history-historia, hour-hora, identify-identificar, identity-identidad, if-si, ignore-ignorar, illustration-ilustración, imaginary-imaginario, imagine-imaginar, immediate-inmediato, immigrant-inmigrante, impatient-impaciente, import-importar, importance-importancia, impose-imponer, include-incluir, independent-independiente, indicate-indicar, industry-industria, influence-influencia, inform-informar, innocent-inocente, insist-insistir, inspire-inspirar, install-instalar, instance-instancia, internal-interno, interpret-interpretar, introduce-introducir, invent-inventar, investigate-investigar, invite invitar, island-isla, July-julio, June-junio, justice-justicia, kilometer-kilómetro, laboratory-laboratorio, lake-lago, language-lenguaje, law-ley, leader-líder, level-nivel, limited-limitado, live-vivo, logical-lógico, March-marzo, mark-marca, massive masivo, mathematics-matemáticas, maximum-máximo, medical-médico, médium-medio, memory-memoria, message-mensaje, meter-metro, method-método, mile-milla, mine-mío, minimum-mínimo, minister-ministro, minority-minoría, mom-mamá, muscle-músculo, museum-museo, nerve-nervio, nor-ni, numerous-numeroso, object-objeto, objective-objetivo, observe-observar, obvious-obvio, offense-ofensa, offensive-ofensivo, offer- oferta, one-uno, opposition-oposición, organize-organizar, organized-organizado, pace-paso, painter-pintor, palace-palacio, participate-participar, passport-pasaporte, patient-paciente, percentage-porcentaje, permit-permiso, pase-fase, photo-foto, photograph-fotografía, phrase-frase, plain-plano, platform-plataforma, pólice-policía, political-político, pollution-polución, poor-pobre, possession-posesión, potato-patata, power-poder, practical-práctico, pray-orar, prefer-preferir, prepare-preparar, prepared- preparado, presence-presencia, preserve-preservar, press-prensa, pretend-pretender, prevent-prevenir, primary-primario, prince-príncipe, principle-principio, prisoner-prisionero, private-privado, produce-producir, producer-productor, project-proyecto, promise-promesa, promote-promover, proper-propio, propose-proponer, prospect-aspecto, race-raza, rank-rango, realistic-realista, reason-razón, reasonable-razonable, recent-reciente, reduce-reducir, reference-referencia, register-registro, relate-relatar, religious-religioso, repair-reparar, represent-representar, rescue-rescate, reserve-reservar, resist-resistir, respect-respeto, respond-responder, reveal-revelar, revise-revisar, route-ruta, routine-rutina, royal-real, salary-salario, scale-escala, scene-escena, scientific-científico, secretary-secretario, secure-seguro, see-ver, senate-senado, sentence-sentencia, separate-separado, serious-serio, serve-servir, service-servicio, seventy-setenta, signal-señal, silence-silencio, site-sitio, six-séis, soup-sopa, specialist- especialista, specific-específico, spirit-espíritu, standard-estándar, station-estación, statue-estatua, stress-estrés, strict-estricto, structure-estructura, stupid-estúpido, sun-sol, sweater-suéter, symbol-símbolo, system-sistema, temporary-temporal, tend-tender, theme-tema, three-tres, title-título, tourism-turismo, tourist-turista, traffic-tráfico, train-tren, transfer-transferir, transform-transformar, trick-truco, type-tipo, urge-urgir, use-usar, used-usado, used-usado, value-valor, vegetable-vegetal, vehicle-vehículo, victory-victoria, violence-violencia, vote-votar, west-oeste, wine-vino, you-tú. (405pairs).

Category 2

A) Terms with distance 3: *accept-aceptar, accuse-acusar, acquire-adquirir, activity-actividad, actress-actriz, affect-afectar, aggressive-agresivo, aid-ayuda, airline-aerolínea, alive-vivo, analyze-analizar, anniversary-aniversario, anxious-ansioso, appeal-apelar, appreciate-apreciar, architect-arquitecto, architecture-arquitectura, armed-armado, assist-asistir, assistant-asistente, associate-asociado, assume-asumir, atmosphere-atmósfera, attack-atacar, attract-atraer, August-agosto, battery-batería, battle-batalla, bee-abeja, benefit-beneficio, bicycle-*

bicicleta, biology-biología, blue-azul, bottle-botella, brief-breve, camp-acampar, campaign-campaña, capable-capaz, capacity-capacidad, celebrity-celebridad, cent-centavo, channel-canal, charge-cargar, circumstance-circunstancia, classical-clásico, clause-cláusula, code-código, coffee-café, collapse-colapsar, college-colegio, committee-comité, communicate-comunicar, comparison-comparación, complicated-complicado, confusing-confuso, connect-conectar, connected-conectado, connection conexión, consequence-consecuencia, consumer-consumidor, contemporary-contemporáneo, content-contenido, convinced-convencido, copy-copiar, corporate-corporativo, count-contar, criticism-crítica, criticize-criticar, cross-cruz, curtain-cortina, dad-papá, debt-deuda, deliberate-delibero, demonstrate-demostrar, desire-deseo, destroy-destruir, detailed-detallado, diamond-diamante, dictionary-diccionario, difference-diferencia, difficult-difícil, difficulty-dificultad, disaster-desastre, doubt-duda, during-durante, each-cada, educated-educado, electricity-electricidad, elephant-elefante, emphasis-énfasis, employ-emplear, encounter-encontrar, end-fin, enemy-enemigo, enormous-enorme, enthusiastic-entusiasta, entrance-entrada, entry-entrada, ethical-ético, example-ejemplo, expect-esperar, explanation-explicación, explode-explotar, fail-fallar, fall-caer, February-febrero, fence-cerca, finance-finanzas, financial-financiero, fine-bien, fixed-fijado, flow-fluir, fry-freír, gain-ganar, gallery-galería, geography-geografía, glass-vaso, god-dios, gold-oro, govern-gobernar, governor-gobernador, hard-duro, hello-hola, her-su, hi-hola, his-su, home-hogar, hurricane-huracán, hurt-herir, illustrate-ilustrar, included-incluido, including-incluido, inner-interno, intelligence-inteligencia, interested-interesado, interrupt-interrumpir, into-en, invest-invertir, judge-juez, juice-jugo, lamp-lámpara, lesson-lección, license-licencia, light-ligero, long-largo, machine-máquina, majority-mayoría, master-maestro, me-mí, means-medio, member-miembro, mention-mencionar, mind-mente, monkey-mono, more-más, mount-montar, mountain-montaña, movement-movimiento, musician-músico, mysterious-misterioso, mystery-misterio, name-nombre, necessary-necesario, nervous-nervioso, new-nuevo, night-noche, nine-nueve, november-noviembre, october-octubre, offend-ofender, online-en línea, only-sólo, organizar-organizador, other-otro, page-página, paint-pintar, pale-pálido, park-parque, past-pasado, pattern-patrón, pay-pagar, peace-paz, permission-permiso, personality-personalidad, phenomenon-fenómeno, poetry-poesía, point-punto, policeman-policía, policy-política, politician-político, popularity-popularidad, population-población, port-puerto, possess-poseer, predict-predecir, price-precio, priority-prioridad, profile-perfil, protect-proteger, prove-probar, provide-proveer, publish-publicar, purple-púrpura, qualification-calificación, quarter-cuarto, read-leer, reality-realidad, receipt-recibo, recipe-receta, recommend-recomendar, recycle-reciclar, red-rojo, reflect-reflejar, relaxed-relajado, repeat-repetir, repeated-repetido, representative-representante, require-requerir, result-resultado, retire-jubilarse, review-revisar, risk-riesgo, river-río, road-route, root-raíz, rule-regla, salad-ensalada, sauce-salsa, science-ciencia, sculpture-escultura, sea-mar, second1 (next after the first)-segundo, second2 (unit of time)-segundo, secondary-secundario, sensitive-sensible, september-septiembre, sequence-secuencia, seven-siete, silk-seda, sir-señor, slave-esclavo, society-sociedad, soil-suelo, soldier-soldado, solve-resolver, sound-sonido, south-sur, space-espacio, sport-deporte, stadium-estadio, state-estado, statistic-estadística, stomach-estómago, strategy-estrategia, style-estilo, subject-sujeto, substance-sustancia, suffer-sufrir, sugar-azúcar, suppose-suponer, sure-seguro, surprise-sorpresa, swear-jurar, talented-talento, tank-tanque, task-tarea, taste-gusto, technical-técnico, technology-tecnología, telephone-teléfono, term-término, theater-teatro, theory-teoría, therapy-terapia, throw-tirar, time-tiempo, tongue-lengua, touch-tocar, towel-toalla, tower-torre, treat-tratar, two-dos, typical-típico, unique-único, unit-unidad, united-unido, university-universidad, vacation-vacaciones, variety-variedad, vary-variación, view-vista, voice-voz (308pairs).

B) Terms with distance 4: *accompany-acompañar, advanced-avanzado, ancient-antiguo, anger-enojo, angry-enojo, appear-aparecer, apply-aplicar, appropriate-apropiado, approve-aprobar, aside-aparte, associated-asociado, authority-autoridad, bag-bolsa, bullet-bala, bus-autobús, certain-cierto, chain-cadena, charity-caridad, chat-charlar, city-ciudad, clock-reloj, colleague-colega, commit-cometer, community-comunidad, computer-computadora, confidence-confianza, confuse-confundir, confused-confundido, conscious-consciente,*

conservative-conservador, contain-contener, contest-concurso, cover-cubrir, daily-diario, December-Diciembre, deny-denegar, design-diseño, desperate-desesperado, destination-destino, discount-descuento, educational-educativo, employee-empleado, engineer-ingeniero, equipment-equipo, establish-establecer, exercise-ejercicio, expected-esperado, experienced-experimentado, explain-explicar, failure-falla, fascinating-fascinante, father-padre, fever-fiebre, fire-fuego, fish-pezu, five-cinco, flame-fuego, faith-fe, focus-enfocar, fold-doblar, food-comida, foot-pie, football-fútbol, force-fuerza, funding-fondos, gentle-amable, give-dar, glove-guante, goal-meta, good-bien, grab-agarrar, grand-grandioso, great-grande, imply-implicar, indoor-interior, interesting-interesante, iron-hierro, joke-broma, justify-justificar, lack-falta, league-liga, leave-dejar, length-longitud, let-dejar, light-luz, living-viviendo, location-ubicación, luxury-lujo, mad-loco, marry-casar, middle-medio, mild-leve, modify-modificar, month-mes, mood-humor, moon-luna, morning-mañana, mother-madre, motorcycle-motocicleta, multiply-multiplicar, nature-naturaleza, never-nunca, ninety-noventa, no one-nadie, novel-novedoso, obtain-obtener, occur-ocurrir, office-oficina, officer-oficial, opportunity-oportunidad, oppose-oponerse, opposed-opuesto, opposite-opuesto, oven-horno, over-sobre, paragraph-párrafo, partner-pareja, passage-paso, pen-pluma, percent-por ciento, pursue-perseguir, physics-física, play-jugar, pleasure-placer, possibility-posibilidad, powder-polvo, powerful-poderoso, pressure-presión, privacy-privacidad, prize-premio, pronounce-pronunciar, proof-prueba, property-propiedad, protest-prueba, psychology-psicología, put-poner, receive-recibir, recognize-reconocer, recover-recuperar, refer-referirse, relaxing-relajante, reservation-reserva, response-respuesta, responsibility-responsabilidad, retain-retener, rhythm-ritmo, rice-arroz, right-recto, round-redondo, row-fila, rude-brusco, same-mismo, sand-arena, satisfied-satisfecho, satisfy-satisfacer, scan-escanear, season-estación, seat-asiento, security-seguridad, sense-sentido, sensible-sensitivo, servant-servidor, shock-choque, sight-vista, significant-significativo, since-desde, skiing-esquiar, skin-piel, soft-suave, status-estado, story-historia, strange-extraño, study-estudiar, suggest-sugerir, surface-superficie, suspect-sospechar, symptom-síntoma, technique-técnica, thirty-treinta, treatment-tratamiento, twenty-veinte, typically-típicamente, ultimately-últimamente, unnecessary-innecesario, useful-útil, user-usuario, usual-normal, visitor-visitante, volunteer-voluntario, wind-viento, wow-guau(195pairs).

C) Terms with distance 5: absolutely-absolutamente, airport-aeropuerto, announce-anunciar, argue-argumentar, automatically-automáticamente, baseball-béisbol, basically-básicamente, bathroom-baño, block-bloquear, café-cafetería, certainly-ciertamente, chemical-químico, cigarette-cigarrillo, collect-recolectar, commonly-comúnmente, competition-competencia, completely-completamente, cook-cocinar, county-condado, covered-cubierto, custom-costumbre, degree-grado, departure-partida, depressed-deprimido, disappear-desaparecer, discover-descubrir, effort-esfuerzo, employment-empleo, engineering-ingeniería, especially-especialmente, eventually-eventualmente, excitement-excitación, factory-fábrica, female-femenino, finally-finalmente, finish-finalizar, generally-generalmente, gradually-gradualmente, guarantee-garantizar, impress-impresionar, impressed-impresionado, income-ingreso, investment-inversión, involve-involucrar, involved-involucrado, jacket-chaqueta, January-Enero, locate-localizar, maintain-mantener, market-mercado, measure-medida, medicine-medicamento, most-mayoría, naturally-naturalmente, need-necesita, Nobody-nadie, none-ninguno, normally-normalmente, originally-originalmente, pants-pantalones, Parent-padre, particularly-particularmente, Passenger-pasajero, personally-personalmente, philosophy-filosofía, photography-fotografía, physical-físico, poverty-pobreza, print-imprimir, proposal-propuesta, Psychologist-psicólogo, pursue-perseguir, qualified-calificado, quality-calidad, quantity-cantidad, quotation-cotización, rarely-raramente, really-realmente, recall-recordar, record-registro, regularly-regularmente, relationship-relación, relatively-relativamente, relax-relajarse, remember-recordar, remind-recordar, replace-reemplazar, Saturday-Sábado, school-escuela, scientist-científico, silent-silencioso, similarly-similarmente, sit-sentar, sixty-sesenta, stranger-extraño, supermarket-supermercado, survive-sobrevivir, ton-tonelada, totally-totalmente, translation-traducción, transportation-transporte, valuable-valioso (112pairs).

D) Terms with distance 6: *Appearance-apariencia, approval-aprobación, assignment-asignación, attempt-intentar, attend-asistir, bike-bicicleta, citizen-ciudadano, collection-colección, commitment-compromiso, consistent-coherente, constantly-constantemente, cooking-cocinando, correctly-correctamente, deliberately-deliberadamente, depressing-deprimente, designer-diseñador, directly-directamente, disadvantage-desventaja, effective-eficaz, effectively-efectivamente, emphasize-enfatizar, entertain-entretener, exactly-exactamente, examination-examen, government-gobierno, guide-guía, gym-gimnasia, humorous-humorístico, immediately-inmediatamente, impressive-impresionante, initially-inicialmente, midnight-medianoche, obviously-obviamente, occasionally-ocasionalmente, package-paquete, perfectly-perfectamente, phone-teléfono, photographer-fotógrafo, probably-probablemente, procedure-procedimiento, properly-propriadamente, rapidly-rápidamente, react-reaccionar, refuse-rechazar, reject-rechazar, related-relacionado, requirement-requisito, retired-retirado, simply-simplemente, specifically-específicamente, stretch-estirar, suggestion-sugerencia, surprised-sorprendido, translate-traducir, unconscious-inconsciente, unexpected-inesperado, unfair-injusto (58 pairs).*

E) Terms with distance 7: *Announcement-anuncio, apparently-aparentemente, approximately-aproximamente, clearly-claramente, entertainment-entretenimiento, extremely-extremadamente, frequently-frecuentemente, incredibly-increíblemente, necessarily-necesariamente, northern-delNorte, pointed-puntiagudo, possibly-posiblemente, recently-recientemente, similarity- semejanza, surely-seguramente, surprising-sorprendente, unemployed-desempleado (17pairs).*

F) Terms with distance 8 and above: *According-deacuerdo, basketball-baloncesto, planning-planificación, significantly-significativamente, disappointing-decepcionante, discovery-descubrimiento (6 pairs).*

Category 3

A) Terms not orthographically similar: he remaining words in the list fall outside the identified similarity thresholds and therefore do not exhibit orthographic proximity to the target items.

Table 3: Similar pairs and their distance

DISTANCE	# PAIRS
0	144
1	347
2	405
3	308
4	195
5	112
6	58
7	17
8	6
Total	1,594

The lexical similarity analysis revealed that 1,594 pairs present certain levels of similarity and are useful for second language acquisition.

Table 4: Total Lexical Similarity Index

S.I. ENGLISH-SPANISH	
LANGUAGES	ENGLISH-SPANISH
Pairs	1,594
S.I.	53.13%

Table 5: Improved Lexical S.I.

	German	French	Russian	Portuguese	Spanish
English	60%	27%	24%	20.40%	53.13%*

*Not genetic similarity, only lexical similarity

5. Conclusion

The study aimed to reveal the SI between Spanish and English and to extract the similar lexical items in both languages. The similarity index calculated can be used as a learning strategy for students of Spanish as an L2 (English L1) or English as an L2 (Spanish as an L1). Hence, a statistically driven algorithm was used to compute lexical similarity among two distantly genetically related languages. The comparison was based on the similarity of the segments. Categories were mapped where the degree of variation augments progressively as the Levenshtein or edit distance augments simultaneously. The higher the number of variations, the higher the edit distance between the string segments. In this case, the genetic relationship of the languages was not calculated as the fundamental goal was to extract common vocabulary in both languages, which in this case comes from loanwords borrowed by English. The results yielded a list of shared terminology in Spanish and English that can be useful for students of both languages.

In this context, the similarity between both languages is not necessarily due to cognacy, but rather to a large number of loanwords, mainly from Latin and to a minor degree from Greek, that have been assimilated into the English language at some point in history.

Funding: This study was not funded in any shape or form by any party.

Conflict of Interest: The author declares that he has no conflict of interest.

Bio-note:

Maria Isabel Maldonado Garcia holds a PhD in General Linguistics from UNED, Spain. Recognized for her outstanding contributions, she has been honored with the prestigious Fatima Jinnah National Pride Award. Currently serving as the Inchargeand Professor at the Institute of Languages and Linguistics, University of the Punjab, Maria Isabel's expertise and dedication continue to enrich the field of linguistics and education. Her remarkable journey and significant achievements make her a trailblazer in her domain, inspiring generations to come.

References

- Allen, D. B., and Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PloS one* 8(8), e72631. doi:10.1371/journal.pone.0072631
- Blair, F. (1990). *Survey on a shoestring: A manual for small-scale language surveys. Publication 98*. Dallas: Summer Institute of Linguistics, University of Texas at Arlington.
- Brew, C., and McKelvie, D. (1996). Word-pair extraction for lexicography. In *Proceedings of the 2nd International Conference on New Methods in Language Processing*, pp. 45-55.
- Campbell, L. (2004). *Historical Linguistics: An Introduction*. MIT Press.
- Centro Virtual Cervantes (2022) El español en el mundo. Anuario 2022. Instituto Cervantes. Accessed 07-02-2024 https://cvc.cervantes.es/lengua/anuario/anuario_22/informes_ic/p01.htm
- Chohan, M. N., & Maldonado García, M. I. (2022). Phonemic Comparison of Majhi and Shahpuri Dialects of Punjabi. *Jahan-e-Tahqeeq*, 5(1), 159-168.
- Chohan, M. N., & Maldonado García, M. I. (2019). Phonemic comparison of English and Punjabi. *International Journal of English Linguistics*, 9(4), 347-357.
- Costa, A., et al. (2005). On the facilitatory effects of cognate words in bilingual speech production." *Brain and language*, 94(1), 94-103.
- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. NY: Routledge.
- Dijkstra, T., Koji M., Brummelhuis, B., Sappelli, M. & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and language*, 62(3): 284-301.
- Dressler, C., Carlo, M. S., Snow, C. E., August, D., & White, C. E. (2011). Spanish-speaking students' use of cognate knowledge to infer the meaning of English words. *Bilingualism: Language and cognition*, 14(2), 243-255.
- Dryer, M. S., & Haspelmath, M. (2011). *The World Atlas of Language Structures online*. Munich: Max Planck Digital Library.
- Eberhard, D. M., G. F. Simons, and Charles D. Fennig (eds.). (2023). *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Hall, P. A.V., & Dowling, G. R. (1980). Approximate string matching." *ACM computing surveys (CSUR)*, 129(4): 381-402.

- Hancin-Bhatt, B., & Nagy, W. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics*, 15(3), 289-310.
- Haensch, G. (2004). *Los diccionarios del español en el siglo XXI (Vol. 10)*. Universidad de Salamanca.
- Holmes, J., & Ramos, R. G. (1993). False friends and reckless guessers: Observing cognate recognition strategies. *Second language reading and vocabulary learning*, 86-108.
- Kavtaria, M. (2011). Greek and Latin Loan Words in English Language (Tendencies of Evolution). *International Journal of Arts & Sciences* 4(18) 255-264.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. In *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, San Francisco: Morgan Kaufmann Publishers Inc., 60-66.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, 10(8): 707-710.
- Lewis, M. P., G. F. Simons, & C. D. Fennig, (Eds.). (2013) *Ethnologue: Languages of the World*. 17th ed. Dallas, TX: SIL International. Ethnologue. <http://www.ethnologue.com/language/spa> ; <http://www.ethnologue.com/language/eng>; <https://www.ethnologue.com/about/language-info>.
- List, J. M. (2012, April). LexStat: Automatic detection of cognates in multilingual wordlists. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH* (pp. 117-125).
- Mackay, W., & Kondrak, G. (2005, June). Computing word similarity and identifying cognates with Pair Hidden Markov Models. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)* 40-47).
- Maldonado García, M. I., & Gavrishyk, E. (2016). False Friends in Urdu and Russian. *Journal of Critical Inquiry*, 48.
- Maldonado García, M. I., & Borges de Souza, A. M. (2014). Lexical similarity level between English and Portuguese. *ELIA*, 14, 145-163.
- Maldonado García, M. I., & Yapici, M. (2014). Common vocabulary in Urdu and Turkish language: A case of historical onomasiology. *PakistanVision*, 15(1), 193.
- Maldonado García, M. I. (2013a). Comparación del léxico básico del español, el inglés y el urdu. Unpublished Doctoral Dissertation. Madrid: Universidad Nacional de Educación a Distancia.
- Maldonado García, M. I. (2013b). Estudio etimológico de cuatro pares de cognados en español y urdu. *Revista Iberoamericana de Lingüística (RIL)*, 8, 61-74.

- Oxford English Dictionary(2024).<http://www.oxforddictionaries.com/definition/english/cognate?q=cognate>. Retrieved on 17-1-2024.
- Rama, Taraka, Kolachina, Prasant & Kolachina Sudheer.(2013).Two Methods for Automatic Identification of Cognates.” In *Proceedings of Quantitative Investigations in Theoretical Linguistics (QITL-5)*. Leuven, Belgium, 76.
- Ramírez, G., Chen, X. & Pasquarella, A. (2013). Cross-linguistic transfer of morphological awareness in Spanish-speaking English language learners: The facilitating effect of cognate knowledge. *Topics in Language Disorders* 33(1): 73-92.
- Real Academia Española (2023).*Diccionario de la lengua española*[Dictionary of the Spanish Language] (23rd ed.). Madrid, Spain.
- Serfaty, J. and Serrano, R. (2024), Practice Makes Perfect, but How Much Is Necessary? The Role of Relearning in Second Language Grammar Acquisition. *Language Learning*, 74. 218-248. <https://doi.org/10.1111/lang.12585>
- Sherkina-Lieber, M. (2004). The cognate facilitation effect in bilingual speech processing; the case of Russian-English bilingualism. *Cahiers linguistics d'Ottawa*, 32: 108-121.
- Sherkina-Lieber, M. (2008). The cognate facilitation effect is a frequency effect: Evidence from Russian-English bilingualism. In G. Zybatow, L. Szucsich, U. Junghanns, & R. Meyer (Eds.), *Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003*. Frankfurt & Main: Peter Lang.192-200.
- Swadesh, M. (1948).*The time value of linguistic diversity*. Paper presented at the Viking Fund Supper Conference for Anthropologists, New York.
- Swadesh, M. (1950). Salish internal relationships. *International Journal of American Linguistics*, 16(4), 157-167.
- Swadesh, M. (1952). Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians And Eskimos. *Proceedings of the American Philosophical Society*, 96(4): 452-463.
- Swadesh, M. (1955). Towards Greater Accuracy in Lexicostatistic Dating.” *International Journal of American Linguistics*, 21(2): 121-137.
- Wagner, R. A., and Michael J. Fischer. (1974). The string-to-string correction problem. *Journal of the ACM (JACM)*219(1): 168-175.
- Wollmann, A. (1993) Early Latin loan-words in Old English. *Anglo-Saxon England*22, 1-26.